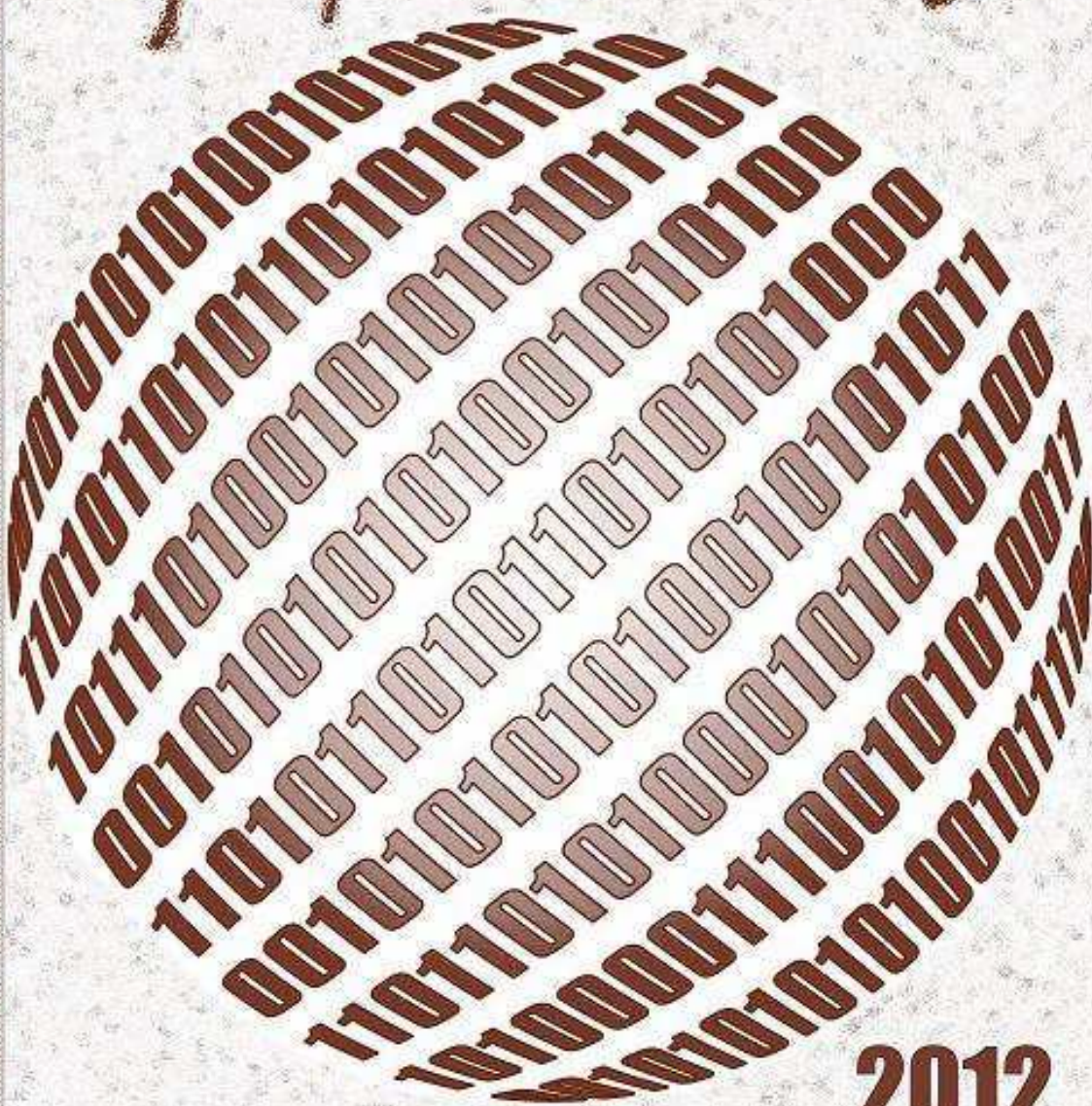


Информационные технологии
в гуманитарных исследованиях



2012

РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ АРХЕОЛОГИИ И ЭТНОГРАФИИ
РОССИЙСКАЯ АКАДЕМИЯ ЕСТЕСТВЕННЫХ НАУК
ЗАПАДНО-СИБИРСКОЕ ОТДЕЛЕНИЕ
НОВОСИБИРСКИЙ НАУЧНЫЙ ЦЕНТР
АССОЦИАЦИЯ "ИСТОРИЯ И КОМПЬЮТЕР"

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ

Выпуск 17

Новосибирск
2012

РЕДКОЛЛЕГИЯ

Главный редактор

академик РАЕН, д.и.н. Ю.П. Холюшкин

Заместитель главного редактора

академик РАЕН, д.и.н., профессор Л.И. Бородкин

Ответственный редактор выпуска (археология, этнография)

академик РАЕН, д.и.н. Ю.П. Холюшкин (ИАЭТ СО РАН, Новосибирск)

Ответственный редактор выпуска (история)

академик РАЕН, д.и.н., профессор Л.И. Бородкин (МГУ, Москва)

Ответственные секретари:

В.С. Костин (ИАЭТ СО РАН, Новосибирск)

к.и.н. И.М. Гарскова (МГУ, Москва)

Редколлегия:

д.ф.-м.н. Е.Е. Витяев (ИМ СО РАН), д.и.н., профессор В.Н. Владимиров (АГУ, Барнаул), доктор П. Доорн (Лейденский университет, Лейден, Нидерланды), д.т.н. О.Л. Жижимов (ИБТ СО РАН, Новосибирск), д.и.н. И.В. Журбин (Физико-технический институт УрО РАН, Ижевск), к.т.н. Ю.А. Загорулько (ИСИ СО РАН, Новосибирск), д.и.н. С.Г. Кащенко (СПбГУ, Санкт-Петербург), к.т.н. Н.А. Мазов (ИНГТ СО РАН), д.ф.-м.н., профессор А.Г. Марчук (ИСИ СО РАН, Новосибирск), д.т.н. В.В. Москвичев (ИВМ СО РАН, Красноярск), к.и.н. В.Л. Носевич (Республиканский архив электронных документов, Минск, Республика Беларусь), чл.-корр. РАЕН, д.и.н. А.Н. Садовой (Институт угля и углехимии СО РАН, Кемерово), чл.-корр. РАН, д.ф.-м.н., профессор А.М. Федотов (ИБТ СО РАН, Новосибирск), ак. РАЕН, д.и.н., профессор Ю.Л. Щапова (МГУ, Москва).

И 74 Информационные технологии в гуманитарных исследованиях: Сборник трудов. Выпуск 17.

Новосибирск: Новосибирский госуниверситет, 2012. – 120 с.

ISSN 1990-9330

Настоящий выпуск представляет материалы исследований по методологии науки, созданию информационных ресурсов, разработкам в области технологии знаний, астроархеологии. Выпуск рассчитан на математиков, археологов, историков, этнографов и на широкий круг исследователей, интересующихся информационными технологиями в гуманитарных исследованиях и образовании.

ББК 60

ISSN 1990-9330

© САТИ СО РАН, 2012

СОДЕРЖАНИЕ

I. ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ	4
Холюшкин Ю.П., Витяев Е.Е., Костин В.С. К вопросу о поисках закономерностей в археологии	4
Костин В.С. Стратегии анализа данных как инструмент математизации археологии	40
II. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ЗНАНИЙ	44
Витяев Е.Е., Москвитин А.А., Подберезный А.А. Инструментальное средство Visual Discovery извлечения информации из данных и решения задач интеллектуального анализа данных	44
Фирсов Н.И. Витяев Е.Е. Сравнение системы «Discovery» с алгоритмами Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services	51
Витяев Е.Е., Неупокоев Н.В. Математическая модель восприятия и образа	63
Демин А.В. Комментарий	73
III. БИБЛИОТЕЧНЫЕ ТЕХНОЛОГИИ	74
Мазов Н.А., Гуреев В.Н. Библиометрия: современное состояние, задачи, инструменты	74
Холюшкин Ю.П. Комментарий	79
Гуреев В.Н., Мазов Н.А. Практическое применение библиометрического анализа при формировании журнального фонда	81
Лобанова Т.П. Комментарий	86
Костин В.С., Холюшкин Ю.П. Некоторые подходы к библиометрическому анализу взаимоцитирования	87
IV. АСТРОАРХЕОЛОГИЯ	93
Ларичев В.Е. Курыканские всадники: космическая охота степных богатырей Прибайкалья на небесных лося или оленя, символов «ВСЕЯ ВСЕЛЕННОЙ»	93
Лыгденова В.В. Комментарий	101
V. ПОЛЕВАЯ АРХЕОЛОГИЯ	103
Зольников И.Д., Постнов А.В. Лямина В.А., Новикова О.И., Глушкова Н.В., Славинский В.С., Чупина Д.А., Кузьмин Я.В., Бондаренко А., Деменьтев В.Н., Селятицкая Н.А. ГИС-моделирование условий обитания древнего человека юга Западной Сибири для определения наиболее благоприятных зон его обитания	103
VI. ЭТНОЛОГИЯ И АНТРОПОЛОГИЯ	111
Кулемзин В.М. Сибиреведение и полевая этнография	111
Тучков А.Г. Комментарий	115
Холюшкин Ю.П. Концепция Г.Чайлда о трех коренных поворотах в истории	116
VII. ИНФОРМАЦИОННЫЕ РЕСУРСЫ	119
Холюшкин Ю.П., Рыбинцев Р.В. Новые ресурсы на сайте «Сибирика»	119

Холюшкин Ю.П., К вопросу о поисках закономерностей в
Витяев Е.Е., археологии¹
Костин В.С.

Статья посвящена решению практических задач археологии: истории поиска и совершенствования методов анализа первичной информации по памятникам, направленных на повышение информативности этих источников, доказуемости выводов, строгости процедуры исследования. Изложены история способов формализации источника и методы обработки эмпирических данных: суммарная характеристика и сравнительный анализ памятников, поиск взаимовстречаемости признаков, а также получение некоторых сопутствующих показателей. Рассмотрено несколько программ конкретно-целевого выделения археологических групп с помощью стратегий анализа данных.

Ключевые слова: математическая статистика, эмпирические данные, анализ данных, Data Mining, признак, тип, археологическая культура.

Введение

В наше время каждый археолог понимает, что собранные им в процессе раскопок данные так или иначе надо «анализировать» (однако лишь незначительная часть их, прибегает к помощи математических методов). Характерной особенностью российской археологии является **безматематичность** – отсутствие числовых расчетов, необходимых для получения стандартности и стандарта надежности. Так, среди отечественных археологов продолжает наблюдаться рост археологического дилетантизма, выражавшийся в злоупотреблении методами, «наиболее распространенной формой, которых было использование метода не потому, что он соответствует задаче, но потому, что они его знают и могут легко применить» [см. Moore, Keene, 1983: p. 4].

Так, Д. Кларк с иронией писал о том, что «многие современные исследования особенно в области палеолита, устраивают большую игру вокруг перехода от коллекции артефактов к коллекции процентных соотношений взятых типов артефактов... Однако вряд ли какое-нибудь из этих исследований определяет свои типы артефактов чем-либо, кроме интуитивных и спорных оснований» [Clarke, 1968: p. 188].

На некоторые ошибки, связанные с применением процентных соотношений, указывал В.А.Ранов (отсутствие в публикациях сведений о базисных цифрах, от которых производились вычисления, проведение процентных сопоставлений, взятых от различных базисных цифр) [Гинзбург, Горенштейн, Ранов, 1980: с.8]. К этому можно добавить внесение ошибок в археологические данные (погрешности арифметических подсчетов в табличных данных, недочеты измерений, неправильные дефиниции, методические просчеты, неверно поставленные задачи математического анализа). Не меньшее значение имеет заблуждение значительной части археологического сообщества в том, что организационные характеристики археологических данных информируют их непосредственно о характере археологической культуры. На самом деле простая процентная демонстрация структурных характеристик и группировок еще не дает полной и достоверной информации о характере процессов в прошлом. Ссылки на различные картины распределения, как на доказательство данной атрибуции значения явления, зачастую представляются неубедительными.

О математической археологии как сформировавшемся научном направлении можно будет говорить лишь тогда, когда произойдет определенный синтез археологии, математики, компьютерной технологии обработки и анализа информации. Тогда и появится нечто иное, специфичное, свойственное только данному научному направлению [Деревянко, Фелингер, Холюшкин, 1989].

Несмотря на то, что значение этого раздела археологии велико в силу самого статистического характера археологических гипотез, в силу значимости отношения релевантности между посылками и заключением объяснения, важность грамотного использования статистических методов до сих пор не осознается российскими археологами.

Поэтому важно попытаться на основе этих высказываний сформулировать доминирующую в

¹ Эта работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026.

отечественной археологии математическую парадигму:

Во-первых, это низкая статистическая культура исследователей. Причины этого достаточно очевидны: на исторических факультетах готовят преподавателей, а не исследователей, и поэтому выпускники этих вузов не имеют необходимой для исследователя математической подготовки. И это вполне разумно, поскольку задачей вузов является подготовка хороших учителей, преподавателей гуманитарных дисциплин в вузах, а не специалистов по археологической и исторической статистике.

Во-вторых, это отсутствие в структурах большинства исторических и археологических НИИ и вузов специализированных лабораторий информатики, призванных обеспечить исследователю квалифицированный статистический анализ наблюдений. В немногочисленных лабораториях в ряде НИИ и вузах (МГУ, АГУ, до недавнего времени и в ИАЭТ СО РАН) имелся лишь немногочисленный штат сотрудников, не позволяющий в полной мере дать основы системного подхода с грамотной формулировкой статистических гипотез научным сотрудникам, аспирантам и докторантам, проводящим свои исследования в этих НИИ и вузах. А наметившаяся тенденция к поголовному сокращению уже имеющихся подготовленных специалистов в этой области (как непрофильных в археологии) лишь ухудшает и без того низкую статистическую культуру археологов.

Третья причина – отсутствие отраслевой нормативной базы (отраслевые археологические стандарты), регламентирующей анализ данных, как завершающий этап кропотливой работы многих специалистов, который в большинстве случаев выполняется самоучками, людьми не имеющими профессиональной подготовки в этой области.

Четвертая причина – отсутствие квалифицированной статистической экспертизы в редакциях журналов, диссертационных и экспертных советах ВАК, что говорит об отсутствии в них специалистов владеющих данными технологиями.

В свое время в СССР было написано учебное пособие Г.А. Федорова – Давыдова [Федоров – Давыдов, 1987], в котором было подчеркнуто, что применение методов математической статистики целесообразно, а иногда и необходимо для решения следующих задач [Федоров – Давыдов, 1987: с.6]:

1. Установление средних размеров археологических объектов как стандартов, которым следовали древние мастера. Установление степени «вариабельности» этих размеров, т.е. их колебаний вокруг средних, определение случайных и закономерных отклонений от этих стандартов. Такие задачи возникают при углубленном исследовании объектов и при сравнении их между собой.

2. Исследование частоты встречаемости объектов и признаков, тех или иных явлений, оценка взаимосвязи между ними. Эти задачи возможны при выявлении типичных видов вещей, в которых признаки закономерно и тесно связаны между собой, при построении хронологических систем, когда нужно установить, какие виды вещей закономерно встречаются вместе при выявлении локальных вариантов или культур, когда требуется установить, что виды вещей и объекты закономерно встречаются на одной территории.

3. Исследование сходства между объектами или их локальными вариантами, полностью не совпадающими друг с другом, но имеющими набор совпадающих и несовпадающих признаков. Эти задачи встают перед исследователем при выделении культур и их локальных вариантов, при построении эволюционных рядов, при разбиении множества каких-либо объектов на группы, т.е. при классификации.

4. Исследование структуры взаимосвязанных признаков археологических объектов с целью выявления более или менее существенных признаков, которое необходимо при построении археологических классификаций при более углубленном исследовании сходства объектов между собой.

Представленные в учебнике методы не учитывали того, что в типичных объектах археологической науки констатируется наличие в большей мере материальной и структурной компонент, чем энергетической. «Археологические объекты являются материально-структурными, а не динамическими системами, в соответствии с сущностной природой которых формируются и методы их исследования» [Деревянко, Фелингер, Холюшкин, 1989: с.74]. Так, Л.Бинфорд, цитируя Пиггота, писал: «Мы должны признать, что в археологии не существует иных фактов, чем...«наблюдаемых данных»... то, что, как доисторики, имеем в нашем распоряжении - это случайно сохранившиеся пережитки материальной культуры, которые мы интерпретируем так, как можем, и неизбежно специфика этого источника определяет тип информации, который мы можем извлечь из него» [Binford, 1972: p.7]. Поэтому задачей археологической науки является реконструкция археологических объектов как определенной системы, множества объединенных в единый целостный комплекс элементов, связанных пространством, временем и контекстом. В этом смысле в понятии «структурная археология» используется полный аналог класса (в терминах

и понятиях OOD) как некоторой структуры, фиксирующей для каждого объекта единство признаков и свойств. Подход структурной археологии к исследованию археологических данных предполагает тяготение к герметическому (закрытому) их анализу, при котором археологические объекты анализируются как самостоятельное целое, а их связи с другими не рассматриваются. Структурализм берет в основу синхронный подход, изучение явлений в их статике, в горизонтальном срезе, в количественном выражении. Дальние истоки структурализма, подхода к его методологии можно встретить уже у пифагорейцев (понимание структуры как целостной, саморегулирующейся системы устойчивых внутренних отношений, не сводимой к сумме составных элементов). Как мы уже подчеркивали ранее, именно целостность отличает структуру от простой совокупности элементов. Поэтому в структурной археологии предполагается собирание и анализ разносторонних фактов и составление их полного перечня; установление взаимосвязей между археологическими фактами, их группирование и выявление внутренней корреляции; построение системы из наличных элементов, создающих единый целостный объект исследования. Основными принципами при этом являются:

а) изучение не отдельных элементов, а выявление системы их взаимосвязей, рассмотрение целостности, состоящей из элементов;

б) не диахронный, а синхронный подход, предполагающий исследование структуры археологических объектов, их функционирования;

в) формализация анализа, абстрагирование от конкретности взаимоотношений и выражение их в формулах, моделях, схемах, позволяющих наглядно сопоставлять явления, которые в реальности своего существования несопоставимы.

В отличие от структурной археологии для процессуальной археологии важен вертикальный разрез явлений, их развитие, их качество. Принципы процессуальной археологии предполагают рассмотрение явлений в их развитии, рассмотрение связей данного явления с другими, и, наконец, исторических высших форм в качестве ключа для понимания предшествующих.

Процессуальная археология как научное направление было создано в американской археологии на рубеже 60-70 годов в связи с движением «новых археологов», лидером которых был Л.Бинфорд [For theory..., 1977]. В применении к археологии это означает, что в ней следует выдвигать гипотезы о поведении людей, устройстве социальных структур и т.п., дедуцировать из них наблюдаемые следствия, касающиеся предполагаемых материальных остатков, соответствующих культур и затем проверять эти следствия на основе собранного при раскопках эмпирического материала.

Эти положения основаны на уверенности ряда «новых археологов» или «процессуалистов» в том, что в археологии должны быть законы, связывающие различные состояния материальных культур и человеческое поведение [Salmon, 1982: р. 33]. В археологических источниках соответственно должны быть отражены процессы адаптации системы древних человеческих популяций к среде и проявлены законы эволюции. Продвижение археологии в этом направлении требует иных подходов, связанных с созданием различных поведенческих моделей.

Другим спорным положением учебника является представление о том, что не следует рассматривать математическую статистику как средство построения некой математической археологии. Археология остается по мнению Г.А. Федорова-Давыдова специфической наукой, а математическая статистика является только инструментом в её руках для выявления скрытых тенденций и оценки степени достоверности тех или иных выводов [Федоров –Давыдов, 1987: с.6-7]. Разумно полагать, что такое представление сохранится до тех пор, пока не будет создан и признан научной археологической общественностью современный учебник и специализированный пакет интеллектуального анализа данных в археологии. Подобный учебник должен раскрывать соответствующие стратегии, включающие приемы и методы как нечто специфичное именно для археологии. Такого учебника пока нет не только у нас в стране, но и за рубежом (на Западе имеется огромное число книг, в которых так или иначе фигурирует словосочетание «анализ данных»; но в этих книгах, по нашему мнению, не достаточно полно и глубоко рассматривается проблема «стыковки» рассматриваемых математических методов именно с археологией).

За исключением немногих учебных заведений, готовящих археологов, предусматривается преподавание предмета, название которого фигурирует в заголовке раздела. Но, на наш взгляд, далеко не всегда совокупность действий, называемая анализом археологических данных, понимается правильно. В первую очередь, мы имеем в виду то, что эта совокупность действий не всегда трактуется как некоторый специфичный процесс, не сводящийся ни к какому набору математических приемов и органично вписывающийся в содержательную ткань археологического

исследования. Непонимание же сути указанного процесса, по нашему мнению, приводит к неэффективному использованию математического аппарата, и, более того, к получению выводов, противоречащих реальности. Неадекватное отношение к процессу анализа данных не является случайным, т.к. большинство археологов сомневаются в существовании дисциплины, именуемой «интеллектуальный анализ данных». Свидетельством этого можно считать, например, то, что упомянутый термин в археологической литературе понимается по-разному. Этот факт тоже существен для археологической практики: чтобы получать корректные выводы, мы должны четко понимать, когда, в каких именно условиях и с какой целью можно использовать анализ данных, а это невозможно без ясного представления о том, что это такое.

Однако, в последние годы стало ясно, что археологи смогут использовать весь потенциал количественных методов и компьютерных технологий, если им будет известны конкретные трудности, ограничения и возможности, присущие археологическим данным и исследованиям процессов в древности на основании сохранившихся археологических остатков. Математическая археология является новой дисциплиной, которая пытается раскрыть, количественно представлять и исследовать специфические свойства и закономерности археологической информации. Фундаментальным исследованиям данных и методов для самостоятельного археологического подхода к обработке информации могут дать количественные методы и компьютерное программное обеспечение, специально направленные для решения археологических проблем и понимание специфики археологических источников.

Математическая археология включает в себя большую часть методов и теорий, разработанных в количественной археологии с 1960 года, исследует способы представления общей археологической информации, проблему структуры данных, люфтов в археологической информации. Количественная составляющая археологических данных, представленных в археологических источниках открывает широкий спектр компьютерной обработки информации, а методы анализа данных подходят для решения задач любой сложности. Это также способствует пониманию формализованных исследований объектов археологии и создает определенную связь между археологией и другими количественными дисциплинами

На повестке дня стоят две основные темы исследований, которые дополняют друг друга:

Фундаментальные исследования (теоретическая составляющая науки) направлены на исследование структуры археологической информации, создание онтологии археологической науки, изучение свойств и знания возможностей археологических данных. Задачами при этом являются выявление нечеткости и неопределенности в археологических данных, определение оптимальной стратегии отбора информации, путей решения возникающих задач и выявления пространственно-временные эффектов (миграции, наследование, заимствование и др.).

Развитие компьютерных алгоритмов и программного обеспечения делают это теоретическое знание доступным для пользователя.

1. Эмпирическая и теоретическая основа для изучения археологических данных

Роль эмпирических данных в изучении явлений, отраженных в археологическом материале, огромна. Достаточно глубокое изучение интересующих археолога закономерностей невозможно без опоры на анализ конкретных фактов, в которых эти закономерности, собственно говоря, и проявляются. «Питательной» средой для теоретических построений чаще всего является эмпирический материал. Именно реальные эмпирические факты, как правило, служат средством проверки теорий, наводят на мысль о необходимости их корректировки, служат почвой для формирования новых теоретических гипотез.

Что же такое археологические эмпирические данные, т.е. данные, характеризующие конкретные археологические факты; данные, в виде которых, собственно говоря, эти факты перед нами и выступают? Данные могут представлять перед исследователем в виде:

- совокупности чисел, характеризующих те или иные объекты (в качестве таких совокупностей могут выступать, например, метрические характеристики артефактов.
- множества индикаторов определенных отношений между рассматриваемыми объектами.
- результатов попарных сравнений каких-либо объектов (такие данные используются в методе парных сравнений - способе построения шкал, отражающих усредненное отношение изучаемой совокупности археологических артефактов к каким-либо объектам.
- совокупность значений каких-либо признаков (характеристик, переменных, величин; будем считать эти термины синонимами), измеренных для каждого из изучаемых объектов.

Мы не будем глубоко анализировать смысл термина «признак», хотя здесь есть о чем поговорить (на наш взгляд, это понятие требует специального обсуждения; здесь мы такой цели перед собой не ставили). Будем считать этот смысл в основном интуитивно ясным. Отметим лишь

некоторые моменты. Признак – это некоторое общее для всех объектов качество, конкретные проявления которого (значения признака; их называют также альтернативами, градациями), вообще говоря, могут меняться от объекта к объекту. Для нас важно, что само введение практически любого признака является моделированием довольно высокого уровня. Признаки не существуют сами по себе, они – плод наших абстрактных рассуждений, идеальные конструкции.

В исторических науках соответствующий процесс абстрагирования является иногда очень непростым. Основными его этапами является выделение понятий (процесс рождения которых уже не прост) и осуществление их т.н. операционализации. Процессу операционализации понятий посвящена обширная литература [Клейн, 1991; Холюшкин, 2010] .

Так, в свое время Р. Даннел сказал, что палеоистория имеет обыкновение специально изобретать для себя термин и потом спорить двадцать лет о том, что он значит, вместо того, чтобы определить этот термин заранее [Dunnell, 1971: р. 4.; Клейн, 1991: с.125]. Такое заявление свидетельствует лишь о том, что археологи не всегда отчетливо осознают тот факт, что «всякое понятие, которым в данной предметной области выражают некоторый объект, свойство, явление или процесс, не является «элементарным кирпичиком мира». Всякое понятие выражает некоторую структуру и системно организованный набор других понятий. Именно через этот системно организованный набор понятий мы осознаем семантику исходного понятия, объясняем, интерпретируем и используем это понятие в некоторых границах, которые также определяются через системно организованный набор понятий» [Деревянко, Фелингер, Холюшкин, 1989: с.12].

В приведенном ниже классификационном фрагменте раскрывается и расширяется понятие классификационной археологии (рис. 1).

Диадная группа классификационной археологии построена по принципу существования регулярности в археологическом материале. Термины были введены Д. Кларком в его книге «Аналитическая археология», согласно которой в монотетических классификациях признаки археологических объектов жестко связаны в сочетания и эти признаки резко разделены. В идеале, как полагает Д.Кларк, все признаки данного набора в каждом данном классе присутствуют.

Согласно Л.С. Клейну: «Монотетизм, Монотетичность, Монотетический – такой характер организованности, при котором каждой группе объектов присуще замкнутое распределение характеристик: разнородные характеристики в материале очень четко сгруппированы, жестко связаны в сочетания, а эти сочетания резко разделены. В одной группе – одни опознавательные характеристики, в другой – другие; могут найтись и общие для обеих групп, отличающие их от третьей. Разграничение таких групп не встречает затруднений. Есть две разновидности монотетической организованности – со сплошным распределением характеристик и с ажурным их распределением» [Клейн, 1991а].

В политетической классификации группировка рассеяна и в каждом объекте присутствует лишь часть набора признаков и ни один признак не представлен у всех объектов. При подобной классификации наборы признаков у объектов различны [Clarke, 1968: р. 35-37].

Общая классификационная археология				
Монотетическая классификация		Политетическая классификация		
Дескриптивная классификационная археология	Сравнительная классификационная археология	Функциональная классификационная археология		
Типологическая археология				
Аналитическая классификационная археология	Партитивная классификационная археология	Морфологическая классификационная археология	Предметная классификационная археология	Генеральная классификационная археология

Рис. 1. Классификационный фрагмент «Классификационная археология» [Деревянко и др., , 1990].

Согласно Л.С.Клейну: «Политетизм, Политетичность, Политетический – такой характер организованности, при котором нет жесткой сопряженности характеристик, привязанности их к одному набору характеристик и к одной группе объектов. Характеристики не только выпадают из набора у членов данной групп, но и появляются, пусть и менее густо, в иных сочетаниях, вне данного набора и данной группы, в неорганизованном периферийном материале или даже у членов смежных групп. Хотя бы в некоторых, а то и у многих объектов представлена лишь та или иная часть некоторого (характерного для данной группы) набора характеристик, и не каждая характеристика представлена у всех объектов. По определению, политетическая организованность не может иметь сплошного распределения характеристик; ей присуще только ажурное распределение» [Клейн, 1991а].

Монотетическая организованность могла бы быть рассмотрена не только как отрицание, но и

как предельный случай политетической. В противоположном пределе должно оказаться, что нет ни одного объекта, у которого были бы представлены только характеристики из одного набора, и нет ни одной характеристики, встречающейся с другими только в одном и том же сочетании или только с характеристиками этого набора [Клейн, 1991a].

Описательная (дескриптивная) классификационная археология отражает проблематику обоснования выбора средств представления археологических объектов, в терминах данного описательного языка (естественного, документального или информационного) [Гарден, 1983: с.52]. Как правило, подобные описательные конструкции, будучи сравнительно простыми, представляют собой последовательности высказываний, относящихся к характеристикам артефактов, и соображений о том, как эти типы могут быть использованы для объяснения археологических данных. Примерами дескриптивной типологии являются списки Борда.

Сравнительная классификационная археология связана с поисками критериев кросс-культурных сравнений с помощью сети конфигураций поведения типов и контекста их нахождения, а также с разработкой мер сходства и различия между различными типами и системами классификаций [Гарден, 1983; Клейн, 1991a].

Функциональная археологическая классификация представляет собой направление, приверженцы которой занимаются разработкой методов по выяснению функциональных назначений древних вещей.

В качестве альтернативно-тождественного понятия в классификационном фрагменте выступает типологическая классификация, содержанием которой является выявление глубинной структуры археологического материала, определение места объектов в системе, их значимости, соотношения типов с культурой.

Пентадная классификационная группа основана на представлениях Д.Кларка [Clarke, 1968], хорошо описанных Л.С.Клейном [Клейн, 1991: с.545], и характеризуется поэтажной, многоуровневой организацией.

Аналитическая классификация, размещённая в пентадной группе, связана с методами анализа на уровне признаков (атрибутов). Согласно Л.С. Клейну: «Группирование аналитическое (анализное), дескриптивный анализ – взаимонезависимые разбиения одного и того же материала (одной и той же совокупности объектов) на группы по разным параметрам – по каждому параметру отдельно, соответственно особой шкале упорядочения характеристик (относящихся к данному параметру). Одна и та же совокупность объектов группируется по каждому параметру особо: по размеру, по цвету, по весу и т.п. Каждая группа получается одномерной, но ведь объекты не одномерны, так что каждый объект входит в разные группы (разных параметров) своими разными характеристиками. Идентификация объекта с определенной позицией во множестве аналитических ячеек требует мысленной увязки затронутых ячеек, принятых за взаимонезависимые, – как бы локализация объекта в n-мерном пространстве координат, где каждая ячейка – отдельная точка привязки. Поэтому такие операции с объектами иногда неудачно называются «координатной группировкой» объектов (название вводит в заблуждение, намекая на общеизвестную Декартову систему координат, которая на самом деле не имеется в виду). Кроме этого наглядного соприкосновения здесь нет иной связи между параметрами, а это соприкосновение показывает только взаимоисключаемость параметров и ничего сверх того [Клейн, 1991: с.545].

В связи с этим рассмотрим классификационный фрагмент «Аналитическая классификационная археология» (Рис.2).

Все артефакты обладают признаками (атрибутами) (рис. 2).

Под признаком (attribute), приведенным в качестве опорного понятия понимается «элементарная характеристика (черта), имеющая культурную значимость (смыслоразличительная). Это любое из таких свойств, различия которых были запрограммированы (сознательно или неосознанно) создателями артефактов и следов (древними мастерами, участниками событий и т.п.) и которые вместе с другими подобными свойствами формировали конструкцию или стиль предмета» [Клейн, 1991: с.377]. Д. Кларк также определял признак, как логически не поддающуюся упрощению особенность двух или более состояний [Clarke, 1968].

Археологи часто небрежно относятся к использованию признаков и их состояний, они иногда не могут провести границу между разными атрибутами и переменными состояниями одного атрибута. Археологов интересуют те признаки, которые являются продуктом человеческой деятельности или же отображены с участием человека. При этом Д.Кларк, И. Рауз рассматривают

атрибуты как основные и фундаментальные единицы информации. Другие археологи сомневаются в существовании таких единиц. Например, Д. Хилл и Р.Эванс утверждают: «Неважно, насколько малы и ничтожны атрибуты, они все равно поддаются дальнейшему членению, т.е. каждый атрибут обладает своими атрибутами» [Hill, Evans, 1972]. Отметим, что, на наш взгляд, признаки надо понимать несколько шире, чем это обычно делается. Так, в него имеет смысл включить, например, различные способы шкалирования (скажем, получение на основе непосредственного анализа значений некоторых вспомогательных признаков и последующий переход к другим, латентным переменным с помощью построения индексов. На практике проблему операционализации чаще всего разделяют на две: выбор признаков, являющихся индикаторами понятий, и выбор набора значений каждого признака. При концептуализации понятий должны решаться вопросы, отнюдь не лежащие на поверхности. Напротив, успешная операционализация предусматривает переход на достаточно глубокий концептуальный уровень рассмотрения предмета исследования, при котором признаки воспринимаются как отражение параметров анализа, релевантных целям исследования, а значения признаков - как результат расчленения каждого параметра на определенные категории, ключевые понятия исследования.

При работе с атрибутами необходимы точность измерений, аккуратность и объективность. Точность означает, что степень измерения является высокой. Аккуратность указывает на то, что измерение верно в рамках указанной точности. Использование компьютера хотя и придает археологии большую точность и аккуратность при работе с археологическими данными, но не обязательно делает археологию объективной. Выбор того, какие атрибуты подлежат фиксации, способа их измерения и методов их обработки остается субъективным решением. Так В.А.Ранов писал о трудностях установления критериев для измерения треугольных сколов. Однако любая обработка, произведенная на компьютере, может быть повторена, и, если атрибуты ясно определены, другие археологи могут произвести те же измерения, осуществить обработку данных и оценить результаты.

Признак (с.1)					
Качественный признак			Количественный признак (показатель)		
Описательный признак		Оцифрованный признак		Нумерованный признак	
Качественный описательный признак	Количественный описательный признак (показатель)	Качественный оцифрованный признак	Количественный оцифрованный признак (показатель)	Качественный нумерованный признак	Количественный нумерованный признак (показатель)
Градации признаков					
Дихотомический признак	Неранжированный признак	Ранжированный признак	Счетно-дискретный признак	Непрерывный признак	

Рис. 2. Классификационный фрагмент «Аналитическая классификационная археология»

Математические приемы обработки требуют определенных форм представления исходных данных, поскольку адекватное применение того или иного метода обработки данных существенным образом зависит от того, в какой «шкале» измерены значения признаков. Согласно предлагаемой классификации признаки могут быть качественными и количественными.

Качественные признаки выражают некоторые сущности (объекты, явления, эталоны), которые не могут быть измерены. Например, виды артефактов, наименования памятников и т.п.

Количественная характеристика (показатель), представляет собой вид научных сведений об объекте, который может быть выражен числом или множеством. Это самая сильная шкала – количественная. Здесь появляется дополнительная информация: насколько велика разность сравниваемых значений. При ее использовании мы от чисто логической информации переходим к информации количественной.

Признак (качественная характеристика) и показатель (количественная характеристика) – это пара диалектически взаимосвязанных категорий. Они неразрывно связаны друг с другом, резкой грани между ними нет.

В качестве первого понятия триады приведены описательные признаки. Описательный признак – это качественный признак, представленный словесно или в виде рисунка. Общим видам описания присущи схематизм и условность. Традиционное описание скорее напоминает результат научной работы, чем беспристрастную констатацию наблюдений [Щапова, 1988: с. 21].

Оцифрованные признаки – это признаки, упорядоченные относительно некоторого качественного эталона, для описания которых были применены цифровые ярлычки.

Нумерованные признаки – это признаки, значения которых всегда получено путем измерения или пересчета.

В пентадной группе в качестве первого понятия приведены дихотомические признаки, которые указывают лишь на присутствие или отсутствие признака (да, нет). Единственная информация, которую можно извлечь из сравнения двух значений в такой номинальной шкале – совпадают они или нет.

Неранжированные признаки выражают какое-либо свойство, качество или состояние артефакта и предполагает наличие списка его значений. Например, список цветов спектра, атрибуты в тип-листах.

Ранжированные (порядковые) признаки – это признаки, которые могут быть выстроены по порядку в соответствии с определенным критерием. Это более сильная (информативная) шкала – порядковая. Здесь кроме совпадения, можно определить, какое из значений больше, а какое меньше.

Счетно-дискретные признаки предполагают возможность измерения с помощью специального анализа. Так, в примере с цветами, красный цвет = 1 квант, оранжевый = 2 кванта и т.д. Т.е. в этом случае единицей измерения является красный цвет. Таким образом, в случае со счетно-дискретными признаками, отдельные значения отличаются на какую-то величину, кратную единице измерения. Их поведение аналогично целым числам в арифметике.

Пятым понятием пентады служат непрерывные (мерные, континуальные) признаки, значения которых могут отличаться на сколь-угодно малую величину. Их поведение аналогично вещественным числам в арифметике. Измерение физических размеров артефактов относятся к этому типу признаков.

Подводя итоги рассмотрения понятия «признак» следует отметить, что уровень сложности исследования атрибутов зависит от рамок исследования. Если это погребальный ритуал, то присутствие/отсутствие артефактов может быть фундаментальным признаком. Если исследование посвящено скреблам, то оно может быть описано посредством десятка атрибутов. Другими словами то, что в одном исследовании считается атрибутом, признаком, в другом исследовании может занимать позицию артефакта. Поэтому при исследовании признаков очень важно определить уровень их анализа [Chenchell, 1968].

Качественные признаки могут быть представлены в следующей описательной схеме (рис. 3). Согласно данной классификации качественные признаки подразделяются на внешние и внутренние [Гарден, 1983]. Под внутренними признаками понимаются непосредственные свойства артефактов (физика, геометрия и семантика). Внешние признаки – это данные, непосредственно в вещах не представленные, но получаемые из изучения «контекста» и сравнительных материалов: «место», «время» и «функция».

Качественный признак (с. 2)					
Внутренний признак			Внешний признак		
Общий признак		Сходный признак		Отличительный признак	
Внутренний общий признак	Внешний общий признак	Внутренний сходный признак	Внешний сходный признак	Внутренний отличительный признак	Внешний отличительный признак
Качество признака					
Главный признак	Существенный признак	Базовый признак	Второстепенный признак	Случайный признак	

Рис. 3. Классификационный фрагмент «Качественный признак»

Тройка понятий указывает на черты сходства и различия сравниваемых артефактов и взятые вместе отражают сущность артефактов.

Качество признаков отражено в пятерке понятий основанных на ранжированной шкале противопоставлений от главных до случайных признаков (Рис. 3).

Классификация следующего уровня (партитивная) предназначена для разработки методики анализа на уровне деталей. Этажом выше морфологическая классификация представляет уровень классификации специфических сторон классифицируемых предметов, таких как морфология, технология, мотивы орнамента, композиция и т.д.

Другим базовым понятием археологии является понятие «Предметная классификационная археология» (рис. 4).

Однако «сложность подхода к постановке задачи предметной классификации, состоит в том, что археология до сих пор не располагает не только общей теорией типологического метода, но и не имеет единого представления о том, что такое тип» [Шер, 1970: с.8]. Имеется достаточно большой разноречивой относительно понимания самого термина [Клейн, 1991]. Мало того, как

отмечали Доран и Ходсон, многие археологи уклоняются от необходимости определять, что они понимают под типом [Doran, Hodson, 1975: p. 158].

В приводимом классификационном фрагменте типы, в соответствии с классификацией Тэйлора делятся на эмпирические и культурные типы. Под первыми понимаются артефакты, складывающиеся из элементарных свойств, присущих материалу и доступных наблюдению и регистрации ... «Выделение эмпирических типов – это вспомогательная, поисковая часть классификации, она относится к непосредственно наблюдаемому миру исследователя и ничего не говорит о культуре» [цит. по Клейн, 1991: с.74].

«Культурные типы это те группировки предметов по сходству, которые были введены в материал самими творцами, создателями культуры, когда они упорядочивали и подчиняли своими нормам материал» [Клейн, 1991: с.74].

В качестве первого понятия триады приведены условные типы, которые в трактовке Форда, Мальмера и Врю являются чистым инструментом исследователя. Такие типы имеют производные рамки, налагаемые на материал для порционного описания его характеристик [Клейн, 1991: с. 48].

Вторым понятием являются так называемые «наблюдаемые» типы, близкие к реальным. И, наконец, естественные типы. Наличие последних отрицается большинством археологов. «Тип, по их мнению, возникает в тот момент, когда для него формулируют точную дефиницию. Без дефиниции нет типа» [Клейн, 1991: с.81].

Тип (с. 1)					
Эмпирический тип			Культурный тип		
Условный тип		Наблюдаемый тип		Естественный тип	
Эмпирический условный тип	Культурный условный тип	Эмпирический наблюдаемый тип	Культурный наблюдаемый тип	Эмпирический естественный тип	Культурный естественный тип
Виды типологических градаций					
Производный тип		Построенный тип		Опознаваемый тип	Функциональный тип
					Исторически-указательный тип

Рис. 4. Классификационный фрагмент «Предметная классификационная археология»

Приводимая градация типологических построений в пентадной группе позволяет несколько уточнить понимание этого вопроса. Первые две позиции занимают произвольные, придуманные или изобретенные типы, которые имеют соответствия с условными, эмпирическими типами. Опознаваемые типы занимают промежуточное положение между эмпирическими и культурными типами. Функциональные типы понимаются в широком культурном смысле по их роли в механизме культуры. Наряду с этим исторические, по И. Раузу, типы имеют также хронологическое и территориально опознавательное значение, служить чем-то вроде «руководящих ископаемых» [Клейн, 1991: с.80].

И, наконец, в качестве пятого пентадного понятия приведена классификация, основанная на учете всех признаков, в том числе и внешних связей (генеральная классификация, типы Кригера) (Рис. 4) [Клейн, 1991: с.55].

Подводя итоги, отметим, что парадигма теоретической археологии только формируется. Несомненно, ее становление и последующее развитие может стать реальным только в недрах информатики.

Приведем новый классификационный фрагмент «Типологическая археология» (рис. 5). Здесь в качестве опорного понятия приведена «Общая типологическая археология», задачей которой является обобщение опыта построения типологий в археологии.

Общая типологическая археология				
Формальная типология		Содержательная типология		
Определение		Анализ		Интерпретация
Этапы типологизации				
Знакомство с материалом	Описание конечным набором признаков	Выбор основания типологизации	Распределение объектов в группы	Использование типологии для получения нового знания

Рис. 5. Классификационный фрагмент «Типологическая археология».

В диадной группе в качестве первого диадного понятия представлена «формальная типология», на основе которой археологи стремятся выявить узловые пункты изменчивости материала, группируя его по сходству параметров вокруг этих пунктов, не обязательно весь, и для нее важны границы между типами» [Клейн, 1987]. По своей природе формальные типологии не

стремятся выявить в материале существующие в нем разграничения, а чаще всего навязывают ему несвойственные ему деления.

Здесь же в качестве второго диадного понятия представлена «содержательная типология», в рамках которой типолог «ставит задачу открыть и охарактеризовать существовавшие в реальной исторической действительности группы» [Грязнов 1969: с.19]. Задачей этого раздела является создание типологий, которые открывают реально существующие в материале типы, т.е. культурные типы.

В качестве первого триадного понятия взято «определение», задачей этой аналитической стадии типологического анализа является разбиение почти безграничного разнообразия признаков на легко различимые группы и попытки выработать отличительные признаки этих групп.

Второе триадное понятие «анализ» отражает этап выявления порядка в соотношении признаков. На этом этапе исследователь должен создать типологию.

Третье триадное понятие «интерпретация». На этом этапе типолог пытается выяснить, как и откуда произошли типы, каковы их «взаимоотношения» и выявить причины их изменений.

Альтернативно-тождественным понятием является понятие «этапы типологизации». Сама процедура создания типологических классификации как определенная последовательность операций, обязательных при любом подходе к классификации, представлена в пентадной группе. Она отражена достаточно ясно в работах разных авторов [например, Clarke 1968; Каменецкий, Маршак, Шер 1975 и др.] и выглядит следующим образом [Колпаков, 1991: с.72]:

1. *Знакомство с материалом, включающее не только его осмотр, но и знакомство с соответствующей литературой, работу с музейными коллекциями, близкими к изучаемому материалу, и т.п. Таким образом, уже самый первый этап классификационной процедуры – выявление свойств и формирование признаков – достаточно сложен и сам включает в себя множество «пробных» классификаций. Результатом первого этапа процедуры является набор признаков и «описания» исследуемых объектов в рамках этого набора.*

2. *Следующий шаг – отбор признаков и их сочетаний, взвешивание признаков.*

3. *Основание типологии, является центральным для традиционной типологии, ибо все существенные характеристики связываются в типологии с особенностями ее основания. Поэтому хорошей типологией является та, которая объединяет в один тип артефакты, максимально сходные друг с другом в существенных признаках, является устойчивой и в то же время достаточно гибкой, чтобы сохраняться в условиях появления все новых и новых объектов исследования.*

4. *Собственно типологизирование, т.е. распределение исследуемых объектов в группы по основанию классификации. В каждой конкретной классификации эти этапы наполняются конкретным содержанием. Например, агломеративная и распределительная стратегии Дорана и Ходсона.*

5. *Использование построенной классификации с целью получения нового знания.*

Следующим уровнем стратегии исследования является изучение поселенческих комплексов. Понятие «Поселенческая археология» как научное направление возникло в 50-60 гг. XX столетия, на основе достижений западной экономической географии. Археология поселений – это изучение изменяющихся структур поселений как части анализа адаптивных взаимодействий между людьми и их внешней средой, как естественной, так и культурной [Chang, 1968]. Структуры поселений, схема распределения жилищ в определенном месте являются результатом взаимоотношения людей, которые решили на основании практических, культурных, социальных, политических и экономических соображений располагать свои дома, хозяйственные постройки и религиозные и погребальные сооружения определенным образом. Таким образом, археология поселений предлагает археологам шанс рассматривать взаимоотношения не только внутри различных коммун, но также в рамках торговых сетей, использования природных ресурсов, социальной организации и культурных тенденций.

Конечной целью является изучение древних систем поселений в аспекте полной картины древнего общества. В связи с этим признаки, находящиеся вне памятников, (off-site features), такие как системы полей, погребальные комплексы играют важную роль в понимании всего диапазона деятельности человека в прошлом [Фаган, ДеКорс, 2007].

В системном виде процесс исследования способов расселения может быть отражен в следующем виде (Рис. 6).

В классификационном фрагменте в качестве опорного понятия приведен «способ человеческого расселения», с помощью которого человек размещает себя на окружающей местности. Это относится к жилищам, их местонахождению и к характеру и расположению других сооружений, связанных с жизнью общины...» [Whitley, 1953: p. 1]. Изучение способов расселения

должно, по мнению сторонников этого направления, сосредоточиться не только на структурном (синхронном), но и на процессуальном (диахронном) аспектах социальных и прочих связей [Trigger, 1967: р. 151]. Данный аспект исследования отражен в диадной группе. Модель заселения (колонизация) есть распределение памятников и заселение человеком местности. Модели заселения определяются многими факторами, включая окружающую среду, виды ведения хозяйства, технические навыки. Археология поселений является частью анализа взаимодействия между людьми и их средой обитания.

Способ человеческого расселения					
Структурные аспекты связей			Процессуальные аспекты связей		
Микроуровень		Уровень памятника		Макроуровень	
Самус	Памус	Сасуп	Пасуп	Самаус	Памаус
Факторы расселения					
Среда	Ресурсы	Технология	Размещение	Взаимодействие	

Рис. 6. Классификационный фрагмент «Поселенческая археология» [Холюшкин, 1994: с.102].

Самус – структурные аспекты микроуровневых связей; Памус – процессуальные аспекты микроуровневых связей; Сасуп – структурные аспекты связей на уровне памятника; Пасуп – процессуальные аспекты связей на уровне памятника; Самаус – структурные аспекты макроуровневых связей; Памаус – процессуальные аспекты макроуровневых связей.

При определении пространственных отношений археологи основывают свое изучение деятельности человеческого общества в целом на моделях и гипотезах, проверяемых с помощью данных, полученных из разных дисциплин. Эти данные имеют отношение к тому, каким образом доисторические общности и остатки их материальной культуры сгруппированы в более крупные единицы в рамках определенного ландшафта. Они также касаются того, как доисторические общности взаимодействовали с постоянно меняющейся средой обитания.

Пространственный контекст является жизненно важным для научной археологии, так как он обеспечивает одно из важнейших измерений археологических данных

Реализуются эти подходы с помощью трех структурных уровней анализа:

- 1) *микроуровень включает в себя анализ отдельных комплексов, выявление их функциональных назначений, структуры и времени бытования;*
- 2) *уровень памятника, включает в себя исследование всего археологического объекта, как единого и сложного социального организма;*
- 3) *макроуровень включает в себя региональные исследования синхронных и асинхронных памятников. Реализация синхронных и асинхронных связей возможна на каждом из приведенных уровней и эти возможности отражены в диадно-триадной классификации.*

Поскольку археология поселений – это изучение изменяющихся адаптивных взаимодействий между человеком и его природным окружением, то необходимо исследовать и факторы такого расселения. Среди этих факторов одним из важных считаются условия окружающей человека среды, которые определяются нами как изменяющийся во времени и пространстве абиотический фактор среды обитания, на который человек реагирует по-разному в зависимости от его силы. В ряде работ, использующих понятия теории систем, подчеркивается обусловленность размещения поселений и направленность экономики охотничье-собираТЕЛЬСКИХ обществ в первую очередь экологическими факторами [Jochim, 1976: р. 76] выступают ресурсы. К ним относятся абиотические и биотические ресурсы. В той же работе М.Джохима обусловленность расположения поселений определяется близостью к «пулу» пищевых ресурсов, определяемых через расстояние до «центра» ресурсов (воды, сырья для изготовления орудий, топлива, пищи).

В качестве третьего пентадного понятия приводится «технология», рассматриваемая как механизм контроля для поддержки стабильности той или иной археологической общности.

Четвертое понятие «размещение» предусматривает исследование факторов и пространственного размещения индивидов человеческих и биологических популяций и сообществ. Исследованием проблем размещения занимается ландшафтная археология.

Представители этого направления исследований поселений обращаются к ландшафтной географии как средству изучения реальных древних ландшафтов, где символическое отношение к окружающей среде, а также к экологии играет важную роль. При этом в западной археологии используются модели гравитации, теория центральных мест, используются ГИС и спутниковые данные. Согласно Б.Фагану и К.ДеКорсу «Многие археологи, занимающиеся исследованием ландшафтов, мыслят об организации ландшафта с позиции трех измерений» [Фаган, ДеКорс, 2007: с.440]:

1. *Физические характеристики и свойства.*
2. *Исторические трансформации во времени.*
3. *Физические и символические отношения людей с окружающей их средой.*

При таком подходе люди организуют свое отношение с социальным миром, как с потенциально жизнеспособным источником информации об идеологии и культурных неосознанных элементах (cultural intangibles). Большая часть такого исследования пополняется информацией из этнографических и исторических материалов. Группа археологов, проводившая крупномасштабное исследование долины Лауер Верде в Аризоне, получила задание изучить структуру использования земель, изменявшуюся в течение длительного времени. Для того чтобы выполнить ее, они зафиксировали современные и исторические ландшафты как европейцев, так и коренных американцев и затем уже работали оттуда в сторону далекого прошлого с теоретической структурой, основанной на теории ландшафта [Фаган, ДеКорс, 2007: с.440].

Такое исследование до сих пор остается новаторской в археологии, но оно делает успехи по мере того, как археологи все более понимают отношения между коренными людьми и их землей. Роберт Макферсон выразительно описал подобное отношение у навахо: «Земля — это не только системы поразительно устойчивых топографических особенностей, которые удивляют людей или подталкивают к ее обработке, она является живым, дышащим организмом в этой неживой Вселенной. Земля с ее водами, растениями и животными является духовным творением, которое привели в движение боги в своей мудрости. Эти составляющие находятся здесь для того, чтобы помогать, учить и защищать нас посредством систем верований, которые объясняют отношение человека к человеку, к природе и к сверхъестественному. Игнорировать эти учения означает игнорировать цель жизни, значение существования» [Фаган, ДеКорс, 2007: с.440].

Пятое понятие «взаимодействие» исследует степень взаимосвязи между поселениями на региональном уровне.

Рассмотрим классификационный фрагмент «Уровень памятника» (Рис. 7). Здесь в качестве опорного понятия приведено «Местонахождение» – любой памятник или совокупность тесно смыкающихся территориально памятников, связанная с определенным местом и заметно отделенная территориально от других конкретных археологических объектов значительным расстоянием (свободным пространством) – так, чтобы заслуживать быть отмеченным отдельным значком (как отдельная точка) на археологической карте [Клейн, 1991: с.376].

В диадной группе приведена классификацияг.-Ю.Эггера. Здесь под отложениями на неизвестных основаниях понимаются остатки, в отношении которых исследователь не может уверенно и однозначно решить, почему они оказались здесь и почему в таком сочетании [Клейн, 1978: с.107-108].

Местонахождение				
Отложения на неизвестном основании		Отложения на известном основании		
Открытые комплексы	Полузакрытые комплексы	Закрытые комплексы		
Типы комплексов				
Комплексы пребывания	Комплексы созидаия	Комплексы обитания	Комплексы хранения	Комплексы упокоения

Рис. 7. Классификационный фрагмент «Уровень памятника» [Холюшкин, 1995: с.102].

Под отложениями на известных основаниях понимаются остатки, причины попадания и отложения которых в культурный слой понятны, очевидны и определены для современных исследователей [Клейн, 1978: с.107-108].

Триадная группа отражает характер отложений комплексов.

В открытые комплексы входят ансамбли артефактов, одновременность отложений которых не может быть гарантирована методами археологического исследования.

В полузакрытые комплексы входят ансамбли артефактов, не определенные по степени длительности функционирования в прошлом.

В закрытые комплексы входят совокупности предметов, одновременность упокоения которых гарантирована условиями обнаружения и свидетельствами их функционального взаимодействия [Клейн, 1978: с.98].

Пентадная группа сформирована нами на основе предложенной Л.С.Клейном классификации с той лишь разницей, что мы разграничили понятия «хранение» и «упокоение» [Клейн, 1978: с.110-114].

Комплексы пребывания. От пребывания остается мало материальных следов (часто вовсе не остается), и они рассеяны по разным местам, которые люди посещали и по которым проходили и

проезжали. Как правило, это остатки изолированные, узкого набора, но безразлично распределяющиеся... [Клейн, 1978: с.111]. К ним относятся: отдельные (случайные находки), дороги, поля древнего земледелия, культовые места, места сражений, затонувшие корабли.

Комплексы созидания. Под созиданием понимаются Л.С.Клейном те преобразовательные действия, в результате которых остаются отдельные сооружения (обычно военно-оборонительного, ирригационного, горнодобывающего и культурного назначения). Это рвы, копи, насыпи, каменные и деревянные конструкции, изображения на скалах. По специальным инструментам для добычи кремня, обсидиана, меди и других металлов идентифицируют памятники-каменоломни (копи, рудники и шахты). Святилища, и культовые сооружения, включая петроглифы и лабиринты, могут быть, а могут и не быть частью поселенческого памятника. Месопотамский зиккурат (храмовый курган) доминировал над своим городом-матерью, а города майя, например, Тикал, отличались впечатляющими ритуальными районами, вокруг которых находились жилые зоны. Другие известные ритуальные памятники, такие как Стоунхендж в Англии, и Грейт Серпент Маунд в штате Огайо, являются отдельно стоящими памятниками. Ритуальные артефакты, такие как шипы морского ската и статуи, могут ассоциироваться со священными памятниками. Торговые памятники идентифицируются по большому количеству экзотических объектов торговли и по их стратегическому положению возле главных городов. Примером может служить процветавший когда-то ассирийский рынок возле города хаттитов Канеше в 1900 году до н.э. [Фаган, ДеКорс, 2007].

Комплексы обитания. Обитание, по Л.С.Клейну, это наиболее обширный, почти всеобъемлющий вид деятельности, втягивающий в себя отдельные операции всех остальных видов, и от него остаются богатые материальные остатки, сосредоточенные на одном месте [Клейн, 1978: с.111]. К комплексам обитания относятся сезонные стоянки охотников — это те места, где доисторические люди забивали животных и селились вокруг туш на время их разделки, а также временные стоянки. Поселенческие памятники являются наиболее важными, поскольку это те места, где люди жили и выполняли множество действий. Артефакты на поселенческих памятниках говорят о домашних работах — приготовлении пищи и изготовлении инструментов. Здесь обычно присутствуют жилища. Положение комплексов обитания говорит о том, что часть их может относиться к отложениям на известных основаниях, в силу внезапного или быстрого погребения. Примером такого захоронения культурных остатков может служить Помпея. Кроме того, как полагают Г.И.Медведев и С.А.Несмеянов, скопления и комплексы, погребенные в аллювиальных и озерных отложениях, так же обладают наиболее высоким характером закрытости [Медведев, Несмеянов, 1988: с.114].

Комплексы хранения. В комплексы включают клады и жертвоприношения. В них обычно находятся специально отобранные вещи. В связи с гибелью хозяина они оставались невостребованными и поэтому до сих пор обнаруживаются эти своеобразные «захоронения» [Мартынов, Шер, 1989: с.10]. Существуют различные классификации комплексов хранения.

Так, Г.Чайлд выделял домашние клады, вотивные клады, клады торговцев и клады литейщиков (Карбунский клад) [Балонов, 1991: с.315]. Клады обычно находят случайно, поэтому они попадают в руки специалистов уже из «вторых рук». Примером может служить найденный в 1877 году Аму-Дарьинский клад [Мартынов, Шер, 1989: с.10].

Комплексы упокоения. Захоронения являются самым важным источником о доисторическом социальном устройстве и социальном ранжировании. Захоронение тела в действительности является минимальной частью погребальной практики в каком-либо обществе. Ритуалы похорон обычно отражаются не только в положении тела в могиле, но также и в украшениях и погребальной утвари (в том числе мебели). Содержимое захоронения, очень богатой или чрезвычайно бедной, является барометром социального статуса [Фаган, ДеКорс, 2007: с.448-449].

Последующим уровнем стратегии исследования является изучение археологических культур. В связи с этим рассмотрим классификационный фрагмент «Культурологическая археология» (Рис.8). Здесь в качестве опорного понятия представлена археологическая культура. Значительная часть археологов, в настоящее время, считает, вслед за Даниелом, что «... Выявление и описание культур не проблема преистории; культуры современного археолога, как и периоды и эпохи его предшественников, — всего лишь понятийные инструменты» [цит. по Клейн, 1991: с.173]. Само понятие имеет ряд семантов:

а) *археологическая культура* - понятие охватывающее процесс перехода от палеокультур к современным. Т.е. под археологической культурой понимаются остатки палеокультур и информация о них, перенесенная в современную культуру;

б) археологическая культура означает комплекс артефактов, обладающий признаковым и типологическим единством. Ее можно назвать типологической культурой.

в) археологическая культура, обозначает иерархический уровень пространственной организации артефактов.

Археологическая культура (с2)					
Материальная культура			Духовная культура		
Техномические элементы культуры		Социотехнические элементы культуры		Идеотехнические элементы культуры	
Техномические элементы материальной культуры	Техномические элементы духовной культуры	Социотехнические элементы материальной культуры	Социотехнические элементы духовной культуры	Идеотехнические элементы материальной культуры	Идеотехнические элементы духовной культуры
Иерархия культур					
Местные образцы	Субкультура	Культура (культурная область)	Культурная группа	Технокомплекс	

Рис. 8. Классификационный фрагмент «Культурологическая археология».

Диадная группа основана на противопоставлении материальной и духовной культуры. Под материальной культурой понимаются вещи, орудия, знания, которые являются продуктами материального производства. К духовной культуре относятся продукты художественной сферы культуры, эстетические ценности и т.д.

Триада построена на основе разработок Л.Бинфорда, заимствованной у Л. Уайта. Здесь в качестве первого понятия приводятся техномические элементы, предназначенные для «совладения» с природной средой. Под вторым понятием понимаются предметы, в которых находят отражение социальная подсистема палеокультуры. Поскольку совладение с природной средой может осуществляться только в процессе осмысления мира, то результаты духовного освоения мира, по Л.Бинфорду, должны определенным образом фиксироваться с тем, чтобы стать достоянием следующих поколений. Один из видов фиксации – отражение идей и представлений в мире вещей. Эти предметы, которые имеют первичный функциональный смысл в «идеологической» подсистеме Л.Бинфорд назвал «идеотехническими». Они символизируют и обозначают определенные идеологические объяснения, которые дает человек явлениям окружающей среды. Л.Бинфорд в качестве примера «идеотехнических предметов» называл фигурки божеств, тотемные символы, символы природных сил [Binford, 1972: p. 216].

Пятерка понятий отражает классификацию Д.Кларка [1968].

Местные образцы – взаимосвязанный набор одновременных типов артефактов.

Субкультура представляет собой часть археологической культуры, соответствующая однородной деятельности субпопуляции и характеризующаяся особой функциональной специализации части общества [Clarke, 1968: p. 669].

Культура – политетический набор специфических и всеобщих категорий типов артефактов, которые неслучайным образом встречаются вместе в комплексах в пределах ограниченного географического ареала.

Культурная группа – семейство трансформированных культур; родственные культуры, характеризующиеся комплексами, которые обладают разными состояниями типов из одного и того же политетического ряда специфических многовариантных типов артефактов.

Технокомплекс – совокупность археологических культур, определяемых политетическим рядом различных типов одной и той же группы типов артефактов, рассматриваемой как широко распространенная и тесно связанная реакция на общие факторы природного окружения, экономики и технологии.

Нетрудно заметить, что этот грандиозный по замыслу опыт формализации градаций подразделений археологической культуры отличается регулярностью на всех уровнях интеграции понятий, когда более сложные ячейки складываются из простых по одним и тем же правилам.

Все вычисления Д.Кларка основаны на этнографических материалах по североамериканским индейцам и племени банту. На основе этих этнографических материалов делаются попытки определить соответствия местных образцов родовой общине (2-10 человек);– группе родовых общин (10-100 человек); культуры – племени (100-1000 человек); группы культур – группе племен (1000-10000 человек); технокомплекса – конфедерации племен (10000-100000 человек). Данные подразделения отличаются определенными пространственными характеристиками: для культуры радиус эксплуатируемой территории составляет – 20-200 миль; для культурной группы – 200-750

милей; для технокомплекса – 750-3000 миль. Кларк также отметил, что, по данным материальной культуры живого этнографического племени, которому должна соответствовать археологическая культура, совпадения типов в комплексах охватывает от 65 до 95% материала, в культурных группах – от 65 до 30% совпадений, в субкультуре уровень совпадений превышает 95% [Clarke, 1968: p. 188, 372].

Привлекаемые Д.Кларком этнографические данные показали, что нельзя безоговорочно отождествлять археологическую культуру с конкретными этнографическими единицами [Деревянко, Фелингер, Холушкин, 1989: с.35]. Нет уверенности и в том, что указанные выше нормативы имеют универсальный характер.

Из этого следует, что термины «культура», «археологическая культура» должны иметь много различных значений, о чем и свидетельствует неоднозначная трактовка этих понятий. То обстоятельство, что с рубежа 40-50-х годов идет сплошной поток статей, посвященных этой проблеме [Клейн, 1991: с.125] свидетельствует, с одной стороны, о важности этой проблематики, а с другой стороны о полном хаосе и неразберихе в ее определении.

Среди попыток наведения порядка в этих представлениях следует отметить работу Л.С.Клейна [1991], выделившего большое число вариантов определения этого понятия, а также известную работу Д.Кларка [1968]. Именно в этих работах наиболее четко осознается сложность понятия археологическая культура.

Данная публикация является продолжением этих разработок и посвящена раскрытию понятия эволюционно-функциональная культура. При ее разработке для нас были наиболее важны развернутые схемы Д.Кларка. Его многоэтажные теоретические конструкции представляют грандиозную попытку формализовать археологическую теорию и установить достаточно жесткую схему классификации. Основные признаковые свойства эволюционных показателей культуры, разработанные Д.Кларком, публикуются в приводимом ниже классификационном фрагменте (рис.9).

Трансформация культуры					
Восходящая стадия			Нисходящая стадия		
Архаическая стадия		Стадия климакса		Стадия деградации	
Техномические элементы материальной культуры	Техномические элементы духовной культуры	Социотехнические элементы материальной культуры	Социотехнические элементы духовной культуры	Идеотехнические элементы материальной культуры	Идеотехнические элементы духовной культуры
Периоды трансформации культуры					
Ювенальный период	Период роста	Пубертатный период	Постпубертатный период	Период умирания	

Рис. 9. Классификационный фрагмент «Эволюционно-функциональная культура»

Здесь в качестве опорного понятия приводится «Трансформация культуры». Под трансформацией мы понимаем пространственно-временные преобразования сочетаний слагающих культуру параметров, из которых из одной конструкции может быть получен ряд других. В археологии различают ряд разновидностей таких трансформаций, среди них можно привести следующие:

Ковариация - параллельная согласованная изменчивость, при которой характеристики вовлеченных в нее параметров изменяются вместе или даже в одном и том же направлении, относительно некоторой общей шкалы или сети координат (Клейн, 1991: с.364). Различают ковариации в пространстве (кодивергенцию) и во времени (коальтерацию).

В качестве диадной группы предлагается эволюционный ряд, имеющий две ветви: восходящую, когда происходит становление культуры, и нисходящую, когда культура деградирует. По мере развития происходит, как считает Д.Кларк, усложнение культуры, т.е. возрастает количество ее типобразующих признаков. На нисходящей ветви происходит ее некоторое упрощение.

В качестве триадной группы приведены: архаическая стадия - стадия климакса- стадия деградации.

Архаическая стадия - ранний этап развития культуры, в котором зафиксированы первичные архаические фрагменты исторического развития.

Стадия климакса - период бурного развития культуры, достигшей своего апогея. По представлениям Д.Кларка, примером такого состояния культуры являются культуры городского типа (Clarke, 1968; Федоров-Давыдов, 1970).

Стадия деградации - заключительная стадия развития культуры, сопровождающаяся процессами угасания и нарушениями в системе типослагающих ее признаков.

В качестве альтернативно-тождественного понятия приведена «Периоды трансформации культуры». Этому понятию можно поставить в соответствие пятиэлементную группу «ювенальный период - период роста - пубертатный период - постпубертатный период - период умирания».

ювенальный период - начальная, нулевая, исходная стадия родившейся культуры. Популяция этой рождающейся культуры состоит из небольшого количества культуuroобразующих признаков. Вся популяция при этом рисуется как неустойчивая и структурно рыхлая;

период роста - характеризуется увеличением числа культуuroобразующих признаков, подъемом уровня коррелированности между культуuroобразующими признаками. На этой стадии происходит рост уровня схожести и однородности популяции. А сама популяция рисуется как более устойчивая, стабилизирующаяся система со сформировавшимся внутренне устойчивым ядром (Clarke, 1968);

пубертатный период - период зрелости. В этот период достигается максимум внутренней однородности культуры, максимум схожести между представителями популяции. При этом темпы эволюции культуры резко возрастает;

постпубертатный период - период угасания культуры, характеризуемый уменьшением тех показателей, которые определяют внутреннюю однородность культуры и обеспечивают схожесть ее представителей между собой. Система признаков, формирующих культуру, приближавшаяся на предшествующей стадии к монотетической, вновь становится рыхлой, политетической;

период умирания - завершающая стадия цикла развития культуры, приводящая в экстремальной ситуации к гибели культуры, а при благоприятных обстоятельствах к той же нулевой стадии, с которой процесс начинался. Вся популяция на этой стадии структурно рыхлая и неустойчивая, а ее ядро представляет ярко выраженную политетическую систему.

Следующим этапом следует считать моделирование процессов имевших место в далеком прошлом.

2. Основные этапы археологических поисков стратегий исследования

2.1. Проблемы хронологии археологических объектов

В поисках стратегии археологических исследований на первый план первоначально ставили выявление хронологии археологических объектов. Периодизация как исследовательский прием является неотъемлемым элементом археологических исследований. Самыми крупными периодами принято считать каменный, бронзовый и железный века. Представление об их существовании возникло еще в античности. Так, уже в I веке до н.э. Тит Лукреций Кар в поэме «О природе вещей» писал: Древним оружием служили руки, когти и зубы, а также камни, обломки кустов и деревьев; затем появились железо и бронза, но бронза появилась раньше; употребление железа стало известно лишь позже [Городцов, 1923: с.24].

Однако археологическое обоснование составленной Лукрецием схеме культурного развития доисторического человека получила лишь в 1836 году в работе датского археолога Кристиана Юргенсена Томсена «Путеводитель по северным древностям» [Монгайт, 1973: с.16-17]. Он систематизировал беспорядочную коллекцию артефактов, распределив их по трем группам, представляющим каменный, бронзовый и железный века.

Классификацию Томсена развил другой датчанин, Й.Я.А. Ворсо, доказавший стратиграфическую целостность этой системы. Изучая археологические находки по всей Европе, Ворсо продемонстрировал обоснованность системы трех веков. Техническое деление доисторического периода давало археологам контекст, в рамках которого они могли помещать свои находки. Система трех веков неоднократно подвергалась модификациям.

Сначала Леббок разделил «каменный век» на палеолит и неолит. Вскоре в 1866 году Уэстроп вставил между ними мезолит. Затем химик Бертло в 1889г. пришел к выводу, что в ряде районов бронзовому веку предшествует период применения чистой меди, без искусственной присадки олова, — «медный век». По отношению к этому периоду были придуманы дополнительные термины — «энеолит» и «халколит», подчеркивающие сохранение важной роли каменных орудий наряду с медными [Черных 1965: с. 109-110].

Палеолит подвергся дальнейшему членению. В 1869г. Г.Мортилье разделил его на четыре эпохи — мустье, ориньяк, солютре, мадлен. Позже он и другие археологи открывали всё новые деления в материале, которые одними рассматривались как эпохи, а другими — как локальные культуры: до-шелль, шелль, клэктон, ашель, леваллуа, мустье, ориньяк, солютре, мадлен и т.д. Такие же прогрессивные стадии были намечены в мезолите, в неолите, в бронзовом и железном веке.

Система перекрестной датировки по импортным вещам была применена в начале XX в. О. Монтелиусом для установления абсолютной хронологии бронзового века Северной Европы. Однако этому предшествовал довольно значительный цикл работ по установлению относительной периодизации эпох, следующих за каменным веком. Еще в 1859г. Ворсо считал, что бронзовый век Северной Германии, Англии и Скандинавии может быть разделен на две фазы, первая из которых характеризуется обрядом ингумации и яркостью форм и украшения бронзового оружия, а вторая — кремацией и упадком бронзовой индустрии. Мортилье в 1875г. также делил французский бронзовый век на две фазы [Монгайт, 1973]. В основе периодизации Монтелиуса лежит технико-типологическая схема. Он пытался перевести эту схему в абсолютные даты, основываясь на синхронизации путем перекрестной датировки по материалам Восточного Средиземноморья и Египта [Монгайт, 1973].

Современник Монтелиуса Софус Мюллер (1846—1934) представил для Дании несколько иную периодизацию бронзового века, поделив его на девять групп.

Несколько позже В. А. Городцов предложил и поныне не потерявшую своего значения схему этапов развития бронзового века Европейской России [Городцов, 1923].

Сходные схемы были созданы и для периодизации железного века. В 1872г. Гильдебранд предложил разделить доримский железный век на два периода, которые он назвал гальштатом и латеном. В 1885г. О. Тишлер разделил латенские материалы на три периода — ранний, средний и поздний, и датировал начало латена 400г. [Монгайт, 1973].

Попытка разработать классификацию и хронологию всех археологических памятников Западной Европы была сделана выдающимся французским археологом Жозефом Дешелеттом (1862—1914) в его четырехтомном «Руководстве по археологии первобытной, кельтской и галло-римской» (1908—1914 гг.) [Монгайт, 1973].

По существу все схемы археологической периодизации, созданные в XIX в., были продолжением и развитием «системы трех веков». «Эпохи», «периоды», «фазы» были подразделениями внутри каждого «века». Эти схемы, основанные на местном материале, не могли быть универсальными, пригодными для всей Европы. Необходимость региональных классификаций, изучения отдельных археологических провинций становилась насущной необходимостью.

При этом отмечались два принципа в подходе к созданию археологических периодизаций: технологический и геолого-палеонтологический. В периодизации палеолита первая разбивка материала во времени, созданная Эдуардом Ларте, включала четыре периода: 1) пещерного медведя; 2) мамонта и шерстистого носорога; 3) северного оленя; 4) зубра и бизона. Гарриго дополнил предшествующий этим четырем периодам период древнего слона, носорога Мерка и гиппопотама [Монгайт, 1973: с.21].

Указанный способ периодизации периодически уступал свое место способу классификации археологических периодов на основании прогрессивного развития разных явлений человеческой культуры. В общей истории культуры еще в XVIII веке получила известность периодизация А. Фергюссона, выделившего три стадии: дикость, варварство и цивилизацию. Схемы Свена Нильсона и Льюиса Моргана, повторили деление истории культуры на те же ступени периодизации. Общекультурная периодизация Моргана, построенная на эволюционном принципе, стала стандартом марксистского подхода к первобытности, и советские археологи старались синхронизировать ее со «схемой трех веков», построенной на технологическом критерии. Как таковая она и представлена в учебнике В.И.Равдоникаса [1939].

В американской антропологии общества делились на несколько типов: а) эгалитарные, не знакомые с имущественным и социальным расслоением; б) стратифицированные, уже затронутые имущественным и социальным расслоением; в) иерархические, которым присуще имущественное и социальное расслоение.

Интересной представляется периодизация Хокса: Средневековая археология — это археология историческая, «сопровождаемую текстами». Раннежелезный век являлся *протоисторией*, освещенной римскими авторами, *пенеисторией*, едва затрагивалась греческими авторами. Бронзовый век назывался *параисторией*, сосуществующей с очагами письменности. Раннебронзовый, медный век и поздний неолит охватывались *телеисторией* (которая, хоть и «далеко от истории», но всё же в связана с ней). А ранний неолит, мезолит и палеолит отводились *антеистории*. Тут в археологи уже отсутствуют тексты [Hawkes 1951: p. 1-9].

По мере накопления археологических материалов обнаружилась территориальная ограниченность нормативной шкалы Мортилье: в Восточной Европе, Азии и особенно Африке,

оказались совершенно другие палеолитические культуры и другие линии развития палеолита, с другими особенностями, с другими делениями шкалы. Такг.П.Григорьевым в ряде работ была высказана идея об отсутствии в Азии и Африке позднего палеолита и о существовании там особой постмустьерской эпохи [Григорьев, 1977: с.44-209].

В этих условиях геологические критерии снова вышли на первый план, и критерием при отнесении памятников к определенной эпохе стали служить не столько технико-типологические особенности индустрий, сколько временные показатели (данные относительной и абсолютной хронологии). Иллюстрацией этого может служить высказывание В.А.Ранова: «... мне казалось, что археологический материал сам по себе обладает достаточной разрешающей способностью, достаточной для создания периодизационных схем». Однако попытки построить такую шкалу окончились неудачей: материал не складывался в единую систему, европейская же система слишком далека и чужда, да и сама по себе дискредитирована как стандартная шкала. Всё это привело Ранова «к отрицанию права за археологическим инвентарем быть основой периодизации» [Ранов 1984].

Различия между рассмотренными периодизациями вызваны произвольным характером критериев периодизации и структур исторических периодов. При этом каждая отдельная периодизация, удобная для исследования тех или иных конкретных вопросов в узких локальных рамках, становится существенным препятствием на пути археологических обобщений. В связи с этим назрела необходимость в построении таких периодизаций, которые отражали бы общие закономерности истории, природы и общества.

По существу все схемы археологической периодизации, созданные в XIX в., были продолжением и развитием «системы трех веков». «Эпохи», «периоды», «фазы» были подразделениями внутри каждого «века». Эти схемы, основанные на местном материале, не могли быть универсальными, пригодными для всей Европы. Необходимость региональных классификаций, изучения отдельных археологических провинций становилась насущной необходимостью.

2.2. Поиски типологии археологических данных

Другим чрезвычайно важным достижением в методике археологических исследований была разработка типологического метода.

Так, английский археолог Питт-Риверс (1827—1900) начал изучать коллекцию оружия и изменение его типов. В процессе исследования истории оружия он пришел к мысли о том, что все археологические материальные объекты развиваются эволюционным путем и могут быть расположены в порядке типологической последовательности. Для этого Питт-Риверс располагал коллекции не по странам и эпохам, а по рядам отдельных вещей, прослеживая путь развития каждой вещи. [Монгайт,1973]. Наконец, Питт-Риверс, подобно тому как это сделал Питри в Египте, показал значение массового материала. Он подчеркивал, что рядовые вещи, обычные и типические, имеют для социологического изучения такое же, если не большее, значение, чем уникальные вещи и произведения искусства, на которые прежде обращали главное внимание. Основным принципом в расположении материала Питт-Риверс считал развитие от более простых форм как более ранних к сложным. Однако он не считал необходимым слепо следовать этому принципу, так как признавал, что существовали периоды упадка и вырождения [Монгайт,1973].

Типологический метод не был изобретением какого-либо одного археолога. Почти одновременно в разных странах начали применять типологию.

Дальнейшее развитие типологический метод нашел в трудах О. Монтелиуса. Изучая древности эпохи бронзы Скандинавии, Италии, Греции, Египта и Вавилонии, он распределил все вещи по типам, исходя из представления, что тип — это группа вещей одного и того же назначения, однородных по внешнему виду, но отличающихся друг от друга в деталях. Внутри каждого типа вещи в зависимости от различия их форм в деталях располагаются в типологические эволюционные ряды в порядке их последовательного изменения от простого к сложному или иногда наоборот — от сложного к простому. Эволюционные ряды разных типов связаны между собой хронологически. Правильность построения рядов Монтелиус проверял по совместным находкам (комплексам) в погребениях. Это — один из важнейших элементов метода Монтелиуса. Он считал типологические ряды лишь рабочей гипотезой до тех пор, пока они не проверены по совместным находкам в «закрытых комплексах» — погребениях, кладах и т. д. Стратиграфия и совместные находки дают возможность установить «параллелизм» серий, т.е. те случаи, когда древнейший тип в одном ряду одновременен древнейшему типу в другом ряду, и более поздний тип одного ряда одновременен более позднему типу другого ряда (например, ранние топоры

одновременны ранним мечам и т. п.). При построении рядов Монтелиус обращал внимание на «типологический рудимент», т. е. на ту деталь вещи, которая первоначально имела практическую функцию, однако позже ее утратила и продолжает существовать в виде орнамента. Свои взгляды Монтелиус изложил в опубликованной в 1903г. книге «Древние культурные периоды на Востоке и в Европе» [Монгайт,1973].

Наиболее отчетливо вопрос о сопоставлении археологической культуры и её носителя - народа и попытка на основании непрерывной генеалогии культуры судить о генеалогии народа — ее носителя поставлен О. Монтелиусом в его статье «О вторжении наших предков на север», опубликованной на немецком языке в 1888г. Монтелиус пытался показать, что нет никакого разрыва между железным и бронзовым веком севера, никакого перерыва в развитии, и делал вывод, что жители севера уже в бронзовом веке были те же самые, что и в последующее время, т. е. принадлежали к германцам. Если удастся перекинуть мост от бронзового века к неолиту, то можно будет утверждать, что германские племена поселились здесь еще тогда. В этих рассуждениях Монтелиуса содержатся важнейшие для археологии выводы: возможность сопоставления археологической культуры и народа и возможность на основании непрерывного развития культуры судить о генеалогии народа — ее носителя. Монтелиус нигде, кроме названной статьи, этот метод не применял и в последующие годы к нему не возвращался. Поэтому неудивительно, что открытие метода было приписано впоследствии германскому археологу Густаву Коссинне (1858—1931) [Монгайт,1973].

Пионерными для своего времени были разработки Франсуа Борда. Он ввел систему измерений и индексов. Им же для среднего палеолита был составлен список стандартных типов — из 63 номеров орудий. Для подготовки комплекса к сопоставлению с другими подсчитывается наличие в нем орудий каждого из стандартных типов, вычисляются проценты ко всему составу, а затем для наглядности это отображалось в виде кумулятивных графиков. Позже по образцу бордовского типлиста его супруга Дениза Сонневиль-Борд и Ж. Перро разработали типлист для французского верхнего палеолита. Этот типлист состоял из 93 номеров. Затем исследователи в разных странах стали применять их с небольшими модификациями. В результате своих сопоставлений Борд пришел к выводу, в мустье есть несколько вариантов. Их можно рассматривать как сосуществовавшие археологические культуры, а культуры эти, полагал Борд, соответствуют этническим группам прошлого.

Заметно усовершенствован метод Борда в работах Анри де ЛЮМЛЕЯ. В книге по исследованиям Грота Ортю были широко использованы различные естественно-научные методы. Типология Борда была отражена и в вышедшей в 1968г. справочнике М. Брезийона «Наименование предметов расщепленного камня» — справочник по терминологии кремневого инвентаря» [Brezillon, 1968]. Итак, в итоге эволюции взглядов исследователей на проблемы вариативности можно выявить следующие тенденции:

- по-прежнему используются на классификационном уровне стандартные тип-листы Ф. Борда, индексы и кумулятивные графики;
- подробно изучается технология обработки камня, включающая археометрию и реконструкции технологических цепочек;
- исследование качественных характеристик сырья;
- привлечение более сложных методов статистического анализа;
- По-прежнему имеет хождение и заблуждение традиционистов в том, что организационные характеристики археологических данных информируют их непосредственно о характере археологической культуры.

На американской почве сериация артефактов имела большее математическое выражение (у Крёбера в 1916г., у Крёбера и Стронга в 1924, у Стронга в 1925 уже с коэффициентами корреляции, затем у Форда в 1936–38гг.) [Ford, 1954;1962]. В 20-е – 30-е годы статистика и корреляция применялись в отдельных работах советских археологов Ефименко, Арциховского и Грязнова. В 1940г. Крёбер опубликовал статью «Статистическая классификация», которая вместе с последовавшей в 1944г. статьей Алекса Кригера «Типологическая концепция» послужила для Сполдинга исходным пунктом для всей его исследовательской деятельности [Клейн, 2004]. Он написал статью «Статистическая техника для открывания типов артефактов». В этой статье Сполдинг рассматривает тип «как группу артефактов, показывающих устойчивый комплекс признаков, сочетание свойств которых дает характерную конфигурацию. Это значит, что даже в контексте очень схожих артефактов классификация в типы есть процесс открывания сочетаний признаков, предпочитавшихся создателями артефактов, а не произвольная процедура классификатора» [Spaulding 1953a: p. 305].

Минимальным требованием для установления типа в построениях Сполдинга является корреляция двух признаков. Приведены примеры применения статистики к конкретным археологическим материалам.

Как видно из предыдущего текста применение математических методов началось до Новой Археологии и проходит также и вне ее. Однако вместо простой статистики или двухмерной корреляции (составления координатных таблиц) ими применялся факторный анализ. Придавая равный вес всем признакам и отбросив культурную иерархию, Бинфорд пришел к тому, что поставил под вопрос существование типов в археологии и выдвинул требование искать более сложные и более тонкие взаимоотношения признаков разных комплексов – с помощью многомерного факторного анализа. Признаки вместо типов вышли на первый план типологии!

Примером применения факторного анализа может служить анализ палеолитических орудий в Бинфордовской статье 1972г. «Построение моделей – парадигм и нынешнее состояние палеолитических исследований». Им было выделено пять факторов в палеолитических местонахождениях Сирии, Израиля и Франции. Которые он сопоставил их с фациями мустье Борда. В отличие от Ф.Борда Бинфорд считал фации мустье не этническими культурами а «структурными позами», сезонно обусловленными. Поскольку процессуалисты отбросили исторический подход и искали в массовом материале только законы, для них отдельные факты потеряли всякое значение.

Более сложные модели были представлены в работах Д.Кларка. Уже в его первой монументальной книге «Аналитическая археология» [1968] содержались «Стратегии» в которых рассматривались возможности графического представления и математического анализа волн диффузии, распространения и густоты сети памятников), обеспеченности районов природными ресурсами, влиятельности и конкуренции центров (построение полигонов Тиссена).

В следующем сборнике «Модели в археологии» вышедшем под его редакцией в 1972г., во вступительной статье Кларка «Модели и парадигмы в археологии» рассмотрены различные [1972].

2.3. Географический детерминизм в стратегиях археологического исследования

Среди других теоретических концепций важное место занимает так называемый диффузионизм, поставивший наряду с понятием эволюции, исторического прогресса понятие культурной диффузии, т. е. пространственного перемещения культурных явлений. Впервые вопрос о роли географического фактора в развитии общества был поставлен в античном мире Геродотом, Гиппократом, Страбоном и Полибием. Культурный переворот VIII – V веков до н. э. вызвал к жизни целый ряд идей, к которым человечество не раз обращалось. В это время человек очень сильно зависел от природной среды и естественно, что главным, и единственным, в географии было то, как природная среда влияет на жизнь людей, их культуру, историю и т.п. Гиппократ считал что тело и дух людей определяется климатом, а Аристотель – что жители холодных стран храбры, но лишены выдумки и технической изобретательности, в отличие от людей, проживающих в теплых краях.

Идеи географического детерминизма вновь получили развитие в XVI в. Одним из представителей этого направления был Жан Боден, который в 1566 г. опубликовал книгу «Метод облегченного изучения истории». Основная идея книги заключается в том, что качества человека во многом зависят от тех природных условий, в которых он проживает. Эта теория, называемая теорией географического детерминизма. В своей работе Жан Боден не только обуславливает особенности народов климатическими и другими географическими условиями, но и делает выводы о роли этих факторов на социально–политическое устройство государств, находящихся в различных точках Земного шара. Он пишет об особенностях развития науки на севере и на юге, на западе и востоке. Наконец, он приходит к определенным выводам о роли различных государств в жизни Европы в период рабовладельческого и феодального строя. При этом он не ограничивается государствами Европы, но переносит свой интерес и на государства Передней Азии и Ближнего Востока.

Можно сказать, что Жан Боден создал с одной стороны первую геополитическую модель нового времени, а с другой – впервые вывел зависимость духовной жизни человека и целой нации от условий их существования. Выдвигая на первый план влияние естественных условий, Боден уделяет особое влияние климату. Боден разделил землю на три пояса:

1. Экваториальный
2. Полярный
3. Умеренный

Каждый из них, он связал с предпосылками для определённого труда человеческой деятельности: «И если правильно изучать историю, увидишь, что величайшие полководцы приходят с севера, а искусство, философия и математика рождаются на юге». Боден даёт в своём труде конкретные рекомендации правителю, желающему установить оптимальный политический режим, связанные с учётом географических факторов. Географический детерминизм Жана Бодена отразился на философских воззрениях многих европейских мыслителей, и особенно на взгляде Монтескье. Монтескье лишь попытался поднять географический детерминизм на уровень естественных наук его времени, сформулировать законы, подобные законам Галилея, Кеплера, Ньютона.

Поклонница французского мыслителя Екатерина II писала: «Российская империя есть столь обширна, что кроме самодержавного государя, всякая другая форма правления вредна ей, ибо все прочие медлительнее в исполнениях». Отсюда – вывод о том, что для таких огромных просторов необходим самодержец – просветитель, вполне совместимый с мыслью о том, что «неудивительно», что «Россия имела среди правителей много тиранов».

Преимущество европейцев над прочими народами в XIX в. было столь очевидно, что Ф. Гегель построил философию истории на принципе мирового прогресса, который должен был быть осуществлен германцами и англосаксами, ибо считал, что все аборигены Азии, Африки, Америки и Австралии – «неисторические народы».

Своё дальнейшее развитие геодетерминизм получил в двух странах. К концу XVIII – началу XIX века уже была создана английская колониальная империя, над которой «никогда не заходило солнце», Германия же представляла конгломерат отдельных государств, частично тяготевших к Пруссии, частично к Австрии. Особенности политического развития оказали влияние на социологические теории, из авторов которых нас более всего будут интересовать Бокль и Риттер. Английский историк Генри Томас Бокль (1821–1862) в своём труде объединил идеи Монтескье о роли климата с демографической концепцией Мальтуса, которого он считал «наиболее крупным авторитетом по вопросам народонаселения». Если сами по себе идеи Мальтуса были выражением глубокого сомнения в том, что мир разумен и создан для человека, то их развитие последователями носило реакционный характер. Так случилось с Боклем, который из мальтузианской идеи ограниченного общественного продукта и теории климатических поясов, сделал вывод о том, что жители тропиков, которым необходимо меньше пищи, чем жителям умеренного, должны делиться своими богатствами с последними. Согласно Боклю, индийский народ, доведённый английскими колонизаторами, разрушившими ирригационные системы, до постоянных голодовок, «осуждён на бедность физическими законами климата».

По сравнению с Боклем, Риттер представлял предыдущий этап развития буржуазной мысли. В классической немецкой философии была поставлена проблема диалектической взаимосвязи развития природы и развития общества, проблема, стоявшая на уровне тех теоретических задач, которые были характерны для философии Гегеля: он «впервые представил весь природный, исторический и духовный мир в виде процесса, то есть в непрерывном движении, изменении, преобразовании и развитии, и сделав попытку раскрыть внутреннюю связь этого движения и развития... Его историческая заслуга состояла в том, что он поставил её». Решить эту задачу попытался в своём труде Карл Риттер (1779 – 1859), впервые внёсший идеалистическую диалектику в анализ конкретного (в данном случае географического), а не исторического материала. В этом отношении показательна фигура Гердера, из которого порой делают какого-то непоследовательного приверженца Монтескье, сделавшего шаг назад сравнительно с его откровениями. А между тем во многом он предвосхитил Риттера: «Единственное и лучшее – это, следуя Гиппократу с его дальновидной наивностью, наблюдать климат отдельных местностей и затем медленно, медленно делать выводы».

Карл Риттер словно разрешал задачу, поставленную Гердером. Однако во главу угла им было поставлено понятие ландшафта, выработанное Гумбольтом и получившее у Риттера логическое завершение в понятии «географического индивидуума». Последнее обозначало «органическую природную область», характеризующуюся как внешними границами, так и внутренними связями, через которую и осуществлялось влияние природы на более или менее компактные массы людей.

Среди студентов, слушавших лекции Риттера, был Карл Маркс. По Марксу географическая среда влияет на человека посредством производственных отношений, возникающих в данной местности на основе данных производительных сил, первым условием развития которых являются свойства этой среды. Механизм этого влияния можно понять, лишь уяснив, что природа и общество не только взаимодействуют друг с другом, но и накладываются друг на друга: «В

понятие экономических отношений включается далее и географическая основа, на которой эти отношения развиваются, и фактически перешедшие от прошлого остатки прежних ступеней экономического развития, которые продолжают сохраняться зачастую только по традиции или благодаря *vis inertiae*, а также, конечно, внешняя среда, окружающая эту общественную форму».

Ф. Энгельс развивая мысль К. Маркса, указывал на прямую связь пищи и уровнем развития разных племен. По его мнению, «обильному мясному и молочному питанию арийцев и семитов и особенно благоприятному влиянию его на развитие детей следует, быть может, приписать более успешное развитие обеих этих рас. Действительно, у индейцев пуэбло Новой Мексики, вынужденные, кормиться почти исключительно растительной пищей, мозг меньше, чем у индейцев, стоящих на низшей ступени варварства больше питающихся мясом и рыбой».

Созданное Марксом учение о роли географического фактора в развитии общества имело огромное идеологическое значение. Маркс показал, что связанная с географическими условиями неравномерность развития различных государств, которая абсолютизировалась приверженцами геодетерминизма, определяется тем, что общество на различных этапах своего развития по-разному использует естественные богатства природы. Последние были разделены Марксом на две группы:

1. Естественные богатства средствами жизни (плодородие почвы, обилие рыбы в водах, дичь, плоды)
2. Естественные богатства средствами труда (действующие водопады, судоходные реки, лес, металлы, уголь, нефть)

«При зачатках культуры, – писал Маркс, – имеет решающее значение первый род, на более высоких ступенях – второй род общественного богатства».

У Маркса диалектика развития природы и общества приобретает законченно-материалистический вид: воздействуя для поддержания своего существования «на внешнюю природу и изменяя её», человек «в то же время изменяет свою собственную природу». Тем самым были заложены основы марксистско-ленинского понимания тех экологических проблем, которые в полном своём объёме встали перед человечеством лишь сегодня.

«Географический поппибилизм», идущий от «географии человека» Видаля де ла Блаша и Л. Февра и ставший концептуальной основой многих современных антропогеографических и историко-географических течений, едва ли может считаться более достойной альтернативой детерминистским и индетерминистским концепциям. Основываясь на мнении о рациональном в своих действиях человеке, поппибилизм исходит в своих построениях из модели о сознательной оптимизации человеком некоторой совокупности альтернативных видов жизнедеятельности с природной средой, выбирая в конечном итоге тот вид жизнедеятельности, который наилучшим образом подходит к данной среде. Иными словами, поппибилизм (как и детерминизм) рассматривает географическую среду как объективную данность, к которой человек в любом случае вынужден приспособливаться.

Становление человечества связано не только с природными воздействиями, как у прочих животных, но и с особым спонтанным развитием техники и социальных институтов. На практике мы наблюдаем интерференцию обеих линий развития. Следовательно, общественно-экономическое развитие через формации не тождественно этногенезам, дискретным процессам, протекающим в географической среде. С.В. Калесник отчетливо показал различие между географической и техногенной средой, в которых люди живут одновременно. Географическая среда возникла без вмешательства человека, и сохранила естественные элементы, обладающие способностью к саморазвитию. Техногенная среда создана трудом и волей человека. Ее элементы не имеют аналогов в девственной природе и к саморазвитию не способны. Они могут только разрушаться. Техно- и социосфера вообще не относятся к географической среде, хотя постоянно взаимодействуют с ней. Отмеченные адаптивные способности человека не просто повышены сравнительно с его предками, а связаны с особенностью, отличающей человека от прочих млекопитающих. Человек не только приспособляется к ландшафту, но и путем труда приспособляет ландшафт к своим нуждам и потребностям. Значит, пути через разные ландшафты ему проложили не адаптивные, а творческие возможности. При этом часто из виду упускалось, что творческие порывы человечества эпизодичны и не всегда приводят к желаемому результату, а следовательно, влияние человека на ландшафт далеко не всегда бывало благотворным. Шумерийцы провели каналы, осушив междуречье Тигра и Евфрата в III тыс. до н. э., – китайцы начали строить дамбы вокруг Хуанхэ 4 тыс. лет тому назад. Восточные иранцы научились использовать грунтовые воды для орошения на рубеже новой эры. Полинезийцы привезли на острова сладкий картофель (кумара) из Америки. Европейцы оттуда же получили картофель, помидоры и табак, а также бледную спирохету – возбудитель сифилиса. Эскимосы расправились

со стеллеровой коровой в Беринговом море; американские колонисты всего за полвека (1830—1880 гг.) перебили бизонов, а австралийские — несколько видов сумчатых.

Бесспорно, что ландшафт промышленных районов и областей с искусственным орошением изменен больше, чем в степи, тайге, тропическом лесу и пустыне, но если мы попытаемся найти здесь социальную закономерность, то столкнемся с непреодолимыми затруднениями. Земледельческая культура майя в Юкатане была создана в V в. до н.э. при господстве родового строя, пришла в упадок при зарождении классовых отношений и не была восстановлена при владычестве Испании, несмотря на внесение европейской техники и покровительство крещеным индейцам. Хозяйство Египта в период феодализма медленно, но неуклонно приходило в упадок, а в Европе в то же время и при тех же социальных взаимоотношениях имел место небывалый подъем земледелия и ремесла, не говоря о торговле.

Классический географический детерминизм послужил исходным пунктом для целого ряда весьма своеобразных концепций, в которых на первое место выдвинулся не отдельный человек, испытывающий влияние среды, а совокупность людей: население определенной территории, народ, наконец, общество как целостное образование, взаимодействующее с природой.

Это направление заметно у Льва Ильича Мечникова (1831—1888), изложившего свои взгляды на всемирную историю в труде «Цивилизация и великие исторические реки (Географическая теория прогресса и социального развития)» (1-е изд. на французском языке, 1889; русск. изд.: СПб., 1898; М., 1924; 1995). Автор формулирует закон «трех фазисов исторического развития». Первые цивилизации возникли в долинах крупных рек: египетская, как выражался Геродот, была «даром» Нила, ассииро-вавилонская возникла на берегах Тигра и Евфрата, китайская — в бассейнах Хуанхэ и Янцзы, индийская — Инда и Ганга. Это были древние века, или *речная* эпоха.

«По прошествии многих веков, — пишет Л.И. Мечников, — поток цивилизации спустился по берегам рек к морю и распространился по его побережью. Так наступила вторая эпоха в истории развития цивилизации, которую можно назвать *морской* эпохой, или *Средиземноморской*, так как цивилизация охватила главным образом берега этого внутреннего морского бассейна, расположенного между Африкой, Азией и Европой». Средние века, или средиземноморский период, охватывал двадцать пять веков — время с основания Карфагена до Карла Великого».

В конце XIX — начале XX веков географический детерминизм уступил свои позиции марксизму и немецкой классической философии, но сейчас, когда роль психологии человека в поведении общества стала играть не последнюю роль, идея географического детерминизма вновь приобретает вес [Мечников, 1995].

Например, Л.Н. Гумилев считал вторым фактором определяющий ход процесса этногенеза — географическую среду [Гумилев, 2002], игнорирование роли которой С.В. Калесник правильно назвал «географическим нигилизмом», но и преувеличение значения географической среды, не приводит к положительным результатам. В 1922г. Л.С. Берг сделал вывод для всех организмов, включая и людей. «Географический ландшафт воздействует на организм принудительно, заставляя все особи варьировать в определенном направлении, насколько это допускает организация вида. Тундра, лес, степь, пустыня, горы, водная среда, жизнь на островах и т.д. — все это накладывает свой отпечаток на организмы. Те виды, которые не в состоянии приспособиться, должны переселиться в другой географический ландшафт или вымереть».

3. Теории диффузионизма в стратегиях археологических исследований.

Появлению диффузионизма предшествовало созданное немецким географом и этнографом Фридрихом Ратцелем (1844—1904) антропогеографическое учение. Ратцель исследовал вызванные природными условиями различия между культурами и формы взаимодействия между народами: переселения, завоевания, обмен, торговлю и т.п. Прямолинейной схеме эволюционистов он противопоставил изучение конкретных условий, в которых эти явления наблюдаются. Собственно, это было не противопоставление, а дополнение, значительно расширившее исторический подход к проблеме эволюции человеческой культуры. Однако из наблюдений Ратцеля некоторые ученые сделали неправильный вывод, что развитие общества целиком определяется географической средой и что из-за разнообразия природных условий человечество не могло развиваться по одинаковым законам. Сходство отдельных явлений культуры у разных народов эти ученые рассматривали только как заимствования одним народом у другого.

В XX в. возникли этнографические школы, создавшие различные концепции культурной и этнической истории. Из области этнографии их учения проникли и в археологию. В 1898г. немецкий этнограф Лео Фробениус сформулировал понятие «культурного круга», под которым он

понимал сочетание целого ряда признаков (главным образом материальной культуры) в определенном географическом районе. Само по себе важное открытие Фробениуса получило в его трудах неправильное толкование. Он представлял себе «культуры» как самостоятельные организмы, развивающиеся независимо от людей и порождаемые главным образом природными условиями.

Еще дальше по пути противопоставления эволюционизму теории о множественности и даже случайности вариантов исторического развития пошел Ф. Гребнер, которого можно считать основателем школы «культурных кругов». Он утверждал, что каждый элемент культуры происходит из одного центра, принадлежит к одному «культурному кругу» и вместе с ним распространяется путем диффузии. Вся история культуры — это история перемещения по земному шару нескольких культурных кругов и их механических соединений друг с другом. Теория Гребнера получила развитие в работах этнографов венской культурно-исторической школы, виднейшими представителями которой были В. Шмидт, Б. Анкерман, В. Копперс.

Важнейшей задачей школы «культурных кругов», как об этом писали сами ее представители, было противопоставить их учение «эволюционизму», т.е. представлению о развитии от простых к более высоким культурам и формам жизни.

Теория «культурных кругов» оказала влияние и на археологию. Австрийский археолог и этнограф О. Менгин в своей монументальной «Всемирной истории каменного века» (1930г.) обрисовал древние культуры, основываясь на типологической классификации археологических находок и сравнении их с вещами, употребляемыми современными примитивными народами. Он пытался доказать, что история первобытного общества — всего лишь результат переселения (миграций) отдельных племен, принадлежащих к разным «культурным кругам». Хотя работа Менгина опиралась на огромный материал, но ошибочные позиции автора, «инвентарно-исторический» схематизм, применение теории «культурных кругов» омертвили исследование, и оно не оказало значительного влияния на развитие науки. У Менгина не было последователей.

В перечисленных здесь теориях заключаются элементы научного течения, которое получило название «диффузионизма». Диффузионисты считают, что все великие открытия и изобретения были сделаны только однажды на протяжении истории и распространялись из единых центров путем заимствований, миграций или просто диффузий элементов культуры или целых культур. Таким образом, все сходное в культурах разных народов не могло возникнуть самостоятельно и параллельно, а было лишь заимствовано. Можно всегда найти, где явление возникло впервые и откуда распространилось.

Сами по себе диффузии, заимствования и миграции — историческая реальность, и никто не может оспорить их роли в развитии человечества. Ошибка диффузионистов в том, что они преувеличивают их значение как исторических факторов и преуменьшают таким образом возможность конвергентного развития сходных элементов культуры в различных обществах.

С. Кроуфорд попытался (в 1912г.) представить распространение в Англии бронзовых топоров, кубков и золотых лунниц в зависимости от географии страны. При этом он не накладывал археологические материалы на современные карты, а стремился реконструировать исторический ландшафт, представить себе, каким он был тысячи лет назад. В 1921г. Кроуфорд издал книгу «Человек и его прошлое», в которой суммировал возможности и технику географического метода в первобытной и древней истории.

Другой английский ученый, К. Фокс, применил идеи Кроуфорда при изучении узкого района вокруг Кембриджа и достиг блестящих результатов, наложив карту с археологическими находками на карты современной топографии и восстановленной древней растительности.

Фокс показал, что археологические объекты связаны с древним, а не с современным ландшафтом. Первичная область расселения в неолите и бронзовом веке была связана с легкими песчаными и лёссовыми почвами; с появлением железных орудий и тяжелого плуга осваиваются жирные глинистые почвы. Важнейшей работой Фокса была впервые изданная в 1932г. и с тех пор неоднократно переиздававшаяся «The Personality of Britain», в которой он еще более ярко показал, что сопоставление археологических карт с картами реконструированного древнего ландшафта представляет собой не только метод иллюстрации, но и орудие научного исследования.

3. Анализ данных в работах сектора археологической теории и информатики

До недавнего времени центром разработки оригинальных методов математической статистики и анализа данных был сектор археологической теории и информатики ИАЭТ СО РАН [Деревянко, Холушкин и др., 1995].

Следует подчеркнуть, что попытка установить какую-то общую схему относительно методов и форм статистической обработки и анализа данных, годную во всех случаях, очевидно, несостоятельна. Однако существует определенная «естественная» логика в отношении использования тех или иных методов и эта логика может служить важным подспорьем в проектировании схемы обработки данных в каждом конкретном случае.

Относительно природы статистических выводов существует целый ряд концепций, из которых, на наш взгляд, наиболее существенны две: концепция рандомизации – принятия решения в условиях неопределенности и концепция редукции (свертки) информации.

Содержательный смысл этих концепций можно пояснить в очень простой форме.

Концепция рандомизации объявляет сутью статистических методов реконструкцию целого на основе неполной информации о нем, а в соответствии с концепцией редукции – представление больших объемов данных на основе выявления внутренних структурированностей, присущей этой совокупности данных.

Всю совокупность статистических методов обработки и анализа данных можно представить в виде нескольких групп. В каждой отдельной группе находится совокупность однородных или близких по структуре и результатам методов. Группы могут быть упорядочены по степени редукции (свертка, сжатие, обобщение) информации [Деревянко, Фелингер, Холушкин, 1989: с. 153-154]. В соответствии с таким подходом могут быть рассмотрены следующие группы методов обработки данных:

- вычисление эмпирических распределений по различным классам событий, формулируемых в терминах значений признаков описания объектов статистической совокупности;
- вычисление характеристик распределений (среднее, мода, медиана, дисперсия, энтропия и т.п.);
- анализ связей и зависимостей между признаками (корреляция парная, частная, множественная, другие меры и коэффициенты связи, регрессионный анализ восстановления зависимостей и т.п.);
- снижение размерности многомерных данных (факторный анализ и методы ему подобные);
- анализ структуры данных (методы и алгоритмы распознавания, кластерный анализ).

Продвинутый статистический анализ археологических данных предполагает взаимосвязанную обработку данных на всех перечисленных уровнях.

Для данных, сущностную основу которых представляют количественные (числовые) значения, имеется большое разнообразие конкретных алгоритмов для каждого уровня.

Привлечение математики коренным образом преобразило статистический подход. Теперь мы можем делать математически точные утверждения, располагая принципиально неточными данными. Это дает нам право рассматривать математическую статистику, как универсальный мост через, казалось бы, непреодолимую пропасть, разделяющую эмпирическую и теоретическую науку. Это становится возможным благодаря тому, что математическая статистика использует не формальную, а вероятностную логику. Формальная логика пригодна только для теоретической науки, а вероятностная, оставаясь предельно строгой, как и положено логике, в то же время позволяет работать с недетерминированными, случайными величинами. Вероятностная логика замыкает на себя противоречие между точностью теоретической формы и размытостью эмпирического содержания.

Предварительный анализ данных

Прежде чем приступить к глубокому статистическому анализу данных, полезно ознакомиться с каждым элементом данных по отдельности. Будем предполагать, что данные представлены таблицей «объект-свойство», то есть каждый объект, будь то отдельная находка или целый слой археологического памятника, характеризуется некоторым набором свойств (признаков), в совокупности составляющих его описание.

Например, свойствами слоя могут быть различные типы артефактов или фаунистические остатки, обнаруженные в этом слое. Для находок, например, каменных артефактов, это может быть материал и количественное и качественное описание артефактов. Каждое свойство обладает своей описательной силой, зависящей от двух моментов: во-первых, насколько разнообразны его значения и, во-вторых, насколько существенные стороны предмета исследования оно отражает. Если для всех объектов рассматриваемое свойство имеет одно и то же значение, то его информативность в точности равна нулю, поскольку анализ данных не позволяет извлекать информацию из констант. Но даже если на всех объектах значения свойства будут различными, это еще не гарантирует его информативности, поскольку не все, что измеряется, имеет какой-либо смысл.

Описательная статистика помогает оценить разброс значений признаков путем построения гистограмм частот, распределений и различных статистик.

В статистике наиболее сильные (т.е. наиболее общие и важные) выводы можно сделать относительно «хороших» переменных. Такими переменными являются количественные, нормально распределенные случайные величины. С помощью описательной статистики исследователь может выяснить, насколько его данные близки к идеалу. Но даже если эти данные далеки от идеала, то в статистике всегда найдутся средства, чтобы сделать обоснованные выводы из их анализа.

Первая задача исследователя, приступающего к статистическому анализу, состоит в определении для каждого признака типа шкалы, в которой он измерен. Для этого достаточно различать три шкалы: номинальную, порядковую и количественную:

Значение переменной в номинальной шкале – это просто имя. Единственная информация, которую можно извлечь из сравнения двух значений в номинальной шкале – совпадают они или нет.

Более сильная (информативная) шкала – порядковая. Здесь кроме совпадения, можно определить, какое из значений больше, а какое меньше.

И, наконец, самая сильная шкала – количественная. Здесь появляется дополнительная информация: насколько велика разность сравниваемых значений. При ее использовании мы от чисто логической информации переходим к информации количественной.

Недостаток информации, содержащейся в значениях, зафиксированной с помощью номинальной и порядковой шкал, приходится компенсировать увеличением числа сравниваемых элементов (увеличением объемов выборки). А при равных объемах выборок статистические выводы для сильной шкалы получаются более определенными.

В задачу предварительного анализа входит проверка корректности данных. Ошибку в данных легче увидеть на графике, чем в таблице. Например, для количественной переменной ошибки (опечатки) часто проявляются в виде выпадающих значений, отстоящих на значительном расстоянии от основной массы значений.

Другой, не менее важной задачей предварительного анализа данных является поиск ответа на вопрос, обладает ли какой-либо (явной или скрытой) структурой анализируемая таблица данных. Достаточно простым и эффективным средством является «серый» (или «спектральный») анализ (Рис.11). Его суть состоит в том, что анализируемая таблица дополняется графической схемой, которая представляет собой образ таблицы в виде прямоугольника, разделенного на ячейки, подобно клеткам исходной таблицы. При «сером» анализе каждая клетка схемы заполняется (заливается) оттенком серого цвета в зависимости от того, какие значения принимает соответствующий признак для данного объекта. Предварительно промежуток, в который попадают числовые значения всех признаков, разбивается на конечное число равных интервалов. Каждому интервалу сопоставляется определенный оттенок серого цвета по правилу – чем больше значения признаков, которые попадают в данный интервал, тем темнее окрашиваются в серый цвет соответствующие клетки таблицы. Результатом серого анализа является наглядный образ данных, где их структура представлена наиболее отчетливо.

Таким образом, графическая схема выглядит как своего рода плоская географическая карта, выполненная оттенками серого цвета, чем и объясняется название соответствующего метода анализа данных. На аналогичном принципе построен метод анализа с помощью оттенков разного цвета («спектральный» анализ), при котором данные таблицы представляются некоторой палитрой разных цветов. Этот метод дает еще более наглядную картину.

Явная структура обычно обнаруживается при взгляде на графическую схему, если на ней контрастно выделяются зоны (области) сгущений и разреженностей. В зонах сгущений (кластерах) концентрируются клетки с заметными (существенными) значениями признаков. В зонах разреженностей значения признаков представлены малыми (или нулевыми) значениями признаков.

Для выявления скрытой структуры требуется соответствующее преобразование исходной таблицы данных, достигаемое с помощью перестановки строк и (или) столбцов.

Рассмотрим пример данных, взятый из [Деревянко, Холушкин, Воронин, Костин, 2004: с.11-15]. На рис. 10 приведена таблица, в которой затененная верхняя строка фиксирует номера орудийных комплексов, а левый затененный столбец – номера археологических памятников, на которых эти орудийные комплексы были найдены. Соответственно на пересечении строк и столбцов указано количество находок. Нулевые ячейки (означающие, что данные пропущены или соответствующие орудия на памятниках не найдены) не заполнены с той целью, чтобы значимые данные были более заметны.

[illegible]

Рис. 10. Орудийные комплексы среднепалеолитических памятников Ближнего и Среднего Востока и Кавказа.

При беглом взгляде на рис. 10 видно, что если эта таблица и имеет какую-либо структуру, то она (эта структура) для наблюдателя представляется скрытой. Поэтому, упорядочивая таблицу перестановкой строк и столбцов, добиваемся того, чтобы эта структура обнаружилась. В наглядной форме эта структура представлена с помощью «серого» анализа (рис. 11).

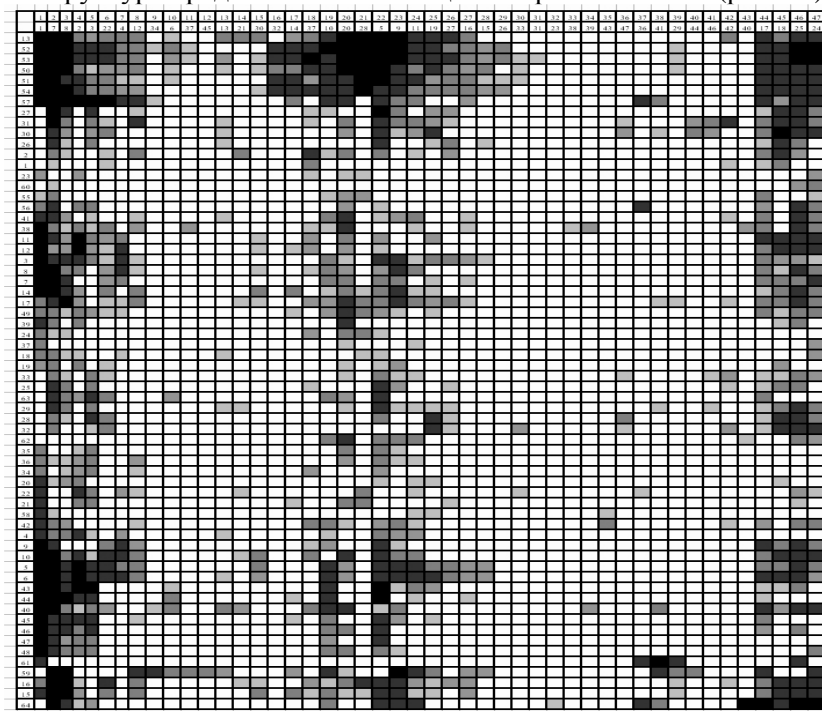


Рис. 11. Выявленная с помощью «серого» анализа структура данных.

Для этой цели перед предварительным упорядочиванием сначала были пронормированы значения признаков таблицы так, чтобы каждому исходному значению был поставлено в

соответствие значение его ранга в таблице (ранговая статистика). Затем построенная таким образом таблица ранговых статистик была упорядочена перестановкой строк и столбцов. Таким образом, было построено наглядное представление структуры данных, выявленное в результате их «серого анализа» (рис. 11).

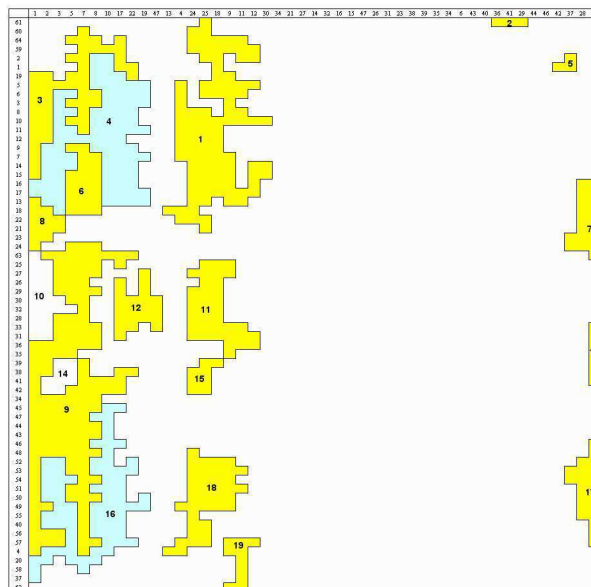


Рис.12. Выделение смежных областей.

Перестановка строк и столбцов при упорядочении таблицы ранговых статистик основывалась на использовании следующей идеи. Как правило, хорошо структурированной таблицей является та, в которой не очень часто происходят скачки по величине значений соседних элементов. Поэтому при перестановке строк и столбцов таблицы ранговых статистик эти данные были упорядочены по строкам и соответственно по столбцам таким образом, чтобы суммы расстояний между соседними элементами стали минимальными. Благодаря такой перестановке, строки, соответствующие памятникам, оказались упорядоченными по близости их распределений по артефактам.

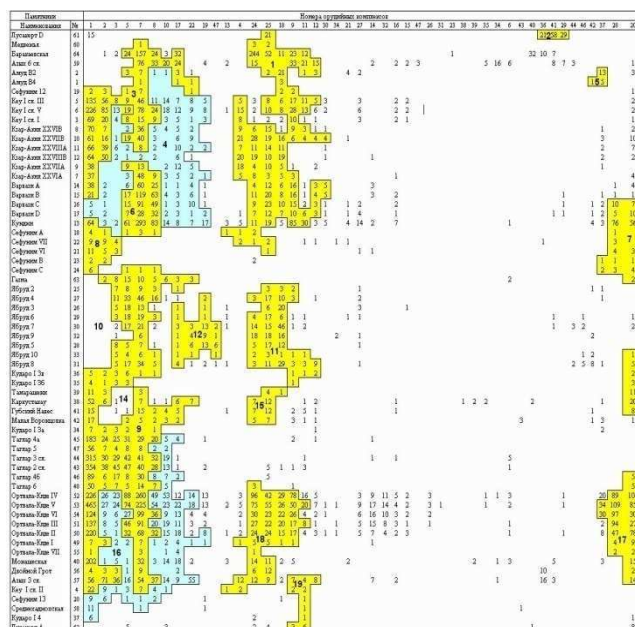


Рис. 13. Наполнение смежных областей

Полученная картина показала, что данные таблицы на самом деле обладают некоторой структурой, и, таким образом, имеет смысл с помощью статистических методов анализа исследовать ее более тщательно.

Для этой цели данные таблицы на рис.10 (а не их ранговые статистики) были вновь упорядочены по строкам и соответственно по столбцам. Причем, упорядочение данных исходной таблицы выполнялось по такой же схеме, что и упорядочение таблицы ранговых статистик.

На основе этого упорядочения было произведено разбиение матрицы на существенные с точки зрения информативности области.

Было выделено 20 связных областей.

Для визуализации результатов выделения каждая из смежных областей была окрашена единым оттенком серого цвета. В итоге структура данных исходной таблицы предстала еще более отчетливой (Рис. 12).

Наконец, завершается предварительный анализ данных визуализацией наполненности связных областей числовыми данными (рис. 13).

Исследование связей и зависимостей

Следующим шагом анализа является обнаружение взаимосвязей и зависимостей между признаками. В статистике разработано множество критериев для проверки наличия связей в данных. Но все они построены по одному принципу: в каждом критерии формулируется своя нулевая гипотеза, которая утверждает, что исследуемые признаки являются независимыми случайными величинами, связь между которыми если и проявляется, то исключительно в силу случайного совпадения. Проверка любого критерия начинается с вычисления своей статистики – величины, характеризующей степень отклонения от независимости. Вычисляемая статистика является количественной переменной и подчиняется в условиях выполнения нулевой гипотезы определенному распределению, которое может быть аналитически рассчитано или аппроксимировано программой. Таким образом, значение статистики переводится в так называемую значимость, которая является не чем иным, как вероятностью наблюдения полученного значения этой статистики при выполнении нулевой гипотезы. Если эта вероятность ниже некоторого заранее выбранного порога, например, 5%, то исследователь имеет основания утверждать, что нулевая гипотеза не подтверждается на его данных, из чего с большой вероятностью следует вывод, что между признаками обнаруживается определенная связь.

Поскольку признаки могут быть измерены в любой из трех шкал (номинальной, порядковой и количественной), то для каждого сочетания шкал надо применять свои критерии. Например, если обе переменные измерены в шкале наименований, то можно применять критерий Хи-квадрат, если одна из них – номинальная, а другая – количественная, то можно пользоваться дисперсионным анализом, а если обе количественные, то подойдет корреляция по Пирсону. При сочетании порядковой и количественной переменных приходится огрублять количественную переменную до порядковой и применять методы ранговой корреляции.

Для обычного исследователя, не владеющего в совершенстве методами анализа данных, одна из основных трудностей при работе с пакетами статистических программ заключается в сложности ориентации среди большого числа предлагаемых методов и критериев, которые в основном носят имена их авторов. Чтобы разобраться в том, а для решения каких именно задач применяется тот или иной метод, необходимо самостоятельно изучать специальную литературу.

В данной разработке для преодоления этого препятствия предполагается упорядочить методы по решаемым задачам и условиям (например, сочетаниям шкал признаков). Кроме того, вместо теоретических соображений в пользу того или иного критерия, мы можем привести практические доводы. А точнее – непосредственно на данных пользователя проверять применимость каждого критерия и рекомендовать лишь те, которые покажут свою работоспособность прямо «на глазах изумленной публики». Для проверки работоспособности критерия связи необходимо смоделировать условия нулевой гипотезы об отсутствии связи между переменными. Это можно сделать, разрушив связь между переменными, для чего достаточно перемешать любую из них, то есть переставить ее значения в случайном порядке, не изменяя самих значений. Многократное применение критерия к перемешанным данным должно давать случайные результаты – значимость при этом должна быть равномерно распределена в интервале от нуля до единицы. Если распределение значимости не будет равномерным, значит, критерий на предоставленных данных не работает. Если все критерии откажутся работать, то такие данные подлежат отбраковке как не пригодные к статистическому анализу связей.

Дополнительным сервисом системы может быть проверка чувствительности критериев к связям разного рода – линейным и нелинейным. Для такой проверки необходимо найти способ моделировать связи между переменными путем неслучайного перемешивания.

Методы снижения размерности

К методам снижения размерностей относятся факторный анализ (метод главных компонент) и многомерное шкалирование.

Эти методы позволяют из многих десятков малоинформативных признаков построить несколько высокоинформативных факторов, содержащих «отжатую» информацию, неравномерно разбросанную по исходным признакам. На примерах можно убедиться, что оба этих метода дают близкие результаты, но метод главных компонент алгоритмически более простой и эффективный, поэтому мы в дальнейшем сосредоточимся на нем.

Факторный анализ, кроме снижения размерности, дает косвенную возможность исследовать связи многих признаков, ибо факторы можно представить в виде линейной комбинации исходных признаков. Те признаки, которые входят с наибольшими коэффициентами в разложение факторов, образуют группы высококоррелированных признаков (см. табл.1 – пример взят из монографии) [Деревянко, Холушкин, Воронин, Костин, 2004: с.25].

Здесь из 47 признаков (типов орудий) предварительно выделены 7 главных компонент, по факторным нагрузкам которых из множества типов орудий были отобраны 19 типов как наиболее значимых. Методом вращения Varimax, при котором максимизируется дисперсия (разброс) факторных нагрузок каждого компонента на исходные переменные, получена матрица нагрузок этих переменных на отобранные главные компоненты (см. табл. 1).

Таким образом, факторный анализ за один проход выделяет все группы связанных переменных. Этот побочный результат данного метода часто оказывается полезен при изучении связей.

Поскольку факторы представляют линейную комбинацию исходных переменных, то полный набор факторов содержит в точности то же количество информации, что и набор исходных признаков. Сущность метода можно понять, представив себе данные как точки в пространстве с размерностью, равной числу исходных переменных. Точки в этом многомерном пространстве сосредотачиваются в некотором компактном облаке рассеяния. Поиск главных компонент сводится к такому вращению системы координат, при котором вдоль первого фактора наблюдается наибольший разброс точек, вдоль второго – меньше и далее по убыванию.

Таблица 1. Факторные нагрузки на типы орудий.

№ типов орудий	Типы орудий	Факторные нагрузки на типы орудий						
		1	2	3	4	5	6	7
7	Продольные скребла	0.880	0.196	0.041	-0.006	0.154	-0.140	0.070
8	Двойные скребла	0.815	-0.209	-0.092	0.055	-0.133	-0.021	0.056
9	Конвергентные скребла	0.671	-0.102	-0.093	-0.045	0.026	0.476	-0.057
25	Зубчатые	-0.050	0.891	0.074	0.027	-0.079	0.003	0.068
24	Выемчатые	-0.096	0.738	-0.146	-0.337	-0.001	-0.083	-0.014
1	Леваллуазские сколы	-0.404	-0.507	-0.459	-0.176	-0.218	0.023	0.145
2	Леваллуазские острья	-0.351	-0.433	-0.229	-0.132	-0.247	-0.132	0.305
18	Резцы	-0.016	0.182	0.838	0.085	-0.116	-0.099	0.036
37	Разные	-0.133	-0.293	0.680	-0.183	0.060	-0.175	-0.289
22	Усеченные отщепы	0.044	-0.040	0.608	-0.024	0.198	-0.008	0.369
19	Проколки	-0.119	0.367	0.432	0.015	-0.287	0.152	0.206
5	Мустьерские острья	0.186	0.004	0.049	0.871	-0.012	0.053	-0.052
3	Леваллуазские ретушированные острья	-0.173	-0.139	-0.088	0.855	-0.067	-0.156	0.049
17	Скребки	-0.027	0.062	0.130	-0.145	0.869	-0.137	-0.023
10	Угловатые скребла	0.112	-0.142	-0.188	0.107	0.673	0.471	-0.037
11	Поперечные скребла	-0.025	0.020	-0.096	-0.092	-0.011	0.885	-0.117
20	Ножи	-0.252	-0.037	-0.141	-0.036	0.185	0.272	-0.654
4	Псевдолеваллуазские острья	-0.158	-0.119	-0.136	-0.259	0.221	-0.178	0.623
28	Сколы, ретушированные со спинки	0.005	-0.142	-0.167	-0.273	0.162	-0.340	-0.613

Но дело в том, что случайные величины всегда содержат шум, вызванный ошибками сбора данных и случайными отклонениями в параметрах объектов. Природа этих ошибок может быть самой разной, но они всегда есть. Значит, и полный набор факторов также содержит шум. Главные компоненты, начиная с фактора номер один, характеризуются наибольшим отношением сигнала к шуму. Чем больше номер фактора, тем меньше полезной информации он содержит. При

некотором критическом номере фактора уровень шума становится выше уровня сигнала. Этот и все последующие факторы должны быть отброшены.

Определить количество факторов, которые надо отбросить, достаточно просто. Для этого воспользуемся тем же подходом, что при проверке применимости критериев связи. Действительно, мы можем сформулировать нулевую гипотезу: все исходные переменные независимы друг от друга. Промоделировать выполнение нулевой гипотезы не представляет труда – достаточно перемешать каждую исходную переменную (может быть, кроме одной) и повторить расчет факторов. Все полученные таким образом факторы, начиная с первого, не содержат никакой информации о связях, а только статистический шум, поскольку это гарантировано выполнением условий нулевой гипотезы. Такой статистический эксперимент можно провести многократно. Если теперь сравнить факторы, полученные по исходным данным с теми, что получились в результате экспериментов с перемешиванием, мы сможем определить количество факторов, действительно содержащих больше полезной информации о связях признаков, чем статистического шума.

Сжатие информации с помощью главных компонент является часто подготовительным этапом для структурного анализа, к рассмотрению которого мы и переходим.

Анализ структур

Если анализ связей выявляет признаки, значения которых согласованно изменяются от объекта к объекту, то анализ структур выявляет объекты, на которых согласованы значения определенного набора признаков.

№	Комплексы	IF	IF _{st}	IL	IL _{am}
1	Семиганч	37.10	14.20	20.20	30.70
24	Оби-Рахмат 15-18	-	-	-	35.50
25	Оби-Рахмат 10-14	-	-	-	34.40
9	Чингиз	54.00	29.00	41.10	0 23.00
11	Актогай	48.60	8.60	44.00	6.60
13	Семизбугу В	57.60	15.20	9.80	2.80
12	Семизбугу А	1 50.80	6.00	9.80	2.80
5	Георгиевский Бугор	44.30	8.80	9.50	10.80
18	Кутурбулак	50.70		2.40-	20.00
19	Зирабулак	36.90	27.30	2.00	3.10
10	Кош-Курган	42.30	13.70	3.90	6.10
16	Хантау	3 35.70	3.60	3.80	2.10
17	Бурма	26.80	9.80	0.80	2.10
6	Кара-Бура	23.60	11.50	4.20	6.30
22	Кызыл-Тау пл.2 сдф	22.00	15.00	2 3.30	0.60
20	Кызыл-Тау пл.1 сдф	15.80	11.80	7.10	0.10
21	Кызыл-Тау пл.1 слдф	9.80	7.50	0.00	0.40
15	Семизбугу D	14.30	4.80	6.80	4.10
14	Семизбугу C	21.40	11.90	11.40	7.00
3	Тоссор	25.60	4.40	12.20	5 20.60
1	Хонако 3	4 24.40	9.00	14.00	31.00
2	Худжи	27.00	6.00	14.50	6 62.00
8	Огзи-Кичик	35.90	4.60	32.70	44.00
26	Оби-Рахмат 6-9	-	-	-	43.40
27	Оби-Рахмат 2-5	-	-	-	7 43.90
23	Оби-Рахмат 19-21				44.20
4	Джар-Кутан	55.00	41.00	37.80	53.00

Рис. 14. Результат работы метода прямого кластерного анализа.

Таблица 2. Дисперсионный анализ областей.

Область	Среднее	Ср.кв.откл.	Объем	Объясняет долю дисперсии (в %)	
				область	Элемент
0	30.920	9.469	10	4.4	0.440
1	51.000	4.546	6	21.7	3.618
2	7.356	4.844	52	32.5	0.624
3	30.657	7.647	7	2.9	0.419
4	26.860	5.459	5	0.9	0.171
5	25.800	7.354	2	0.2	0.121
6	62.000	0.000	1	6.7	6.661
7	43.889	6.889	9	19.3	2.141

Комментарий к таблице 2: Среднеквадратичное отклонение по таблице составляет 16.99 %. Объясненная дисперсия составляет 88.5 %.

Промежуточное место между анализом связей и анализом структур занимает метод прямого кластерного анализа, разработанный П.С.Ростовцевым [Ростовцев, 1985]. Этот метод позволяет непосредственно на таблице объект-признак выделить, после переупорядочения строк и столбцов, области неправильной формы, выделяющиеся близкими значениями признаков. Каждая такая область объединяет несколько объектов и несколько признаков в «пятно». Попытки проинтерпретировать наблюдаемую картину областей могут натолкнуть исследователя на новые, интересные гипотезы. Пример применения прямого кластерного анализа взят из [Деревянко, Холюшкин, Воронин, Костин, 2005б: с.45-46] (см. табл. 2 и рис. 14).

В частности, рассмотрение структуры областей на рис. 14 приводит к выводу о том, что хорошими признаками для выделения структурной информации являются IF и I_{Lam}. Как для всех вышерассмотренных методов, здесь также имеется возможность сформулировать и проверить нулевую гипотезу об отсутствии выделенных областей. В качестве статистики для измерения отклонений от независимости можно выбрать долю объясненной дисперсии (в нашем примере 88.5%), также называемой коэффициентом детерминации R^2 . Такое дополнение превращает метод из эвристического в статистический метод. А это позволяет использовать полученные результаты не только для научного поиска, но и делать вполне обоснованные утверждения.

К методам анализа структур относятся в первую очередь методы автоматической классификации, среди которых наиболее распространен кластерный анализ. Для проведения кластерного анализа необходимо выбрать несколько признаков – построить так называемое признаковое пространство. Задача состоит в том, чтобы выделить в этом пространстве отдельные сгущения точек – кластеры.

Разнообразные типы кластерного анализа активно применялись в археологических исследованиях. В них авторами осознавались недостатки процедур кластерного анализа, главными из которых являются два: отсутствие четких рекомендаций по выбору числа классов и невозможность индивидуального учета отдельных элементов при объединении классов [Деревянко, Холюшкин, Воронин, Костин, 2005а: с.5].

В ходе таких исследований было обнаружено, что кластеры, замечательным образом найденные в первый раз и разумно описанные исследователем, после повторного сбора информации (новых раскопок и нового применения кластерного анализа) могут «рассыпаться» из-за случайности выявленной кластерной структуры (при малых выборках, ненормальных распределениях, плохо обусловленных моделях и т.д.).

С этой целью используется метод повторной выборки с возвращением, известный как метод boot-strap [Efron, 1986]. К сожалению, этот метод мало затронул археологию.

Так Кинти [Kintigh, 1984] использовал выборку методом Монте-Карло, чтобы генерировать псевдодоверительные интервалы для результатов анализа многообразий k-значной кластеризации пространственных данных. Рингроуз [Ringrouse, 1992] использовал boot-strap для оценки подобным способом результатов анализа соответствия. Суть метода состоит в имитации повторного сбора данных, в ходе которой генерируется выборка, совпадающая с исходными данными [Деревянко, Холюшкин, Воронин, Костин, 2005а: с.5-6].

Существует множество алгоритмов для выделения сгущений, основанных на различных подходах. Мы выбрали один из простейших и вычислительно наиболее эффективных методов кластерного анализа – метод k-средних. Согласно этому методу, принадлежность объекта к кластеру определяется евклидовым расстоянием между объектом и центром кластера. Объект приписывается к ближайшему кластеру. Процедура начинается с некоторого начального приближения, а затем запускается итерационный процесс, на каждом шаге которого объекты перемещаются между кластерами, что приводит к изменению координат центров кластеров.

Итерации продолжаются до тех пор, пока объекты не перестанут перебегать из одного кластера в другие. При этом достигает своего минимального значения оптимизируемый функционал – остаточная дисперсия, которая вычисляется как сумма квадратов отклонений координат объектов от центров своих кластеров.

И в этом случае имеется возможность сформулировать и проверить нулевую гипотезу, которая звучит так: в признаковом пространстве точки рассеяны так, что образуют единственный кластер [Жданов, Костин, 2002]. Правда, в этом случае дело обстоит сложнее, чем во всех предыдущих, поскольку разрушить кластерную структуру, оставив в то же время в нетронutom виде связи между переменными, намного сложнее, чем просто разрушить связи. Простое перемешивание признаков здесь не подходит. Приходится вводить дополнительное предположение о том, что точки в пространстве признаков имеют многомерное нормальное

распределение, которое не всегда выполняется. Но если мы принимаем такое допущение, то далее обычным путем генерируем данные с разрушенной кластерной структурой, получая экспериментальным путем распределение значений оптимизируемого функционала в условиях нулевой гипотезы об отсутствии многокластерной структуры. По значимости нулевой гипотезы можно обоснованно сказать, наблюдается ли кластерная структура на нашей выборке. Более того, по наименьшей значимости можно даже определить наиболее вероятное количество кластеров.

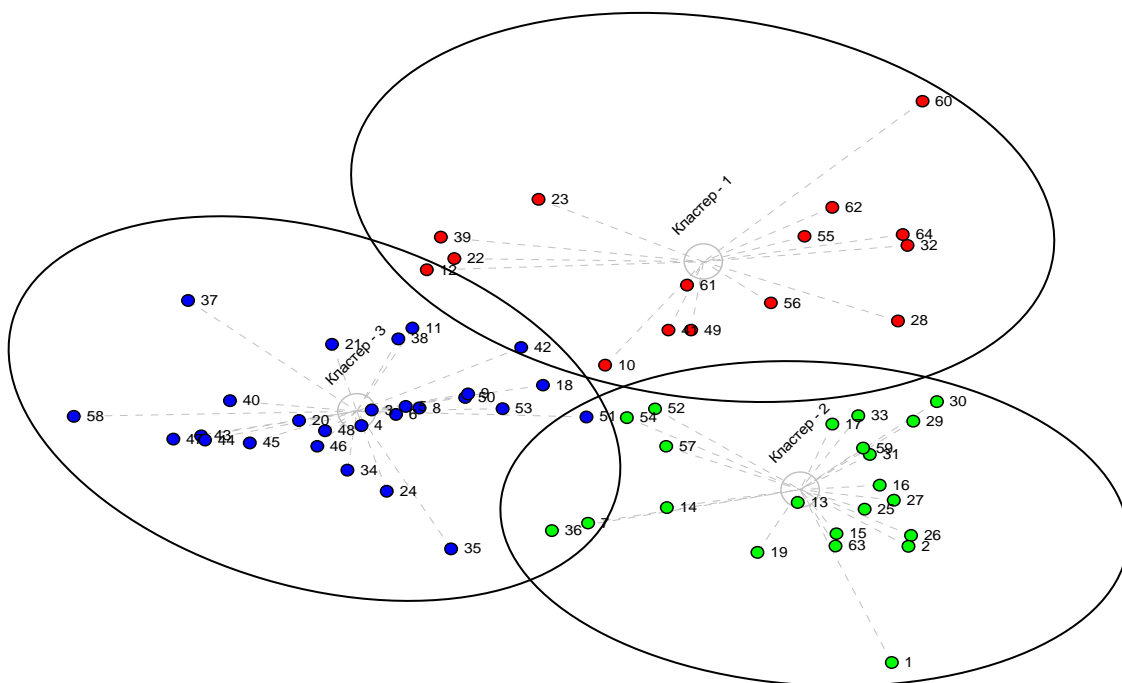


Рис. 15. Результат работы кластерного анализа методом k-средних

Но даже возможность получить оптимальную кластерную структуру не позволяет говорить о том, что мы действительно можем извлечь из данных содержащуюся там структурную информацию. Трудность состоит в том, что методы кластерного анализа хорошо работают при небольшой размерности признакового пространства (2-3), а выбор наиболее информативного подпространства признаков превращается в неподъемную переборную задачу. К тому же появляется проблема сравнения результатов классификаций и выбора наилучшей из них.

Задача сравнения классификаций была поставлена при анализе совпадения классификаций (см. рис. 15), построенных на основе данных по типологии орудий среднепалеолитических индустрий Ближнего и Среднего Востока и Кавказа [Деревянко, Холушкин, Воронин, Костин, 2004: с.25-66].

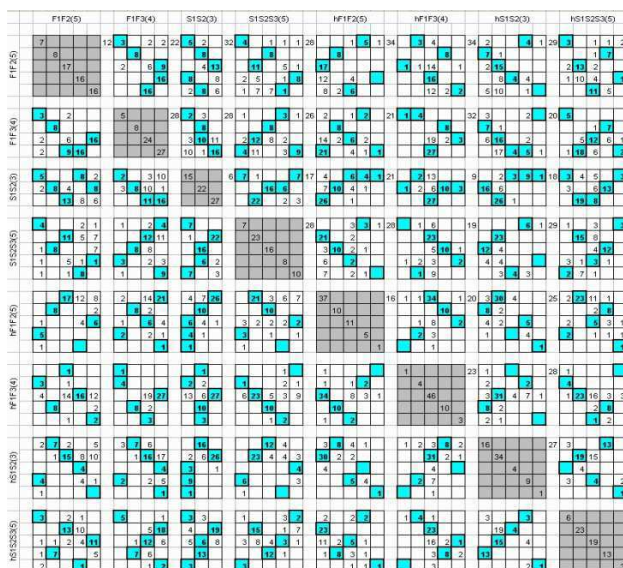


Рис. 15. Покластерное совпадение кластеров, построенных по разным классификациям.

Для каждой классификации были выбраны наиболее значимые переменные, позволившие построить наиболее отчетливые группировки:

1) классификация, построенная методами *k*-средних в пространстве факторов 1, 2 – выделено 5 кластеров;

2) классификация, построенная методами *k*-средних в пространстве факторов 1, 3 – выделено 4 кластера;

3) классификация, построенная методами *k*-средних в пространстве шкал 1, 2 [Деревянко, Холушкин, Воронин, Костин, 2004: с.36] – выделено 3 кластера;

4) классификация, построенная методами *k*-средних в пространстве шкал 1, 2, 3 [Деревянко, Холушкин, Воронин, Костин, 2004: там же] – выделено 5 кластеров;

5) классификация, построенная методами иерархического кластерного анализа в пространстве факторов 1, 2 (выделено 5 кластеров);

6) классификация, построенная методами иерархического кластерного анализа в пространстве факторов 1, 3 (выделено 4 кластера);

7) классификация, построенная методами иерархического кластерного анализа в пространстве шкал 1, 2 (выделено 3 кластера);

8) классификация, построенная методами иерархического кластерного анализа в пространстве шкал 1, 2, 3 (выделено 5 кластеров).

По этим классификациям построены попарные классификации для перечисленных разбиений. Результаты анализа совпадений кластеров при попарном сравнении состава их элементов приведены в таблицах 3, 4.

Таблица 3. Совпадение кластеров, построенных по разным классификациям (в %).

Исходные классификации	Сопряженные классификации							
	1	2	3	4	5	6	7	8
1	100.0	81.3	65.6	50.0	56.3	46.9	46.9	54.7
2	81.3	100.0	56.3	56.3	59.4	67.2	50.0	68.8
3	65.6	56.3	100.0	90.6	73.4	67.2	85.9	71.9
4	50.0	56.3	90.6	100.0	56.3	56.3	70.3	54.7
5	56.3	59.4	73.4	56.3	100.0	75.0	68.8	60.9
6	46.9	67.2	67.2	56.3	75.0	100.0	64.1	56.3
7	46.9	50.0	85.9	70.3	68.8	64.1	100.0	57.8
8	54.7	68.8	71.9	54.7	60.9	56.3	57.8	100.0

Таблица 4. Несовпадение кластеров, построенных по разным классификациям (количество объектов, не попавших в общие кластеры).

Исходные классификации	Сопряженные классификации							
	1	2	3	4	5	6	7	8
1		12	22	32	28	34	34	29
2	12		28	28	26	21	32	20
3	22	28		6	17	21	9	18
4	32	28	6		28	28	19	29
5	28	26	17	28		16	20	25
6	34	21	21	28	16		23	28
7	34	32	9	19	20	23		27
8	29	20	18	29	25	28	27	

В.С.Костиным был предложен вариант решения, где в качестве статистики, измеряющей степень отклонения от независимости классификаций, выбрана максимальная доля совпадающих объектов при оптимальном соответствии кластеров [Костин, 2003]. А задачу выбора наилучшей классификации была трансформирована в задачу объединения результатов большого количества независимо построенных классификаций [Костин, Корнюхин, 2003] и построения на основе этого объединения классификации обобщенной. Наглядное представление обобщенной классификации предлагается на рис. 16. В клетках представленной на этом рисунке таблицы указана степень согласованности включения объектов в одни и те же кластеры. Более темным оттенкам серого соответствует высокая степень согласованности, более светлым – менее высокая.

Оценивая весь оригинальный (не полностью описанный из-за ограничений на объемы статьи в настоящем издании) инструментарий, следует указать наиболее важную особенность методологии и методики, на которой он выстроен: все процедуры и методы завершаются обязательной проверкой статистической значимости полученных результатов.

Следует заметить, что непрерывное пополнение описанного нами выше инструментального арсенала для статистического анализа археологических данных продолжается.

Дальнейшим шагом в этом направлении будет получение не одной, а нескольких обобщенных кластерных структур. Проблема состоит в том, что весь набор признаков, описывающий исследуемые объекты, как правило, отражает их с самых разных сторон, выделяя разные предметы исследования одного и того же объекта, разные уровни и формы организации и развития мира [Витяев, Костин, 1992].

За каждым предметом исследования стоит свой класс аспектов. Если же мы смешаем все классификации аспектов в одну, то получим либо необоснованно переусложненную картину кластеров (пересечение аспектных или предметных классификаций), либо не получим отчетливой картины вовсе.

Решение следует искать в разделении всего набора предварительно проведенных аспектно-предметных классификаций на группы. Внутри каждой из подобных групп попарная близость классификаций друг к другу должна быть в среднем больше, чем близость классификаций из разных групп. Тогда на базе каждой из этих групп классификаций уже можно строить «чистую» обобщенную классификацию.

31	34	36	38	41	43	45	47	49	51	53	55	57	59	61	63	65	67	69	71	73	75	77	79	81	83	85	87	89	91	93	95	97	99	101	103	105	107	109	111	113	115	117	119	121	123	125	127	129	131	133	135	137	139	141	143	145	147	149	151	153	155	157	159	161	163	165	167	169	171	173	175	177	179	181	183	185	187	189	191	193	195	197	199	201	203	205	207	209	211	213	215	217	219	221	223	225	227	229	231	233	235	237	239	241	243	245	247	249	251	253	255	257	259	261	263	265	267	269	271	273	275	277	279	281	283	285	287	289	291	293	295	297	299	301	303	305	307	309	311	313	315	317	319	321	323	325	327	329	331	333	335	337	339	341	343	345	347	349	351	353	355	357	359	361	363	365	367	369	371	373	375	377	379	381	383	385	387	389	391	393	395	397	399	401	403	405	407	409	411	413	415	417	419	421	423	425	427	429	431	433	435	437	439	441	443	445	447	449	451	453	455	457	459	461	463	465	467	469	471	473	475	477	479	481	483	485	487	489	491	493	495	497	499	501	503	505	507	509	511	513	515	517	519	521	523	525	527	529	531	533	535	537	539	541	543	545	547	549	551	553	555	557	559	561	563	565	567	569	571	573	575	577	579	581	583	585	587	589	591	593	595	597	599	601	603	605	607	609	611	613	615	617	619	621	623	625	627	629	631	633	635	637	639	641	643	645	647	649	651	653	655	657	659	661	663	665	667	669	671	673	675	677	679	681	683	685	687	689	691	693	695	697	699	701	703	705	707	709	711	713	715	717	719	721	723	725	727	729	731	733	735	737	739	741	743	745	747	749	751	753	755	757	759	761	763	765	767	769	771	773	775	777	779	781	783	785	787	789	791	793	795	797	799	801	803	805	807	809	811	813	815	817	819	821	823	825	827	829	831	833	835	837	839	841	843	845	847	849	851	853	855	857	859	861	863	865	867	869	871	873	875	877	879	881	883	885	887	889	891	893	895	897	899	901	903	905	907	909	911	913	915	917	919	921	923	925	927	929	931	933	935	937	939	941	943	945	947	949	951	953	955	957	959	961	963	965	967	969	971	973	975	977	979	981	983	985	987	989	991	993	995	997	999	1001	1003	1005	1007	1009	1011	1013	1015	1017	1019	1021	1023	1025	1027	1029	1031	1033	1035	1037	1039	1041	1043	1045	1047	1049	1051	1053	1055	1057	1059	1061	1063	1065	1067	1069	1071	1073	1075	1077	1079	1081	1083	1085	1087	1089	1091	1093	1095	1097	1099	1101	1103	1105	1107	1109	1111	1113	1115	1117	1119	1121	1123	1125	1127	1129	1131	1133	1135	1137	1139	1141	1143	1145	1147	1149	1151	1153	1155	1157	1159	1161	1163	1165	1167	1169	1171	1173	1175	1177	1179	1181	1183	1185	1187	1189	1191	1193	1195	1197	1199	1201	1203	1205	1207	1209	1211	1213	1215	1217	1219	1221	1223	1225	1227	1229	1231	1233	1235	1237	1239	1241	1243	1245	1247	1249	1251	1253	1255	1257	1259	1261	1263	1265	1267	1269	1271	1273	1275	1277	1279	1281	1283	1285	1287	1289	1291	1293	1295	1297	1299	1301	1303	1305	1307	1309	1311	1313	1315	1317	1319	1321	1323	1325	1327	1329	1331	1333	1335	1337	1339	1341	1343	1345	1347	1349	1351	1353	1355	1357	1359	1361	1363	1365	1367	1369	1371	1373	1375	1377	1379	1381	1383	1385	1387	1389	1391	1393	1395	1397	1399	1401	1403	1405	1407	1409	1411	1413	1415	1417	1419	1421	1423	1425	1427	1429	1431	1433	1435	1437	1439	1441	1443	1445	1447	1449	1451	1453	1455	1457	1459	1461	1463	1465	1467	1469	1471	1473	1475	1477	1479	1481	1483	1485	1487	1489	1491	1493	1495	1497	1499	1501	1503	1505	1507	1509	1511	1513	1515	1517	1519	1521	1523	1525	1527	1529	1531	1533	1535	1537	1539	1541	1543	1545	1547	1549	1551	1553	1555	1557	1559	1561	1563	1565	1567	1569	1571	1573	1575	1577	1579	1581	1583	1585	1587	1589	1591	1593	1595	1597	1599	1601	1603	1605	1607	1609	1611	1613	1615	1617	1619	1621	1623	1625	1627	1629	1631	1633	1635	1637	1639	1641	1643	1645	1647	1649	1651	1653	1655	1657	1659	1661	1663	1665	1667	1669	1671	1673	1675	1677	1679	1681	1683	1685	1687	1689	1691	1693	1695	1697	1699	1701	1703	1705	1707	1709	1711	1713	1715	1717	1719	1721	1723	1725	1727	1729	1731	1733	1735	1737	1739	1741	1743	1745	1747	1749	1751	1753	1755	1757	1759	1761	1763	1765	1767	1769	1771	1773	1775	1777	1779	1781	1783	1785	1787	1789	1791	1793	1795	1797	1799	1801	1803	1805	1807	1809	1811	1813	1815	1817	1819	1821	1823	1825	1827	1829	1831	1833	1835	1837	1839	1841	1843	1845	1847	1849	1851	1853	1855	1857	1859	1861	1863	1865	1867	1869	1871	1873	1875	1877	1879	1881	1883	1885	1887	1889	1891	1893	1895	1897	1899	1901	1903	1905	1907	1909	1911	1913	1915	1917	1919	1921	1923	1925	1927	1929	1931	1933	1935	1937	1939	1941	1943	1945	1947	1949	1951	1953	1955	1957	1959	1961	1963	1965	1967	1969	1971	1973	1975	1977	1979	1981	1983	1985	1987	1989	1991	1993	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017	2019	2021	2023	2025	2027	2029	2031	2033	2035	2037	2039	2041	2043	2045	2047	2049	2051	2053	2055	2057	2059	2061	2063	2065	2067	2069	2071	2073	2075	2077	2079	2081	2083	2085	2087	2089	2091	2093	2095	2097	2099	2101	2103	2105	2107	2109	2111	2113	2115	2117	2119	2121	2123	2125	2127	2129	2131	2133	2135	2137	2139	2141	2143	2145	2147	2149	2151	2153	2155	2157	2159	2161	2163	2165	2167	2169	2171	2173	2175	2177	2179	2181	2183	2185	2187	2189	2191	2193	2195	2197	2199	2201	2203	2205	2207	2209	2211	2213	2215	2217	2219	2221	2223	2225	2227	2229	2231	2233	2235	2237	2239	2241	2243	2245	2247	2249	2251	2253	2255	2257	2259	2261	2263	2265	2267	2269	2271	2273	2275	2277	2279	2281	2283	2285	2287	2289	2291	2293	2295	2297	2299	2301	2303	2305	2307	2309	2311	2313	2315	2317	2319	2321	2323	2325	2327	2329	2331	2333	2335	2337	2339	2341	2343	2345	2347	2349	2351	2353	2355	2357	2359	2361	2363	2365	2367	2369	2371	2373	2375	2377	2379	2381	2383	2385	2387	2389	2391	2393	2395	2397	2399	2401	2403	2405	2407	2409	2411	2413	2415	2417	2419	2421	2423	2425	2427	2429	2431	2433	2435	2437	2439	2441	2443	2445	2447	2449	2451	2453	2455	2457	2459	2461	2463	2465	2467	2469	2471	2473	2475	2477	2479	2481	2483	2485	2487	2489	2491	2493	2495	2497	2499	2501	2503	2505	2507	2509	2511	2513	2515	2517	2519	2521	2523	2525	2527	2529	2531	2533	2535	2537	2539	2541	2543	2545	2547	2549	2551	2553	2555	2557	2559	2561	2563	2565	2567	2569	2571	2573	2575	2577	2579	2581	2583	2585	2587	2589	2591	2593	2595	2597	2599	2601	2603	2605	2607	2609	2611	2613	2615	2617	2619	2621	2623	2625	2627	2629	2631	2633	2635	2637	2639	2641	2643	2645	2647	2649	2651	2653	2655	2657	2659	2661	2663	2665	2667	2669	2671	2673	2675	2677	2679	2681	2683	2685	2687	2689	2691	2693	2695	2697	2699	2701	2703	2705	2707	2709	2711	2713	2715	2717	2719	2721	2723	2725	2727	2729	2731	2733	2735	2737	2739	2741	2743	2745	2747	2749	2751	2753	2755	2757	2759	2761	2763	2765	2767	2769	2771	2773	
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	--

- Городцов В.А.** Типологический метод в археологии. – Рязань, 1927: 125 с.
- Гражданников Е.Д.** Метод построения системной классификации наук. – Новосибирск, 1987.
- Григорьев П.П.** Палеолит Африки. // Палеолит Мира - Л., 1977: с.45-197.
- Грязнов М.П.** Классификация, тип, культура // Теоретические основы советской археологии. – Л., 1969: с.18-22.
- Гумилев Л.Н.** Этногенез и биосфера Земли. – СПб.: Азбука – Классика, 2002.
- Деревянко А.П., Фелингер А.Ф., Холушкин Ю.П.** Методы информатики в археологии каменного века. – Новосибирск, 1989а: 268 с..
- Каменецкий И.С., Маршак Б.И., Шер Я.А.** Анализ археологических источников. – М., 1975.
- Клейн Л.С.** Археологические источники. – Л.: ЛГУ, 1978: 119 с.
- Клейн Л.С.** Классификация типология // Исторические чтения памяти Михаила Петровича Грязнова. Тез. Докл. – Омск. 1987 : с.33-35.
- Клейн Л.С.** Археологическая типология. – Л., 1991: 448 с.
- Клейн Л.С.** Введение в теоретическую археологию: учебное пособие. Кн.1. – СПб: Бельведер, 2004: 469 с.
- Колпаков Е.М.** Теория археологической классификации. – СПб, 1991: 112 с.
- Костин В.С.** Статистика для сравнения классификаций // Информационные технологии в гуманитарных исследованиях. Вып. 6. – Новосибирск, 2003: с.57-65.
- Костин В.С., Корнюхин Ю.Г.** Построение обобщенной классификации // Информационные технологии в гуманитарных исследованиях. Вып. 6. – Новосибирск, 2003: с.65-72.
- Мартынов А.И., Шер Я.А.** Методы археологического исследования. – М., 1989: 224 с.
- Медведев И.И., Несмеянов С.А.** Типизация «культурных отложений» и местонахождений каменного века // Методические проблемы археологии Сибири. – Новосибирск: Наука, 1988.
- Мечников Л.И.** Цивилизация и великие исторические реки (Географическая теория прогресса и социального развития) – М., 1995.
- Монгайт А.Л.** Археология Западной Европы. Каменный век. – М., 1976.
- Равдоникас В.И.** За марксистскую историю материальной культуры // Известия ГАИМК. – Л., 1930. – Т. 7, вып.3–4: 94 с.
- Ростовцев П.С.** Черно-белый анализ связи переменных // Анализ и моделирование экономических процессов переходного периода в России. Новосибирск: ИЭиОПП, 1996. с.264-286
- Ростовцев П.С., Костин В.С., Корнюхин Ю.Г., Смирнова Н.Ю.** Анализ структур социологических данных. Устойчивость // Анализ и моделирование экономических процессов переходного периода в России. – Новосибирск: ИЭиОПП: 1997.С.174-208.
- Фаган Б., ДеКорс К.** Археология. В начале. – М.: Техносфера, 2007:–592 с.
- Федоров-Давыдов Г.А.** Понятие «археологический тип» и «археологическая культура» в «Аналитической археологии» Дэвида Кларка // СА, №3, 1970: с.258-270.
- Федоров-Давыдов Г.А.** Статистические методы в археологии. – М.: Высшая школа, 1987: 216 с.
- Фелингер А.Ф.** Статистические алгоритмы в социологических исследованиях. Новосибирск: Наука, СО АН СССР, 1985.
- Холушкин Ю.П.** Системная классификация понятия «поселенческая археология» // Методология и методика археологических реконструкций. – Новосибирск, 1994: с.102.
- Холушкин Ю.П.** Системная археология. Учебное пособие. – Новосибирск, 2010: 554 с.
- Шер Я.А.** Интуиция и логика в археологическом исследовании (к формализации типологического метода в археологии). // Статистико-комбинаторные методы в археологии. – М., 1970: 8–24.
- Шапова Ю.Л.** Естественно-научные методы в археологии. - М., 1988.
- Шапова Ю.Л.** Введение в вещеведение. – М.: Издательство МГУ, 2000: 144 с.
- Binford L. R.** An archaeological perspective. – NY, London, 1972: 464 p.
- Breizillon M.** La denomination des objets de pierre taillees. // Matériaux pour un vocabulaire de préhistoriens de langue française. – Paris, 1968: 413 p.
- For theory building in archaeology: Essays on faunal remains, aquatic resources, spatial analysis, and Systematic modeling** // Ed. by Binford L.R. – N.Y., 1977 – XVII: 419 p.
- Chang K.C.** Rethinking archaeology – N.Y.: Random House, 1967.
- Chenhall R.G.** The impact of computers on archaeological theory: an appraisal and projection // Computers and the Humanities, 1968, 3(1): p. 15-24.
- Clarke D.L.** Analytical archaeology. – L: Methuen, 1968: 684 p.
- Clarke D.L. (ed).** Models in archaeology. – L: Methuen, 1972: 1055 p.
- Doran, J.E. & Hodson, F.R.** Mathematics and Computer in archaeology.– Edinburg, 1975: 381 p.
- Dunnell R.C.** Systematics in prehistory. - N.Y., 1971.
- Efron B.** Better bootstrap confidence intervals.//J. Amer. Statist. Ass., 81, 1986
- Ford J.A.** The type concept revised // American Anthropologist. Vol. 56, 1954: p. 42-54.
- Ford J.A.** A quantitative method for deriving cultural chronology. – Washington, 1962.
- Hawkes C. S.C.** British prehistory half-way through the century. – Proceedings of the Prehistoric Society, V. 17: 1951: p. 1–15.
- Hill J.N, Evans R.K.** A model for classification and typology. // Models in archaeology. – L., 1972: p. 231–273.
- Jochim M.A.** Hunter-gatherer subsistence and Settlement. A predictive model. –NY, 1976.
- Kintigh K.** Measuring archaeological diversity by comparison with Simulated assemblages // American Antiquity, vol. 49, 1984: p. 44-54.
- Moore J.A., Keene A.S.** Archaeology and the Law of the Hummer // Archaeological hummers and theories. – N.Y., 1983, v.8. - (Studies in archaeology).
- Ringrouse T.,** Bootstaping and correspondenting analysis in archaeology // Journal of Archaeological Sciences. Vol. 19, 1992: p.615-629.
- Salmon M.H.** Philosophy and archaeology. – N.Y., 1982 – XI: 203.
- Spoulding A.S.** Statistical techniques for discovery of artifact types // American antiquity, 1953, v.18, № 4: p. 305–313.
- Trigger B.C.** Settlement archaeology – its goals and promise // American antiquity – 1967, V. 32, № 2.
- Whitley G.R,** Prehistoric settlement pattern in the Valley // Bureau of American Ethnology, Bulletin. – 1953, № 155.

Статья посвящена развитию представлений об одном из инструментов математизации отечественной археологии. Проведено краткое обсуждение современного состояния археологии и проблем, затрудняющих применение математических методов. Вводится понятие стратегии анализа данных как сетевого общедоступного информационного ресурса в виде интерактивного программного комплекса для автоматизации решения задач этого класса, не требующего от исследователя специальной математической подготовки.

Ключевые слова: эмпирические данные, анализ данных, признаковое пространство, математическая статистика, археология.

Введение. Любая наука, включая гуманитарные, проходя путь от описания эмпирических объектов до построения работоспособной теории, вынуждена обращаться к математическим методам анализа и синтеза информации. И чем активнее происходит внедрение математических методов, тем быстрее и эффективнее идет становление и созревание научной дисциплины.

В эмпирических науках исследователь работает с материальными предметами и их описаниями. Создание таких описаний – информационных копий предметов позволяет другим людям полноценно осмысливать и перерабатывать отраженный в них эмпирический материал. Информационные копии предметов, для хранения которых, в отличие от самих предметов, не требуются большие помещения и оборудование, кроме разве что книг и электронных носителей, могут быть легко сосредоточены в памяти одного человека или компьютера, который в состоянии провести обобщения и построить информационные модели достаточно широкого класса объектов и процессов.

От адекватности отражения информационными копиями эмпирических объектов в решающей степени зависит продуктивность всей дальнейшей теоретической деятельности по построению информационных моделей и теорий. Революция в мышлении, произведенная в эпоху возрождения Галилео Галилеем и породившая современную науку и технику, обязана идее измерения, открывшей “ящик Пандоры”. Именно возможность характеризовать материальные объекты и явления количественно определенными признаками вызвала цепную реакцию идей, структурировавших все человеческие представления о материальном мире. Человек получил в свои руки внешний инструмент для отражения материального мира в мире идеальном. Если до этого момента внешний мир отражался лишь в субъективном внутреннем мире человеческого сознания, то после этого появилась возможность моделировать реальность с помощью идеальных объектов – физических величин, имеющих количественную определенность и связанных между собой не просто ноль – единичными логическими правилами, а намного более богатыми количественными соотношениями, которые были предметом уже достаточно развитой к тому времени науки – математики. Отдельные физические законы соединялись в модели физических явлений и процессов, порождая новую физическую картину мира, в которой постепенно все более отчетливо и верно отражался окружающий человека материальный мир. Постепенно к механике, термодинамике и электричеству подключились другие разделы теоретической физики. Проникнув с помощью математических моделей от поверхностных, механических явлений в глубины материи, физика революционизировала химию и биологию.

Справедливости ради надо отметить, что даже самое точное измерение любого множества несущественных характеристик несколько не улучшит понимания предмета, зато после такого продвижения история неудачных попыток позволит четко разделить “зерна от плевел”. Получается, что выработка методики построения информационных копий, “оцифровки” объектов, закладывает фундамент теории. Мощность теории определяется прочностью основания. Пока не построена действительная теория, мы не можем быть уверены, что правильно видим предмет науки. В частности, поскольку археология к настоящему времени не вышла за пределы эмпирической науки, мы не имеем права терять исходный эмпирический материал, поскольку, утратив его, останемся только с теми бедными содержательно и весьма поверхностными отражениями, которые успели сконструировать, но на базе которых еще не успели построить теории³.

² Эта работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026.

³ Вполне возможно, что существенная, если не основная, часть археологической информации о памятнике хранится в буквальном смысле в земле. Раскапывая памятник, археолог его безвозвратно уничтожает. Если практическая археология уничтожит памятники до того, как эта наука пройдет этап математизации с последующим формированием теории, то она может опоздать уже навсегда. Это вопрос динамики – что произойдет раньше...

Итак, для продуктивного развития науки в рамках существующих научных парадигм требуется эмпирический материал в виде информационных копий объектов, включающих количественные измерения некоторого набора качественно определенных характеристик объекта или процесса. Но это только с одной стороны, – с эмпирической. С теоретической же стороны требуется наличие определенного преобразователя информации, позволяющего путем логических переходов получать из сырого эмпирического материала теоретические конструкты – элементы моделей и теорий.

Точка встречи эмпирического материала с преобразователями информации является точкой роста теории. Но, поскольку производители (носители) эмпирического материала, с одной стороны, и информационных преобразователей, с другой, редко совмещаются в одном лице, рост теоретического знания в большинстве гуманитарных областей без приложения специальных усилий протекает весьма вяло, тем более, что, в отличие от представителей естественных наук, гуманитарии недостаточно ориентированы на применение точных математических методов. Стоит обратить внимание, что развитие естественных наук в значительной степени стимулировалось потребностями технического воплощения их идей, а технику, в свою очередь, стимулировало капиталистическое производство в эпоху индустриальной революции, а в дальнейшем также государственные оборонные и космические программы. Включившись в эту связку, естественные науки были вынуждены привлечь мощнейший ресурс построения теорий – математические модели и методы. В гуманитарных науках сопоставимого по масштабам процесса в этом направлении пока не наблюдается.

Стратегии анализа данных. Одним из способов решения проблемы математизации археологии нам представляется создание общедоступного сетевого информационного ресурса, основной целью которого будет создание условий для продуктивного взаимодействия представителей гуманитарных наук со специалистами в области анализа данных. Их взаимодействие в упомянутой точке встречи должно быть взаимовыгодным. Для эмпирика эффект взаимодействия определяется возможностью содержательного анализа своего эмпирического материала, а для разработчика методов – возможностью убедиться в практической полезности новых методов и их комбинаций.

Анализируя практику применения методов анализа данных в гуманитарных науках, мы пришли к выводу, что крайне низкую эффективность взаимодействия поставщика данных с поставщиком методов можно объяснить тем, что в качестве основных элементов, опосредующих это взаимодействие, выступают сами методы, вырванные из контекста и технологий своего применения. Действительно, при внимательном рассмотрении процесса применения методов выясняется, что сам по себе любой метод анализа данных не является достаточным для проведения анализа, а требует некоторой предварительной подготовки эмпирического материала, приведения его к пригодному для обработки виду. Такая предварительная подготовка может быть весьма разнообразной, в зависимости от конкретного вида эмпирических данных и степени их “сырости”. Кроме того, доведение анализа до конца требует интерпретации полученных результатов, что само может быть представлено в виде процедуры, выходящей за рамки метода. Включать в каждый метод процедуры подготовки и интерпретации данных неразумно с точки зрения идейной чистоты, а отрывать от них – значит изолировать от приложений. Но разработчик методов подходит к этому вопросу, как правило, со своей, теоретической точки зрения, обрекая пользователя на проблемы в процессе практического применения его разработок.

Существующие пакеты статистического и интеллектуального анализа данных содержат достаточно полный набор методов как для самого анализа, так и для предварительной подготовки данных, но беда в том, что соединить методы в полноценный преобразователь информации в состоянии только эксперт в области анализа. А поскольку анализом данных вынуждены в основном заниматься сами исследователи – гуманитарии, то перед ними вырастает значительный входной барьер в виде необходимости получения дополнительного образования - изучения чуждой им по идеологии специальной предметной области или самостоятельного поиска выходов из математических лабиринтов.

Очевидным решением в этой ситуации нам представляется создание общедоступных, расположенных в глобальной сети, информационных комплексов, предназначенных для проведения определенных видов анализа от начала до конца, что называется под ключ, но без непосредственного участия специалистов в области анализа данных. Их участие будет опосредованным – на предварительном этапе разработки таких информационных комплексов, которые мы в дальнейшем будем называть стратегиями анализа данных.

Попробуем разобраться, каким требованиям должна удовлетворять стратегия, чтобы продуктивно выполнять свою функцию. Прежде всего определим основную цель создания стратегии, учитывая два момента:

Во-первых, мы вводим стратегии как инструмент математизации эмпирической науки преимущественно гуманитарного направления и, прежде всего, археологии, которая все еще находится на этапе, весьма далеком не только от продвинутых методов математической обработки данных, но даже и от формализованного представления эмпирического материала в виде пригодных для анализа массивов данных.

Во-вторых, явно недостаточное математическое образование основной массы научных сотрудников гуманитарного профиля и сам способ их мышления делает практически невозможным самостоятельное применение ими методов точных наук.

Отсюда следует, что стратегии должны быть рассчитаны на плохо подготовленный фактический материал и таких же пользователей. А основные цели, преследуемые созданием стратегий:

Во-первых, служить индикатором качества формализации признаков археологических объектов.

Во-вторых, структурировать эмпирический материал путем применения к нему повторяемых вычислительных процедур, которые позволяют делать выводы, проверяемые на корректность.

Учитывая особенности археологического материала, можно представить возможную картину применения стратегии для решения конкретной задачи.

Допустим, археолог пытается упорядочить керамические сосуды по форме, размеру, материалу, узорам на поверхности, чтобы в дальнейшем связать это с назначением сосуда, археологическим временем, типом памятника и т.п. Он вводит некоторую систему обозначений признаков и их значений для формализованного описания сосудов, применяет к полученному массиву данных стратегию и получает результат в виде анализа связей между признаками и классификации объектов. Кроме этого, стратегия проверяет стандартные статистические гипотезы, которые, допустим, показывают, что обнаруженные связи и структуры статистической значимостью не обладают. Исследователь в нашем примере понимает, что сам эмпирический материал просто обязан содержать искомую информацию, но первая проверка дала нулевой результат. Он принимает объяснение, что стратегия не может ошибаться, поскольку в других случаях она доказывала свою эффективность. Значит, информация была потеряна на этапе формализации. А сам математический метод просто преобразует входные данные в выходные. Если на входе было пусто, то будет пусто и на выходе. Поняв, что сконструированные им признаки описывают предмет слишком поверхностно, не отражая внутренней структуры, он углубляет описание и пропускает данные через стратегию еще раз. Повторяя такую процедуру многократно, он добивается, наконец, извлечения ожидаемой информации. Кстати, представляется весьма интересным публиковать не только конечный результат, но и процесс движения к нему путем постепенного построения знаковой системы для информационного отображения предмета с анализом недостатков промежуточных вариантов. Это поможет другим исследователям подключаться к процессу конструирования и во многих случаях будет способствовать не только улучшению качества формализации, но и повышению методологического уровня научного сообщества.

Принципы построения стратегий. Теперь рассмотрим основные принципы, которыми следует руководствоваться при проектировании и реализации стратегий анализа данных:

Во-первых, стратегия должна предъявлять минимально строгие требования к подготовке входных данных, чтобы исследователь мог сосредоточиться на решении задачи построения признакового пространства по существу, а не по форме. В частности, стратегия должна включать методы приведения разнотипных входных признаков к виду, требуемому основным методом анализа. Такие процедуры, называемые шкалированием или оцифровкой, превращают признаки, измеренные в шкале наименований или порядка, в количественные переменные путем присвоения их значениям числовых меток. Для каждого вида анализа конкретные значения числовых меток получаются в результате оптимизации, исходя из соответствующего функционала с учетом конкретного статистического распределения значений.

Во-вторых, стратегия должна по возможности исключать необходимость вмешательства человека в процесс получения результатов анализа. В крайнем случае, такое вмешательство можно ограничиваться предварительной настройкой стратегии. Соблюдение этого требования обеспечит повторяемость и воспроизводимость анализа и даст устойчивую базу для сравнения.

В третьих, в состав стратегии должна входить проверка на статистическую значимость полученных результатов для количественной оценки качества формализации эмпирического материала.

Что касается реализации, то стратегия внешне должна выглядеть как типичная электронная публикация, описывающая в виде журнальной статьи определенную технологию подготовки и анализа данных, но при этом обладающая элементами интерактивного поведения. В частности, в соответствующей среде разработки, стратегия позволяет подставлять вместо данных, приведенных в качестве примера, данные пользователя. Сами же данные будут выглядеть, как обычная таблица. При большом количестве строк или столбцов таблица будет автоматически умещаться в заданном прямоугольнике, показывая многоточиями опущенные строки и столбцы. Все промежуточные результаты расчетов, которые могут быть скрыты или выведены в виде таблиц или диаграмм, будут корректно передаваться по цепочке последующим блокам алгоритма. Естественно, что фиксированное в тексте описание результатов расчета и выводы могут разойтись с вновь полученными результатами. Для предотвращения таких несоответствий стратегия должна содержать условные описания, которые становятся видимыми в зависимости от индикаторов, встроенных в тело стратегии.

Становится понятно, что стратегия, как любой электронный документ, имеет две стороны – внутреннюю и внешнюю. Внешне она представляется как статья, а внутри – как программа.

Внешнее представление отвечает за то, чтобы пользователь, не знакомый с применяемыми методами, смог определить, дадут ли они нужный ему результат, и сможет ли он подготовить данные требуемого объема и качества. Кроме того, описание метода в форме живого примера дает возможность снять все неопределенности понимания и отследить логику процесса преобразования данных не только по формулам, что вряд ли понравится гуманитариям, но и наглядно – по числам и графикам.

Внутреннее представление, в виде набора вычислительных модулей, собирается как обычная программа, запускаемая по событию, например – нажатию кнопки или гиперссылки. Современные интегрированные среды разработки программ предоставляют разработчику наряду с текстовым редактором для вставки программного кода также и визуальную среду конструирования внешнего вида оконного интерфейса пользователя, в которой разработчик редактирует элементы интерфейса в режиме *wysiwyg* (что ты видишь, то ты и получишь).

Витяев Е.Е., Инструментальное средство Visual Discovery извлечения ин-
Москвитин А.А. формации из данных и решения задач интеллектуального
Подберезный А.А анализа данных¹

В предыдущей работе, посвященной извлечению информации из данных, аргументировалось, что для получения осмысленного и интерпретируемого результата методами интеллектуального анализа данных необходимо, чтобы онтология предметной области была согласована с онтологией применяемого метода. Для такого согласования необходимо, во-первых, извлечь информацию из данных в терминах онтологии предметной области, а, во-вторых, применить такой метод, который может обнаружить на данных закономерности, интерпретируемые в терминах этой информации. Таким методом является разработанная нами в рамках реляционного подхода система Discovery. Для того, чтобы пользователь мог в этой системе удобно и визуально извлекать информацию из данных и формировать гипотезы в терминах этой информации, нами была разработана система Visual Discovery, которая позволяет это делать в режиме визуального конструктора. В работе дано описание данной системы и пример её использования для решения задачи из области медицины.

Ключевые слова: Data Mining, KDD&DM, интеллектуальный анализ данных.

1. Онтология предметной области. Как отмечалось в работе [Витяев Е.Е., 2006, 2010], для корректного применения методов интеллектуального анализа данных необходимо определить онтологию предметной области. Кратко напомним основные понятия предметной области и онтологии [Витяев Е.Е., 2010]. *Предметная область* – это совокупность объектов предметной области, рассматриваемых с точки зрения некоторого предмета исследования – совокупности существенных свойств (атрибутов) и отношений объектов исследования, описываемых в некоторой системе понятий предметной области. Предмет исследования может быть задан онтологией предметной области – специфицирующей в некотором формальном языке множество рассматриваемых объектов, связи между ними, систему понятий, и свойства объектов. Предмет исследования и онтология определяют «взгляд», «точку зрения», с которой рассматриваются (описываются в системе понятий) объекты предметной области, отношения и их свойства. Предметная область может быть задана эмпирической системой $\mathfrak{S} = \langle A, \Omega \rangle$, где A – объекты ПО, а Ω – онтология ПО (система понятий онтологии, заданная одноместными предикатами, и множество отношений и операций, интерпретируемых в системе понятий ПО).

2. Онтология методов интеллектуального анализа данных. Как отмечалось в работе [Витяев Е.Е., 2010], для того чтобы методы интеллектуального анализа данных позволяли получать осмысленные и интерпретируемые результаты, необходимо, чтобы они правильно использовали содержащуюся в данных информацию. Методы имеют свою онтологию, которая включает:

1. типы данных, с которыми работает метод;
2. язык оперирования и интерпретации данных;
3. класс гипотез, проверяемый методом и сформулированный в языке интерпретации данных.

Для получения осмысленных и интерпретируемых результатов, необходимо, чтобы онтология метода и онтология ПО были согласованы между собой следующим образом:

1. типы данных, с которыми работает метод, должны интерпретироваться в онтологии Ω предметной области. Поэтому атрибуты, свойства и признаки, используемые в данных метода, должны быть интерпретируемы в онтологии Ω . Тем самым определяется информация, извлекаемая из данных этим методом, которая представляется множеством интерпретируемых в онтологии Ω математических отношений и операций;
2. класс проверяемых методом гипотез должен интерпретироваться в онтологии ПО. Это означает, что он должен выражаться через информацию, извлекаемую из данных.

Для того, чтобы знать какая информация содержится в данных, нам необходимо извлечь информацию из данных. Для этого надо представить информацию, содержащуюся в данных, множеством отношений и операций, интерпретируемых в онтологии предметной области.

¹Работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026, грантом РФФИ 08-07-00272-а, интеграционными проектами СО РАН № 3,86,136 НШ-276.2012.1 и программой президента Российской Федерации поддержки научных школ НШ-276.2012.1..

Для этого нами была разработана программная система Visual Discovery, которая позволяет визуально извлекать информацию из данных в режиме конструктора. Описание системы приведено ниже.

3. Извлечение информации из данных. Проанализируем, как следует задавать свойства и атрибуты объектов ПО в терминах онтологии Ω . Чтобы правильно извлекать информацию и знания из свойств и атрибутов, необходимо их интерпретировать в системе понятий ПО. Как говорилось в [Витяев Е.Е., 2010], сами по себе числовые значения величин смысла и информацию не содержат, смысл величин указывается в их интерпретации, например, 5 метров, 5 литров, 5 килограмм и т.д. Интерпретация числовых значений – метры, литры, килограммы и т.д. привязана к онтологии. Для извлечения информации из атрибутов, свойств, признаков и величин ПО нужно определить множество интерпретируемых в онтологии Ω математических отношений и операций и включить их в онтологию Ω .

Для работы непосредственно с информацией, извлеченной из данных и представленной множеством Ω отношений и операций, нами разработан специальный реляционный подход (Relational Data Mining) к методам извлечения знаний и система «Discovery», реализующая его [Витяев, 2006; Витяев, Москвитин, 1993; Kovalerchuk, Vityaev, 2000; Vityaev, Kovalerchuk, 2008]. Однако в этой системе предполагается, что информация уже извлечена и представлена множеством Ω отношений и операций. Для удобства пользователей необходимо, чтобы и сам процесс извлечения информации тоже сопровождался системой. С этой целью и разработано инструментальное средство Visual Discovery, описанное далее.

4. Инструментальное средство Visual Discovery извлечения информации из данных и решения задач интеллектуального анализа данных.

Нами разработано инструментальное средство Visual Discovery, позволяющее специалисту ПО работать с онтологией ПО, извлекать информацию из данных, вручную формировать шаблоны предикатов (онтологию) и классы гипотез, и решать следующие три основные задачи:

1. Предсказывать значения признаков объектов на основе проверенных гипотез;
2. Выбирать информативную подсистему признаков на основе гипотез, которые подтвердились на данных;
3. Классифицировать объекты в соответствии с обнаруженными группами подтвержденных гипотез (закономерностей).

Для простоты, работу системы проиллюстрируем на следующем конкретном примере. В области распространения медицинских препаратов было проведено исследование компаний и численно измерены признаки, представленные на рис. 1:

Номер компании	Организационная культура	Товарооборот на 1 специалиста	Средний доход	Прибыль на 1 специалиста	Издержки на 1 специалиста	Доля издержек на весь персонал	Среднее число профессионального обучения, часов	Доля издержек на обучение	Коэффициент текучести	Доля часов на обучение в общем балансе времени
Компания № 1	ПОК (1+4)	3301,5	734,4	163,5	282,3	8,2	25,6	0,05	123,3	1,5
Компания № 2	БОК (2+3)	1851,6	548,3	144,3	191,5	10,3	10,3	0,03	10,6	0,6
Компания № 3	БОК (2+1)	1934,5	648,3	154,6	153,4	9,3	5,4	0,01	8,0	0,3
Компания № 4	ООК (3+1)	2635,4	528,4	126,4	104,3	6,2	15,4	0,05	5,6	0,8
Компания № 5	ПОК (1+2)	2031,2	526,4	113,4	107,4	6,3	12,3	0,01	17,6	0,6
Компания № 6	ПОК (1+2)	1935,4	583,4	139,4	144,3	9,2	30,2	0,07	19,3	1,7
Компания № 7	ПОК (1+3)	3142,4	675,3	159,3	218,3	11,3	12,4	0,04	15,3	0,6
Компания № 8	БОК (2+1)	2023,4	426,4	114,3	123,4	7,9	5,8	0,01	8,6	0,3
Компания № 9	ПОК (1+3)	2673,3	429,3	123,4	98,6	6,4	6,2	0,01	12,6	0,3
Компания № 10	ООК (3+1)	2515,4	335,6	153,4	97,4	8,6	10,6	0,03	10,4	0,5
Компания № 11	БОК (2+1)	2134,7	235,4	104,3	83,4	8,3	35,6	0,05	7,4	2,0
Компания № 12	ООК (3+1)	2435,7	506,7	118,3	84,6	7,1	16,7	0,06	5,3	0,9
Компания № 13	ПОК (1+2)	3004,5	683,2	159,6	243,4	10,6	5,2	0,01	16,3	0,3
Компания № 14	ПОК (1+3)	2947,3	675,4	153,2	204,3	9,4	6,8	0,01	10,2	0,4
Компания № 15	БОК (2+1)	1543,5	254,6	54,6	72,3	4,6	10,3	0,01	2,0	0,6
Компания № 16	ООК (1+2)	3124,6	653,4	160,4	234,3	10,4	20,6	0,05	1,0	1,1

Рис. 1. Таблица Объекты/признаки, характеризующие медицинские компании.

Необходимо было найти закономерности между признаками (выбирать информативную подсистему признаков), характеризующие максимальный товарооборот на одного специалиста.

4.1. Процесс решения задач. В общем случае, процесс решения задач 1-3, можно представить

в виде диаграммы концептуальной модели деятельности (рис. 2) на языке UML. Эта модель может быть спецификацией прецедента использования, а также, многократно используемых функциональных свойств.

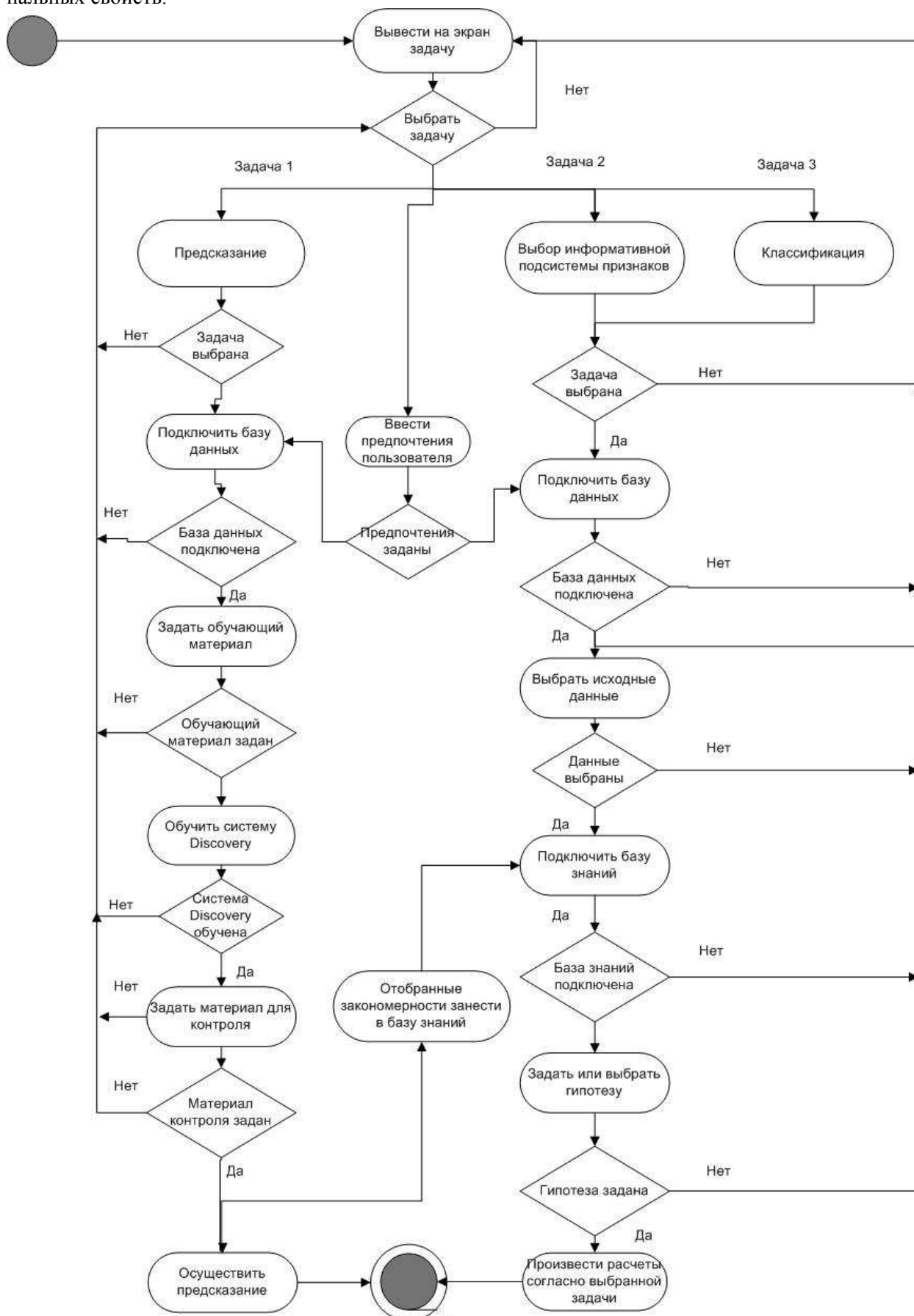


Рис. 2. Диаграмма прецедентов в системе Visual Discovery

Модель деятельности описывает поведение, в которое вовлечено несколько элементов системы. В системе Visual Discovery такими элементами являются: основные данные, шкала, справочник по шкалам, рабочие гипотезы, описание предпочтений пользователя и знания экспертов. Диаграмма деятельности показывает шаги вычисления в системе Visual Discovery. За каждым классом деятельности скрывается определенная процедура вычислений.

Передача управления от одного состояния вида деятельности к другому задает поток управления, который определяет, как выполнение действия в одном узле влияет на выполнение действия в других узлах и, одновременно, испытывает их влияние.

Рассмотрим поток управления для нашей задачи – поток №2 (задача 2 в диаграмме) – выбор информативной подсистемы признаков. Для этого необходимо последовательно выполнить следующие действия (см. диаграмму):

1. Выбрать исходные данные – объекты/признаки;
2. Задать онтологию ПО в виде шаблонов предикатов;
3. На основе онтологии и исходных данных сформировать класс проверяемых гипотез;
4. Задать основные параметры работы системы;
5. Получить найденные закономерности;
6. Проинтерпретировать найденные закономерности и, тем самым, получить результат.

4.2. Выбор исходных данных (блок «выбор исходных данных» в потоке №2 диаграммы). Данные в системе Visual Discovery представляются виде таблицы, объекты в которой представлены строками, а признаки столбцами (см. screen shot на рис. 3). Система позволяет загружать данные из следующих источников данных:

1. MS Excel. Данные берутся из выбранной таблицы пользователем при открытии файла;
2. MS Access. Данные извлекаются с помощью соответствующего SQL запроса из файла;
3. MS SQL Server. Данные извлекаются из БД сервера с помощью соответствующего SQL запроса.

Для поставленной задачи данные были внесены в Excel файл. Признаки A1-A11 имеют следующую интерпретацию:

1. A1 и A2 – организационная культура;
2. A3 – товарооборот на 1 специалиста;
3. A4 – средний доход;
4. A5 – прибыль на 1 специалиста;
5. A6 – издержки на 1 специалиста;
6. A7 – доля издержек на весь персонал
7. A8 – среднее число профессионального обучения, часов;
8. A9 – доля издержек на обучение;
9. A10 – коэффициент текучести;
10. A11 – доля часов на обучение в общем балансе времени.

Обучающее множество													
Поиск закономерностей													
Классы гипотез													
Solution Explorer													
Clear.mydsl1*													
	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	R3	
1	1	4	3301.5	734.4	163.5	282.3	8.2	25.6	0.05	123.3	1.5	1	
2	2	3	1851.6	548.3	144.3	191.5	10.3	10.3	0.03	10.6	0.6	0	
3	2	1	1934.5	648.3	154.6	153.4	9.3	5.4	0.01	8	0.3	0	
4	3	1	2635.4	528.4	126.4	104.3	6.2	15.4	0.05	5.6	0.8	0	
5	1	2	2031.2	526.4	113.4	107.4	6.3	12.3	0.01	17.6	0.6	0	
6	1	2	1935.4	583.4	139.4	144.3	9.2	30.2	0.07	19.3	1.7	0	
7	1	3	3142.4	675.3	159.3	218.3	11.3	12.4	0.04	15.3	0.6	1	
8	2	1	2023.4	426	114.3	123.4	7.9	5.8	0.01	8.6	0.3	0	
9	1	3	2673.3	429.3	123.4	98.6	6.4	6.2	0.01	12.6	0.3	0	
10	3	1	2515.4	335.6	153.4	97.4	8.6	10.6	0.03	10.4	0.5	0	
11	2	1	2134.7	235.4	104.3	83.4	8.3	35.6	0.05	7.4	2	0	
12	3	1	2435.7	506.7	118.3	84.6	7.1	16.7	0.06	5.3	0.9	0	
13	1	2	3004.5	683.2	159.6	243.4	10.6	5.2	0.01	16.3	0.3	1	
14	1	3	2947.3	675.4	153.2	204.3	9.4	6.8	0.01	10.2	0.4	1	
15	2	1	1543.5	254.6	54.6	72.3	4.6	10.3	0.01	2	0.6	0	
16	1	2	3124.6	653.4	160.4	234.3	10.4	20.6	0.05	1	1.1	1	

Рис. 3. Исходная таблица объектов/признаков в Visual Discovery.

4.3. Задание онтологии (см. блок «подключить базу знаний» в потоке №2 диаграммы). Отличительной особенностью Visual Discovery от других систем интеллектуального анализа данных

является графическая модель задания онтологии (задания отношений и операций на исходных данных).

Самая сложная часть работы специалиста предметной области сводится к заданию онтологии для информации, извлекаемой из данных путем создания диаграммы шаблонов предикатов в интуитивно понятном графическом интерфейсе. Эта задача решается визуальным интерфейсом системы Visual Discovery рис. 4.

Шаблоны предикатов могут быть заданы двумя способами:

1. Загружены из файла;
2. Созданы непосредственно в системе путем создания диаграммы шаблонов предикатов;
3. Получены системой путем решения одной из 3-х задач (п. 4).

Диаграмма шаблонов предикатов разбита на три поля, которые содержат:

- Шаблоны предикатов;
- Функции от переменных;
- Исходные данные.

На поле «Шаблоны предикатов» помещаются предикаты с соответствующими термами и отношениями между ними. В данный момент поддерживаются следующее множество отношений, соответствующее шкале порядка:

- отношения сравнения $=$, $<$, $>$, \leq , \geq ;
- отношения принадлежности предиката к множеству или интервалу значений ($T1$ in $[T2,T3]$), ($T1$ in $(T2,T3)$), ($T1$ in $(T2;T3]$), ($T1$ in $[T2;T3)$).

На поле «Функции от переменных» помещаются функции, которые являются интерпретацией термов из поля «Шаблоны предикатов». Функции определяют переменные и операции над ними. Функция может быть задана арифметическим выражением или любой другой математической функцией.

На поле «Исходные данные» помещаются признаки объектов или константы, на которые ссылаются переменные из функций или термов предикатов.

Для решения поставленной задачи была создана диаграмма шаблонов предикатов, представленная на рис. 4. Каждый признак был разбит на интервалы некоторым алгоритмом, выделяющим «сгустки» значений и по этим интервалам были сформированы предикаты, отвечающие за принадлежность признака некоторому интервалу значений.

Например, признак A3 (товарооборот на 1 специалиста) был разбит на три интервала:

1. $T3$ in $[1851,6;2023,4)$;
2. $T3$ in $[2435,7;2947,3)$;
3. $T3 \geq 2947,3$

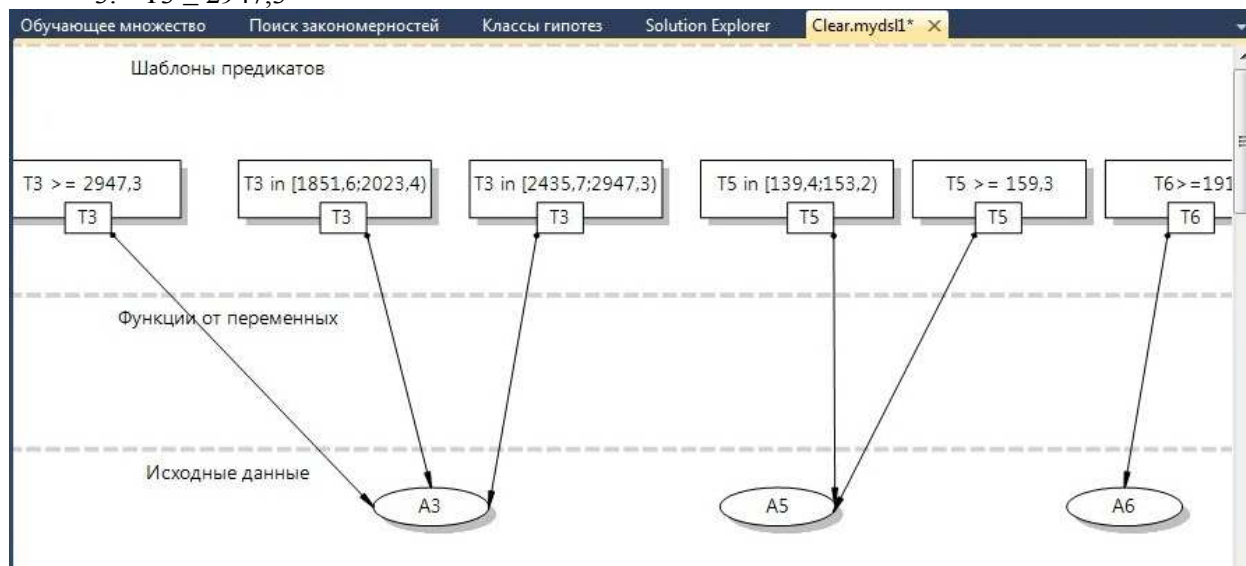


Рис. 4. Формирование шаблонов предикатов.

Соответствующие шаблоны предикатов для признака A3 введены в раздел «Шаблоны предикатов» (рис.4). Каждый предикат интерпретируется в некотором признаке и связывается с ним направленной стрелкой. Аналогично для признаков A5 и A6 выбираем другие шаблоны. Интервалы значений могут находиться автоматически программой, вводиться пользователем в соответствии с интерпретацией признаков, редактироваться и удаляться. Шаблоны предикатов фиксируют ин-

формацию, извлекаемую из этих признаков. Гипотезы будут формироваться с использованием только этой информации. Поэтому нужно определить, столько шаблонов предикатов, сколько нужно для выражения всей интересующей нас информации.

4.4. Формирование классов гипотез (блок «задать или выбрать гипотезу» рис.2.). Гипотезы задаются на основе шаблонов предикатов и определяют то знание, которое мы бы хотели получить в результате анализа данных. Гипотезы, и тем самым будущее знание, задается правилами, содержащими посылки и следствия. Задание гипотез также осуществляется визуально и представлено на рис.5.

В нашей задаче классы гипотез задаются по шаблонам предикатов и целевому предикату, выбранному из шаблонов предикатов. Например, целевой предикат «A3 in [2435,7; 2947,3]» выбирается из множества предикатов, задающихся шаблоном предиката «T3 in [2435,7; 2947,3]»;

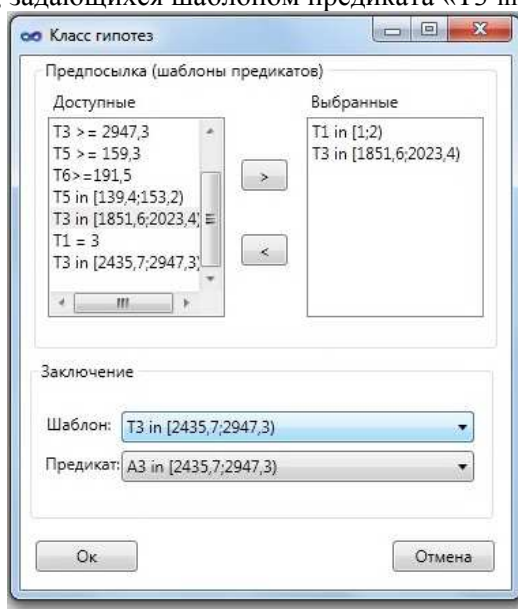


Рис. 5. Формирование класса гипотез.

В нашей задаче были сформированы классы гипотез, представленные на рис.6.

Обучающее множество Поиск закономерностей Классы гипотез Solution Explorer Clear.mydsl1			
Предпосылка	Исход		
(T5 in [139,4;153,2))	A3 in [1851,6;2023,4)		Добавить
(T1 = 3)	A3 in [2435,7;2947,3)		Удалить
(T1 in [1;2)) and (T4 >= 648,3) and (T5 >= 159,3)	A3 >= 2947,3		Свойства

Рис.6 Заданные классы гипотез.

4.5. Проверка классов гипотез (блок «произвести расчеты согласно выбранной задаче», рис. 2.). Далее сформированные классы закономерностей проходят проверку на данных. Если какая-то гипотеза подтверждается на данных, пройдя серию статистических тестов, то она фиксируется как закономерность в данных и выдается программной системой Visual Discovery. Для обнаружения закономерностей программной системой Visual Discovery нужно задать следующие параметры [Kovalerchuk, Vityaev, 2000; Витяев, 2006]:

- доверительный уровень для критерия Фишера;
- доверительный уровень для критерия Юла;
- количество объектов обучения;
- глубина базового перебора.

После чего были получены результаты, представленные на рис. 7. На этом рисунке:

В поле «Правило» записаны закономерности, связывающие признаки объектов.

В поле «Вероятность» приведена условная вероятность правила.

В поле «Фишер» указан критерий Фишера для предиката, содержащегося в правиле.

В поле «Юла» указан критерий Юла для предиката, содержащего в правиле.

В поле «Список объектов» приведены номера объектов, на которых выполняется правило.

Знак «+» после номера объекта означает положительный исход правила, знак «-» означает отрицательный исход соответственно.

Обучающее множество Поиск закономерностей X Классы гипотез Solution Explorer Clear.myds1

Критерий Фишера: 0,05

Критерий Юла: 0

Количество объектов: 16

Глубина базового перебора: 1

Найти закономерности

Правило	Вероятность	Фишер	Юла	Список объектов	
(A5 in [139,4;153,2]) -> (A3 in [1851,6;2023,4])	1			1+ 5+	
A5 in [139,4;153,2]		0,0249809460432507	1		
(A1 = 3) -> (A3 in [2435,7;2947,3])	1			3+ 9+ 11+	
A1 = 3		0,00713529839065878	1		
(A1 in [1;2]) -> (A3 >= 2947,3)	0,625			0+ 4- 5- 6+ 8- 12+ 13+ 15+	
A1 in [1;2]		0,01282238408501	1		
(A4 >= 648,3) -> (A3 >= 2947,3)	0,8333333			0+ 2- 6+ 12+ 13+ 15+	
A4 >= 648,3		0,00137326664423067	1		
(A5 >= 159,3) -> (A3 >= 2947,3)	1			0+ 6+ 12+ 15+	
A5 >= 159,3		0,00274588200143116	1		

Рис. 7. Полученные результаты.

4.6. Получение результатов. Закономерности на рис.7, полученные в системе Visual Discovery, были проинтерпретированы и проанализированы специалистом и, в результате, были сделаны следующие заключения о связи между признаками:

- Из закономерности $(A1 \in [1;2]) \Rightarrow (A3 \geq 2947,3)$ следует, что между типом организационной культуры и величиной товарооборота существует закономерность. Так, максимальный товарооборот 2950-3300 тыс. руб. на одного человека в год обеспечивает рыночная культура, ориентированная на стабильность. Иерархическая культура обеспечивает товарооборот на одного специалиста на уровне 2435-2950 тыс. руб.
- Из закономерности $(A5 \geq 159,3) \Rightarrow (A3 \geq 2947,3)$ следует, что между максимальным товарооборотом и прибылью на одного человека существует закономерность. Максимальный товарооборот 2950-3300 тыс. руб. на одного человека в год обеспечивает максимальную прибыль – от 159 до 163,5 тыс. руб. на одного специалиста. При товарообороте на уровне 1850-2020 тыс. руб. максимальная прибыль составит 139-153 тыс. руб. на одного специалиста.
- Из закономерности $(A4 \geq 648,3) \Rightarrow (A3 \geq 2947,3)$ следует связь между средним доходом и максимальным товарооборотом. Максимальный средний доход от 648,3 до 734 тыс. руб. обеспечивается при максимальном товарообороте 2950-3300 тыс. руб. на одного специалиста.

Литература

- Витяев Е.Е.** Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. Новосибирск, 2006. 293с.
- Витяев Е.Е.** Извлечение информации из данных // Информационные технологии в гуманитарных исследованиях, Вып. 15, ИАЭТ СО РАН, Новосибирск, 2010, 9-16.
- Витяев Е.Е., Москвитин А.А.** Введение в теорию открытий. Программная система DISCOVERY. // Логические методы в информатике (Вычислительные системы, вып. 148), Новосибирск, 1993, с.117-163
- Е. Vityaev, B.Y. Kovalerchuk**, Relational Methodology for Data Mining and Knowledge Discovery. *Intelligent Data Analysis*. Special issue on "Philosophies and Methodologies for Knowledge Discovery and Intelligent Data Analysis" eds. Keith Rennolls, Evgenii Vityaev. v.12(2), IOS Press, 2008, pp. 189-210
- Kovalerchuk B., Vityaev E.** Data Mining in Finance: Advances in Relational and Hybrid methods. (Kluwer international series in engineering and computer science; SECS 547), Kluwer Academic Publishers, 2000, p.308.

В работе проводится сравнение системы «Discovery» с алгоритмами Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services. Показывается, что система «Discovery», во-первых, обладает теоретическими преимуществами перед этими алгоритмами, и, во-вторых, практически работает лучше на данных, где эти преимущества проявляются явно. Эти данные, в том числе, хорошо демонстрируют и иллюстрируют преимущества системы Discovery.

Ключевые слова: Интеллектуальный анализ данных, извлечение знаний, предсказание, обнаружение закономерностей.

1. Введение. В последнее время получили широкое развитие и активно применяются на практике различные KDD&DM-методы (Knowledge Discovery in Data Bases and Data Mining). Однако используемые сейчас KDD&DM-методы имеют серьезные ограничения [Витяев, 2006; Kovalerchuk, Vityaev, 2000]: каждый метод может работать только с определенными типами данных, имеет свой язык оперирования и интерпретации данных, и обнаруживает только определенный класс гипотез. Таким образом, они не способны извлекать из данных все знания в полном объеме, а также могут получать результаты, не интерпретируемые в терминах предметной области.

Система «Discovery» реализует реляционный подход к методам извлечения знаний [Витяев, 2006; Vityaev, Kovalerchuk, 2006; Kovalerchuk, Vityaev, 2000], снимающий практически все ограничения, свойственные KDD&DM-методам.

Система «Discovery» обладает следующими важными теоретическими свойствами: может обнаруживать теорию предметной области, может обнаруживать все правила, имеющие максимальные условные вероятности, может обнаруживать непротиворечивую вероятностную аппроксимацию теории предметной области [Vityaev, Kovalerchuk, 2006], обнаруживает все максимально специфические правила, позволяющие предсказывать без противоречий [Vityaev, 2005].

Наиболее близкими к системе «Discovery» методами можно считать поиск ассоциативных правил (Microsoft Association Rules) [Tang, MacLennan, 2005] и Decision Trees, в виду того, что закономерности в этих методах также представляются в форме логических правил. В данной работе ставится задача сравнения системы «Discovery» с Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services. Мы покажем, что система «Discovery» обладает теоретическими преимуществами перед этими методами и практически работает лучше на данных, где эти преимущества явно проявляются.

Система «Discovery» была реализована в виде плагина, подключаемого к службам Microsoft SQL Server 2005 Analysis Services (SSAS). Это позволяет использовать для сравнения алгоритмов единую среду разработки Business Intelligence Development Studio, единые средства визуализации Data Mining моделей, а также стандартные средства сравнения качества Data Mining моделей: диаграмму роста (Lift Chart) и классификационную матрицу (Classification Matrix).

2. Association Rules. Алгоритм Microsoft Association Rules состоит из двух шагов. Первый шаг – это ресурсоемкая фаза нахождения часто встречающихся наборов. Второй шаг – это генерация ассоциативных правил с использованием множества часто встречающихся наборов.

Нахождение часто встречающихся наборов. Под набором (itemset) мы понимаем набор предикатов. Например, $\{A = 1; B = 0; C = 1\}$ – это набор длины 3. Запись таблицы содержит некоторый набор, если на этой записи выполнены все предикаты данного набора. Поддержка набора – это количество записей таблицы, которые содержат данный набор.

Основным параметром, участвующим в нахождении часто встречающихся наборов, является параметр Minimum Support, который определяет, в каком минимальном количестве записей анализируемой таблицы должен содержаться некоторый набор, чтобы он являлся часто встречающимся.

На первой итерации находятся все часто встречающиеся наборы длиной 1. Алгоритм просто сканирует таблицу и подсчитывает поддержку каждого возможного предиката. Предикаты с поддержкой большей, чем Minimum Support, добавляются во множество часто встречающихся наборов длины 1. На второй итерации из часто встречающихся наборов, найденных на первой итерации, строятся всевозможные наборы длины 2, подсчитываются поддержки этих наборов, те набо-

² Работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026, грантом РФФИ 08-07-00272-а и интеграционными проектами СО РАН № 3,86,136.

ры, которые проходят критерий Minimum Support, добавляются во множество часто встречающихся наборов длины 2. Далее из предикатов, входящих в часто встречающиеся наборы длины 2, строятся всевозможные наборы длины 3 и т.д. Алгоритм повторяется для наборов длины 3, 4, 5 и т.д., пока находятся наборы удовлетворяющие критерию Minimum Support.

Далее проверяется условие, что каждый поднабор часто встречающегося набора, также должен являться часто встречающимся набором.

Генерация ассоциативных правил. Следующая процедура генерирует ассоциативные правила:

Для любого часто встречающегося набора f , генерируем все поднаборы x и их дополнения $y = f - x$.

Если Поддержка(f)/Поддержка(x) > Minimum Probability, тогда $x \Rightarrow y$ является ассоциативным правилом с условной вероятностью Prob = Поддержка(f)/Поддержка(x).

Параметр Minimum Probability задается перед началом обучения модели.

Прогнозирование. Следующий алгоритм по набору предикатов, поданных на вход, предсказывает значение целевого признака, либо выдает множество (n штук) наиболее вероятных значений целевого признака:

1. На вход подается некоторый набор предикатов. Ищутся все правила, условная часть которых совпадает либо с данным набором, либо с некоторым поднабором данного набора, а целевая часть содержит целевой признак. Найденные правила (k штук) применяются: целевые части правил и соответствующие условные вероятности добавляются в список рекомендаций.
2. Если подходящих правил не найдено, или их слишком мало ($k < n$), находятся $n-k$ наиболее популярных значений целевого признака. То есть, среди всех правил вида $\Rightarrow P = a_i$ (с пустой условной частью и целевым признаком в правой части) находятся $n-k$ правил с наибольшей условной вероятностью.
3. Предикаты, полученные на первых двух шагах, сортируются по вероятности.

3. Decision Trees. Основная идея алгоритма решающих деревьев состоит в рекурсивном разделении данных на подмножества, содержащие более или менее однородные состояния целевого (прогнозируемого) атрибута. При каждом разделении, все входные атрибуты оцениваются по их влиянию на целевой атрибут. Когда этот рекурсивный процесс заканчивается, решающее дерево сформировано.

Построение таблицы подсчета корреляций. Пусть для анализа с помощью алгоритма решающих деревьев дана некоторая таблица, с входными колонками F0, F1, F2, F3, и целевой колонкой P, размером 3000 записей, содержащую в своих ячейках только 0 или 1.

Тогда таблица подсчета корреляций на первом шаге алгоритма будет выглядеть следующим образом:

		F0		F1		F2		F3	
		0	1	0	1	0	1	0	1
P	0	300	700	700	300	400	600	500	500
	1	200	1800	400	1600	400	1600	1100	900

Каждая колонка таблицы подсчета корреляций соответствует паре атрибут-значение одного из входных атрибутов. Каждая строка соответствует значению целевого атрибута.

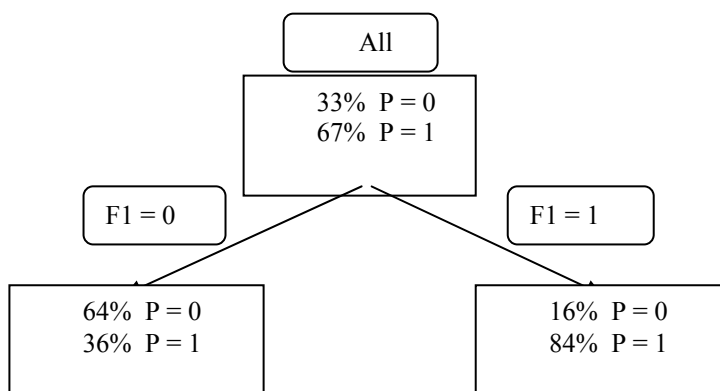
Ячейки таблицы содержат количество корреляций соответствующих пар: входной атрибут-значение, целевой атрибут-значение. Например, пересечение первой строки и первого столбца таблицы подсчета корреляций содержит число 300, т.е. в анализируемой таблице есть 300 записей, у которых одновременно F0 = 0 и P = 0.

Нахождение наиболее подходящего для разбиения атрибута. Одним из широко известных критериев, с помощью которого можно найти необходимый атрибут, является *Энтропия*. Энтропия (H) определяется следующим образом:

$$H(p_1, p_2, \dots, p_n) = -p_1 \cdot \log(p_1) - p_2 \cdot \log(p_2) - \dots - p_n \cdot \log(p_n), \quad p_1 + p_2 + \dots + p_n = 1$$

Например, энтропия атрибута F1 равна: $H(F1) = H(700, 400) + H(300, 1600) = 0.946 + 0.629 = 1.571$

Атрибут с наименьшей энтропией является наиболее подходящим для разбиения. В данном примере первое разбиение будет по атрибуту F1. Решающее дерево после первого разбиения выглядит следующим образом:



Таким образом, разделив данные входной таблицы на два подмножества, далее рекурсивно применяем первый шаг алгоритма к новообразованным вершинам решающего дерева. Например, новая таблица подсчета корреляций для вершины решающего дерева $F1 = 0$ выглядит следующим образом:

		F0		F1		F2		F3	
		0	1	0	1	0	1	0	1
P	0	300	700	700	0	400	600	500	500
	1	200	1800	400	0	400	1600	1100	900

Прогнозирование. Алгоритм предсказания решающих деревьев достаточно простой и эффективен. Поданный на вход предсказания набор предикатов, например $(F1 = 1, F2 = 0, F3 = 0, F4 = 1)$, спускается по дереву от корня к листьям по соответствующему этому набору пути, вершина дерева, находящаяся в конце этого пути и определяет предсказанное значение целевого признака. Таким образом, количество шагов в процессе прогнозирования не превышает максимальной длины пути от корня к листьям решающего дерева.

4. Система «Discovery». Алгоритм поиска закономерностей системы «Discovery» реализует метод семантического вероятностного вывода [Витяев, 2006], позволяющего находить все максимально специфические и максимально вероятные закономерности в данных [Витяев, 2006]. Определим на высказываниях языка первого порядка вероятность, как описано в [Halpern, 1990].

Семантическим вероятностным выводом (СВВ) некоторого атома/литерала P является такая последовательность правил C_1, C_2, \dots, C_n , что:

- 1) $C_i = (A_1^i \& \dots \& A_{k_i}^i \Rightarrow P)$, $i = 1, \dots, n$;
- 2) C_i является подправилом правила C_{i+1} , т.е. $\{A_1^i, \dots, A_{k_i}^i\} \subset \{A_1^{i+1}, \dots, A_{k_{i+1}}^{i+1}\}$;
- 3) $\text{Prob}(C_i) < \text{Prob}(C_{i+1})$, $i = 1, 2, \dots, n-1$, где $\text{Prob}(C_i)$ – условная Вероятность (УВ) правила, $\text{Prob}(C_i) = \text{Prob}(P/A_1^i \& \dots \& A_{k_i}^i) = \text{Prob}(P \& A_1^i \& \dots \& A_{k_i}^i) / \text{Prob}(A_1^i \& \dots \& A_{k_i}^i)$;
- 4) C_i – Вероятностные Законы (ВЗ), т.е. для любого подправила $C' = (A_1 \& \dots \& A_j \Rightarrow P)$ правила C_i , $\{A_1, \dots, A_j\} \subset \{A_1^i, \dots, A_{k_i}^i\}$ выполнено неравенство $\text{Prob}(C') < \text{Prob}(C_i)$;
- 5) C_n – Сильнейший Вероятностный Закон (СВЗ), т.е. правило C_n не является подправилом никакого другого вероятностного закона.

Вероятностные неравенства в пунктах 3-4 проверяются на данных с помощью точного критерия независимости Фишера и критерия Юла [Кендалл, Стьюарт, 1973; Закс, 1975].

Множество всех цепочек СВВ предиката P образуют дерево СВВ предиката P .

Реализовать семантический вероятностный вывод в чистом виде не представляется возможным ввиду требований к производительности алгоритма, т.к. пункты 2 и 4 определения СВВ подразумевает большое пространство перебора. Для уменьшения перебора применяется следующие упрощения.

Во-первых, положим, что при построении цепочки СВВ правило C_{i+1} получается из правила C_i добавлением к его условной части только одного предиката. Эксперименты показывают, что крайне редка ситуация, когда добавление в условную часть правила сразу двух предикатов дает ВЗ, а добавление любого из этих двух признаков по отдельности не дает ВЗ. Следовательно, мы можем значительно уменьшить пространство перебора, почти не снижая количество и качество

извлеченных из данных закономерностей.

Во-вторых, для того чтобы уменьшить перебор при проверке условия в пункте 4, используется поуровневая схема генерация правил: сначала генерируются все ВЗ с одним предикатом в условной части и заключением Р, затем с двумя предикатами, тремя и т.д. Таким образом, для проверки, является ли некоторое правило ВЗ, достаточно просмотреть все его подправила, находящиеся на предыдущем уровне дерева СВВ.

Перед началом обучения колонки входной таблицы помечаются атрибутами Input, PredictOnly и Predict, которые указывают, каким образом та или иная колонка участвует в обучении: в качестве входного признака, целевого признака, или в качестве обоих одновременно. Также в качестве параметров модели могут задаваться пороговые величины: условная частота правила, уровни значимости критериев Фишера и Юла, максимальное число интервалов значений для признака и др.

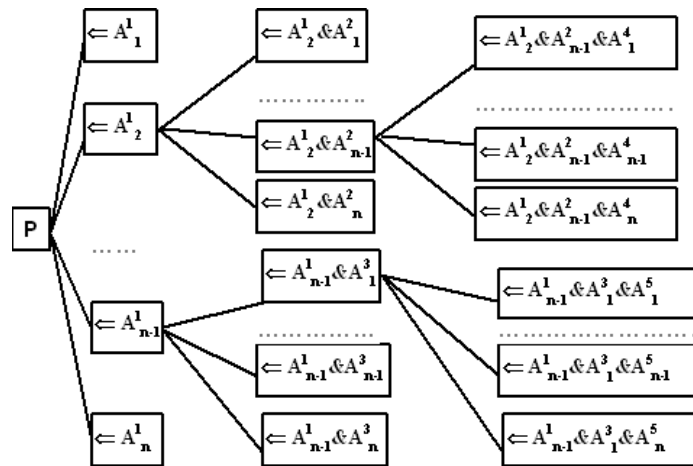


Рис. 1. дерево СВВ, включающее все СВВ, содержащие в заключении атом Р.

Результатом работы алгоритма является:

- 1) дерево СВВ для каждого целевого предиката;
- 2) множество ВЗ и СВЗ этих деревьев;
- 3) *Максимально Специфический Закон* (МСЗ) для каждого целевого предиката, определяемый как СВЗ, обладающий наибольшей условной вероятностью среди других СВЗ дерева вывода этого предиката.

Множество всех МСЗ обладает таким важным свойством как потенциальная непротиворечивость [Витяев, 2006].

Алгоритм поиска закономерностей:

1. Generate_First_Tree_Level (Queue Q, Tree T)

– Создаются все возможные правила, состоящие только из целевой части (они все по определению ВЗ). Для них сразу рассчитывается вся необходимая статистика на основе входных данных. Все элементы добавляются в корень дерева Т; элементы, содержащие Predict предикаты, добавляются в очередь Q.

2. Generate_Subsequent_Tree_Level (Queue Q, Tree T)

Обход дерева в ширину:

– Берем элемент из начала очереди Q, для него генерируем потомков, т.е. уточняем правило путем добавления 1-го нового предиката в условную часть.

– Проверяем, является ли новое правило ВЗ (см. Check_If_Probability_Low). Если является ВЗ, добавляем в дерево Т, добавляем в очередь Q.

– Процесс повторяется, пока очередь не пуста, т.е. еще можно получить новые ВЗ, путем добавления 1-го предиката в условную часть правила. Правила, соответствующие элементам которых не имеют потомков, являются СВЗ.

3. Check_If_Probability_Low (Rule R)

– Проверяем, является ли статистически значимым правило R с помощью критериев Фишера и Юла [Витяев, 2006, с. 117-120]. Если нет, то R не является ВЗ;

– просматриваем все подправила Sr длины Length(R) - 1

– Если Prob(Sr) > Prob(R), то R не является ВЗ;

– иначе R является ВЗ.

4. Extract_MSL (Tree T)

- Для каждого целевого предиката просматриваем его дерево СВВ. Сортируем множество СВЗ этого дерева по вероятности. Правила с наибольшей условной вероятностью являются Максимально Специфическими Законами (МСЗ) рассматриваемого целевого предиката.

Прогнозирование. Следующий алгоритм по набору предикатов, поданных на вход и множествам обнаруженных закономерностей ВЗ, СВВ и МСЗ, предсказывает значение целевого признака:

1. На вход подается некоторый набор предикатов. Ищутся все максимально специфичные закономерности, условная часть которых совпадает с данным набором либо с некоторым поднабором данного набора, а целевая часть содержит целевой признак. Максимально специфичная закономерность с наибольшей УВ определяет предсказанное значение целевого признака.
2. Если подходящих МСЗ не найдено, то рассматриваются все ВЗ с дерева СВВ целевого признака. Среди рассмотренных ВЗ ищутся те правила, условная часть которых совпадает с входным набором либо с некоторым поднабором входного набора. Правило с наибольшей УВ определяет предсказанное значение целевого признака.

5. Теоретическое сравнение с Association Rules и Decision Trees. В силу определения семантического вероятностного вывода в посылку правила всегда добавляются только такие предикаты, которые статистически значимо строго увеличивают условную вероятность правила. Такая проверка в других методах не проводится. В этом случае в правила могут включаться признаки, которые не имеют отношения к предсказанию и в том числе случайные. Опора на случайные признаки в предсказании может значительно ухудшить его точность.

Важность отсева случайных признаков критична, если нам надо не просто получить предсказание, а еще и проинтерпретировать полученные результаты. Специалист предметной области, интерпретируя правило, должен быть уверен, что признаки, входящие в правило действительно имеют отношение к предсказываемому признаку. Иначе интерпретация будет невозможна, либо специалист скажет, что правила бессмысленны и полученные предсказания – гадание на кофейной гуще и отчасти будет прав.

В Decision Trees для выбора разделяющих признаков используется энтропия, но на малых данных или в концах веток дерева могут включаться признаки, которые случайны и минимум энтропии признака получается чисто случайно. В Decision Trees нет статистического критерия проверки случайности добавляемых признаков. Поэтому правила, получаемые в Decision Trees, могут содержать случайные признаки, не имеющие отношение к предсказываемому признаку.

Для сравнения с алгоритмами Association Rules в качестве модельных данных возьмем следующие тестовые таблицы. Тестовая таблица 1 имеет три значимых колонки F_1, F_2, F_3 определяемые следующим выражением:

$$F_1(i) = \begin{cases} 1, i \in (1, 512k) \\ 0, i \in (512k+1, 1024k) \end{cases}, \quad 1024 - \text{количество записей в таблице.}$$

$$F_2(i) = \begin{cases} 1, i \in (1, 256k) \cup (512k+1, 768k) \\ 0, i \in (256k+1, 512k) \cup (768k+1, 1024k) \end{cases}$$

$$F_3(i) = \begin{cases} 1, i \in (1, 128k) \cup (256k+1, 384k) \cup (512k+1, 640k) \cup (768k+1, 896k) \\ 0, i \in (128k+1, 256k) \cup (384k+1, 512k) \cup (640k+1, 768k) \cup (896k, 1024k) \end{cases}$$

и колонку P, используемую в качестве целевого признака:

$$P(i) = \begin{cases} 1, i \in (1, 128k) \\ 0, i \in (128k+1, 1024k) \end{cases},$$

а также 5 колонок $R_1 - R_5$ с независимыми Бернуллиевскими случайными значениями.

Тестовая таблица 2 также имеет три значимых колонки F_1, F_2, F_3 определяемые следующим выражением:

$$F_1(i) = \begin{cases} 1, i \in (1, 729k) \\ 2, i \in (729k+1, 1458k) \\ 3, i \in (1458k+1, 2187k) \end{cases}$$

$$F_2(i) = \begin{cases} 1, i \in (1, 243k) \cup (729k+1, 972k) \cup (1458k+1, 1701k) \\ 2, i \in (243k+1, 486k) \cup (972k+1, 1215k) \cup (1701k+1, 1944k) \\ 3, i \in (486k+1, 729k) \cup (1215k+1, 1458k) \cup (1944k+1, 2187k) \end{cases}$$

$$F_3(i) = \begin{cases} 1, i \in \bigcup_{j=0}^2 \left((1+729kj, 81k+729kj) \cup (1+243k+729kj, 324k+729kj) \right) \\ \quad \cup (1+486k+729kj, 567k+729kj) \\ 2, i \in \bigcup_{j=0}^2 \left((1+81k+729kj, 162k+729kj) \cup (1+324k+729kj, 405k+729kj) \right) \\ \quad \cup (1+567k+729kj, 648k+729kj) \\ 3, i \in \bigcup_{j=0}^2 \left((1+162k+729kj, 243k+729kj) \cup (1+405k+729kj, 486k+729kj) \right) \\ \quad \cup (1+648k+729kj, 729k+729kj) \end{cases}$$

(2187 – количество записей в таблице) и колонку Р, используемую в качестве целевого признака:

$$P(i) = \begin{cases} 1, i \in (1, 81k) \\ 0, i \in (1+81k, 2187k) \end{cases},$$

а также 5 колонок $R_1 - R_5$ с независимыми Бернуллиевскими случайными значениями.

На тестовых таблицах можно увидеть две простые закономерности:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1) \Rightarrow P = 1, \quad (F_1 \neq 1 \wedge F_2 \neq 1 \wedge F_3 \neq 1) \Rightarrow P = 0.$$

Под тестовой таблицей с n-процентным шумом мы подразумеваем тестовую таблицу, в которой в n процентах ячеек, выбранных случайным образом, значение заменено на противоположное.

Хотя система «Discovery» и алгоритм Microsoft Association Rules достаточно похожи, в виду того, что в обоих подходах закономерности представляются в форме логических правил, тем не менее, между ними существуют принципиальные отличия.

В детерминированном случае, когда нет шума в данных, система «Discovery» обнаружит одно правило $A \& B \Rightarrow C$, истинное на данных. В то же время алгоритм, обнаруживающий ассоциативные правила, обнаружит все правила вида $A \& B \& \dots \& D \Rightarrow C$, которые получаются из правила $A \& B \Rightarrow C$ добавлением дополнительных условий D, F, \dots : $A \& B \& D \Rightarrow C$, $A \& B \& F \Rightarrow C$;

Например, при анализе тестовой таблицы 1, где в качестве входных колонок использовались F_1, F_2, F_3 , а также колонка R_1 со случайными данными, алгоритмом Association Rules было обнаружено правило $(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1) \Rightarrow P = 1$, а также следующие 2 правила с УВ = 1:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 1) \Rightarrow P = 1,$$

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 0) \Rightarrow P = 1.$$

Таким образом, в случае, когда цель анализа – найти закономерности в данных, эксперт, использующий алгоритм Association Rules, получит три противоречивых правила с УВ = 1. Доверия к таким результатам не будет.

Кроме того, алгоритмом Association Rules было обнаружено множество правил следующего вида:

$$(F_1 = 1 \wedge R_1 = 1) \Rightarrow P = 1, \quad (F_1 = 1 \wedge R_1 = 0) \Rightarrow P = 1,$$

$$(F_2 = 1 \wedge R_1 = 1) \Rightarrow P = 1, \quad (F_2 = 1 \wedge R_1 = 0) \Rightarrow P = 1,$$

$$(F_1 = 1 \wedge F_2 = 1 \wedge R_1 = 1) \Rightarrow P = 1, \quad (F_1 = 1 \wedge F_2 = 1 \wedge R_1 = 0) \Rightarrow P = 1.$$

Последние правила могут иметь приоритет над правилами с целевым предикатом $P = 0$, и, соответственно, ложно предсказывать 1, когда колонка Р содержит 0. Например, в случае, когда на вход подаются колонки F_1, F_2, F_3 и только одна колонка со случайными данными R_1 , процент правильно предсказанных алгоритмом Association Rules значений составит около 87%, когда на

вход подаются F_1, F_2, F_3 плюс две колонки R_1 и R_2 точность падает до 70% (подробнее см. параграф 5, таб. 1).

Система «Discovery» в данном случае обнаружила только следующие правила:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1) \Rightarrow P = 1, (F_1 = 0) \Rightarrow P = 0,$$

$$(F_2 = 0) \Rightarrow P = 0, (F_3 = 0) \Rightarrow P = 0,$$

т.к. они уже имеют $УВ = 1$, и добавление каких-либо предикатов в условную часть правила не может увеличить условную вероятность правила. Предсказание, основанное на этих правилах, будет 100% точным.

Когда есть шум в данных, система «Discovery» может обнаружить одно правило $A \& B \Rightarrow C$, представляющее собой вероятностный закон с определенным уровнем статистической значимости. В то же время алгоритм, обнаруживающий ассоциативные правила, должен обнаружить все детерминированные правила вида $A \& B \& D \Rightarrow C$, $A \& B \& F \Rightarrow C$, включающие случайные признаки, что приведет к ухудшению предсказания.

Например, при анализе тестовой таблицы 1 с 3% шумом, и колонками F_1, F_2, F_3, R_1 в качестве входных колонок, системой «Discovery» были обнаружены только 4 правила, являющиеся СВЗ:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1) \Rightarrow P = 1, (F_1 = 0) \Rightarrow P = 0,$$

$$(F_2 = 0) \Rightarrow P = 0, (F_3 = 0) \Rightarrow P = 0.$$

Эти правила не содержат колонку со случайными данными, т.к. любое правило, имеющее в условной части колонку R_1 не пройдет проверку критерием Юла-Фишера и будет удалено.

Алгоритм Association Rules в данном примере обнаружил правило $(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1) \Rightarrow P = 1$ с $УВ = p$, а также правила:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 1) \Rightarrow P = 1, (F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 0) \Rightarrow P = 1,$$

причем одно из них имеет $УВ > p$. Таким образом, в случае, когда цель анализа – найти закономерности в данных, эксперт, использующий алгоритм Association Rules, получит три противоречивых правила. При этом правило с наибольшей $УВ$ может содержать колонку со случайными данными.

Кроме того, из-за шума в данных алгоритм Association Rules обнаруживает множество правил следующего вида:

$$(F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 0 \wedge R_1 = 1) \Rightarrow P = 1, (F_1 = 1 \wedge F_2 = 1 \wedge F_3 = 0 \wedge R_1 = 0) \Rightarrow P = 1,$$

$$(F_1 = 1 \wedge F_2 = 0 \wedge F_3 = 1 \wedge R_1 = 1) \Rightarrow P = 1, (F_1 = 1 \wedge F_2 = 0 \wedge F_3 = 1 \wedge R_1 = 0) \Rightarrow P = 1,$$

$$(F_1 = 0 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 1) \Rightarrow P = 1, (F_1 = 0 \wedge F_2 = 1 \wedge F_3 = 1 \wedge R_1 = 0) \Rightarrow P = 1,$$

которые могут иметь приоритет над правилами с целевым предикатом $P = 0$ и ложно предсказывать 1, когда колонка P содержит 0. Это приводит к ухудшению предсказания (что показано на рис. 6 в приложении 4).

6. Экспериментальное сравнение Discovery с Association Rules. В качестве анализируемых данных используются тестовые таблицы, описанные выше. Для анализа данных таблиц применим систему «Discovery». В качестве входных колонок используем колонки $F_1, F_2, F_3, R_1 - R_5$, в качестве целевого признака – колонку P . В качестве критериев статистической значимости используемая реализация применяет точный критерий Фишера с пороговым значением 0,05 и критерий Юла с пороговым значением 0,1.

Система «Discovery» обнаруживает следующие правила с условной вероятностью ($УВ$), равной 1.

На тестовой таблице 1:

УВ 1: IF (F1 = 1) AND (F2 = 1) AND (F3 = 1) THEN (P = 1);

УВ 1: IF (F3 = 0) THEN (P = 0);

УВ 1: IF (F2 = 0) THEN (P = 0);

УВ 1: IF (F1 = 0) THEN (P = 0);

На тестовой таблице 2:

УВ 1: IF (F1 = 1) AND (F2 = 1) AND (F3 = 1) THEN (P = 1);

УВ 1: IF (F3 = 2) THEN (P = 0);

УВ 1: IF (F3 = 3) THEN (P = 0);

УВ 1: IF (F2 = 2) THEN (P = 0);
 УВ 1: IF (F2 = 3) THEN (P = 0);
 УВ 1: IF (F1 = 2) THEN (P = 0);
 УВ 1: IF (F1 = 3) THEN (P = 0);

Теперь проанализируем тестовые таблицы с помощью Association Rules. В качестве входных колонок используем колонки F_1, F_2, F_3 , в качестве целевого признака – колонку P.

Алгоритм Association Rules обнаруживает 20 правил с УВ, равной 1, на тестовой таблице 1 (57 на тестовой таблице 2), в том числе и все правила найденные системой «Discovery». При добавлении колонки R_1 ко входным колонкам Association Rules обнаруживает уже 60 правил с УВ, равной 1, на тестовой таблице 1 (228 на тестовой таблице 2). В Приложении 1 на рисунке 2 показано как растет количество правил с УВ, равной 1, обнаруженных алгоритмом Association Rules, при добавлении к F_1, F_2, F_3 колонок $R_1 - R_5$ в качестве входных колонок.

Далее посмотрим, как добавление в модель колонок со случайными данными ухудшает качество предсказания «Ассоциативных правил», а также покажем, что количество записей в таблице незначительно влияет на качество предсказания.

Таблица 1. Процент правильно предсказанных алгоритмом Association Rules значений на тестовой таблице 1

	0	1	2	3	4	5
Association Rules, n = 256	100%	87.5%	69.53%	45.70%	29.29%	19.53%
Association Rules, n = 1024	100%	87.59%	70.41%	45.31%	30.56%	23.92%
Association Rules, n = 4096	100%	88.15%	69.14%	47.07%	32.86%	24.43%

Таблица 2. Процент правильно предсказанных алгоритмом Association Rules значений на тестовой таблице 2

	0	1	2	3	4	5
Association Rules, n = 243	100%	91.81%	74.16%	58.46%	25.68%	16.26%
Association Rules, n = 2187	100%	91.95%	75.76%	55.05%	28.30%	17.92%
Association Rules, n = 19683	100%	91.06%	76.85%	55.19%	29.71%	18.46%

Отметим, что на тестовых таблицах 1 и 2 без шума система «Discovery» имеет 100% правильно предсказанных значений целевой колонки P при любом количестве случайных колонок $R_1 - R_5$, участвующих в обучении модели.

В Приложении 2 на рис. 3,4 проводится сравнение качества предсказания алгоритма Association Rules и системы «Discovery».

Рассмотрим тестовые таблицы 1 и 2 с наложением 3% шума. В качестве входных колонок используем колонки $F_1, F_2, F_3, R_1 - R_5$, в качестве целевого признака – колонку P.

В результате система «Discovery» обнаруживает следующие правила, являющиеся СВЗ.

На тестовой таблице 1:

УВ 0,854: IF (F1 = 1) AND (F2 = 1) AND (F3 = 1) THEN (P = 1)

УВ 0,959: IF (F2 = 0) THEN (P = 0)

УВ 0,944: IF (F1 = 0) THEN (P = 0)

УВ 0,945: IF (F3 = 0) THEN (P = 0)

Первые два из них являются МСЗ.

На тестовой таблице 2:

УВ 0,910: IF (F1 = 1) AND (F2 = 1) AND (F3 = 1) THEN (P = 1)

УВ 0,988: IF (F1 = 2) AND (F2 = 3) AND (F3 = 2) THEN (P = 0)

УВ 0,962: IF (F2 = 2) AND (F3 = 3) THEN (P = 0)

УВ 0,967: IF (F1 = 3) AND (F2 = 2) THEN (P = 0)

УВ 0,971: IF (F1 = 3) AND (F3 = 3) THEN (P = 0)

УВ 0,977: IF (F1 = 3) AND (F2 = 3) THEN (P = 0)

Первые два из них также являются МСЗ.

Проанализируем тестовые таблицы 1 и 2 с наложением 3% шума с помощью Association Rules. В качестве входных колонок используем колонки F_1, F_2, F_3 , в качестве целевого признака – колонку P. В результате, на тестовой таблице 1 Association Rules обнаруживает 29 правил, из них 20 правил с УВ > 0.85, в том числе и все правила, найденные системой «Discovery». На тестовой таблице 2 Association Rules обнаруживает 60 правил с УВ > 0.85, в том числе и все правила, найденные системой «Discovery». В Приложении 3 на рис. 5 показано, как растет количество правил с УВ > 0.85, обнаруженных алгоритмом Association Rules, при добавлении колонок $R_1 - R_5$ в качестве входных колонок.

В Приложении 4 проводится сравнение качества предсказания алгоритма Association Rules и системы Discovery на тестовой таблице 2 с шумом 0%, 2% и 3%.

7. Экспериментальное сравнение Discovery с Decision Trees. В качестве анализируемых данных используются таблицы следующего вида:

F0	F1	F2	F3	F4	P
1	1	0	0	0	1
1	1	1	0	0	1
1	1	1	1	0	1
0	1	1	1	1	1
0	0	1	1	1	1
0	0	0	1	1	1
1	0	1	0	0	0
1	0	1	0	0	0
1	0	1	0	0	0
1	0	1	0	0	0
1	0	0	1	0	0
1	0	0	1	0	0
0	1	0	1	0	0
0	1	0	1	0	0
0	1	0	1	0	0
0	1	0	1	0	0
0	1	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0
0	0	1	0	1	0
0	0	1	0	1	0
0	0	1	0	1	0
0	0	0	1	0	0
0	0	0	1	0	0
0	0	0	0	1	0

Для обучения алгоритмов Discovery и Decision Trees использовались таблица1, полученная из таблицы, приведенной выше, масштабированием в 8 раз (и удалением нескольких записей в конце таблицы), так, что итоговая длина тестовой таблицы1 равна 197.

Сначала для анализа таблицы1 применим систему «Discovery». В качестве входных колонок используем колонки $F_0 - F_4$, в качестве целевого признака – колонку P. В качестве критериев статистической значимости применим точный критерий Фишера с пороговым значением 0,13 и критерий Юла с пороговым значением минус 0,03.

Система «Discovery» обнаруживает следующие правила с условной вероятностью (УВ) равной 1 предсказывающие $P = 1$

УВ 1: IF (F0 = 1) AND (F1 = 1) THEN (P = 1);

УВ 1: IF (F1 = 1) AND (F2 = 1) THEN (P = 1);

УВ 1: IF (F2 = 1) AND (F3 = 1) THEN (P = 1);

УВ 1: IF (F3 = 1) AND (F4 = 1) THEN (P = 1).

А также несколько правил, предсказывающие $P = 0$. Далее проанализируем таблицу1 с помощью Decision Trees. В качестве входных колонок используем колонки $F_0 - F_4$, в качестве целевой – колонку P.

Параметр COMPLEXITY_PENALTY для Decision Trees установим равным 0.001. Алгоритм Decision Trees обнаружил следующие правила, предсказывающие $P = 1$:

УВ 0.965: IF (F0 = 1) AND (F1 = 1) THEN (P = 1);

УВ 0.982: IF (F1 = 0) AND (F3 = 1) AND (F4 = 1) THEN (P = 1);

А также несколько правил, предсказывающие $P = 0$. Нетрудно заметить, что найденных Decision Trees правил не достаточно для точного предсказания $P = 1$.

Для сравнения качества предсказания Discovery и Decision Trees будем использовать тестовую таблицу2, полученную масштабированием таблицы1 в 4 раза. Введем следующую меру ошибки: если алгоритм ошибочно предсказывает 1 вместо 0, то стоимость ошибки равна 1, если ошибочно предсказывает 0 вместо 1, то стоимость ошибки равна 3. Данную меру ошибки можно представить в виде таблицы:

Actual\Predicted	0	1
0	0	1
1	3	0

Максимальная ошибка, возможная на тестовой таблице2 имеет стоимость 1172. В результате выполнения предсказания для записей тестовой таблицы 2, Discovery не допустила ни одной ошибки, т.е. предсказала абсолютно точно. Decision Trees на тестовой таблице2 допустила ошибку общей стоимостью 96, или 8.2% от стоимости максимальной ошибки.

Проанализируем, как добавление шума в таблицы влияет на обнаруживаемые закономерности

и ухудшает качество предсказания алгоритмов Discovery и Decision Trees.

Под таблицей с n -процентным шумом мы подразумеваем таблицу, в которой в n процентах ячеек, выбранных случайным образом, значение признака заменено на противоположное.

На тестовой таблице 1 с шумом 1% система «Discovery» обнаруживает следующие правила, предсказывающие $P = 1$:

УВ 0.958: IF ($F_0 = 1$) AND ($F_1 = 1$) THEN ($P = 1$);

УВ 0.962: IF ($F_1 = 1$) AND ($F_2 = 1$) THEN ($P = 1$);

УВ 0.960: IF ($F_2 = 1$) AND ($F_3 = 1$) THEN ($P = 1$);

УВ 1.000: IF ($F_3 = 1$) AND ($F_4 = 1$) THEN ($P = 1$);

А также правила, предсказывающие $P = 0$. На этой же таблице Decision Trees обнаруживает следующие правила, предсказывающие $P = 1$:

УВ 0.982: IF ($F_1 = 0$) AND ($F_3 = 1$) AND ($F_4 = 1$) THEN ($P = 1$);

УВ 0.982: IF ($F_0 = 1$) AND ($F_1 = 1$) AND ($F_2 = 1$) THEN ($P = 1$);

УВ 0.878: IF ($F_0 = 0$) AND ($F_1 = 1$) AND ($F_2 = 1$) THEN ($P = 1$);

А также несколько правил, предсказывающие $P = 0$. В следующей таблице показаны результаты предсказаний, выполненных алгоритмами Discovery и Decision Trees. Алгоритмы обучались на таблице 1 с шумами в 1, 2 и 3 процента. Предсказание выполнялось для записей таблицы 2 с теми же шумами 1, 2, 3 процента.

	Discovery		Decision Trees	
	Стоимость	% от max	Стоимость	% от max
шум 0%	0	0	96	8.19
шум 1%	40	3.38	128	10.83
шум 2%	92	7.7	179	14.99
шум 3%	123	10.12	198	16.28

Во всех приведенных случаях система Discovery работала лучше Decision Trees.

ЛИТЕРАТУРА

Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов. – НГУ, Новосибирск, 2006. – 293 с.

Закс Ш. Теория статистических выводов – М.: Мир, 1975. – 776 с.

Кендалл М. Дж., Стьюарт А. Статистические выводы и связи – М.: Наука, 1973. – 899 с.

Halpern J. Y. An analysis of first-order logic of probability. – Artificial Intelligence. 1990. – С. 311–350.

Kovalerchuk, B.Y., Vityaev E.E. Data mining in finance: Relational and hybrid methods. Kluwer, 2000. 308 p.

Tang Z., MacLennan J. Data Mining with SQL Server 2005. Wiley Publishing, Inc., 2005. 483 p.

Vityaev E., Kovalerchuk B. Empirical Theories Discovery based on the Measurement Theory // Mind and Machine, 2006. V.14. N4, P. 551-573.

Vityaev E. The logic of prediction. In: Mathematical Logic in Asia // Proceedings of the 9th Asian Logic Conference (August 16-19, 2005, Novosibirsk, Russia). World Scientific, Singapore, 2006. P. 263-276.

Приложение 1

На следующем графике показано, как растет количество правил с УВ, равной 1, обнаруженных алгоритмом Association Rules, при добавлении к F_1, F_2, F_3 колонок $R_1 - R_5$ в качестве входных с колонок.

Здесь и далее на горизонтальной оси отмечено количество колонок $R_1 - R_5$ со случайными данными, используемых (в дополнение к F_1, F_2, F_3) в качестве входных данных.

Как видим, количество правил, найденных Association Rules, экспоненциально растет при добавлении новых колонок.

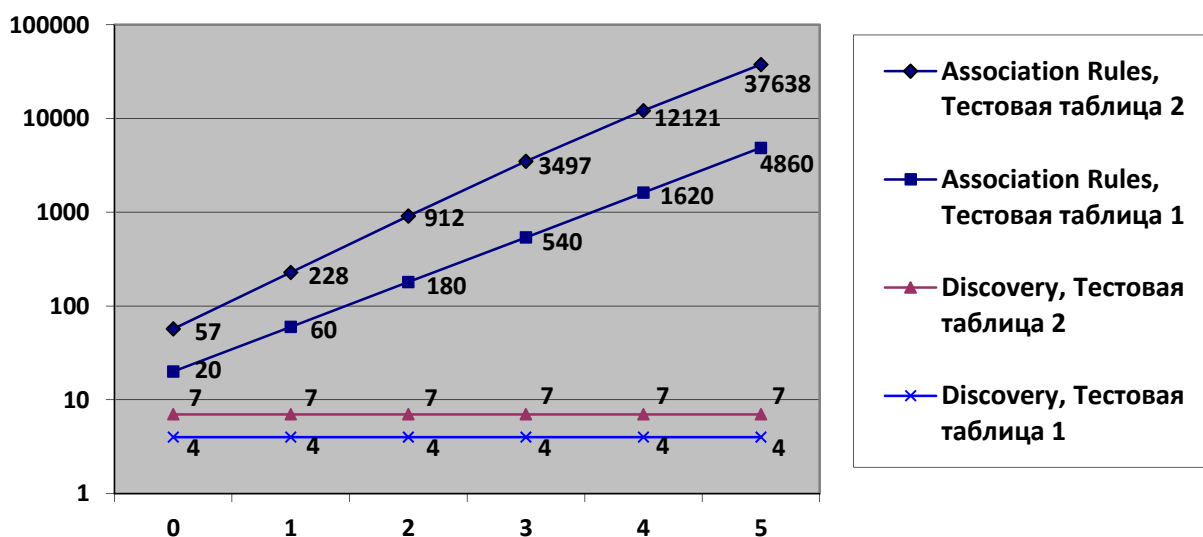


Рис. 2. Количество правил с УВ = 1, обнаруженных алгоритмом Association Rule

Приложение 2

На следующем графике показан процент правильно предсказанных значений алгоритма Association Rules и системы «Discovery» в зависимости от количества колонок $R_1 - R_5$ со случайными данными, используемых в качестве входных данных, а также размера тестовой таблицы.

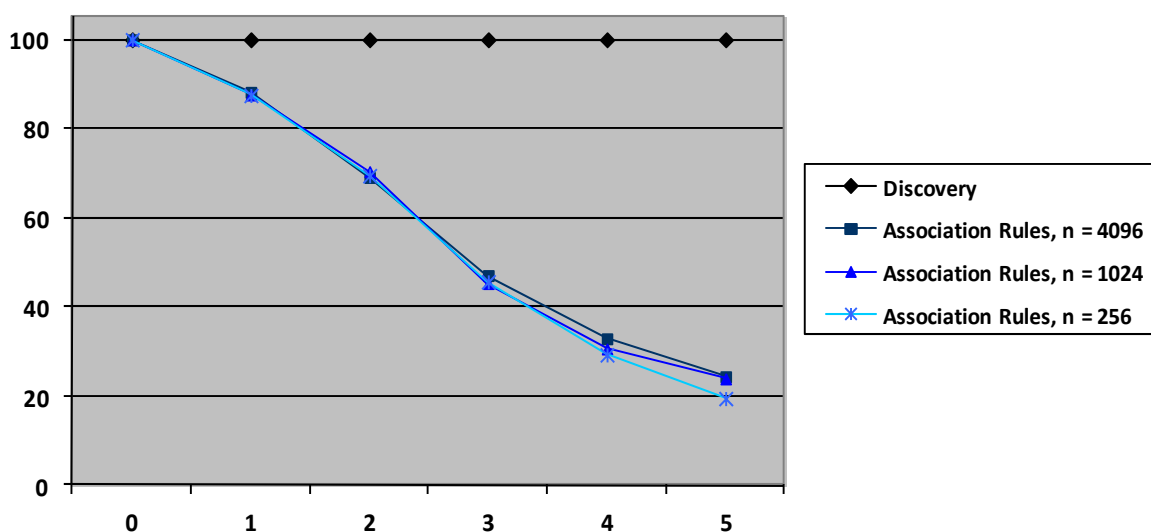


Рис. 3. Процент правильно предсказанных значений при анализе тестовой таблицы 1.

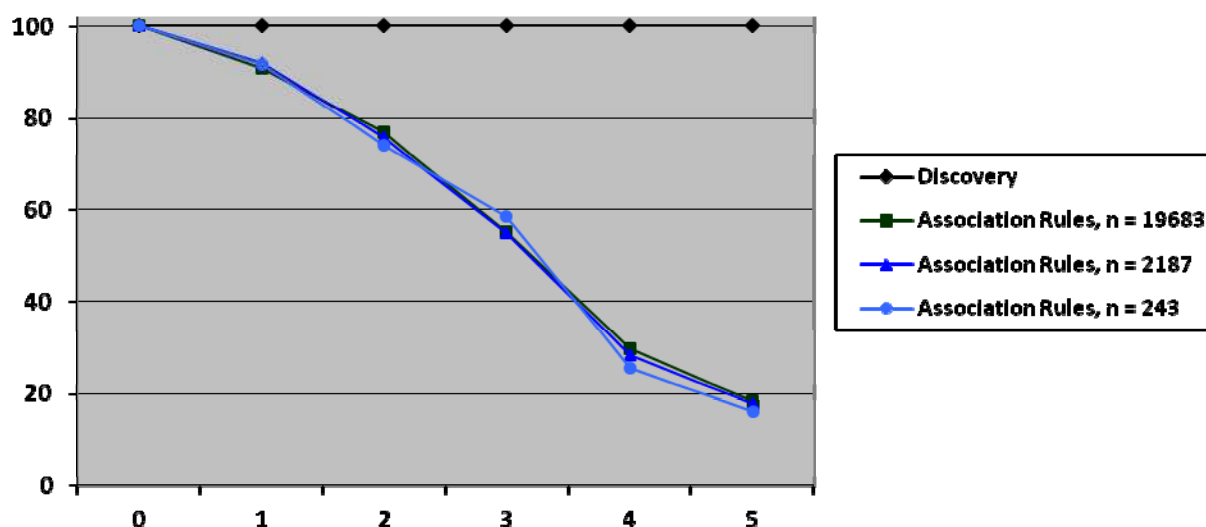


Рис. 4. Процент правильно предсказанных значений при анализе тестовой таблицы 2.

Как видим, при увеличении числа колонок $R_1 - R_5$ со случайными данными, используемых в качестве входных данных, качество предсказания алгоритма Association Rules значительно падает. Размер тестовой таблицы незначительно влияет на результат.

Приложение 3

На следующем графике показано, как растет количество правил с $UB > 0.85$, обнаруженных алгоритмом Association Rules, при добавлении колонок $R_1 - R_5$ в качестве входных колонок. Анализируются тестовые таблицы с наложением 3% шума.

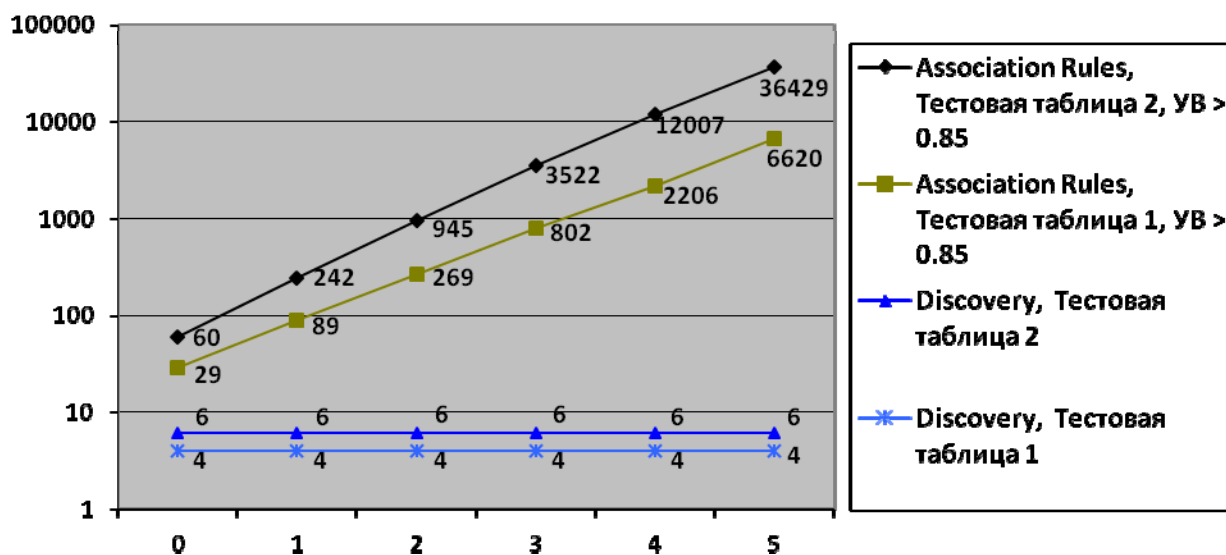


Рис. 5. Количество правил, обнаруженных алгоритмом Association Rules и системой Discovery.

Приложение 4

Сравним качество предсказания системы “Discovery” и алгоритма Association Rules на тестовой таблице 2 с шумом 0%, 2% и 3%. Как видим, система “Discovery” дает наиболее близкое к идеальному предсказание, причем качество прогнозирования не зависит от количества колонок со случайными данными, используемых в качестве входных данных, а зависит только от величины шума. Заметим, что модель, обученная с помощью алгоритма “Discovery” на данных с шумом, будет давать 100% верные предсказания на данных без шума. Алгоритм Association Rules дает аналогичное Discovery качество предсказания в случае, когда случайные колонки не участвуют в

обучении модели. При добавлении в модель случайных колонок $R_1 - R_5$ качество прогнозирования Association Rules значительно падает.

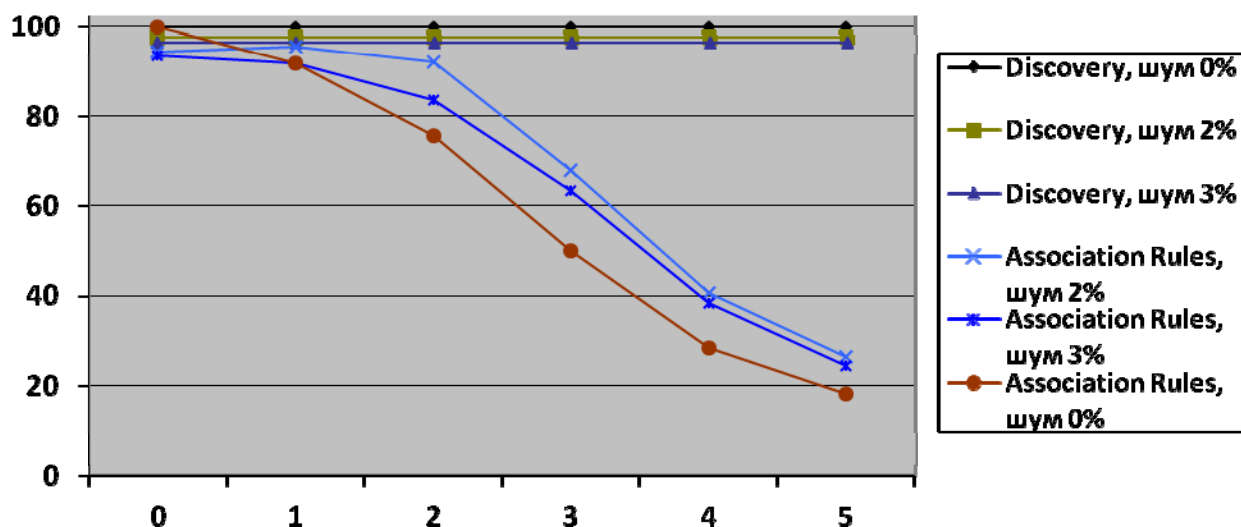


Рис. 6. Процент правильно предсказанных значений при анализе тестовой таблицы 2

Заметим, что качество предсказания Association Rules на данных без шума при добавлении 2 и более колонок $R_1 - R_5$ заметно хуже, чем на данных с шумом 2% или 3%. Это объясняется тем, что на данных без шума Association Rules обнаруживает огромное количество «равноправных» правил с $UB = 1$, выбрать верное из которых не представляется возможным.

Витяев Е.Е.³,
Неупокоев Н.В.

Математическая модель восприятия и образа

Существующие формализации образов и зрительного восприятия никак не связаны с психологией восприятия и с когнитивными процессами порождения образов. Тем самым нет связи между воспринимаемыми объектами и теми образами, которые они порождают. Для моделирования процессов порождения образов, их изменений и развития в процессе восприятия внешнего мира нужна адекватная формализация образа и восприятия, которая бы основывалась на психологии восприятия. В данной работе предлагается такая формализация, где образ и восприятие рассматриваются, в соответствии с существующими представлениями, как непрерывный процесс предвосхищения (предсказания) образом поступающих стимулов и проверка предсказаний на соответствие реальным стимулам. Показывается, что формализацией этого процесса являются неподвижные точки предсказаний. В работе приведена математическая модель неподвижных точек, алгоритм их обнаружения и эксперименты, демонстрирующие возможности данной модели. Проведено сравнение с сетями Хопфилда, которые являются наиболее похожими на неподвижные точки нейронными сетями.

Ключевые слова: Образ, восприятие, предсказание, интеллектуальный анализ данных, Data Mining.

1. Введение. Теория Функциональных Систем П.К.Анохина начинается с принципа опережающего отражения действительности. Мозг непрерывно во времени предвосхищает события окружающей среды и одновременно контролирует акцептором результатов действия правильность сделанных предсказаний.

В восприятии предвосхищение (антиципация) непрерывно во времени сравнивает «образ» («образ мира») с наличной стимуляцией и является процессом активного движения от «образа» к внешнему миру – непрерывным во времени процессом проверки предсказаний «образа» на соответствие стимулам внешнего мира. Только если все многочисленные предсказания будут совпадать с реальными стимулами непрерывно во времени, только тогда есть восприятие [1]. «Все это позволяет нарисовать следующую картину хода познавательной деятельности на уровне воспри-

³ Эта работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026, Российским Фондом Фундаментальных Исследований № 11-07-00560-а, интеграционными проектами СО РАН № 3, 87, 136 и программой президента Российской Федерации поддержки научных школ НШ-276.2012.1.

ятия. Индивид всегда имеет некоторый образ или модель окружения, которая непрерывна во времени и пространстве и носит прогностический характер, т.е. в ней экстраполируются и воспроизводятся на языке чувственных модальностей ожидаемые результаты воздействия источника стимула на наши органы чувств» [1].

У. Найсером восприятие описано схемами и перцептивными циклами. «Если образы суть предвосхищения, они должны облегчать последующее восприятие. Перцептивная готовность это не скромный побочный продукт визуализации, это её суть. Иметь перцептивную установку в отношении чего-либо это значит иметь образ». «По моему мнению, важнейшими для зрения когнитивными структурами являются предвосхищающие схемы, подготавливающие индивида к принятию информации строго определенного, а не любого вида и, таким образом, управляющие зрительной активностью. ... В каждый момент воспринимающим конструируются предвосхищения некоторой информации, делающие возможным для него принятие её. ... Эта исследовательская активность направляется все теми же предвосхищающими схемами, представляющими собой своего рода планы перцептивных действий. ... Термин «восприятие» относится ко всему циклу, а не к какой-то отдельной его части» [2].

Насколько известно авторам, в настоящее время нет формализации восприятия, как непрерывного во времени процесса предвосхищения стимулов воспринимаемого объекта и проверки этих предвосхищений на соответствие стимулам внешнего мира. В данной работе мы предлагаем такую формализацию в виде неподвижных точек предвосхищений.

Наиболее похожими подходами, в которых также есть неподвижная точка или «резонанс», являются работы Гроссберга и Хопфилда. Сети Хопфилда и экспериментальное сравнение с ними приведено далее в разделе 6.

В работах Гроссберга впервые говорится о «резонансе» между имеющейся моделью и поступающими стимулами – идентификация и распознавание объекта получаются в результате взаимодействия («резонанса») ожиданий «сверху-вниз» и сенсорной информацией «снизу-вверх» [3]. Принципиальное отличие нашего подхода от подхода Гроссберга состоит в том, что в нашей модели «резонанс» в виде неподвижной точки рассматривается внутри модели, а не вне её.

2. Формальная модель образа как неподвижной точки предвосхищений. Введем некоторые определения. Далее под предвосхищением будем понимать предсказание, а под схемой – совокупность закономерностей, предсказывающих, что будет воспринято в следующий момент времени при выполнении определенных перцептивных действий. Перефразируем высказывание Найсера в терминах закономерностей: в каждый момент времени воспринимающим извлекается из памяти весь опыт по восприятию данного объекта в виде совокупности закономерностей $Mem = \{P_1 \& \dots \& P_k \& A \Rightarrow P_0\}$. Эти закономерности означают, что, если мы воспринимаем признаки $P_1 \& \dots \& P_k$, то после осуществления перцептивного действия A , переводящего взгляд на признак P_0 , мы воспримем значение признака P_0 .

Предположим, что объекты восприятия определяются набором значений признаков (стимулов) x_1, \dots, x_n . Каждый признак x_i принимает некоторое множество значений $I_i = \{x_1^i, \dots, x_{k_i}^i\}$, $i = 1, \dots, n$. Будем предполагать, что признаки на объектах могут принимать по несколько значений. Тогда воспринимаемый объект, a описывается совокупностью подмножеств значений воспринимаемых признаков $X(a) = \{X_{j_1}(a), \dots, X_{j_m}(a)\}$, $X_{j_1}(a) \subset I_{j_1}, \dots, X_{j_m}(a) \subset I_{j_m}$, $X_{j_s}(a) \neq \emptyset$, $s = 1, \dots, m$. Восприниматься могут не все n признаков, а только m из них. Для каждого стимула (некоторого значения признака) определим предикат $P_j^i(a) \Leftrightarrow (x_j^i \in X_i(a))$. Предикат может быть с отрицанием $\bar{P}_j^i(a)$ или без него $P_j^i(a)$. Отрицание предиката в посылке правила означает, что нет данного стимула в поступающей информации. Предсказание отрицания предиката означает торможение соответствующего стимула. Предикат, который может быть, как с отрицанием, так и без него, обозначим через $\hat{P}_j^i(a)$.

Закономерности определим как высказывания вида $\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow \hat{P}_{j_0}^{i_0}$, в котором действие опущено, поскольку по значениям i_0, j_0 предиката $\hat{P}_{j_0}^{i_0}$ всегда ясно, на что надо обратить внимание и куда перевести взгляд (на признак i_0 и его значение j_0). Смысл действия в том, чтобы изменить воспринимаемую стимуляцию. В закономерности фактически записано, какое действие должно быть осуществлено – надо от восприятия признаков $\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k}$ перейти к восприятию признака

$\hat{P}_{j_0}^{i_0}$. Закономерность представляет собой формализацию действия и одновременно предвосхищение результата действия. Все закономерности из Мет обнаруживаются в процессе самообучения семантическим вероятностным выводом, рассмотренным в разделе 4.

Будем говорить, что закономерность $\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow \hat{P}_{j_0}^{i_0}$ *извлекается из памяти* при восприятии объекта a , если посылка $\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k}$ закономерности и, значит, все предикаты, входящие в нее, становятся истинными на объекте a , т.е. когда $x_{j_s}^{i_s} \in X_{i_s}(a)$, если предикат $P_{j_s}^{i_s}$ не имеет отрицания и $x_{j_s}^{i_s} \notin X_{i_s}(a)$, если предикат $\bar{P}_{j_s}^{i_s}$ имеет отрицание, $s = 1, \dots, k$. Обозначим через $LP(X(a)) \subseteq \text{Мет}$ множество закономерностей, извлекаемых из памяти при восприятии объекта a . Если закономерность $\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow \hat{P}_{j_0}^{i_0}$ применима к воспринимаемому объекту a и её заключение $\hat{P}_{j_0}^{i_0}$ истинно в указанном смысле ($x_{j_0}^{i_0} \in X_{i_0}(a)$, если предикат $P_{j_0}^{i_0}$ не имеет отрицания и $x_{j_0}^{i_0} \notin X_{i_0}(a)$, если предикат $\bar{P}_{j_0}^{i_0}$ имеет отрицание), то будем говорить, что предвосхищение, осуществляемое закономерностью, *подтвердилось* на объекте a , в противном случае *опроверглось*.

Восприятие объекта a – это непрерывный цикл предсказаний одних свойств объекта по другим свойствам посредством всех извлечённых из памяти закономерностей и проверка того, что все эти закономерности подтвердились. В этом случае перцептивный цикл завершен, и все предсказания стимулов $\hat{P}_{j_0}^{i_0}$ совпали с наличной стимуляцией. Если предсказывается отрицание стимула, то его не должно быть в образе – это процесс вытормаживания стимулов. Возможные противоречия в предсказаниях закономерностей решаются на основании специального критерия Крит максимальной согласованности предсказаний, приведённого в разделе 4.

Восприятие, как непрерывный цикл перцептивных действий и предсказаний вместе с проверкой предсказаний на совпадение с реальными стимулами, формально может быть описан неподвижной точкой предсказаний (своеобразным «резонансом» закономерных связей).

Определим оператор предсказания Pr , применённый к некоторому множеству стимулов $X = \{X_{i_1} \cup \dots \cup X_{i_m}\}$ (это могут быть признаки $X(a)$ объекта a , либо целая картина или сцена, описываемая множеством признаков X). Используя все «извлечённые из памяти» закономерности $(\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow P_{j_0}^{i_0}) \in LP(X)$ этот оператор предсказывает наличие стимулов $x_{j_0}^{i_0}$ в восприятии, если предикат $P_{j_0}^{i_0}$ не имеет отрицания и предсказывает отсутствие стимула $\bar{x}_{j_0}^{i_0}$, если предикат $\bar{P}_{j_0}^{i_0}$ имеет отрицание. Тогда оператор предсказания Pr может быть записан следующим образом:

$$Pr(X) = \Phi_{\text{Крит}}(X \cup \{x_{j_0}^{i_0} \mid (\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow P_{j_0}^{i_0}) \in LP(X)\} \cup \{\bar{x}_{j_0}^{i_0} \mid (\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow \bar{P}_{j_0}^{i_0}) \in LP(X)\}),$$

где функция $\Phi_{\text{Крит}}$ минимизирует возможные противоречия в предсказаниях, используя специальный критерий согласованности Крит закономерностей по предсказанию. Подробное определение Крит и $\Phi_{\text{Крит}}$ приведено в разделе 4.

В восприятии осуществляется не один цикл предвосхищений, как показывают эксперименты с тахистоскопом. Циклы предвосхищений должны пройти несколько раз, чтобы исчезли противоречия между предвосхищениями и реальными стимулами. Когда это достигнуто, то восприятие объекта a завершено и мы имеем неподвижную точку оператора Pr . Если восприятие началось с восприятия стимулов $X(a)$ объекта a , то после нескольких итераций предвосхищения оператором Pr получим неподвижную точку – образ объекта a . Обозначим n -кратное применение оператора Pr через Pr^n . Тогда восприятие стимулов объекта a в виде неподвижной точки оператора Pr будет определяться равенством $Pr^{n+1}(X(a)) = Pr^n(X(a))$, где n – этап стабилизации предвосхищений. Более детальное описание этого процесса дано в разделе 4.

Приведём машинные эксперименты, иллюстрирующие введенные понятия на примере закодированных цифр.

3. Эксперименты по автоматическому обнаружению неподвижных точек цифр. Проиллюстрируем нахождение неподвижных точек восприятия с помощью цифр. Закодируем цифры горизонтальными, вертикальными и диагональными линиями, расположенными в квадратах от 1 до 24 как указано в табл. 1. Вместе с белым пространством они образуют 7 признаков клетки. На рис. 1 видно, что есть некоторые разрывы в цифрах, но на результат восприятия это никак не по-

влияет. Цифры с добавленной сеткой демонстрируют, что некоторые цифры кодируются однозначно, а другие, как например 4 или 1, имеют избыточную кодировку (вертикальная линия состоит из двух кодов).

Нами было проведено два эксперимента.

3.1. **Первый эксперимент** был проведен на обучающем множестве **A**, состоящем из 360 цифр (12 цифр рис. 1 дублировались в 30-ти экземплярах), которое представлялось как единое множество объектов без указания, где какая цифра. Эксперимент состоял в том, чтобы найти все неподвижные точки на этом множестве и убедиться, что они порождают образы, соответствующие нашим 12 цифрам.

Таблица 1. Кодировка полей цифр.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24

На обучающем множестве **A** было обнаружено 55089 закономерностей. Алгоритм обнаружения закономерностей приведен в разделе 4. По этим закономерностям оператором предсказания **P_г** было построено 12 неподвижных точек, которые в точности соответствовали нашим цифрам. Алгоритм обнаружения неподвижных точек также приведен в разделе 4. Таким образом, на данном множестве образов алгоритм обнаружения неподвижных точек в точности решает задачу обнаружения заданных образов.

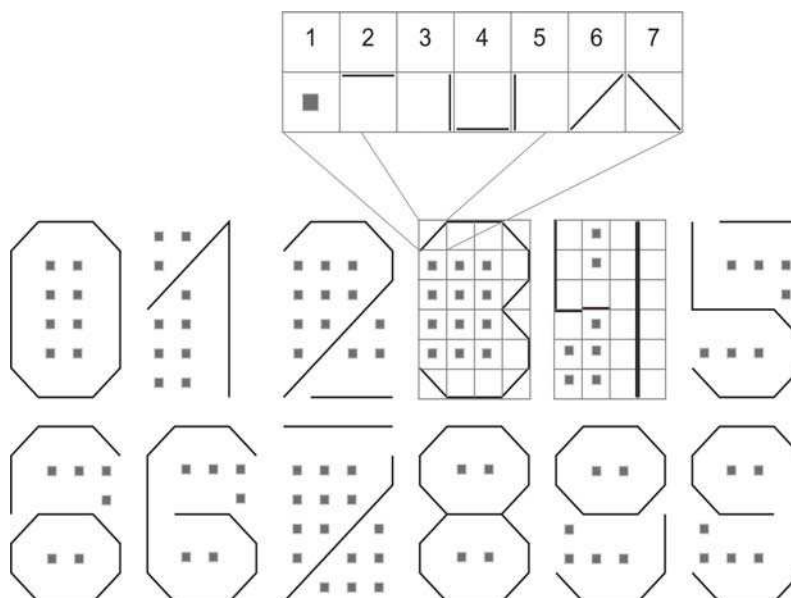


Рис. 1. Кодировка цифр.

Рассмотрим на примере цифры 6 (рис. 2), что представляет собой неподвижная точка. Первая закономерность цифры 6, представленная в первом прямоугольнике после фигурной скобки, говорит, что, если в квадрате 13 (табл. 1) стоит признак 6 (обозначим это как 13-6), то в квадрате 3 должен стоять признак 2 (обозначим также как (3-2)). Предсказываемый признак обозначается точечной линией. Запишем эту закономерность как $(13-6 \Rightarrow 3-2)$. Нетрудно проверить по цифрам, что эта закономерность действительно имеет место на множестве **A**. Вторая закономерность говорит, что из признака (9-5) и отрицания значения 5 первого признака $\neg(1-5)$ (первый признак не должен быть равен 5) следует признак (4-7). Отрицание $\neg|$ значения $|$ с номером 5 некоторого признака или отрицание $\neg/$ значения $/$ с номером 6 некоторого другого признака обозначается пунктирной линией (точка-пунктир), как показано в нижней части рисунка 2. В результате получаем закономерность $(9-5 \& \neg(1-5) \Rightarrow 4-7)$. Последующие 3 закономерности в первой строке цифры 6 на рис. 2 будут соответственно закономерности $(13-6 \Rightarrow 4-7)$, $(17-5 \& \neg(13-5) \Rightarrow 4-7)$, $(13-6 \Rightarrow 16-7)$.

На рис. 2 видно, что признаки цифры 6 взаимно предсказывают друг друга, образуя неподвижную точку.

3.2. **Второй эксперимент** состоял в том, чтобы обнаружить эти же образы в условиях недостатка информации. В качестве данных было взято множество **B**, включающее те же 12 цифр в 30-ти экземплярах каждая, но с пробелами в информации – в каждой цифре отсутствовал один слу-

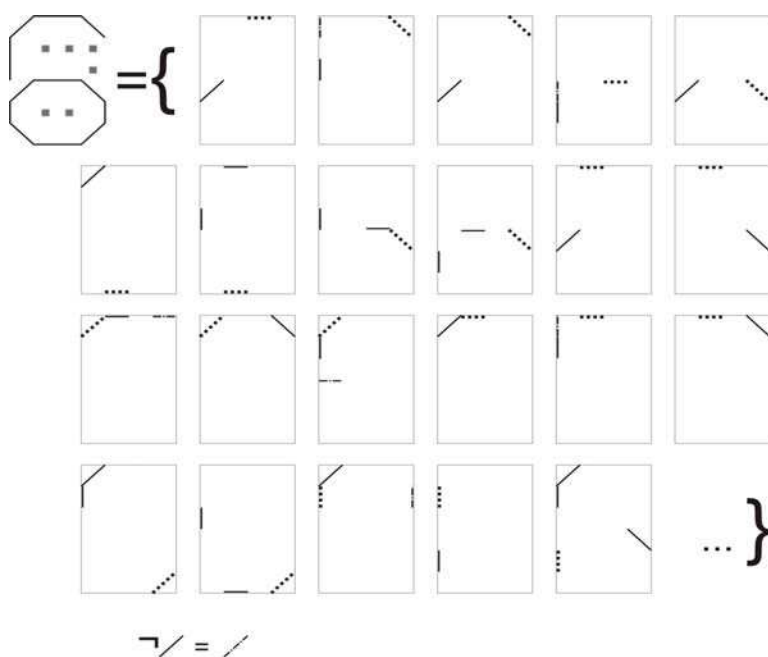


Рис. 2. Неподвижная точка цифры 6.

взаимного предсказания признаков цифр требуются закономерности в более точном условии, содержащем больше предикатов в посылке правила.

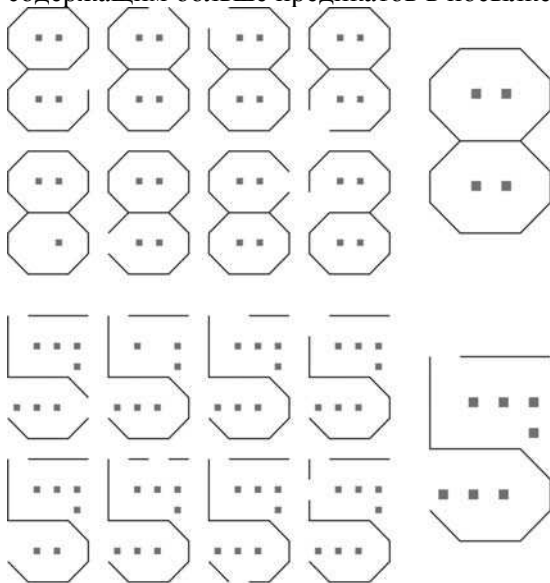


Рис. 3. Цифры с отсутствующими признаками.

предсказывающие различные значения признака в квадрате с недостающей информацией. Эти противоречия разрешаются критерием максимальной согласованности предсказаний и оператором Φ_{Krit} модификации стимулов (см. определения в разделе 4). Критерий согласованности закономерностей в этом случае отдает предпочтение более сильному предсказанию (по четырем закономерностям второй строки) вертикальной линии, игнорируя (вытормаживая) предсказание косой линии. При этом критерий Φ_{Krit} учитывает и другие закономерности, которые взаимно предсказывают признаки цифры 5, как например, третья и четвертая закономерности первой строки рис. 6. Тогда цифра 5 восстанавливается полностью и правильно. В нашем эксперименте все 360 цифр с пробелами в информации были распознаны правильно.

4. Математическое определение неподвижной точки. Для полного определения оператора предсказания Pr , приведенного в конце 2-го раздела, необходимо определить множество закономерностей Mem , критерий максимальной согласованности предсказаний Krit и функцию Φ_{Krit} .

случайно выбранный признак, как показано на рис. 3. К полученным 360 цифрам была добавлена отрицательная выборка, состоящая из 1050 объектов со случайными признаками. На этих данных было обнаружено 73458 закономерностей. По этим закономерностям оператором предсказания Pr также были обнаружены неподвижные точки, которых оказалось 14. Из них 12 цифр это цифры рис. 1 и для цифр 6 и 9 ещё по одной неподвижной точке, содержащих пробелы в 13-м признаке для цифры 6 и в 12-м признаке для цифры 9.

На рис. 4 показана неподвижная точка цифры 6, полученная по этим данным. На рис. 4 видно, что закономерности здесь более сложные, чем на рис. 2 так как данные более сложные и для более надёжного

Отнесение некоторой цифры с недостающим признаком к некоторому образу (распознавание цифры) осуществляется тем же оператором предсказания Pr , путем построения неподвижной точки, отправляясь от признаков X этой цифры.

Например, для цифры 5 первыми двумя закономерностями верхней строки рис. 5 предсказывается недостающий признак – хвостик в верхнем правом углу, которого нет в исходной цифре.

На второй и третьей строке рис. 5 представлены закономерности, предсказывающие признак в квадрате 1. Во второй строке представлены закономерности, которые предсказывают значение признака 5 (вертикальная линия), а в третьей строке приведена единственная обнаруженная закономерность, предсказывающая значение признака 6 (косая линия) для первого квадрата. В процессе распознавания и восстановления недостающей информации в цифре, могут возникать противоречия, как в данном случае,

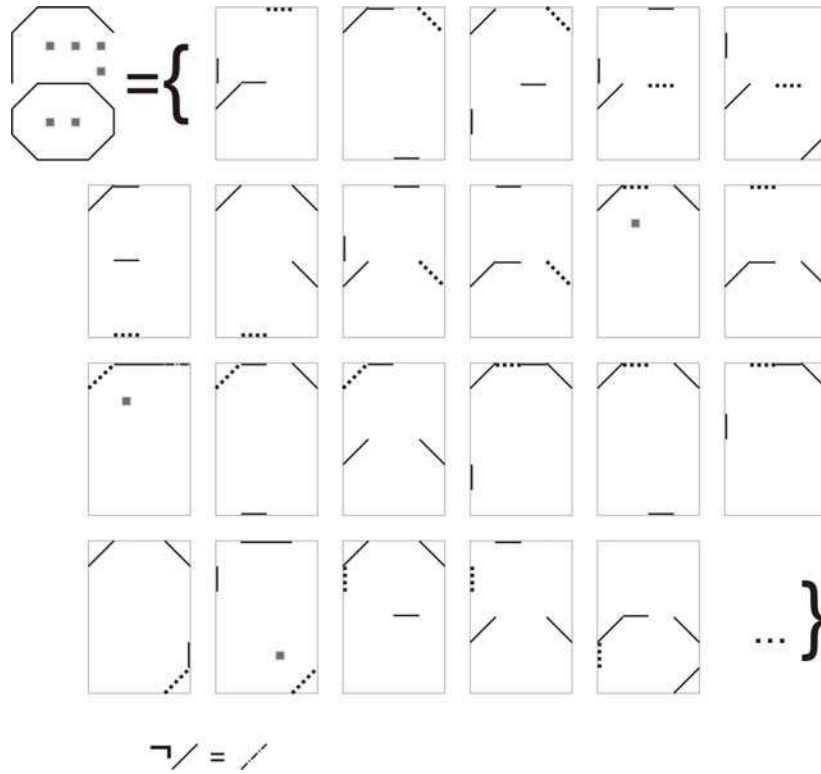


Рис. 4. Неподвижная точка цифры 6 по данным с неполной информацией.

Обозначим через $\Pi = \{\hat{P}_j^i, i = 1, \dots, n; j = 1, \dots, k_i\}$ множество всех предикатов, фиксирующих поступающие стимулы (напомним, что под обозначением \hat{P}_j^i мы понимаем либо сам предикат P_j^i , либо его отрицание \bar{P}_j^i).

Все закономерности из Мем получаются *семантическим вероятностным выводом* (СВВ) [4-6]. Описательно СВВ рассмотрен в [6], где показано, что он формализует правило Хебба образования условных связей на уровне нейрона. Семантический вероятностный вывод устроен так, что он автоматически включает в закономерность все стимулы, которые могут усилить предсказание (увеличить условную вероятность) интересующего нас стимула (например, того, на который будет переведён взгляд).

Формально СВВ определяется как последовательность правил $R_1 \sqsubset R_2 \sqsubset \dots \sqsubset R_m$, которые от номера к номеру все сильнее предсказывают интересующий нас стимул, заданный предикатом $P_0 \in \Pi$. Эта должна удовлетворять следующим условиям:

1. $R_i = (P_1^i \& \dots \& P_{k_i}^i \Rightarrow P_0), P_j^i \in \Pi, i = 1, \dots, m$;
2. R_i – подправило правила R_{i+1} , т.е. $\{P_1^i, \dots, P_{k_i}^i\} \subset \{P_1^{i+1}, \dots, P_{k_{i+1}}^{i+1}\}$;
3. $\text{Prob}(R_i) < \text{Prob}(R_{i+1}), i = 1, 2, \dots, m-1$, где $\text{Prob}(R_i) = \text{Prob}(P_0^i / P_1^i \& \dots \& P_{k_i}^i)$ – условная вероятность правила;
4. R_i – *вероятностные законы*, т.е. для любого подправила $R' = (P_1 \& \dots \& P_k \Rightarrow P_0)$ правила R_i , $\{P_1, \dots, P_k\} \subset \{P_1^i, \dots, P_{k_i}^i\}$ выполнено неравенство $\text{Prob}(R') < \text{Prob}(R_i)$;
5. R_m – *максимально специфический закон*, для которого цепочка правил $R_1 \sqsubset R_2 \sqsubset \dots \sqsubset R_m$ не может быть продолжена, для R_m не существует правила R_{m+1} удовлетворяющего условиям 1-4.

Предикат $P_0 \in \Pi$ может предсказываться различными семантическими вероятностными выводами, поэтому полное множество правил предсказывающих предикат $P_0 \in \Pi$ образует решетку $\text{Lat}(P_0)$ семантических вероятностных выводов. Полное множество закономерностей $\text{Mem}(P_0)$, которые участвуют в предсказании предиката $P_0 \in \Pi$ состоит из всех вероятностных законов, входящих в $\text{Lat}(P_0)$. Вся память Мем есть объединение всех закономерностей $\text{Mem}(P_0)$ для всех предсказываемых предикатов $P_0 \in \Pi$.

Алгоритм обнаружения закономерностей из Мем в точности следует данному определению и состоит из следующих шагов:

1. На вход алгоритма подается обучающее множество воспринятых объектов $A = \{a_1, \dots, a_m\}$, для которых известны все стимулы в виде значений всех предикатов из Π ;
2. для обнаружения всех вероятностных законов, входящих в $\text{Lat}(P_0)$, фиксируем некоторый предикат $P_0 \in \Pi$;
3. для предиката $P_0 \in \Pi$ осуществляем базовый перебор глубины d правил множества $\text{Base}_d(P_0) = B_1(P_0) \cup \dots \cup B_d(P_0)$, где $B_1(P_0) = \{(P \Rightarrow P_0), P \in \Pi\}$, ..., $B_d(P_0) = \{(P_1 \& \dots \& P_d \Rightarrow P_0), P_1, \dots, P_d \in \Pi\}$;
4. перебираем все правила из каждого множества $B_i(P_0)$, $i = 1, \dots, d$ и удаляем те из них, которые не являются вероятностными законами. Для проверки, является ли некоторое правило вероятностным законом, применяем специальный оператор LP , проверяющий выполнение пункта 4 определения СВВ. При этом вероятностные неравенства $\text{Pr ob}(R') < \text{Pr ob}(R_i)$ из пункта 4 проверяем по имеющимся данным точным критерием независимости Фишера [7]. В результате получим множества $BP_i(P_0) = LP(B_i(P_0))$, $i = 1, \dots, d$ правил содержащих только вероятностные законы. В результате базового перебора глубины d получим множество $\text{Mem}(P_0) = BP_1(P_0) \cup \dots \cup BP_d(P_0)$;
5. Далее осуществляем направленный перебор правил, который состоит в уточнении правил из $BP_d(P_0)$. Для этого правила наращиваются оператором уточнения T следующим образом:

$$T(BP_d(P_0)) = \{(P'_1 \& \dots \& P'_{d+1} \Rightarrow P_0) \mid (P_1 \& \dots \& P_d \Rightarrow P_0) \in BP_d(P_0), \{P_1, \dots, P_d\} \subset \{P'_1, \dots, P'_{d+1}\}, P_1, \dots, P_{d+1} \in \Pi\}$$
;
6. К уточнённым правилам снова применяем оператор проверки на вероятностные законы LP и получим новое множество вероятностных законов $P_{d+1}(P_0) = LP(T(BP_d(P_0)))$, выходящее за базовый перебор.
7. Дальнейшее наращивание множества вероятностных законов $\text{Mem}(P_0)$ происходит рекурсивно применением операторов T и LP , $P_{d+1}(P_0) = LP(T(P_{d+1-1}(P_0)))$, $i > 1$. Наращивание заканчивается на этапе t , когда уже не хватает данных для статистической проверки критерия Фишера. Тогда $P_{d+t}(P_0) = LP(T(BP_{d+t-1}(P_0))) = P_{d+t-1}(P_0)$;
8. Все множество вероятностных законов $\text{Mem}(P_0)$ получается объединением всех вероятностных законов, полученных как в результате базового перебора, так и направленного:

$$\text{Mem}(P_0) = BP_1(P_0) \cup \dots \cup BP_d(P_0) \cup P_{d+1}(P_0) \cup \dots \cup P_{d+t-1}(P_0)$$
;
9. Для определения всего множества Мем, переходим к пункту 2 и выбираем новый, не рассмотренный ранее предикат $P_0 \in \Pi$.

Определим функцию Φ_{Krit} минимизации возможных противоречий в предсказаниях и крите-

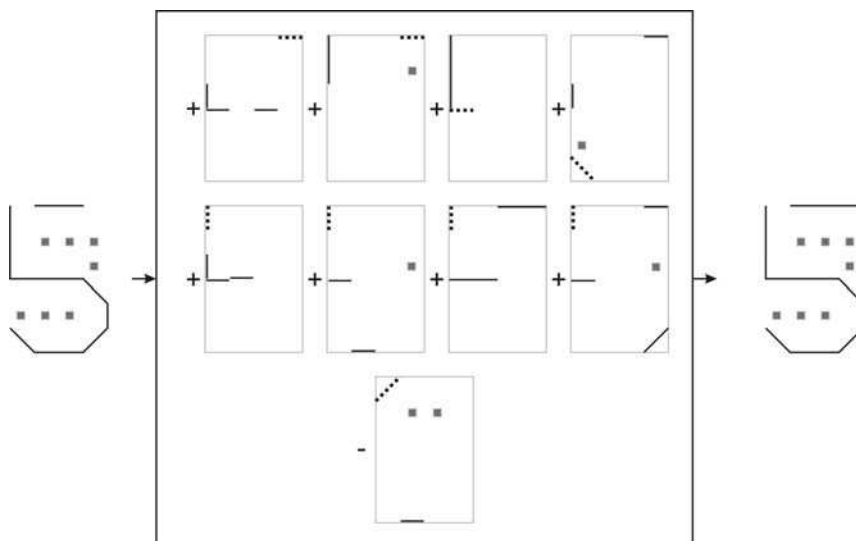


Рис. 5. Восстановление образа посредством неподвижной точки восприятия.

рий согласованности предсказаний K_{rit}.

Оператор предсказания Pr предсказывает два множества стимулов, которые должны присутствовать $x_{j_0}^{i_0}$ и которые должны отсутствовать $\bar{x}_{j_0}^{i_0}$:

$$\text{Pr}^+(X) = \{x_{j_0}^{i_0} \mid (\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow P_{j_0}^{i_0}) \in \text{LP}(X)\}, \text{Pr}^-(X) = \{\bar{x}_{j_0}^{i_0} \mid (\hat{P}_{j_1}^{i_1} \& \dots \& \hat{P}_{j_k}^{i_k} \Rightarrow \bar{P}_{j_0}^{i_0}) \in \text{LP}(X)\}.$$

Тогда оператор предсказания Pr примет вид: $\text{Pr}(X) = \Phi_{\text{Krit}}(X \cup \text{Pr}^+(X) \cup \text{Pr}^-(X))$

Определим функцию Φ_{Krit} на множестве стимулов $Y = X \cup \text{Pr}^+(X) \cup \text{Pr}^-(X)$. Она либо добавляет один элемент $x_{j_0}^{i_0}$ во множество X , либо удаляет один элемент $\bar{x}_{j_0}^{i_0}$ из множества X . При этом, она учитывает предсказания не отдельных правил из $\text{LP}(X)$, а их взаимную согласованность по предсказаниям (их закономерный «резонанс»).

Для этого используется специальный критерий K_{rit} взаимной согласованности закономерностей по предсказанию, который определяется следующим образом. Пусть $S(X) \subset \text{LP}(X)$ – множество закономерностей, подтверждающихся на интересующем нас наборе стимулов X , а $F(X) \subset \text{LP}(X)$ – множество закономерностей, опровергающихся на наборе X . Тогда критерий K_{rit} есть сумма весов подтверждающихся закономерностей минус сумма весов опровергающихся закономерностей:

$$\text{Krit}(X) = \sum_{R \in S(X)} \mu(R) - \sum_{R \in F(X)} \mu(R), \text{ где } \mu(R) = -\log(1 - \text{Pr ob}(R)).$$

Функция $-\log(1 - \text{Pr ob}(R))$ учитывает не саму вероятность, а её близость к 1. Логарифм берется потому, что рассматривается логарифм критерия и величины суммируются, а не перемножаются.

Функция Φ_{Krit} , при добавлении/удалении какого-то элемента $x_{j_0}^{i_0} / \bar{x}_{j_0}^{i_0}$ во множество X , должна строго увеличивать взаимную согласованность всех применимых к $X \cup x_{j_0}^{i_0}$ или к $X \setminus \bar{x}_{j_0}^{i_0}$ закономерностей и должно выполняться либо неравенство $\text{Krit}(X) < \text{Krit}(X \cup x_{j_0}^{i_0})$, либо неравенство $\text{Krit}(X) < \text{Krit}(X \setminus \bar{x}_{j_0}^{i_0})$. В противном случае множество X остается без изменений. В обоих случаях нас интересует такое добавление/удаление элемента, которое максимально увеличивает критерий. Эти величины равны соответственно:

$$\delta^+(X) = \max_{x_{j_0}^{i_0} \in \text{Pr}^+(X)} \{\text{Krit}(X \cup x_{j_0}^{i_0}) - \text{Krit}(X)\}, \delta^-(X) = \max_{\bar{x}_{j_0}^{i_0} \in \text{Pr}^-(X)} \{\text{Krit}(X \setminus \bar{x}_{j_0}^{i_0}) - \text{Krit}(X)\}.$$

Функция Φ_{Krit} добавляет/удаляет элемент из множества X , который максимизирует соответствующее значение. Эти элементы определяются следующим образом:

$$x_{j_0}^{i_0}(X) = \arg \max_{x_{j_0}^{i_0} \in \text{Pr}^+(X)} (\text{Krit}(X \cup x_{j_0}^{i_0})), \bar{x}_{j_0}^{i_0}(X) = \arg \max_{\bar{x}_{j_0}^{i_0} \in \text{Pr}^-(X)} (\text{Krit}(X \setminus \bar{x}_{j_0}^{i_0}))$$

При каждом применении оператора предсказания Pr функция Φ_{Krit} не одновременно добавляет/удаляет элемент из множества X , а выбирает тот из них, который максимально увеличивает критерий, т.е. добавляет элемент $x_{j_0}^{i_0}(X)$, если $\delta^+(X) > \delta^-(X)$, $\delta^+(X) > 0$ и удаляет элемент $\bar{x}_{j_0}^{i_0}(X)$, если $\delta^-(X) > \delta^+(X)$, $\delta^-(X) > 0$.

Итак, функция модификации Φ_{Krit} определяется следующим образом:

$$\Phi_{\text{Krit}}(X) = \begin{cases} X \cup x_{j_0}^{i_0}(X), & \text{если } \delta^+(X) > \delta^-(X), \delta^+(X) > 0, \\ \quad x_{j_0}^{i_0}(X) = \arg \max_{x_{j_0}^{i_0} \in \text{Pr}^+(X)} (\text{Krit}(X \cup x_{j_0}^{i_0})) \\ X \setminus \bar{x}_{j_0}^{i_0}(X), & \text{если } \delta^-(X) \geq \delta^+(X), \delta^-(X) > 0, \\ \quad \bar{x}_{j_0}^{i_0}(X) = \arg \max_{\bar{x}_{j_0}^{i_0} \in \text{Pr}^-(X)} (\text{Krit}(X \setminus \bar{x}_{j_0}^{i_0})) \\ X, & \text{если } \delta^+(X) \leq 0 \text{ и } \delta^-(X) \leq 0. \end{cases}$$

Неподвижная точка $\text{Pr}^{n+1}(X) = \text{Pr}^n(X)$ получается в третьем случае, когда добавление/удаление элемента не увеличивает критерий.

Алгоритм обнаружения неподвижных точек точно следует данным определениям:

1. на вход алгоритма подается множество воспринятых в данный момент стимулов X и множество закономерностей $\text{LP}(X)$;

2. обнаруживаются множества $Pr^+(X)$ и $Pr^-(X)$;
3. применяется оператор предсказания $Pr(X) = \Phi_{\text{крит}}(X \cup Pr^+(X) \cup Pr^-(X))$ к множеству X , в результате чего оно изменяется на один элемент;
4. применяем оператор предсказания до тех пор пока не получим неподвижную точку $Pr^{n+1}(X) = Pr^n(X)$.
5. переходим к пункту 1 и выбираем новое множество X . Выбор различных совокупностей множеств X зависит от решаемой задачи. Нами брались последовательно множества $X(a) \in A$ для экспериментов со множеством A и $X(a) \in B$ для экспериментов со множеством B .

5. Сравнение с сетями Хопфилда. Нами выбрана нейронная рекуррентная сеть Хопфилда, в силу того, что она обладает цикличностью, т.е. неоднократной обработкой входного сигнала, в

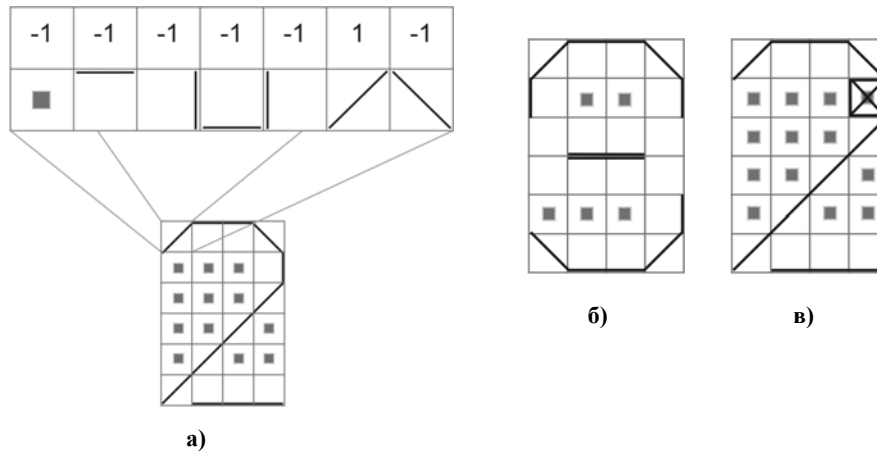


Рис. 6. а) Замена каждого символа на соответствующий биполярный блок;
 б) распознавание сети Хопфилда со сведением всех объектов к одному классу и
 в) созданием классов со специфическим признаком

результате которой сигнал сводится к неподвижной точке или точке равновесия [8]. Сеть в режиме обучения перебирает все тренировочные объекты, осуществляя вычисление матрицы весов

$$w_{ij}^1 = \frac{1}{n} x_i^1 x_j^1 (i \neq j), \text{ где } n - \text{длина входного сигнала, равная числу нейронов в сети.}$$

Для каждого следующего объекта вычисляется матрица и складывается с полученной ранее: $W = W^1 + W^2 + \dots + W^l$, где l – число объектов в тренировочном множестве. Таким образом, матрица весов формируется по формуле $w_{ij} = (1 - \delta_{ij}) \frac{1}{n} \sum_{k=1}^l x_i^k x_j^k$.

Все вычисления по распознаванию сводятся к простому умножению входного вектора на матрицу весов $h(0) = y, h(1) = \text{sign}(Wy), h(2) = \text{sign}(Wh(1))$, пока не будет найдена неподвижная точка $\text{Output} = h(t+1) = h(t)$. Существование неподвижной точки гарантируется уменьшением энергии сети $E(t) = -\frac{1}{2} \sum w_{ij} h_i(t) h_j(t)$. Для каждой итерации d верно $E(d+1) - E(d) \leq 0$, откуда следует, что существует t такое, что $h(t+1) = h(t) = \text{sign}(Wh(t))$.

Для сетей Хопфилда возможна катастрофа памяти, которая происходит в результате переполнения сети, когда обучение сети состоит из большого числа классов, и поэтому на некотором этапе после обучения матрица связей “разрушается”, т.е. теряется суперпозиция предыдущих шагов обучения, которую хранила матрица W . В [9] указывается, что число различных объектов, которые может запомнить сеть без катастрофы памяти равно $m(n) = n / 2 \log n$.

В работе [10] исследован способ обхода катастрофы памяти. Этим метод память продолжает функционировать и за пределами границы, но имеет свойство “забывать” паттерны с низким порогом, и для восстановления забытых паттернов их нужно предъявлять заново, чтобы установить новый порог. Такой способ поведения сети неэффективен.

Покажем, что и в нашем примере происходит катастрофа памяти. Проведем соответствующий эксперимент. Для этого проинтерпретируем 7-значный вектор признаков, кодирующий наши буквы, как биполярный (см. рис. 6). Для этого каждое значение признака представим набором из семи

единиц и минус единиц, где (-1) означает отсутствие значения, (1) – наличие значения (рис. 6.а). Тогда, например, вместо признака 6 получим набор (-1 -1 -1 -1 -1 1 -1), вместо 2 получим набор (-1 1 -1 -1 -1 -1 -1). В итоге вместо, например, вектора (6 2 2 7 1 1 1 3 1 1 1 6 1 1 6 1 1 6 1 1 6 4 4 4), описывающую цифру в порядке номеров клеток в соответствии с таблицей 1, получим вектор длины 168 вида

(-1 -1 -1 -1 -1 1 -1 | -1 1 -1 -1 -1 -1 -1 | -1 1 -1 -1 -1 -1 -1 | -1 -1 -1 -1 -1 -1 1...)

Преобразуем наши обучающие данные **A** и **B** в соответствии с данной кодировкой. Мы провели два эксперимента при $n=168$ на полученных обучающих данных:

1. Классический вариант сети сводит все объекты, как обучающего множества **A**, так и обучающего множества **B** к одному классу, не совпадающему ни с одним объектом этих множеств (рис. 6.б). Пустая клетка означает, что все значения признаков равны -1;

2. Сеть из пакета MATLAB [11] справляется с обучающими данными **A** и даёт для них 12 классов, как и алгоритм неподвижных точек, но не справляется с данными **B**. Для данных **B** все пропуски информации превращаются алгоритмом в специфические признаки, когда все признаки равны 1, т.е. в клетке с пропуском одновременно присутствуют все семь признаков (рис. 6.в). Так как пропуски есть в каждой цифре, то во втором эксперименте сеть пакета MATLAB даёт столько классов, сколько цифр во множестве **B**.

Результаты сведены нами в таблицу 2.

Таблица 2. Результаты сравнения с сетью Хопфилда

Тип	Обучение	Результат
Классическая сеть Хопфилда	A	1 класс
	B	1 класс
newhor из MATLAB	A	12 классов
	B	Для каждого объекта порождается свой класс
Алгоритм неподвижных точек	A	12 классов
	B	14 классов

6. Выводы

Рассмотренный в работе пример с цифрами, хотя и является модельным, но тем не менее является достаточно сложным, что позволяет говорить о возможности решения нетривиальных прикладных задач алгоритмом неподвижных точек. Первоначально данный алгоритм рассматривался нами как алгоритм «естественной» классификации и имел самостоятельное обоснование (см. работы по «естественной» классификации на сайте [12]). В таком виде алгоритм успешно применялся для решения задач биоинформатики [13].

ЛИТЕРАТУРА

- Витяев Е.Е.** Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов // Новосибирский гос. ун-т. Новосибирск, 2006. с.293.
- Витяев Е.Е., Перловский Л.И., Ковалерчук Б.Я., Сперанский С.О.** Вероятностная динамическая логика мышления. Нейроинформатика, 2011, том 5, № 1, стр. 1-20
- Карандашев Я.М., Крыжановский Б.В., Литинский Л.Б.** Взвешенные паттерны и устранение «катастрофы памяти» в модели Хопфилда. //XIV Всероссийская научно-техническая конференция “Нейроинформатика-2012” : Лекции по нейроинформатике. - М.: НИЯУ МИФИ, 2012. с.152-189.
- Кендал М., Стьюарт А.** Статистические выводы и связи. М.: Наука, 1973. С. 899.
- Найсер У.** Познание и реальность. “Прогресс”, М. 1981, с. 229.
- Смирнов С.Д.** Психология образа. МГУ, М., 1985, с.231
- Carpenter, G.A. & Grossberg, S.** Adaptive Resonance Theory, In Michael A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, Second Edition, Cambridge, MA: MIT Press, 2003, pp. 87-90.
- Evgenii Vityaev.** The logic of prediction. In: Mathematical Logic in Asia. Proceedings of the 9th Asian Logic Conference (August 16-19, 2005, Novosibirsk, Russia), edited by S.S. Goncharov, R. Downey, H. Ono, World Scientific, Singapore, 2006, pp.263-276
- John Hertz, Anders Krogh, Richard G. Palmer** Introduction To The Theory Of Neural Computation. - Addison-Wesley Pub. Co., 1991 P.327.
- Li J., A.N. Michel, W. Porod** Analysis and synthesis of a class of neural networks: linear systems operating on a closed hypercube. IEEE Transactions on Circuits and Systems, Vol. 36, No. 11, November 1989, pp. 1405-1422.
- McEliece, Robert J. and Posner, Edward C. and Rodemich, Eugene R. and Venkatesh, Santosh S.** The capacity of the Hopfield associative memory. IEEE Transactions on Information Theory, 33 (4). 1987 pp. 461-482.
- Scientific Discovery:** <http://www.math.nsc.ru/LBRT/logic/vityaev>
- Vityaev E.E., Lapardin K.A., Khomicheva I.V., Proskura A.L.** Transcription factor binding site recognition by regularity matrices based on the natural classification method. *Intelligent Data Analysis*. v.12(5), IOS Press, 2008, 495-512.

КОММЕНТАРИЙ

К статье Витяева Е.Е., Неупокоева Н.В. «Математическая модель восприятия и образа»

В работе приводится формализация психологической модели активного восприятия и осуществлена практическая проверка предложенной формализации в серии машинных экспериментов. Предложена достаточно оригинальная формализация, основанная на неподвижной точке предсказаний, представляющей собой процесс «непрерывного во времени и пространстве» процесса предсказания признаков воспринимаемого объекта.

Проведен достаточно нетривиальный машинный эксперимент по проверке предложенной модели и показано, что она точно классифицирует определенным образом закодированные цифры.

Существующие модели восприятия основаны на нейронных сетях. Поэтому в работе проведено сравнение с нейронными сетями, которые также в определенном смысле основаны на неподвижной точке – рекуррентными сетями Хопфилда. Говорится, что сети Хопфилда страдают так называемой «катастрофой памяти», которая проявилась и на примере с закодированными цифрами – все цифры были сведены в один класс. Поэтому пример, на котором была продемонстрирована работоспособность модели, оказался недоступен для сетей Хопфилда, что достаточно нетривиально. Сравнение с сетями Хопфилда важно, поскольку в настоящее время, например, в проекте «Россия 2045» и рабочих совещаниях на конференциях по «Нейроинформатике», они рассматриваются как одни из наиболее перспективных моделей «обратного конструирования мозга». Поэтому проведенный эксперимент показывает, что эти ожидания могут не оправдаться.

В целом, данная работа достаточно интересна и своевременна. В качестве замечания следует отметить, что, в силу важности сравнения с сетями Хопфилда, его стоит провести многосторонне – для различных значений параметров и вариаций сетей Хопфилда, а также для других примеров задач. Разработанный метод также надо проверить на множестве прикладных задач.

Демин А.В., м.н.с. ИСИ СО РАН

Мазов Н.А.,
Гуреев В.Н.

Библиометрия: современное состояние, задачи, инструменты

Применение современных методов объективной оценки результатов деятельности ученых и специалистов приобретает сегодня особое значение для российской науки. В условиях прогрессирующего развития научной информации становятся все более актуальными разработка и использование новых механизмов для её структурирования, анализа и оценки, одним из которых является библиометрия. В настоящей статье дана характеристика современного состояния, решаемых задач и инструментариев науки библиометрии.

Ключевые слова: библиометрия, библиометрический анализ, анализ цитирования, практическое использование библиометрического анализа.

Общая характеристика

Молодую науку библиометрию можно охарактеризовать как приложение математических и статистических методов к публикациям [Egghe, Rousseau, 1990; Бредихин, Кузнецов, 2012; Bibliometrics..., 2008]. Библиометрическая оценка основана на допущении, что большинство научных открытий и результатов исследований в итоге публикуются в международных научных журналах, где их могут прочесть и процитировать другие ученые. Коммуникация, т. е. обмен результатами исследований, является движущей силой науки, и научные публикации выступают как очень важные элементы в этом процессе обмена знаниями. Работа высокого качества вызывает реакции коллег ученых. Количество цитат журнальной статьи можно рассматривать как отражение влияния статьи на научное сообщество.

Прикладная библиометрия анализирует количество научных статей, опубликованных выбранным количеством авторов, цитаты к этим статьям и связь между статьями, авторами и дисциплинами.

Место библиометрии среди других инструментов оценки

Влияние научного исследования, а соответственно, автора и организации, можно измерить многими способами: количественные подходы включают подсчет публикаций, количество прибыли от исследования, число аспирантов, размер исследовательской группы, количество просмотров и загрузок результатов исследования в режиме онлайн, число полученных патентов и лицензий, выигранных грантов и пр. Результаты библиометрического и цитатного анализа – лишь часть из множества количественных показателей. При этом возможность применять этот анализ и его значимость в общей оценке исследования изменяется от одной области к другой [Игра, 2011].

Отдельно стоит обозначить острый вопрос взаимоотношений между библиометрией, с одной стороны, и рецензированием, экспертной оценкой – с другой. Баланс между библиометрией и рецензированием в оценке научных результатов, как на индивидуальном, так и на общем уровне в настоящее время является одной из наиболее обсуждаемых тем в научном и библиотечном сообществе [Raaij, 2003]. Учитывая количественную природу результатов, у библиометрии могут быть некоторые преимущества, поскольку она дает возможность получить хорошие результаты путем достаточно быстрого генерирования различных видов статистических данных в сравнении с ресурсоемкой природой рецензирования качества и инноваций интеллектуальной работы. Библиометрия, являясь направлением с очень краткой историей, достаточно быстро стала альтернативой отработанному столетиями научному рецензированию.

Недаром в последние годы вопросам библиометрического (наукометрического, вебометрического и т.п.) анализа научной литературы различного рода (статьи из журналов, журналы, материалы конференций, книги, диссертации и др.) уделяется колоссальное внимание как у нас в стране, так и за рубежом. Достаточно, сказать, что поток публикаций по данной тематике за последнее десятилетие вырос более чем в четыре раза, о чем свидетельствуют данные, полученные нами на основе исследования динамики роста публикаций по информетрии из базы данных (БД) Web of Science (WoS). Аналогичную картину можно увидеть при обращении с соответствующим запросом к библиографической БД Scopus, к БД Российский индекс научного цитирования (РИНЦ) или к отечественной БД РЖ ВИНТИ «Информатика».

Развитию библиометрических исследований, на наш взгляд, существенно способствовало то, что увеличилась доступность наукометрических библиографических БД, содержащих информацию о публикациях и о цитировании. В настоящее время ко всем ведущим наукометрическим БД (WoS, Scopus, Google Scholar (GS), РИНЦ и др.) предоставляется доступ в сети Интернет. При наличии сервисов поиска и сохранения результатов и соответствующего программного обеспечения становится возможным проведение анализа публикаций по различным библиографическим переменным, таким как автор(ы), их место работы, ключевые слова, цитирования и др.

Основные задачи

В настоящее время библиометрические данные с успехом используются для того, чтобы проследить развитие соответствующей области знаний в течение длительного периода времени, для определения основных исследователей и их сотрудничества, степени влияния различных журналов, публикующих и цитирующих статьи, касающихся данной предметной области, а также для того, чтобы обеспечить объективные индикаторы работы, касающиеся качества исследований конкретного человека, исследовательской группы или организации.

Своевременное информирование ученых о росте влияния того или иного научного направления, о рейтингах тех или иных научных изданий позволяет исследователям выбрать наиболее авторитетное издание для последующих публикаций, обратить более пристальное внимание на наиболее цитируемую научную проблему. Для академических библиотек результаты библиометрических исследований могут эффективно использоваться в процессе управления репертуаром подписки периодических изданий [Мазов, 2011].

Структурные элементы

1. Источниковый массив данных

Сбор информации по цитированию – масштабная задача, и все источники проходят тщательный отбор. Основные массивы данных по источникам – это WoK, Scopus, GS и РИНЦ, а также базы для отраслевых специалистов в ряде областей (например, Spire, Medline, CiteSeer, ArXiv).

Каждый из перечисленных массивов собирает информацию по цитированию из статей только в выбранной сфере публикаций. Отметим, что перекрытие в содержании этих источников весьма невелико. Основные библиометрические инструментариумы выдают автоматические метрические показатели для отдельных исследователей, а также необработанные данные, которые можно использовать для ручного подсчета или проверки метрических показателей.

1.1. Web of Knowledge

WoK – набор программ Института научной информации, входящего в медиакомпанию Thomson Reuters, современный лидер на рынке библиометрии. В WoK входят БД Web of Science, BIOSIS, Journal Citation Reports, Essential Science Indicators, Science Watch и др. В платформу WoK интегрирована бесплатная БД Medline.

Основным рабочим инструментом является доступная по подписке БД Web of Science, на сегодняшний день расписывающая более 11 тысяч научных журналов, 80% которых относятся к точным наукам. Данные по цитированиям восходят к 1900 г., что является наиболее протяженным диапазоном на рынке аналогичных продуктов. Осуществим поиск по цитируемому источнику (Cited Reference Search) для автора или группы. Возможно найти статьи, в которых цитируется работа ученого; цитирующие материалы анализируются по географическому принципу, дисциплине, типу документа. Отличительной особенностью БД является то, что в нее включены все цитирования к работе автора, а не только те, которые отобраны для индексирования Институтom научной информации. В отчет по цитированию (Citation Report) автора включены публикации и цитирования по годам, индекс Хирша для автора. Этот отчет использует для вычисления цитирования материалов только те источники, которые были отобраны для индексирования в Web of Science. Проверка цитирований отдельной статьи позволяет получить подсчет цитирований для отдельной статьи, посмотреть цитирующие статьи и карту цитирований. Можно получить отчет по цитированию для содержания журнала, включая публикации и цитирования по годам, индекс Хирша журнала.

Недостатки Web of Knowledge.

- Журналы открытого доступа и материалы конференций охвачены слабо, несмотря на ряд обновлений базы данных.
- Монографии не включены в принципе.
- Прослеживается смещенность в сторону английского языка и американской науки.

- Отсутствие стандартных авторских имен – все группы результатов следует проверять и сокращать. Программа выдает все формы фамилии автора.
- Отсутствие стандартных сведений об организации, к тому же только 3 организации могут быть приписаны каждому отдельному автору.

1.2. Scopus

Реферативная БД Scopus создана в 2004 г. издательством Эльзевир. База данных расписывает 19500 научных реферируемых журналов, что значительно превышает количество журналов из БД Web of Knowledge. Диапазон включенных материалов считается лучшим для гуманитарных наук, и в 2009 г. произошло значительное увеличение числа расписываемых журналов. В базе данных больше европейского контента, и в нее включено больше языков, чем в продуктах Института научной информации: 60% источников – это неамериканская литература. Значительно лучше представлен сектор российской науки. Включены журналы открытого доступа, материалы конференций, веб-страницы, патенты и продолжающиеся издания.

Scopus позволяет проводить поиск по автору (Author Search) для идентификации отдельного автора, причем эта опция значительно превосходит по возможностям поиска аналогичную из WoK. Поиск содержит полезные инструменты для снятия многозначности по авторам путем ввода страны, организации и пр. Можно вывести список публикаций каждого автора, включая цитирования. Можно использовать функцию Системы отслеживания цитат (Citation tracker) для создания Обзора цитирований (Citation Overview) публикаций автора, куда включен индекс Хирша и анализ цитирования для каждой статьи по годам. Присутствует инструмент для анализа журналов (Journal Analyzer), используемый для выбора названия журнала и создания группы названий. Анализ включает представление в виде графиков и карт.

Недостатки Scopus.

- Цитирования из публикаций индексируются только с 1996 г., что является недостатком в сравнении с продуктами Института научной информации.
- Основные показатели, как, например, индекс Хирша для автора, вычисляются лишь для публикаций, написанных после 1996 г., отчего для авторов, пик карьеры которых пришелся на более ранние годы, этот показатель не отражает действительной информации.
- Переводные журналы часто расписываются вместе с оригинальными, отчего в БД попадают дублиеты, что затрудняет подсчет числа цитирований и публикаций.

1.3. Google Scholar совместно с Publish or Perish

Бесплатный продукт GS совместно с надстройкой Publish or Perish (PoP) разработан в 2004 г. для создания метрических показателей по опубликованным материалам. GS обращается к веб-сайтам издателей, репозитариям препринтов, веб-сайтам университетов, книгам, техническим отчетам и пр. для поиска научной информации. Охвачено множество материалов (в том числе монографии и «серая» литература), которые неудовлетворительно представлены в других библиометрических инструментариях. PoP создает широкий диапазон метрических показателей для отдельных ученых, число которых намного больше в сравнении с другими системами.

Недостатки GS.

- Не приводится список охваченных журналов.
- Не отображена охваченная временная шкала.
- Охвачены некоторые сомнительные материалы, например, студенческие проекты.
- Слабое покрытие изначально бумажных материалов.
- Опции расширенного поиска достаточно ограниченные: нет способов провести разграничение между авторами с одинаковыми инициалами; невозможно провести исчерпывающий поиск для журналов, у которых различные сокращения названий.
- В результатах часто встречаются дублиеты некоторых статей (обычно в случае препринтов и опубликованных материалов) и даже ложные результаты.

1.4. Российский индекс научного цитирования

Продукт компании «Научная электронная библиотека», разработанный в 2005 г. для более широкого отражения информации об отечественной науке. Функционально напоминает БД Web of Science и располагает схожим инструментарием для оценки журналов и отдельных ученых. Имеются опции поиска профиля по организации, автору, журналу. Авторский поиск, с одной стороны, позволяет однозначно найти нужного автора, с другой же стороны у каждого автора до половины его публикаций оказываются непривязанными к профилю, что аналогично авторским множествам в зарубежных БД.

Недостатки РИНЦ.

- Статьи из переводных журналов индексируются дважды.

- Не определены принципы включения журналов в список РИНЦ.
- Глубина архива невелика, и нет тенденции к ее увеличению.
- Журналы расписываются с большим опозданием.
- Неудовлетворительна идентификация организаций.
- Передача прав на редактирование профиля авторам статей преждевременна ввиду технических недоработок.

2. Метрические инструменты и методы, применяемые к данным по источникам

Разработано огромное количество метрических показателей для способствования оценке результатов исследования ученых (организаций) и влияния журналов. Они варьируются от простого подсчета публикаций и цитирований до математических формул, принимающих в расчет как результаты, так и влияние работы исследователя. Условно их можно разбить на два больших класса: количественные и структурные. Количественные показатели, в свою очередь, делятся на абсолютные показатели и показатели, прошедшие нормирование, – для обеспечения возможности междисциплинарных сравнений. В точки зрения оценки показатели делятся на показатели, вычисляющие влияние ученых и их объединений, и показатели, вычисляющие влияние журналов. И те и другие могут быть абсолютными и нормированными.

2.1. Количественные показатели

2.1.1. Количественные показатели, оценивающие отдельных ученых и их объединения

Общее количество статей обозначает простой подсчет числа опубликованных работ исследователя.

Общее число цитирований предполагает подсчет всех цитирований, полученных опубликованными работами исследователя.

Коронный показатель сравнивает среднее количество цитирований к публикациям исследуемого элемента со средним числом цитирований к международным публикациям за тот же год, по той же дисциплине и того же типа документов.

Верхние 5% указывают на долю публикаций, отнесенных к группе авторов, которые принадлежат к 5% наиболее цитируемых публикаций в мире за тот же год, по той же дисциплине и того же типа документов.

1. *Нецитируемые публикации.* Ту часть публикаций, которая не получает цитирований в течение определенного периода времени, можно рассматривать как противоположную показателю верхних 5%. В последнее время нецитируемости публикаций начинает уделяться все больше внимания [Hsu, Huang, 2012; Eigenfactor, 2012].

Со времени разработки в 2005 г. *индекс Хирша* стал наиболее популярным метрическим показателем оценки научных результатов отдельных ученых. Индекс Хирша представляет собой количество работ ученого, которые были процитированы хотя бы h раз. Появился ряд вариантов индекса Хирша:

- G-индекс Эгга, присваивающий больший вес наиболее цитируемым статьям;
- индекс Хирша отдельного ученого, который учитывается при подсчете влияния соавторства;
- современный индекс Хирша придает меньший вес старым цитируемым статьям;
- взвешенный по возрасту уровень цитирования, который также принимает в расчет возраст статей.

2.1.2. Количественные показатели, оценивающие журналы

Журнальный импакт-фактор (JIF, разработанный WoK). Представляет собой среднее количество цитирований, полученное за год статьями, опубликованными в журнале за предыдущие 2 года. Аналогичный алгоритм используется в отечественной БД РИНЦ. Разновидностью является пятилетний импакт-фактор.

Индикатор журнального рейтинга (SJR, разработанный издательством Эльзевир). SCImago Journal Rank Indicator (SJR). Его принцип в целом схож с JIF. Основное отличие заключается в имитации алгоритма Google PageRank. По существу он присваивает более высокое значение тем цитатам, которые были получены из более престижных журналов. SJR охватывает период цитирования за 3 года.

2. *Эйгенфактор* (или собственный фактор) также использует алгоритм Google PageRank для классификации журналов, так что цитирования из более рейтинговых журналов имеют больший вес в сравнении с прочими. Значение эйгенфактора принимает в расчет и другие величины, такие как отношения внутри дисциплины между цитирующими и цитируемыми журналами. Охватывается 5-летний период цитирования [Eigenfactor: 2012; West, Bergstrom, 2012].

Влияние статей (Article Influence от создателей Эйгенфактора). Учитывая, что Эйгенфактор предусматривает, скорее, измерение общего влияния, которое имеет журнал, чем измерение

влияния каждой статьи, то для измерения постатейного влияния журнала (как в случае с JIF), разработан показатель Влияние статей. Напротив, импакт-фактор измеряет постатейное влияние конкретного журнала. При подсчете вычисленное влияние журнала (Эйгенфактор) делится на число опубликованных статей [West, Bergstrom, 2012].

Уникальным на сегодняшний день показателем является *Влияние в расчете на статью* (SNIP, разработанный издательством Эльзевир). Этот показатель проводит нормирование по дисциплинам, таким образом, позволяя сравнивать ученых, работающих в разных областях. SNIP измеряет контекстуальное влияние цитирования с помощью утяжеленных цитирований на основе общего количества цитирований в предметной области. Влияние одиночного цитирования получает большее значение в предметных областях, в которых цитирования случаются реже, и наоборот. Обозначим ключевые характеристики этого показателя более детально:

- измеряется контекстуальное влияние цитирования при помощи нормирования значений цитирований;
- берется в расчет частота цитирований в той или иной области исследований;
- рассматривается оперативность – как часто статья вероятнее всего будет иметь влияние в заданной области;
- принимается в расчет, насколько хорошо охватывается область в соответствующей БД;
- подсчет ведется без использования классификации журналов по предметным областям, чтобы избежать необходимости разграничения;
- вычисляется любая возможность, когда редколлегия может манипулировать.

Индекс Хирша для журналов (разработка GS) представляет собой количество статей из журнала, процитированных как минимум h раз. Временной промежуток для статей динамичный. Данный метрический показатель в большей мере подходит для некоторых гуманитарных и общественных наук, поскольку GS охватывает больше материалов в этих областях. В нем также шире охват материалов конференций, что может положительно отразиться на результатах библиометрических исследований таких предметов, как компьютерные науки. Гибкость временных границ индекса Хирша также может подойти для дисциплин, где опубликованные исследования медленнее воздействуют на последующие публикации.

2.2. Структурные показатели: публикационные модели

С помощью анализа публикуемости и цитируемости также можно определить связь между публикациями, авторами и областями исследований. Одним из примеров выступает использование карт связи для демонстрации того, сколько различных элементов публикуется совместно или как отобранное количество публикаций связано посредством общей области исследований.

2.3. Проблемные моменты и ограничения

При построении или рассмотрении рейтинга журналов и оценке ученых всегда следует учитывать ряд особенностей и присущих различным показателям недостатков.

Метрические показатели журналов можно сравнивать только в пределах одной и той же дисциплины из-за различных традиций цитирования (исключение составляет лишь SNIP).

В настоящий момент ни один из инструментов журнального рейтинга не оценивает удовлетворительно мультидисциплинарные журналы.

В некоторых областях нет традиции усиленно цитировать работу, поэтому использование традиционных инструментов построения рейтинга журналов в принципе не подходит для оценки некоторых дисциплин (например, искусства и культуры).

В тех областях, где цитирование общепринято, источники данных часто не индексируют публикации, в которых традиционно публикуются исследования, – слабо индексируются местные публикации, «серая» литература, работы не на английском, монографии, конференции и рабочие отчеты.

Новаторские исследования часто не цитируются.

Отрицательные цитирования подсчитываются как зачетные.

Возможно манипулирование системой с помощью таких средств, как самоцитирование, множественное авторство, разбиение результатов на несколько статей, политика журналов, когда для публикации предпочитают высокоцитируемые обзорные статьи.

Определение поля и уровня детализации, до которых производится оценка. Это существенным образом меняет результат для отдельного ученого или исследовательской группы при использовании нормированных показателей сравнения.

Несоответствующее использование метрических показателей цитирования, например, использование импакт-фактора журнала для оценки результатов работы отдельного исследователя

или сравнение индекса Хирша между дисциплинами, игнорирование выявленных вариаций моделей цитирования.

Использование различных данных об источниках и различных инструментов метрических показателей означает, что в различных продуктах журналы могут оцениваться лучше или хуже.

ЛИТЕРАТУРА

- Бредихин С.В., Кузнецов А.Ю.** Методы библиометрии и рынок электронной научной периодики. Новосибирск, Москва: ИВМиМГ СО РАН, НЭИКОН, 2012. — 256 с.
- Мазов Н.А.** Оценка потока научных публикаций академического института на основе библиометрического анализа цитирования // Информационные технологии в гуманитарных исследованиях. — 2011. — № 16. — С. 25—30.
- Bibliometrics** Publication Analysis as a Tool for Science Mapping and Research Assessment 2008 [Электронный ресурс]. — Режим доступа: http://ki.se/content/1/c6/01/79/31/introduction_to_bibliometrics_v1.3.pdf (Дата обращения: 12.08.2012).
- Игра в цифирь**, или как теперь оценивают труд ученого (сборник статей о библиометрике). М.: МЦНМО, 2011: 72 с.
- Egghe L.** The mathematical relation between the impact factor and the uncitedness factor // *Scientometrics*, Vol. 76, No. 1 (2008) 117–123.
- Egghe L., Rousseau R.** Introduction to informetrics. Amsterdam: Elsevier, 1990. — 420 p.
- Eigenfactor:** Detailed methods [Электронный ресурс]. — Режим доступа: www.eigenfactor.org/methods.pdf (Дата обращения: 12.10.2012).
- Hsu J.-W., Huang D.-W.** A scaling between Impact Factor and uncitedness // *Physica A* 391 (2012) 2129–2134.
- Raan A. F. J. van.** The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments // *Technikfolgenabschätzung – Theorie und Praxis*. — 2003. — V.1(12). — P. 20—29.
- West J., Bergstrom C.T.** Pseudocode for calculating EigenfactorTM Score and Article InfluenceTM Score using data from Thomson-Reuters Journal Citations Reports. — 2008. — [Электронный ресурс]. — Режим доступа: http://octavia.zoology.washington.edu/people/jevin/Documents/JournalPseudocode_EF.pdf (Дата обращения: 12.10.2012).

КОММЕНТАРИЙ

К статье Н.А. Мазова и В.Н. Гуреева «Библиометрия: современное состояние, задачи, инструменты»

В последние десятилетия XX в., когда занятие наукой стало достаточно массовым явлением, все более настоятельным требованием времени становится необходимость оценки вклада ученых с помощью количественных и качественных методов, характеризующих научную деятельность и не зависящих от каких-либо субъективных факторов. Особую важность приобретает объективная оценка, когда речь идет о «знаках отличия» отдельного ученого или научного коллектива при получении бюджетного или грантового финансирования научных исследований, а также поощрении отдельных исследователей премиями, медалями, учеными степенями и званиями. Однако, по моему субъективному мнению, на данном этапе у научного сообщества, нет никаких объективных количественных критериев оценки качества научной деятельности, и применительно к конкретному исследователю она носит исключительно субъективный характер – в виде волевых решений, голосований на ученых советах и в других аналогичных формах. Такой сугубо цифирный подход носит волюнтаристский субъективный характер.

Для оценки продуктивности и взаимосвязи авторов научных публикаций сравнительно давно пытаются использовать различные параметры системы научных коммуникаций. Это вполне объяснимо: через систему коммуникаций поступают в хранилище «отпечатки» научной продукции – публикации. Показатель цитирования действительно важный показатель, но позволяет ли измерение цитируемости верно оценить научный вклад? Действительно, цитирование отражает связь между работами ученых, говорит об использовании научных результатов – но оно отражает не все связи. Цитат-индексы имеют малую точность в оценках работ. Для них характерно наличие субъективных ошибок, связанных с экспертом. Как подчеркивают некоторые науковеды, «цитат-индекс» – завуалированный метод экспертных оценок, причем не в лучшем варианте, ибо экспертизу производит непосредственный исполнитель. Естественно, что это порождает субъективизм метода цитирования, и картина связей научных работ искажается вследствие либо личных причин, либо пристрастия и «местного патриотизма», либо языковых барьеров и т.д. и т.п. [Хайтун, 1983: с. 89].

В данной статье А.Н. Мазов и его соавтор В.Н. Гуреев обращаются к решению сугубо практической задачи – ознакомлению специалистов в области гуманитарных наук с основными библиометрическими инструментариями (WoK, Scopus, GS и РИНЦ, а также с базой для отраслевых специалистов в ряде областей (например, Spire, Medline, CiteSeer, ArXiv).

При этом они подчёркивают не только достоинства указанных инструментариев, но отмечают и их серьезные недостатки. Часть этих замечаний никак не влияет на качественные характеристики журналов и публикуемых в них статей, и несут лишь количественные характеристики. Другие могут серьезно влиять на количественные показатели. Приведем некоторые из них:

1. Возможность манипулирования системой с помощью таких средств, как самоцитирование, множественное авторство, разбиение результатов на несколько статей, политика журналов, когда для публикации предпочитают высокоцитируемые обзорные статьи.

2. Несоответствующее использование метрических показателей цитирования, например, использование импакт-фактора журнала для оценки результатов работы отдельного исследователя или сравнение индекса Хирша между дисциплинами, игнорирование выявленных вариаций моделей цитирования. Действительно, когда неуместное акцентирование важности импакт-фактора становится целью, он перестает быть хорошим показателем [Арнольд, Фаулер, 2011: с. 53].

Не определены принципы включения журналов в список РИНЦ. В России для ранжирования археологических и этнографических журналов по их значимости используется включенность в список ВАК (достаточно субъективный фактор). Так включение в списки РИНЦ археологических ежегодников «Археологические вести (ИИМК РАН), «Краткие сообщения Института археологии РАН» (ИА РАН) и «Проблемы археологии, этнографии и антропологии Сибири и сопредельных территорий» (ИАЭТ СО РАН) из которых два последних включены в списки ВАК является сомнительным из-за низкого уровня цитирования (Табл.1). При этом считается, что публикация статьи в журнале, включенном в списки ВАК, лучше характеризуют специалиста и значительно повышают его рейтинг, давая возможность защитить кандидатскую или докторскую диссертации.

Таблица 1. Показатели некоторых периодических изданий включенных в РИНЦ на 29.10.2012 г.

Журналы	ИФ ISI 2008	ИФ РИНЦ 2010	Включенность в списки ВАК
Проблемы археологии, этнографии и антропологии Сибири и сопредельных территорий	нет	нет	Нет
КСИА	нет	0.027	Да
Археологические вести	нет	нет	Да

Оригинальная сущность работы проявляется и в том, что в ней показано: цитирование в различных научных направлениях имеет свою специфику и в малой степени подходит для таких специфических областей, как культура и искусство. Поэтому можно пока лишь говорить о возможности ее применения, а не о перспективности. Для определения перспективности требуется, на наш взгляд, проведение не только сравнительного количественного анализа с помощью существующих методик цитатного анализа, учитывающих общее и специфическое в цитировании в различных областях знаний, но и качественного анализа.

Многие ученые предлагают учитывать не только частоту цитирования, но и её качество. Детальная качественная классификация была предложена в 1976 году Е.Д. Гражданниковым и Сорокиной [1976: с. 14] и усовершенствована в 1987 г. [Гражданников, 1987: с. 39-46]. Согласно Е.Д. Гражданникову ранжированная классификация видов цитирования основана на пяти принципах классификации:

- принцип определенности;
- принцип независимости от автора;
- по объему;
- положительная оценка;
- отрицательная оценка.

К сожалению, эта классификация мало известна среди специалистов в области наукометрии. Мною с соавторами была предложены реальные возможности, показывающие примеры полезности применения анализа цитирования для решения ряда научных проблем в археологии. Используемая методология имеет весьма общий характер и вполне переносима на другие научные дисциплины [Костин, Холушкин, 2010].

Имеются и другие примеры качественного анализа цитирования [Гарскова, 2010]. Исследование И.М.Гарсковой включает некоторые результаты количественного анализа полной тематической библиографии работ, опубликованных под эгидой отечественной Ассоциации «История и компьютер», в том числе и периодического издания «Информационные технологии в гуманитарных исследованиях» [Гарскова, 2010: с. 138]. В работе приведены графы, отражающие связи между авторами в форме «незримых научных коллективов». Эти графы были построены на основе таблиц сопряженности между авторами по количеству совместно написанных работ. По существу, эти графы показывают отдельные региональные центры и школы исторической информатики и позволяют изучить концентрацию виртуальных научных коллективов на основе сведений о соавторстве внутри сетевой структуры, формировании отдельных направлений внутри сети вокруг центральных фигур.

Интересные направления по анализу публикаций и БД развиваются у генетиков. В работах П.С. Дименкова из Института цитологии и генетики СО РАН описана компьютерно-информационная система для автоматического извлечения и интеграции ассоциативных знаний из фактографических баз данных и текстовых источников информации [Деменков, 2008]. Под ассоциацией между молекулярно-генетическими объектами понимается прямое или опосредованное их взаимодействие, а также следственно-причинные связи между генами, белками и заболеваниями. Эта система AND позволяет пользователю быстро получать и анализировать большие объемы данных в форме графически визуализированных сетей молекулярно-генетических взаимодействий и их ассоциаций с заболеваниями. Это направление является новаторским и в перспективе может быть использовано в других областях знаний. В свое время в ИАЭТ СО

РАН предпринималась попытка создания археологического портала знаний. Но из-за отсутствия финансирования и отсутствия интереса к подобным разработкам работа была прекращена.

В настоящее время отечественные разработки в области извлечения и интеграции знаний при одновременной работе с текстовыми и фактографическими базами данных, ориентированные на фармакологию, биотехнологию и биомедицину, библиометрию, археологию отсутствуют, а имеющиеся за рубежом характеризуются низкой эффективностью. Остается только надеяться, что рано или поздно такие новаторские разработки появятся в СО РАН. Хотелось бы пожелать Н.А. Мазову и его соавтору решать в будущем и эти задачи.

ЛИТЕРАТУРА

- Арнольд Д., Фаулер К. Гнусные цифры цитирования // Игра в цифирь, или как теперь оценивают труд ученого (сборник статей о библиометрике). – М.:МЦИМО, 2011: 53-62.
- Дименков П.С. Математическое и программное обеспечение для реконструкции ассоциативных сетей молекулярно-генетических взаимодействий. Автореф. дис. на соиск. уч. степени канд. технических наук. – Томск, 2008: 24 с.
- Гарскова И.М. Анализ историографии исторической информатики как научного направления // Харківський історіографічний збірник – ХНУ імені Каразіна, 2010. Вип.10: с. 171.
- Гражданников Е.Д. Проблема критериальной оценки научных результатов // Проблемы развития научно-образовательного потенциала. – Новосибирск: Наука, 1987: с. 24-47.
- Гражданников Е.Д., Сорокина Т.В. Наукометрические методы библиографического поиска. - Новосибирск, 1976.
- Деревянко А.П., Холюшкин Ю.П. Проблема качественного анализа археологических публикаций // Методология и методика археологических реконструкций. - Новосибирск, 1994: 24-32.
- Костин В.С., Холюшкин Ю.П. Некоторые подходы к библиометрическому анализу взаимцитирования // Информационные технологии в гуманитарных исследованиях. Вып. 16. Новосибирск, 2010 :
- Хайтун С.Д. Наукометрия. – М., 1974.
- Холюшкин Ю.П. Системная археология. - Новосибирск: НГУ, 2010.

Холюшкин Ю.П., д.и.н.

Гуреев В.Н.
Мазов Н.А.,

Практическое применение библиометрического анализа при формировании журнального фонда

Использование библиометрического анализа имеет широкий спектр применения во многих областях науки, образования и библиотечного дела. К рассмотрению предлагается опыт приложения цитатного анализа к формированию фонда периодических изданий в научной библиотеке. Разработанный на базе ИНГТ СО РАН и апробированный в ГНЦ ВБ «Вектор», метод позволяет за короткий срок определить ядро журнального фонда и составить оптимальный репертуар подписки на научную периодику.

Ключевые слова: библиометрический анализ, анализ цитирования, комплектование библиотек, научная периодика.

Быстрое развитие информационных технологий в последние годы отразилось практически на всех направлениях библиотечной деятельности и позволило с новых позиций оценивать эффективность работы библиотек. В статье рассмотрены возможности применения цитатного анализа в изучении информационных потребностей ученых научных институтов и в оценке удовлетворенности читателей имеющейся подпиской. Значимость проводимого анализа обусловлена рядом факторов, среди которых не всегда достаточное финансирование научных библиотек при росте цен на ресурсы, увеличение числа электронных ресурсов, особенно открытого доступа. Библиометрический анализ, став доступным и менее трудоемким с появлением специализированных ресурсов и программ, может способствовать более успешной работе при комплектовании библиотеки НИИ, а его результаты могут стать объективной основой для формирования репертуара журнальной подписки.

Следует отметить, что принципы комплектования за последние годы не претерпели кардинальных изменений, хотя при переходе на электронную подписку и были скорректированы и дополнены новыми. Главным принципом была и остается селективность, поскольку комплектователю необходимо отобрать лишь значимые документы, согласующиеся с функциональными установками библиотеки и потребностями ее пользователей.

Библиометрический анализ сегодня является наиболее прогрессивным методом, который позволяет объективно и точно, в отличие от устаревших традиционных методов, выявлять уровень удовлетворенности ученых текущей подпиской, отчего в большей мере может способствовать оптимизации комплектования. Немаловажно, что акцент на объективности библиометрического анализа все чаще делается даже при его сравнении с научным рецензированием, что подчеркивает его возросшую за последние годы значимость [Raaij, 2003]. Базы данных (БД), содержащие необходимую для анализа информацию, доступны в большинстве научных организаций, причем зарегистрированные пользователи имеют к ним удаленный доступ. Крупнейшие компании Томсон

и Эльзевир популяризируют свои БД Web of Science (WoS) и Scopus, проводят обычные и онлайн-семинары, что способствует росту технически грамотных пользователей.

Библиометрический анализ производится специальными программами по заданным алгоритмам (реже производится вручную) и основан на различных исходных данных. Такими данными могут выступать импакт-фактор журнала, Эйгенфактор [Eigenfactor, 2012, West, Bergstrom, 2008], спрос на издания, выявленный на основе анализа книговыдачи (количества обращений), отказов или цитирования; индекс Хирша, f-индекс [Katsaros, Akritizis, Bozanis, 2009], представленность в авторитетных БД и пр. Используются математические формулы и методы статистического анализа. Для библиотекаря анализ пристатейной библиографии позволяет оценивать эффективность использования журналов фонда, их информационную значимость и определять минимально необходимую глубину архива [Слащёва, Мохначёва, Харыбина, 2008]. Метод хорошо подходит для формирования фонда научной периодики, позволяя решать как задачу включения новых изданий в репертуар подписки, так и задачу исключения тех журналов, которые перестали отвечать требованиям пользователей [Мазов, 2011].

В нашей работе мы применили метод анализа пристатейной литературы на примере публикаций ученых ГНЦ ВБ «Вектор» в области биологии и медицины, а также публикаций сотрудников ИНГГ СО РАН в области наук о Земле. Также проанализированы кластеры журналов, в которых были опубликованы статьи сотрудников двух институтов в указанный период, и проведено сравнение ядерной зоны ссылок с ядерной зоной источников этих ссылок.

Тематические профили двух институтов – медико-биологические исследования и геофизика – занимают значительное место в журнальной периодике, а обслуживающие эти отрасли журналы обладают наибольшими импакт-факторами, наблюдается их количественный рост и параллельная научным разработкам конкуренция. Поэтому наши методы и результаты могут оказаться полезными применительно к научным библиотекам других смежных профилей.

Источниками ссылок стали статьи научных сотрудников ГНЦ ВБ «Вектор» за 2007-2011 и ИНГГ СО РАН за 2006-2010 гг. в журналах и продолжающихся изданиях. К рассмотрению не принимались монографии, тезисы и материалы научных конференций (в том числе опубликованные в журналах), поскольку наша конечная цель – рассмотрение комплектования журнальной периодики.

Затем анализировались ссылки из пристатейных списков литературы. Из пристатейных списков литературы не анализировались ссылки на монографии, тезисы и материалы научных конференций. Адаптированные для англоязычных читателей ссылки на переводные источники (например, на журнал *Biophysics* вместо «Биофизика» или *Russian Geology and Geophysics* вместо «Геология и геофизика») рассматривались как ссылки на оригинальную русскоязычную статью. Журналы, изменившие свое название, рассматривались как один журнал. Журналы, впоследствии разделенные на несколько журналов, рассматривались как разные. Для анализа выписывались заглавие журнала и год выхода цитируемой статьи. Впоследствии эти данные дополнялись данными об издательстве и доступности журнала.

В среднем сотрудники каждого из институтов публиковали порядка 150 статей в год, в которых ежегодно цитировалось примерно 1600 источников. Таким образом, наши результаты могут рассматриваться как типичные для организаций аналогичных тематических профилей. По полученным нами данным для выявления статей сотрудников достаточно использовать БД РИНЦ, аккумулирующую информацию о подавляющей части (95-98%) как российских, так и зарубежных публикаций сотрудников. Важно отметить, что поиск должен проводиться по публикациям каждого научного работника.

Ссылки из журналов, отсутствующих в БД WoS, выписывались вручную, что потребовало длительных временных затрат и большого объема труда. В ходе работы выяснилось, что в $\approx 35\%$ статей сотрудников, проиндексированных в WoS, было сделано до 80% всех ссылок. По-видимому, это можно объяснить: а) более тщательной подготовкой статей в рейтинговые журналы; б) выполнением правил западной культуры цитирования, где делается больше ссылок в сравнении со статьями в отечественных журналах [Варшавский, Маркусова, 2009]. В целом наше наблюдение позволяет сделать вывод о том, что можно избежать трудоемкой рутинной работы, воспользовавшись автоматическим сервисом EndNote от Thomson Reuters, позволяющим выявить и проанализировать в полуавтоматическом режиме до большинство ссылок.

Всего за пять лет сотрудниками геологического профиля было процитировано более 1000 наименований журналов (750 зарубежных и 250 отечественных), а медико-биологического профиля – более 1300 (1168 зарубежных и 171 отечественных). Было проанализировано порядка

8000 ссылок по каждому институту. В каждой из этих двух групп ссылки были разбиты на три равные части. Ядро цитируемых отечественных журналов и периодические издания второй группы показаны в табл. 1.

Таблица 1. Ядерные отечественные журналы и журналы второй группы, выявленные на основе цитирования

№	Медико-биологический профиль			Науки о Земле		
	Журнал	% к-ва ссылок	Всего %	Журнал	% к-ва ссылок	Всего %
1.	Вопросы вирусологии	16,94	16,94	Геология и геофизика	35	35
2.	Молекулярная биология	6,44	23,38	Доклады РАН	19	54
3.	Биотехнология	6,07	29,45	Геохимия	7	61
4.	Доклады Академии наук	5,09	34,54	Петрология	4	65
5.	Оптика атмосферы и океана	4,42	38,96	Геотектоника	2	67
6.	Вестник РАМН	4,04	43	Стратиграфия и геологическая корреляция	2	69
7.	Журнал микробиологии, эпидемиологии и иммунобиологии	3,82	46,82	Физика Земли	2	71
8.	Молекулярная генетика, микробиология и вирусология	2,92	49,74	Геология рудных месторождений	2	73
9.	Бюллетень экспериментальной биологии и медицины	2,77	52,51	Литология и полезные ископаемые	2	75
10.	Биоорганическая химия	2,69	55,2	Записки ВМО	1	76
11.	Проблемы туберкулеза и болезней легких	2,62	57,82	Известия АН. Сер. Геологическая	1	77
12.	Биохимия	1,79	59,61	Отечественная геология	1	78
13.	Химико-фармацевтический журнал	1,72	61,33	Тихоокеанская геология	1	79
14.	Инфекционные болезни	1,57	62,9	Геология нефти и газа	1	80
15.	Бюллетень СО РАМН	1,2	64,1	Физическая мезомеханика	1	81
16.	Генетика	1,12	65,22	Химия в интересах устойчивого развития	1	82

Из данных таблицы видно, что лишь 9% (16 из 171) от общего числа цитируемых российских журналов медико-биологического профиля обеспечивают 2/3 всех ссылок, и всего на 6% (16 из 250) журналов приходится 80% всех цитирований в области геологии и геофизики. Отметим также, что данные двух институтов по цитируемости хорошо согласуются друг с другом, с той лишь разницей, что в статьях в области геологии больше цитируются отечественные журналы, тогда как в области биологии предпочитается цитирование зарубежных статей.

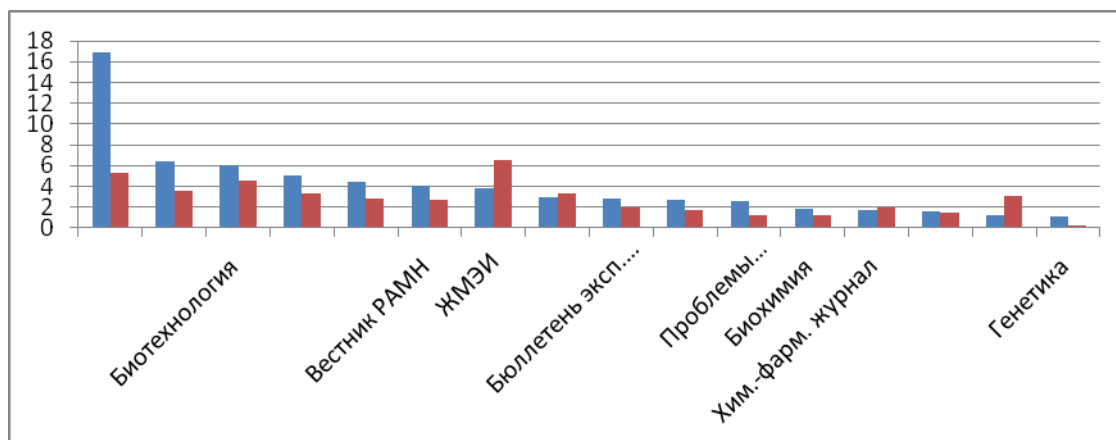


Рис. 1. Данные по публикуемости (красный) и цитируемости (синий) в области биологии и медицины в процентном соотношении к общему количеству (по данным табл. 1).

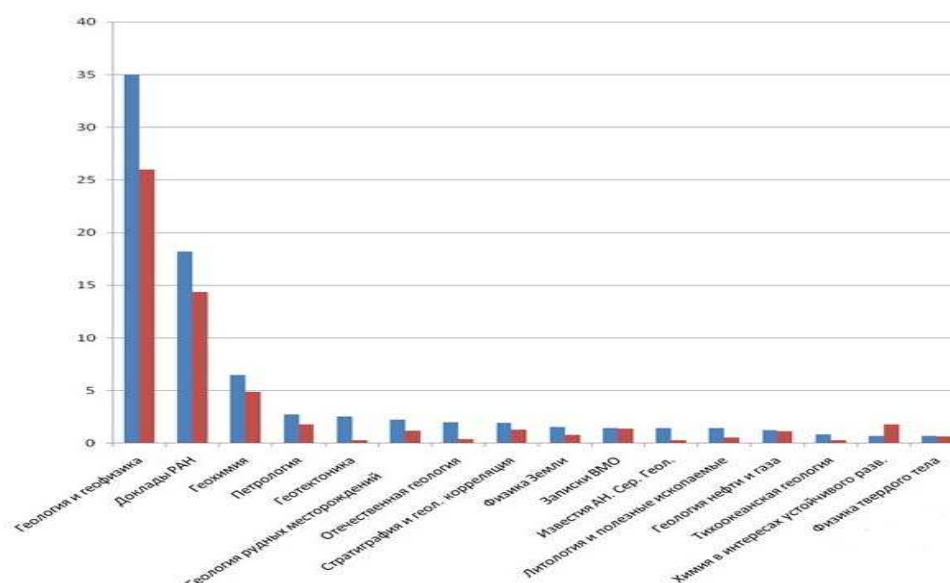


Рис. 2. Данные по публикуемости (красный цвет) и цитируемости (синий) в области наук о Земле в процентном соотношении к общему количеству (по данным табл. 1).

Таблица 2. Ядерные зарубежные журналы и наиболее востребованные журналы второй группы, выявленные на основе цитирования

Медико-биологический профиль				Науки о Земле		
№	Журнал	% к-ва ссылок	Всего %	Журнал	% к-ва ссылок	Всего %
1.	Journal of Virology	6,03	6,03	Earth Planetary Science Letters	7	7
2.	Proceedings of the National Academy of Sciences of the United States of America	3,36	9,39	Contributions to Mineralogy and Petrology	6	13
3.	Virology	3,15	12,54	Geochimica et Cosmochimica Acta	5	18
4.	Vaccine	2,77	15,31	Journal of Geophysical Research	5	23
5.	Journal of Biological Chemistry	2,4	17,71	Nature	4	27
6.	Journal of General Virology	2,4	20,11	American Mineralogist	3	30
7.	Nature	2,04	22,15	Lithos	3,5	33,5
8.	Science	1,94	24,09	Journal of Petrology	3,5	37
9.	Journal of Clinical Microbiology	1,94	26,03	Chemical Geology	2,5	39,5
10.	Emerging Infectious Diseases	1,89	27,92	Geology	2,5	42
11.	Nucleic Acids Research	1,59	29,51	Tectonophysics	2	44
12.	Journal of Infectious Diseases	1,38	30,89	Science	1,5	45,5
13.	Journal of Molecular Biology	1,35	32,24	Economic Geology	1,5	47
14.	Journal of Immunology	1,24	33,48	European Journal of Mineralogy	1,5	48,5
15.	Journal of Medical Virology	1,19	34,67	Canadian Mineralogist	1,5	50
16.	Virus Research	0,86	35,53	Geophysics	1,5	51,5
17.	New England Journal of Medicine	0,86	36,39	Precambrian Research	1,5	53
18.	Archives of Virology	0,83	37,22	Journal of Crystal Growth	1	54
19.	Antiviral Research	0,75	37,97	Physical Review. Ser.B	1	55
20.	Lancet	0,73	38,7	Russian Geology and Geophysics*	1	56
21.	Cell	0,65	39,35	Diamond and Related Materials	1	57
22.	Applied and Environmental	0,6	39,95	Journal of Geology	1	58

	Microbiology					
23.	Biochemistry	0,59	40,54	International Geology Review	1	59
24.	Journal of Experimental Medicine	0,59	41,13	Geophysical Research Letters	0,5	59,5
25.	FEBS Letters	0,56	41,69	Physics of the Earth and Planetary Interiors	0,5	60

Большая концентрация ссылок на меньшее количество журналов в области геологии также может объясняться общим меньшим количеством ссылок.

В статьях же биологической и медицинской направленности наблюдается тенденция к более обильному цитированию (чем подтверждаются высокие импакт-факторы биологических, и в особенности медицинских журналов), что ведет к большему распределению ссылок. Корреляция между цитируемостью и публикуемостью показана на рис. 1 и 2.

В целом данные по цитируемости и публикуемости согласуются между собой, причем в большей мере это очевидно на примере статей и цитирований в области наук о Земле, где лидирует журнал «Геология и геофизика». Ссылки на него составляют более трети всех ссылок. С одной стороны, это объясняется тем, что настоящий журнал издается в ИНГГ, а с другой, что это один из основных политематических отечественных журналов в области наук о Земле.

Зависимость между цитируемостью и публикуемостью для зарубежной периодики показана на рис. 3 и 4.

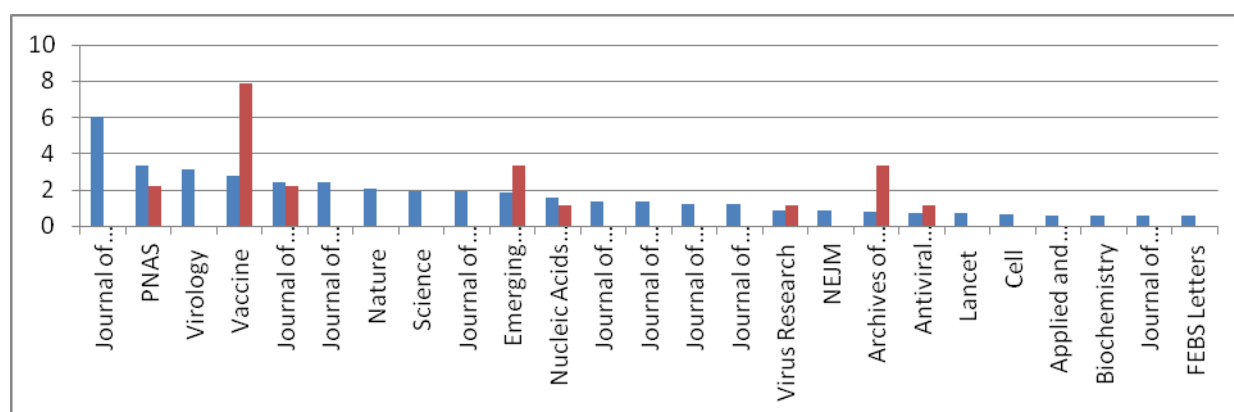


Рис. 3 Данные по публикуемости (красный) и цитируемости (синий) в области биологии и медицины в процентном соотношении к общему количеству (по данным табл. 2).

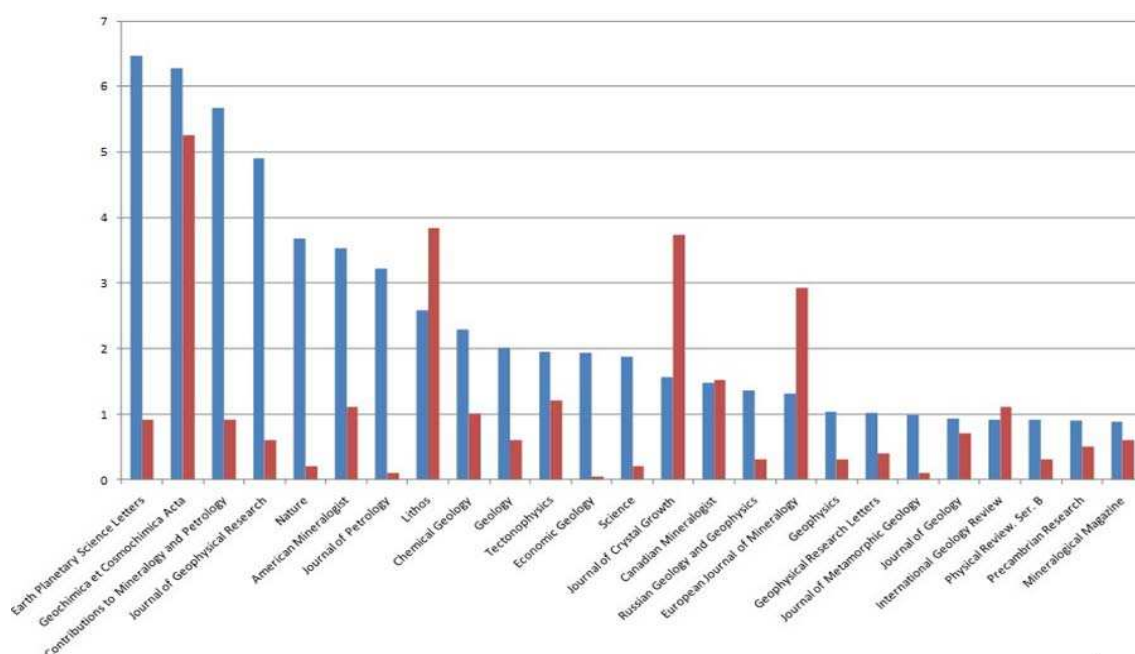


Рис. 4. Данные по публикуемости (красный) и цитируемости (синий) в области наук о Земле в процентном соотношении к общему количеству (на основе табл. 2).

Полученные данные по цитируемости зарубежных журналов приводятся в табл. 2.

Ссылки на иностранные журналы показывают распределение, схожее с отечественными: большая концентрация ссылок представлена в журналах геологического профиля, притом что на 25 журналов приходится 2/3 всех ссылок. В области же биомедицины 2/3 всех ссылок представлены в ста журналах. Лишь 1,3% (15 из 1168) процитированных журналов обеспечивает треть всех ссылок, сделанных из статей медико-биологического профиля; для наук о Земле эта цифра составляет 0,9% (7 из 750).

Из графиков видно, что данные по цитируемости зарубежных журналов распределяются более гладко в сравнении с отечественной периодикой, где счет ядерных журналов идет на единицы. Это можно объяснить меньшим количеством российских журналов в целом. Также графики демонстрируют разницу между двумя научными областями: если в науках о Земле во всех цитируемых зарубежных журналах ядерной зоны сотрудники также публиковались, то в области биомедицины в большей части цитируемых журналов статьи сотрудников не публиковались. Причиной тому может быть, во-первых, большее количество цитируемых журналов в целом, а также, возможно, более низкий общий уровень статей.

Таким образом, на основе данных по цитированию было выделено ядро фонда научной периодики, в первую очередь необходимое для информационного обеспечения научных исследований. В области биологии и медицины в первые две зоны (66% всех ссылок) вошли 16 отечественных и 100 зарубежных журналов; а в области наук о Земле – 4 российских и 25 иностранных.

Данные по публикуемости лишь частично согласовались с данными по цитированию, причем в меньшей степени корреляция наблюдалась в области биомедицины.

Важно отметить, что вся работа проводилась без привлечения сторонних трудовых или материальных ресурсов. Полученные результаты и разработанный метод, на наш взгляд, являются наиболее объективными и доступными в использовании в библиотеках научных организаций. Авторы использовали полученные данные при уточнении репертуара подписки на научную периодку.

Полученные данные также могут быть использованы другими службами научной организации для решения иных задач:

- 1) на их основе можно, в частности, выявлять эффективность научных исследований организации;
- 2) можно определить наиболее эффективную научную группу или отдел;
- 3) можно выявить время полужизни цитируемых журналов, сравнить его с аналогичным параметром в Journal Citation Reports и узнать, насколько свежей литературой пользуются сотрудники;
- 4) можно уточнить, насколько тематика чтения сотрудников соответствует профильной тематике организации.

ЛИТЕРАТУРА

- Варшавский А.Е., Маркусова В.А. Оценку эффективности российских фундаментальных ученых следует скорректировать [Электронный ресурс] // Наука и технологии РФ. – 2009 (11 января). – Режим доступа: http://strf.ru/organization.aspx?CatalogId=221&d_no=17296 (Дата обращения: 14.10.2012).
- Мазов Н.А. Оценка потока научных публикаций академического института на основе библиометрического анализа цитирования // Информационные технологии в гуманитарных исследованиях. – 2011. – № 16. – С. 25–30.
- Слащева Н.А., Мохначева Ю. В., Харыбина Т. Н. Изучение информационных потребностей пользователей Пущинского научного центра РАН в Центральной библиотеке Центра (отдел БЕН РАН) // Библиотеки национальных академий наук: проблемы функционирования, тенденции развития: Науч.-практ. и теорет. сб. – Киев, 2008. – Вып. 6. – С. 247–264.
- Eigenfactor: Detailed methods [Электронный ресурс]. – Режим доступа: www.eigenfactor.org/methods.pdf (Дата обращения: 14.10.2012).
- Katsaros D., Akritidis L., Bozanis P. The f Index: Quantifying the Impact of Coterminal Citations on Scientists' Ranking // Journal of the American society for information science and technology. – 2009. – V. 60(5). – P. 1051–1056.
- Raan A.F.J. van. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments // Technikfolgenabschätzung – Theorie und Praxis. – 2003. – V.1(12). – P. 20–29.
- West J., Bergstrom C.T. Pseudocode for calculating EigenfactorTM Score and Article InfluenceTM Score using data from Thomson-Reuters Journal Citations Reports. – 2008. - [Электронный ресурс]. (Дата обращения: 14.10.2012). – Режим доступа: http://octavia.zoology.washington.edu/people/jevin/Documents/JournalPseudocode_EF.pdf

КОММЕНТАРИЙ

К статье Гуреева В.Н. и Мазова Н.А. «Практическое применение библиометрического анализа при формировании журнального фонда».

В статье рассматривается практический опыт использования цитатного анализа к формированию фонда периодических изданий в научной библиотеке. Убедительно показано, что предлагаемый метод позволяет

оперативно определить ядро журнального фонда и оптимизировать репертуар на научную периодику, что чрезвычайно важно в условиях хронического недофинансирования текущей подписки.

Анализ пристатейной библиографии позволяет оценивать эффективность использования журнального фонда научной библиотеки, информационную значимость каждого журнала для исследователей конкретного научного учреждения и определять необходимую глубину архива. Практика показывает, что метод хорошо подходит для оптимального формирования фонда научной периодики, позволяя решать как задачу включения новых изданий в репертуар подписки, так и задачу исключения тех журналов, которые перестали пользоваться спросом исследователей.

Представляется важным тот факт, что для выявления статей сотрудников достаточно использовать отечественную базу данных РИНЦ, наиболее полно аккумулирующую информацию о подавляющей части (95-98%) как российских, так и зарубежных публикаций сотрудников.

Сравнение ядерной зоны ссылок с ядерной зоной источников этих ссылок демонстрирует разницу между двумя научными областями: в то время как в науках о Земле во всех цитируемых зарубежных журналах ядерной зоны также размещены и публикации сотрудников, в области биомедицины в большей части цитируемых журналов статьи сотрудников не публиковались. Для объяснения причин этой разницы хорошо бы апробировать предлагаемый метод на более широкий круг областей исследования.

Лобанова Т.П., кандидат химических наук, зав.
информационно-аналитическим отделом ФБУН
ГНЦ ВБ «Вектор»

**Костин В.С.,
Холюшкин Ю.П.,**

Практическое применение библиометрического анализа цитирования и самоцитирования журналов археологического и этнографического профиля¹

В статье приводятся примеры использования библиометрического анализа в археологии и этнографии. Проводится статистический анализ ведущих археологических и этнографических журналов России.

Ключевые слова: библиометрия, библиометрический анализ, анализ цитирования, *h*-индекс

Проблема, которую обсуждают в гуманитарном мире, о которой много говорят – измерение цитируемости и количество публикаций в рейтинговых журналах, включенных в списки ВАК. Мы не призываем к отказу от подобной статистики цитирования как способа оценки качества научного исследования. Мы призываем лишь к правильному его использованию. Об этом же говорится в докладе «Международного математического союза», в котором подчеркивается, что индекс цитирования может быть лишь частью оценки и представляет собой ограниченный и неполный взгляд на качество исследований. Научные исследования достаточно сложная вещь, чтобы измерять её достижения таким грубым средством [Адлер, Эванс, Тейлор, 2011].

Важно знать, что индекс цитирования, а особенно импакт-фактор журнала сильно зависит от специальности. В гуманитарных науках он будет всегда ниже, чем в естественных науках.

Библиометрия широко используется для оценки вклада отдельных ученых, статей, институтов и даже государств. Наукометрическая база данных ключевых научных показателей (Essential Science Indicator) поддерживаемая медиакомпанией Thomson Reuters рекламируется как средство ранжирования «ведущих стран, журналов, учёных, статей и институтов в разных исследовательских областях» [Адлер, Фаулер, 2011: с. 60]. По мнению авторов этой статьи, данные указатели основываются на тех же данных о ссылках, которые используются для подсчета импакт-фактора, и такими показателями так же легко манипулировать. В качестве примера авторы статьи привели специальный выпуск журнал «Journal of Physics: Conference Series», который редактировался Хэ и собрал 243 ссылки на его журнал, там также содержатся 353 ссылки на самого Хэ. Он утверждал, что общее число ссылок на его статьи превышает 6800. По данным недавнего анализа научных показателей от Thomson Reuters, профессор Хэ Цзыхуань назван восходящей звездой компьютерной науки [Адлер, Фаулер, 2011: с. 60].

Другим популярным показателем продуктивности исследователей является *h*-индекс (индекс Хирша). Хорхе Хирш писал: учёный имеет индекс *h*, если *h* из его *N_p* статей цитируются как минимум *h* раз каждая, в то время как оставшиеся (*N_p* — *h*) статей цитируются не более, чем *h* раз каждая. Иными словами, учёный с индексом *h* опубликовал *h* статей, на каждую из которых сослались как минимум *h* раз. Так, если у данного исследователя опубликовано 100 статей, на

¹ Эта работа поддержана Российским Гуманитарным научным фондом, проект № 12-01-12026

каждую из которых имеется лишь одна ссылка, то его h-индекс равен 1. Таким же будет h-индекс исследователя, опубликовавшего одну статью, на которую сослались 100 раз. В то же время (более реалистический случай), если среди публикаций исследователя имеется 1 статья с 9 цитированиями, 2 статьи с не менее, чем 8 цитированиями (включая уже упомянутую статью с 9 цитированиями), 3 статьи с не менее, чем 7 цитированиями, ..., 9 статей с не менее, чем 1 цитированием каждой из них, то его h-индекс равен 5 (т.к. на 5 его статей сослались как минимум по 5 раз). Обычно распределение количества публикации $N(q)$ в зависимости от числа их цитирований q в очень грубом приближении соответствует гиперболе: $N(q) \approx \text{const} \times q^{-1}$. Координата точки пересечения этой кривой с прямой $N(q) = q$ и будет равна индексу Хирша. Этот индекс хорошо работает лишь при сравнении учёных, работающих в одной области исследований, поскольку традиции, связанные с цитированием, отличаются в разных отраслях науки (например, в биологии и медицине h-индекс намного выше, чем в физике).

В норме h-индекс физика примерно равен продолжительности его научной карьеры в годах, тогда как у выдающегося физика он вдвое выше. Хирш считал, что в физике (а в реалиях в США) h-индекс, равный 10-12, может служить одним из определяющих факторов для решения о предоставлении исследователю постоянного места работы в крупном исследовательском университете; уровень исследователя с h-индексом, равным 15-20, соответствует членству в Американском физическом обществе; индекс 45 и выше может означать членство в Национальной академии наук США [<http://ru.wikipedia.org/wiki/H-%E8%ED%E4%E5%EA%F1>].

Таблица 1. Показатели рейтинга ведущих журналов по археологии и этнографии (данные 05.11.12)

	Место в общем рейтинге SCIENCE INDEX за 2010 год	Место в рейтинге SCIENCE INDEX за 2010 год по тематике "Исторические науки".	ИФ ISI	ИФ РИНЦ
Археология, этнография и антропология Евразии	211	2	нет	0.459
Российская археология	831	7	нет	0.106
Этнографическое обозрение	1081	9	нет	0.110
Вестник НГУ	1098	12	нет	0.105
Вестник археологии, антропологии и этнографии	1365	17	нет	0.229

Индекс Хирша, разумеется, не идеален. Нетрудно привести ситуацию, когда h-индекс даёт совершенно неверную оценку значимости исследователя. В частности, короткая карьера учёного приводит к недооценке значимости его работ. Так, h-индекс выдающегося французского математика, основателя современной высшей алгебры Эвариста Галуа, убитого на дуэли в возрасте 20 лет равен 2 и останется таким навсегда. Если бы Альберт Эйнштейн умер в начале 1906 г., его h-индекс остановился бы на 4 или 5, несмотря на чрезвычайно высокую значимость статей, опубликованных им в 1905. [<http://ru.wikipedia.org/wiki/H-%E8%ED%E4%E5%EA%F1>]. Приведенные выше данные о профессоре Хэ дополняются его утверждением о том, что его h-индекс составляет 39, в то время как сам Хирш оценивал среднее значение своего индекса для Нобелевских лауреатов в области физики как 35 [Адлер, Фаулер, 2011: с. 61].

Если говорить об индексе Хирша у ведущих российских археологов и этнографов, то хиршерами (имеющим индекс больше единицы) являются академики РАН В.А. Тишков (h-index 6), А.П. Деревянко ((h-index 2) и В.А.Молодин (h-index 1). Относительно маленькая величина данного показателя объясняется тем, что в БД Web of Science, на сегодняшний день 80% журналов относятся к точным наукам, а публикаций указанных авторов по данной тематике сравнительно мало.

Таблица 2. Показатели цитирования ведущих журналов по археологии и этнографии (данные 05.11.12)

Иностранные журналы		Отечественные журналы	
Название	Число цитирований	Название	Число цитирований
Current Anthropology	4478	Российская археология	4333
American Anthropologist	4299	Этнографическое обозрение	3819
American Antiquity	1934	Археология, этнография и антропология Евразии	986
American Ethnologist	1278	Вестник археологии, антропологии и этнографии	427

Обратимся к сравнительным показателям ведущих отечественных археолого-этнографических журналов (Табл. 1). Так занимающий 2 место в рейтинге SCIENCE INDEX за 2010 год по тематике «Исторические науки» журнал «Археология, этнография и антропология Евразии» имеет ИФ РИНЦ почти вдвое превосходящий ИФ других журналов. Однако в общем рейтинге SCIENCE INDEX за 2010 год – лишь 211 место.

Таблица 3. Показатели цитирования и самоцитирования журналов по истории, археологии и этнографии, включенных в список ВАК

Название журнала	2008				2009				2010				2011			
	Публикаций за предыдущие 2 года	Число цитат за предыдущие 2 года	Число самоцитирований	Цитирование в русской версии	Публикаций за предыдущие 2 года	Число цитат за предыдущие 2 года	Число самоцитирований	Цитирование в русской версии	Публикаций за предыдущие 2 года	Число цитат за предыдущие 2 года	Число самоцитирований	Цитирование в русской версии	Публикаций за предыдущие 2 года	Число цитат за предыдущие 2 года	Число самоцитирований	Цитирование в русской версии
Археология, этнография и антропология Евразии	122	23	7	19	123	31	16	27	111	51	18	39	118	33	14	31
Российская археология	178	22	8		200	16	4		198	21	8		159	24	6	
Этнографическое обозрение	48	19	1		123	17	4		164	18	2		156	24	2	
Вестник НГУ	180	13	9		172	18	16		239	25	21		350	55	51	
Вестник Древней истории	67	23	16		95	9	6		126	20	2		109	8	0	
Традиционная культура	104	11	2		99	9	0		119	7	3		138	8	4	
Вестник МГУ	51	6	1		75	0	0		96	2	1		101	6	2	
Вестник археологии, антропологии и этнографии	43	10	8		39	9	8		70	16	13		101	26	16	
Уральский исторический вестник									148	6	1		140	17	6	
Вестник СПбГУ	172	6	3		255	5	4		237	8	1		196	2	1	

При обращении к сравнительным показателям ведущих отечественных и зарубежных археолого-этнографических журналов видно, что только два российских журнала имеют

близкие к журналам «Current Anthropology» и «American Anthropologist» показатели цитирования (Таблица 2).

Анализ табл. 3-4 показывает, что среднее число цитат к числу публикаций журнала «Археология, этнография и антропология Евразии» составил за период с 2008 по 2011 гг. 29,5% (19+25+46+28), а средний показатель самоцитирования составил за этот же период 39,75% (30+52+35+42). Интересен средний показатель цитирования в англоязычной версии – 14,96% к общему числу цитирований журнала. Данное обстоятельство свидетельствует либо о низкой активности читателей этой версии, либо о слабой востребованности журнала за рубежом.

Таблица 4. Показатели цитирования российских археологических и этнографических журналов (%)

Название журнала	2008				2009				2010				2011			
	Публикаций за предыдущие 2 года	Число цитат к числу публикаций	Число самоцитирований к числу цитат	Число внешних цитат к числу публикаций	Публикаций за предыдущие 2 года	Число цитат к числу публикаций	Число самоцитирований к числу цитат	Число внешних цитат к числу публикаций	Публикаций за предыдущие 2 года	Число цитат к числу публикаций	Число самоцитирований к числу цитат	Число внешних цитат к числу публикаций	Публикаций за предыдущие 2 года	Число цитат к числу публикаций	Число самоцитирований к числу цитат	Число внешних цитат к числу публикаций
Археология, этнография и антропология Евразии	122	19%	30%	13%	123	25%	52%	12%	111	46%	35%	30%	118	28%	42%	16%
Российская археология	178	12%	36%	8%	200	8%	25%	6%	198	11%	38%	7%	159	15%	25%	11%
Этнографическое обозрение	48	40%	5%	38%	123	14%	24%	11%	164	11%	11%	10%	156	15%	8%	14%
Вестник НГУ	180	7%	69%	2%	172	10%	89%	1%	239	10%	84%	2%	350	16%	93%	1%
Вестник Древней истории	67	34%	70%	10%	95	9%	67%	3%	126	16%	10%	14%	109	7%	0%	7%
Традиционная культура	104	11%	18%	9%	99	9%	0%	9%	119	6%	43%	3%	138	6%	50%	3%
Вестник МГУ	51	12%	17%	10%	75				96	2%	50%	1%	101	6%	33%	4%
Вестник археологии, антропологии и этнографии	43	23%	80%	5%	39	23%	89%	3%	70	23%	81%	4%	101	26%	62%	10%
Уральский исторический вестник									148	4%	17%	3%	140	12%	35%	8%
Вестник СПбГУ	172	3%	50%	2%	255	2%	80%	0%	237	3%	13%	3%	196	1%	50%	1%

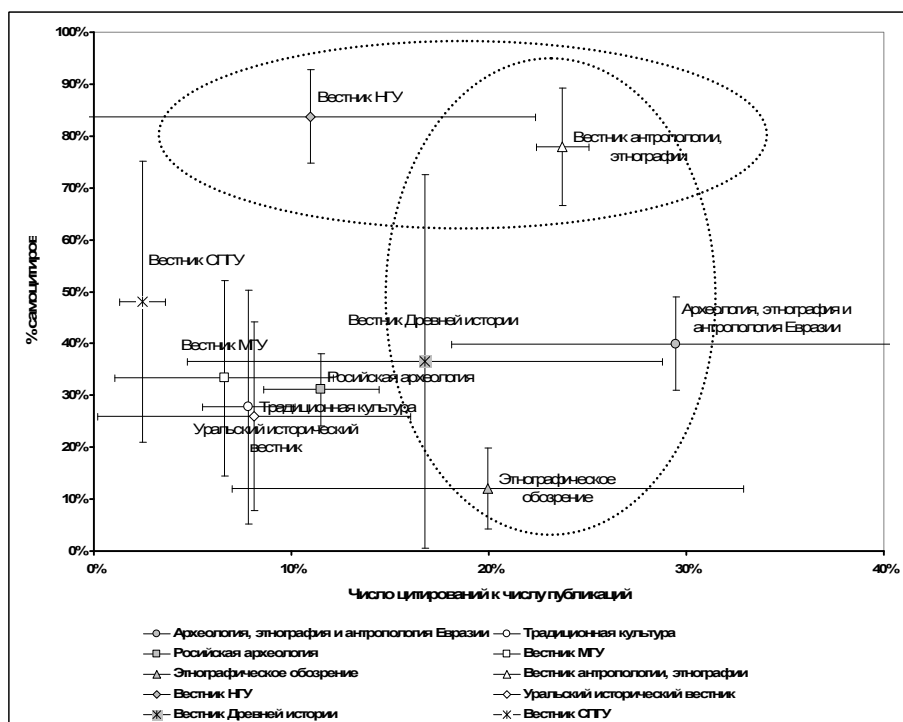


Рис. 1. Средние доли цитирования и самоцитирования за 2008-2011 годы и их доверительные интервалы.

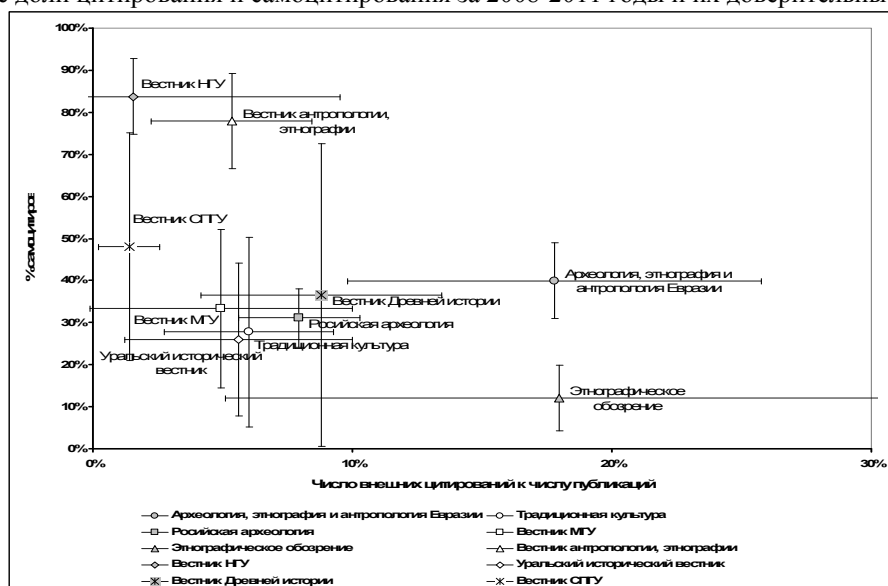


Рис. 2. Средние доли внешнего цитирования и самоцитирования за 2008-2011 годы и их доверительные интервалы.

На рис. 1 те же издания представлены в графическом виде. По оси X отложено среднее значение за 4 года (2008-2011) отношения показателя цитирования к числу публикаций. Видно, что самое высокое значение не превышает 30%. По оси Y отложено среднее за те же 4 года отношение числа самоцитирований к числу цитирований, которое не может превышать единицы. На графике эллипсами выделены 2 группы изданий, первая из которых отличается высоким уровнем самоцитирования («Вестник НГУ» и «Вестник археологии, антропологии, этнографии»), а вторая – высоким уровнем цитирования (тот же «Вестник археологии, антропологии, этнографии», «Вестник Древней истории», «Этнографическое обозрение» и «Археология, этнография и антропология Евразии»).

Заметность издания для внешнего мира тем больше, чем больше цитирование и меньше самоцитирование, то есть чем правее и ниже расположен журнал на графике.

На рис. 2 по оси X вместо общего цитирования приведено внешнее цитирование, из которого исключено самоцитирование. Здесь уже видно, что «Этнографическое обозрение» по внешнему

цитированию не отстает от «Археологии, этнографии и антропологии Евразии», но отстает от нее по самоцитированию.

Заключение

Нас не удивляют невысокие показатели российских археолого-этнографических журналов, т.к. бесконечное «реформирование» российской науки, связанное с внедрением негибких бюрократических правил и сопровождаемое грудой многочисленных и бесполезных отчетных документов, плодящихся как грибы, могут в конечном итоге убить науку. Стремление к опубликованию работы в так называемых рейтинговых журналах «настолько повсеместно приняты, что определяют большинство значимых вещей: иметь постоянную должность или быть безработным, получить грант или нет, быть успешным и защитить диссертацию или быть провальным... Как следствие, научная деятельность деформировалась, а полезность, качество и объективность статей ухудшилась» [Лоуренс, 2011: с. 40]. На наш взгляд творчество в науке не может быть организовано. Оно возникает спонтанно из индивидуального таланта. Об этом говорит пример с Г.Я. Перельманом, который не отвечал принятым стандартам в нынешнем научном мире: ушел с поста ведущего научного сотрудника лаборатории математической физики, уволился из Математического института и практически полностью прервал контакты с коллегами. Но при этом доказал знаменитую гипотезу Пуанкаре. Для настоящего учёного всегда было важно опубликовать свою работу с целью распространения новой информации, способствующую усвоению коллегами и подлежащую проверке при помощи экспериментов или дополнительных наблюдений, а не для того, чтобы получить «талон на питание» [Лоуренс, 2011: с. 40]. Если же редколлегия журналов решат повышать импакт-факторы своих изданий, то для этого следовало бы использовать опыт журнала «Current Anthropology». Публикации совместно с комментариями к ним, несомненно, повысило бы цитируемость статей, помещенных в журнале. Это повысило бы и ответственность рецензентов, относящихся порой недобросовестно к своей задачеⁱ. При этом автор должен иметь возможность ответить на критику, помещенную в этих комментариях. Негативной практикой является отказ редсовета и редколлегии журнала «Российская археология» опубликовать ответ А.А.Формозоваⁱⁱ на критику его книги «Человек и наука», поскольку считали, что дальнейшая дискуссия по поднятым вопросам нецелесообразна — обе стороны высказались с предельной ясностью [Российская, 2006, с. 165–181]. А.А.Формозова было за что критиковать, однако рецензии были банальными. В одних содержались оправдания о неточностях в биографии (Шер). «Надо признать, что в ряде случаев отзывы А.А. о коллегах, особенно провинциальных, были несправедливы. Как видно, в этих случаях обвинения в коммерциализации, некомпетентности он воспринимал с чужого голоса, причём необъективного. Эти обвинения ложились в его схему целой когорты рвачей и мошенников в науке» [Клейн, Щавелев. <http://statehistory.ru...>]. В других содержались утверждения о том, что Формозов давно не ездил в поле и т.д. Я не увидел глубокой конструктивной критики в этих рецензиях. Кроме того, таким отказом редколлегия ничего не добилась. Свой ответ А.А.Формозов разместил в Интернет, а это вызвало ещё большую дискуссию и не всегда положительную для редколлегии журнала.

ЛИТЕРАТУРА

- Адлер Р., Эвинг Дж., Тейлор П. Статистика цитирования // Игра в цифры, или как теперь оценивают труд ученого (сборник статей о библиометрике). – М.:МЦИМО, 2011: с.6-38.
- Арнольд Д., Фаулер К. Гнусные цифры цитирования // Игра в цифры, или как теперь оценивают труд ученого (сборник статей о библиометрике). – М.:МЦИМО, 2011: 53-62.
- Википедия. <http://ru.wikipedia.org/wiki/H-%E8%ED%E4%E5%EA%F1>
- Л. С. Клейн, С. П. Щавелёв. Александр Александрович Формозов (1928–2009). Послесловие // <http://statehistory.ru/books/Aleksandr-Aleksandrovich-Formozov--1928-2009---Posleslovie/3>
- Лоуренс П.А. Потерянное при публикации: как измерение вредит науке // // Игра в цифры, или как теперь оценивают труд ученого (сборник статей о библиометрике). – М.:МЦИМО, 2011: с.39-45.
- Майнстринг Дж. Гениальный «Хиршер» // Terra economicus, 2011, Т.9, №4: с.142-144.
- Российская археология, 2006, № 3, С. 165–181.
- Холошкин Ю.П. Системная археология. – Новосибирск, 2010:

ⁱ В свое время мне удалось прочитать рецензию на статью Р.С.Васильевского и Д.Ю.Березина, из которой следовало, что рецензент статью не читал. Возможно, это был единичный пример недобросовестности рецензента.

ⁱⁱ Свое отношение к последним работам А.А.Формозова я высказал в учебном пособии «Системная археология» [2010].

Ларичев В.Е.

Курыканские всадники: космическая охота степных богатырей Прибайкалья на небесных лося или оленя, символов «ВСЕЯ ВСЕЛЕННОЙ» (опыт реконструкции космогонического мифа¹ кочевников средневековой Сибири)*Памяти Игоря Ивановича Кириллова посвящаю.*

Мемориальные заметки: пара страниц размышлений о былом и ушедших в «Страну невозврата». Удачный разворот жизненного пути молодого учёного в провинциальной Сибири середины прошлого века определяло множество случайностей. Они обычно оставались непредсказуемыми для того, кто пытался сам по себе ориентировать ход событий по заранее продуманному маршруту. Окончательные коррективы всё же вносило время. Оно, да, конечно же, место течения жизни и действующие в ней персонажи диктовали закономерности бытия в прошлом. Всё это и теперь выписывает сюжетные коллизии драм, трагедий и комедий большого числа судеб тех, кто решил посвятить свою жизнь науке

Ключевые слова. Космогонический миф, астроархеология, лунные циклы.

Пишу о том вот к чему: Игорю Ивановичу Кириллову, с кем судьбе угодно было свести меня более полувека назад, повезло и со временем, и с местом благодаря двум случайным обстоятельствам. Первое связано с тем, что в начале второй половины прошлого века в Читинском педагогическом институте историю Греции, Рима, ранних государств Ближнего Востока и Северной Африки преподавал Михаил Иосифович Рижский, редкого дара учёный и талантливый педагог, тонкий знаток латыни, Библии и первоисточников по истории, культуре и философии государств Средиземноморья эпох античности и эллинизма [подробности см. Пиков, Мякин, 2002]. Он принадлежал к элитному кругу учёного мира столицы и однажды не по своей воле оказался далеко от него – на востоке Сибири, в Забайкалье, крае ссыльных декабристов.

Зная педагогическую элиту столицы Забайкалья середины прошлого века, Читы, легко вообразить, насколько важна была эта «случайность» для тех, кто учился в 50-е годы в пединституте, готовился стать учителем в школе и кто из молодых преподавателей читал курсы по истории Отечества и стран зарубежья. М.И. Рижский на каждой лекции, в частных беседах, на заседаниях Учёного совета и консультациях для студентов давал мастер-классы приёмов передачи знаний и обретения их в большем объёме и широте. Его в учебном и учёном закулисье города почтительно именовали «папа Рижский», отдавая тем самым должную дань уважения его мудрости, доброте, самоотверженной увлечённости просвещением и служению науке.

Перелистывая книги Игоря Ивановича по случаю организации в столице Забайкальского края первых «Кирилловских чтений», я замечаю в них отблеск влияния М.И. Рижского на творчество его. Оно, влияние это, было обоюдным, ибо он, в свою очередь, вводил столичного невольника в курс проблем забайкальской и прибайкальской археологии. Результатом сотрудничества стало сотворчество – написание учебного пособия для студентов пединститута. Оно же стало тогда и своего рода путеводителем по древностям Восточной Сибири для каждого горожанина и селянина Забайкалья, кто любопытствовал по части первобытных культур и краеведения [см. Кириллов, Рижский, 1973]. Не могло обойтись без бесед с И.И. Кирилловым и написание М.И. Рижским книги «Из глубины веков. Рассказы археолога о древнем Забайкалье» [Рижский, 1965], расширенного и дополненного варианта научно-популярных очерков древней истории того же региона Северной Азии [краткие реплики об археологических исследованиях М.И. Рижского см. Ларичев, 2005; 2006].

Второе обстоятельство связано с началом обретения сибирской археологией иного статуса – академического. Тому способствовали возобновление работы экспедиций А.П. Окладникова в Забайкалье, Амурской области, на Дальнем Востоке и на территории Монгольской Народной республики, а затем организация в конце 50-х годов Сибирского отделения Академии наук СССР и включение в структуры его в начале 60-х годов первого гуманитарного подразделения – Института истории, филологии и философии, специализированной частью которого стал Отдел археологии и этнографии (позже преобразован в Институт археологии и этнографии Сибирского отделения Российской Академии наук).

¹ Здесь в смысле учения (понятий о Мироздании) жречества древних культур и этнографической современности Сибири.

И.И. Кириллов начал в ту пору выполнять программу исследований древних культур Забайкалья при поддержке А.П. Окладникова, который по переезду из Ленинграда в Новосибирск стал главным куратором сибирской археологии. То была весьма специфическая школа научного поиска, нацеленная на решение проблемных задач с охватом источников от эпохи палеолита до палеометалла и средневековья. Допустимый объём статьи не позволяет представить в должном формате итоговые результаты работы И.И. Кириллова. Поэтому ограничусь кратким резюме: *его перу принадлежит, помимо прочего, фундаментального разряда труд, который стал основой докторской диссертации, а организационная деятельность привела к формированию в Чите того, что принято называть теперь «школой» – самодостаточного коллектива археологов*, способного вести изыскания по плану, нацеленному на актуальное не только в текущий момент, но, главное, на стратегическое и на дальнюю перспективу [см. Окладников, Кириллов, 1980].

Научная деятельность И.И. Кириллова проходила в исключительно важный момент сибирской археологии. В 60 – 80-е годы она начала освобождаться от пут статуса, двусмысленно названного академиком РАЕН А.И. Мартыновым «колониальным» (имеется в виду исполнение исследований древностей Сибири командированными из столиц специалистами, которые использовали местные кадры «на подхват», а в последующие годы взяли за правило поучать провинциалов по всякому поводу да и без повода тоже. Стремление к монополизму столичных авторитетов теперь кануло в лету, ибо в пределах Сибири, от Омска и Тобольска на Западе, а также до Хабаровска, Сахалина и Владивостока на Востоке сложились самостоятельные центры региональной археологии, сотрудники которых, более бесценные не нуждаются в опеке. Одним из таких центров стало Забайкалье, бесценное наследие научной и организационной деятельности И.И. Кириллова.

В заключение беглых заметок скажу несколько слов о характере Игоря Ивановича: *для него были примечательны обострённое чувство справедливости, нацеленность на борьбу с негативом в археологическом сообществе, а также стремление содействовать благу в науке*. Меня он без колебаний поддержал в годы унижительных поучений столичных авторитетов, а когда мне потребовались его уникальные материалы по раннему неолиту Забайкалья, то по первой же просьбе переслал их в Новосибирск, нарушая притом препятствующие такому делу административные каноны [см. Ларичев, 2007 а; 2007 б]. По всему этому я избрал для статьи, написанной в память о нём, сюжет из набора тех, по которым приходилось 30 лет назад вести противоборство с завзятыми «колонистами», представляющими северную столицу.

Вводные замечания. Среди множества специфических тем сибирской астроархеологии, нацеленной на раскрытие мировоззренческих установок интеллектуальных элит архаических культур, не владеющих письменностью, самыми сложными остаются сюжеты, связанные с космогонией и космологией – *реконструкциями идей возникновения Мировоздания, устройства его и бытия высокого ранга божеств, творцов «окружающей действительности», т.е. Природы*; выяснения места, которое занимали в её структурах обитатели Земли – сообщества живых и отошедших в инобытие людей, животные и растительные миры.

Представители традиционной (вещеведческой) археологии), для которых палеоастрономия и палеокалендаристика – табу, решают эти проблемы прямолинейно, «по аналогии», с привлечением этнографо-мифологических материалов, посредством совмещения их с археологическими источниками. В них они высматривали отблески стародавних мироведческих концепций. В наибольшей степени успех в том сопутствовал в середине прошлого века историку и археологу А.П. Окладникову (он решал эти вопросы с широким привлечением наскальных изображений эпох неолита и палеометалла Прибайкалья и Забайкалья) и А.Ф. Анисимову, превосходному знатоку религий коренных народов Сибири. Его привлекли семантические реконструкции А.П. Окладникова [см. Окладников 1950; Анисимов, 1958; 1959; там же обширные списки литературы вопроса; см. также Титов, 1923].

Постановка проблемы и программная цель исследования. Тому и другому авторам недоставало, однако, одного – *должной меры доказательности идеи отражения информации о Космосе (т.е. о Небе и светилах) посредством отдельных образов и многофигурных композиций писаниц*, размещённых на скальных плоскостях храмов каменного и бронзового веков. Она, этой великой ценности информация, рассматривалась ими, большей частью, в рамках общего плана понятий о мифологическом сознании и мышлении, но (и в этом была определяющая слабость идеи) *вне учёта хорошей осведомлённости сибирского жречества эпохи первобытности и средневековья в астрономии, знания ими систем счисления времени, порядка чередования сезонов солнечного года, закономерностей смен фаз Луны, умения совмещать в единый поток время лунное, солнечное, звёздное, драконическое и аномалистическое*. Сказано это вовсе не в упрек А.П. Окладникову и А.Ф. Анисимову, ибо метод астроархеологии стал достоянием науки о древностях лишь в последние десятилетия прошлого века, да и то только для тех представителей её, кто не пугался стать сторонником «непопулярной научной традиции». Истоки разработки метода восходят к последним десятилетиям XIX века [обзор проблемы см. Ларичев, 2008].

Представлю вариант астроархеологического стиля интерпретации средневековых наскальных изображений из Шишкино (Прибайкалье, верховья р. Лены). В них запечатлены три варианта космического мифа об охоте на похитителя Солнца то ли тотемного (в образе животного) предка, то ли земного (т.е. реального) богатыря-охотника, который сподобился вознестись на Небо и совершить подвиг во благо погружённых во тьму смертельно перепуганных соплеменников [см. первые монографические публикации: Окладников, 1959; Окладников. Запорожская, 1959]. Речь идёт о варианте эвенкийского мифа, повествующего об охоте космического медведя Манги (или антропоморфного облича героя Маина; у якутов он – Халлан Уола; у алтайцев – Кочутэ) на самого великого по мощи животного-небожителя Хэглэна (Харги), похитителя Солнца, на рогатую «Мать Вселенной» или «Мать-Землю», носительницу небесных светил, воплощение всего Мироздания от Преисподней внизу до звёзд вверх. В *космичности содержания этого мифа* нет причин сомневаться. Он полон *указаний на созвездия*: Хэглэн – Большая Медведица (лось или олень), а телёнок животного – Малая Медведица; охотник – медведь Манги (великан, чудовище, сказочный богатырь, дух предка, первопредок, «Хозяин Нижнего мира») – созвездие Волопас и его ярчайшая звезда Арктур, именуемая астрономами из-за яркости «Северным Сириусом», *первое утреннее появление которого в средних широтах Сибири предвещает восход Солнца в день весеннего равноденствия, а заход – в равноденствие осеннее*.

Средневековые образы и композиции выбраны для анализа вот почему: *они иллюстрируют миф, датированный временем, далеко отстоящим от этнографической современности, когда записывались изложенные выше архаические варианты его*. Сделано это с надеждой, что в рисунках средневековья содержатся детали, выпавшие из памяти информаторов XIX–XX веков, знатоков старинных преданий, многое в повествовании которых остаётся порой непонятным или недосказанным.

Источники. Персонажи композиций. Методические установки исследования. Среди множества сцен охоты курыканских воинов, обнаруженных на Шишкинских скалах, особое внимание привлекают композиции на плоскостях 2 первого яруса скалы XIII и 60 второго яруса скалы XIX (рис. 1 и 3); копии Л.С. Мельникова и С. Николаева [см. Мельников, Николаев, Демьянович, 2011; табл. 39, рис. 8, 9; табл. 149, рис. 1–6]. Обращусь сначала к первой из выбранных композиций. Характерные для стиля искусства курыкан подчеркнуто изящная фигура всадника и «небесной красоты» лошадь под ним (см. рис. 1) изображены в близком соседстве со странным, явно «не от мира сего», гибридного облика существом, туловище которого и голова с «серьгой» на подбородке – *лося*, а краса этого таёжного гиганта, рога, – *оленья*. Мало того, от лба озадачивающего облича существа выдвинут вперёд едва заметный, прямой, шпаговидно тонкий рог, знаковый признак персонажа сказочно-мифических повествований – *единого рога*, а одна из передних ног – *непомерно длинная, серповидно изогнутая в нижней половине* – странная деталь, не поддающаяся объяснению.

Разгадать канву сюжета двухфигурной композиции не составляет труда: персонажи двигаются один за другим, и, понятно, что *охотник преследует пугающе экзотического вида добычу*. Необычна в каноничной, кажется, сцене ещё одна заставляющая задуматься деталь: человек, видимо, не намеревался (опасался? не мог?) убить странного зверя, ибо в руке его находится *не орудие смерти* – копьё, кинжал, нож или лук со стрелой, а *аркан*. Надо полагать, цель изначально, при подготовке к преследованию, была у него иная – *становить бег чудовища, быть может, обуздать его и подчинить своей воле*.

Но для чего «остановить и подчинить»? И, наконец, главное – *какая причина спровоцировала начало погони*, чтобы этот, вобщем-то, банальный сюжет оказался достойным воспроизведения на поверхности камня с такой старательностью и безупречно высоким художественным вкусом?

Напрашиваются три варианта ответа, контент каждого из которых *зависит от выбора методики интерпретационно-семантических реконструкций образов архаического искусства*:

1 – *традиционный, аналоговый, метод*, использованный А.П. Окладниковым и А.Ф. Анисимовым. Метод такой предполагает *совмещение этнографо-мифологических повествований с образами и композициями древнего искусства* – своего рода иллюстрациями к изложенным в мифе сюжетам;

2 – *пифагорейский, числовой, метод* семантических реконструкций – «*Всё (Вселенная) происходит из числа и объясняется посредством числа*». Этот *естественно-научный*, т.е. отличающийся определённой (точностью) установок, а не *гуманитарной расплывчатостью* канон, в традиционном искусствоведении Сибири относится к «непопулярной научной традиции». Она, как таковая, не воспринимается, не используется и «резко критикуется». Надо думать, отвергается по прискорбной неосведомлённости, ибо он-то, метод этот, как раз и *позволяет привести в искусствоведческие интерпретации то фундаментально*

желанное, чего всегда недоставало (и всегда будет недоставать) методу аналоговому – доказательности предлагаемым «прочтениям» сакрального (не лежащего на поверхности, неочевидного, скрытого) смысла образов и композиций писаниц храмов и предметов искусства малых форм. Это вариант методики использовался неоднократно при интерпретациях образов и композиций искусства Евразии. Для сюжетов статьи «Курыканские всадники...» полезным будет предварительное знакомство со статьями: [Ларичев, 1996; 1998; 2010; 2011]. В них анализируются источники, подтверждающие осведомлённость палеолитического жречества о 19-летнем лунно-солнечном цикле Метона, о фиксации его посредством количества отростков рогов оленя (пещера Ляско) и о знании того же цикла представителями жречества культур раннего железного века юга Западной и Восточной Сибири, а также центральной Азии;



Рис. 1. Всадник с арканом преследует гибридного облика существо.

3 – комбинационный, аналогово-пифагорейский метод, когда мифическое повествование, связанное с Верхним миром Вселенной, т.е. с Небом и светилами (а значит, с астрономией и, быть может, с календаристикой), следует сопровождать числовым подтверждением верности будто бы фантастически сказочного, придуманного, отрешённого от реальностей, а то и абсурдного – порождений «мифологического сознания», не способного объяснить Природу научно (по невежеству своему, простительному, видите ли, для варваров первобытности и средневековья).

Воспользуюсь комбинационным методом для «прочтения» (расшифровки) сюжета первого яруса скалы XIII (рис. 1). Выполню эту задачу в рамках содержания мифа о космической охоте Манги (Маина, Халлан Уола или Кочуктэ) на Хэглэна (Харги), похитителя Солнца, владыку Буни, Нижнего мира, и Угу буни, Верхнего Мира.

Внеземное многообразное чудище жречества курыкан. Его знаково-символическая информатика: календарно-астрономический контекст «записей» чисел. Мифологичность (фантастичность, сказочность, внушающее ужас невероятностью обличья) существа, объекта преследования всадника, очевидна. *То не лось или олень, а слитые воедино одно с другим*, что по естеству Природы невозможно, но в «мифологическом сознании» – почему бы и нет?

Но в чудище есть столь же очевидная реальность Природы, зафиксированная вне «придумок» мифологического мышления. Речь идёт о числах, представленных своеобразными счётными единицами: 16-ю отростками двух ветвей оленьих рогов, шпаговидным рогом на лбу (1), подшейной серьгой (1) и самим носителем рогов – головой лося (1). Эти знаки составляют при суммировании одно из самых великих чисел астрономии и лунно-солнечной календаристики (см. рис. 2 а и б):

$$(16+1+1+1) \text{ лет} = 19 \text{ лет, если каждую единицу принять за символ года.}$$

Честь открытия этого числа историки точных наук традиционно приписывают греческому астроному Метону. Ему удалось (так уверяют историки астрономии и календаристики) первому из античных наблюдателей Неба почти с идеальной точностью (а идеальная попросту невозможна из-за неуничтожимой дробности календарно-астрономических величин) слить воедино лунный и солнечный потоки времени. Это не удавалось сделать ни при отслеживании лунного времени трёхлетиями с однократным месячным дополнением в конце цикла (самый популярный в древности период многолетия), ни при использовании 8-летнего лунного периода с интеркаляцией дополнений в 3, 5 и 8-ой годы (рациональность цикла заслуживает особого разговора, и он будет проведён далее). Метон предложил считать время лунными 19-летиями, в которых интеркаляционными (дополненными) предложил считать 7 лет, а простыми (без дополнений) – остальные 12. Итог впечатляет: длительность 235 синодических (фазовых: смещение Луны относительно Солнца) месяцев, которые как раз и составляют такое 19-летие, оказывается предельно близкой длительности 19-и солнечных лет:

$$235 \text{ синодических месяцев} \times 29,5306 \text{ сут.} = 6939,691 \text{ сут.};$$

6939,691 сут.:365,242 сут.=19,000254≈19 солнечных лет.

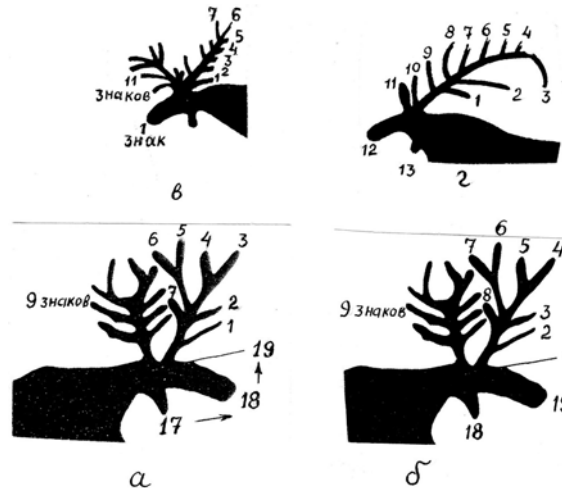


Рис. 2. Голова гибридного облика существа с рогами как элементами счётных систем: а – если за начало отсчёта принять нижний правый отросток правой ветви рогов, то получим на этой ветви число 7, а остальные элементы счётной системы – левая ветвь (9 знаков), 1 серьга, 1 голова и 1 шпаговидный рог лба составят число 12; в 7-и годах дополнительный месяц вводился, а в 12-и – не вводился; б – если за начало отсчёта принять шпаговидный рог, то получим «записи» следующих циклов: № 1-3 – трёхлетие; № 1-8 – восьмилетие; № 1-18 – восемнадцатилетия; № 1-19 – девятнадцатилетия (цикл Метона); в – голова гибридного облика существа, покорённого всадником; ветви рогов и голова – «запись» 19-летнего цикла (серьга не принималась в расчёт); г – голова гибридного облика существа – «запись» месяцев лунного и солнечного годов (см. рис. 3 б).

Возникает вопрос – можно ли выстроить порядок считывания знаков так, чтобы они оказались чётко подразделёнными на две группы – 7 лет и 12 лет? Да, если начинать счисления с нижнего отростка правой ветви рогов и смещаться далее вверх, а затем вниз, в результате чего получим число 7 (см. рис. 2 а) правой ветви оленьих рогов (7 знаков), а после считывать 9 отростков левой ветви рогов, подбородочную «серьгу» (17-й знак), голову лося (18-й знак) и шпаговидный рог (19-й знак). Отростки второго рога и прочие знаки как раз и составят число 12, вторую часть 19-летия. Такая система проверяема на точность: если, допустим, сутки начала 19-летия определяло полнолуние, совпавшее с днём летнего солнцестояния, то того же самого явления следовало ожидать в конце метонова периода, и, случись такое, значит календарист без потерь точно отслеживал время на протяжении всех 19 лет.

Столь скрупулёзная мера точности была необходима не только потому, что лунное время, фазы ночного светила, начала и окончания месяцев, полугодия, т.н. «сезонов» лунного года (3 цикла по 118 суток), а также начало и окончание каждого годового периода, (трёхлетия, восьмилетия и т.д.), определяли даты религиозных праздников, в чём ошибка для служителей богам считалась недопустимой и наказывалась разжалованием, а то и смертью. Она требовалась тем, кто должен был предсказывать время наступления затмений. А всё дело в том, что 19-летие составляло базовую часть Большого Лунного Сароса:

$(19+18+19)$ лунных лет = 56 лунных лет,

периода, по истечении которого повторялось затмение, случившееся более полувека назад в месте наблюдения того же события; Сарос – повтор местоположения в небесном пространстве Луны, Солнца и узлов лунной орбиты, когда происходят затмения. Разделяющий 19-летия период 18 лет мог отсчитываться по той же счётной системе головы чудовища с учётом одной единицы (допустим, серьги или, что вероятнее, – шпаговидного рога).

Такой приём открывает возможность счисления Большого Солнечного Сароса:

$(18+18+18)$ солнечных лет = 54 солнечных года,

периода, по истечению которого повторялось затмение в месте наблюдения самого ошеломляющего небесного явления – исчезновения Солнца и появление на его месте чёрной дыры, бездны, лишённой звёзд.

Изложенным информационный контент водружённых на голову лося шпаговидного рога и «серьги», не исчерпывается. Если за точку начала счисления знаков принять шпаговидный, тонкий, длинный, едва заметный рог (хвала и честь Л.С. Мельникову и В. С. Николаеву за редкостную точность копирования чудовища!), а затем считывать отростки правой ветви оленьего рога, после чего – левой (см. нумерацию знаков на рис. 2 б), то можно выделить следующие «записи»:

1 – *трёхлетнего лунного цикла* (см. на рис. 2, б № 1–3; далее следует изменение вида знаков – появляются «рогульки», символы двойцы); по окончании такого периода в счётную структуру вводился интеркаляционный «месяц» длительностью 34 суток²:

$$(354 \text{ сут.} \times 3) + 34 \text{ сут.} = 1096 \text{ сут.};$$

$$1096 \text{ сут.} : 365,242 \text{ сут.} = 3,0007501 \approx 3 \text{ солнечных года};$$

2 – *восьмилетнего лунного цикла* (см. на рис. 2 б № 1–8); включаются все отростки правой ветви рога; при счислении времени такого цикла *дополнительные лунные периоды* подключались к третьему (34 сут.), пятому (22 сут., $\frac{3}{4}$ синодического месяца) и восьмому (34 сут.) годам [см., соответственно, № 3 (позиция и знак заметны), № 5 (позиция и знак заметны) и № 8 (последний знак правой ветви оленьего рога)]. Такой расклад лунно-солнечного *восьмилетия* – столь же прекрасен, как тот, что использовался жрецами для отсчета времени календарно-астрономического периода длительностью 19 лет. В *восьмилетии* целое число синодических лунных месяцев (99) \approx равно по количеству суток в целом числе солнечных лет (8) и целом числе (5) синодических (относительно Солнца) оборотов величайшей (по яркости и почитанию человеком из-за связи её с двумя циклами беременности женщины) планеты – Венеры:

$$99 \text{ синодических месяцев} \times 29,5306 \text{ сут.} = 2923,5294 \text{ сут.};$$

$$8 \text{ солнечных лет} \times 365,242 \text{ сут.} = 2921,936 \text{ сут.};$$

$$5 \text{ синодических оборотов Венеры} \times 583,9 \text{ сут.} = 2919,5 \text{ сут.}$$

Далее осуществлялся выход на рубеж окончания 18 и 19-летия (считывались 9 отростков второго рога, серьга (17-й год), носитель знаковой системы, головы (18-й год) и шпаговидный рог лба (19-й год).

Прежде чем перейти ко второй части исследования – раскрытию мотива преследования чудовища и семантической значимости столь странного образа (в свете результатов «прочтения» связанного с ним числового «текста») сначала развею сомнение в осведомлённости кочевников Центральной Азии и Южной Сибири о цикле Метона. Допустимый объём статьи исключает пространное опровержение скептицизма, и потому ограничусь репликой с указаниями на литературу вопроса: *открытие великого цикла уходит в глубины ранней поры верхнего палеолита* [см. Ларичев, 1996]. Цикл Метона фиксируют 1 ухо и 18 отростков 2-х ветвей рогов оленей из Ляско, а в ожерелье из Лабастю – 1 голову бизона и 18 голов горного козла [см. Ларичев, 1998]. У ранних кочевников востока Азии факты знания его зафиксированы в знаково-образной форме, по меньшей мере, в эпоху поздней бронзы и раннего железного века за полтора-два тысячелетия до появления курыкан в Прибайкалье и за несколько веков до времени жизни Метона в Афинах [подробности см. Ларичев, 2010; 2011].

Сюжетная канва космогонического мифа об олене или лосе, похитителях Солнца, в записях этнографов не богата событиями. Гигантское копытное, которое скрывается в чащах «небесной тайги» (среди невидимых в лучах дневного светила созвездий), появляется на горах западного горизонта вечером, подхватывает Солнце на рога и скрывается за край Земли. Наступает темень. Ради появления света на следующий день, охотник пускается в погоню, убивает оленя и утром возвращает похищенное на восточную окраину небосвода. В «небесной тайге» прячется, однако, оленёнок. Он, превратившись в конце дня во взрослого, вновь подхватывает Солнце на рога и погружает Землю в темноту, что вынуждает охотника опять пуститься в погоню. Так будет продолжаться до скончания века...

Предлагаю внести коррективы в изложенное, учитывая исключительно высокий уровень астрономических знаний, совершенство и вариантность годовых и многолетних систем счисления времени тех, кто запечатлел на плоскости скалы XIII Шишкино композицию погони за чудовищем.

Его голова несёт на рогах не только Солнце, но также Луну, по крайней мере, одну из планет – Венеру, а, быть может, и другие «блуждающие звёзды» (что потребует продолжения «прочтения», числового «текста», связанного с головой странного существа). Если так, то становится понятной идея восприятия его воплощением Мироздания с тремя главными структурами – Верхним Миром и светилами, Средним Миром, Землёй, обителью человека, а

² Период от последней видимости серпа первого синодического месяца + весь цикл второго месяца + серп первой видимости третьего месяца.. Это и время, кратное сидерическому (цикл смещения Луны на фоне звёзд) месяцу (34 сут.: $27,32 \text{ сут.} = 1,2445 \approx 1\frac{1}{4}$ сидерического м. месяца), времени реальной длительности оборота ночного светила вокруг Земли, что нужно знать при расчётах наступления затмений; 34 сут. – эпоха солнечного затмения, цикл прохода Луны в зоне узла её орбиты, т.е. точки пересечения лунного пути с эклиптической – затменной «дорогой» Солнца).

также того, кто погонится за чудовищем, и Нижним Миром, обителью предков. Можно ли после такого истолкования чудовища удивляться странностям облика его?

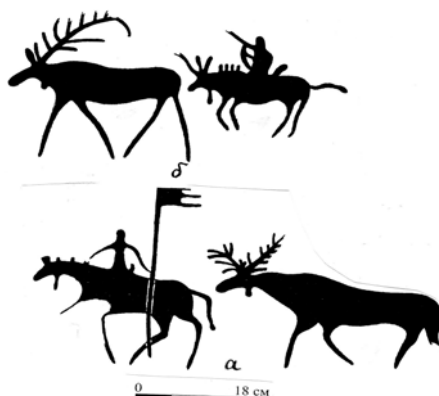


Рис. 3: *а* – всадник со знаменем и покорно следующее за ним гибридного облика существо. Композиция, символизирующая победоносное возвращение богатыря в северную сферу Мироздания (Солнце оказывается в точке *весеннего равноденствия*, пересекает Небесный экватор и оказывается в Верхнем мире, обители человечества); *б* – всадник преследует гибридного облика существо – воплощение лунно-солнечного цикла (10 отростков рога, 1 ухо, 1 голова и 13-й месяц, выравнивающий лунное трёхлетие с трёхлетием солнечным (см. 2, *г*)).

А теперь о погоне всадника – что вызвало необходимость её, по какому маршруту и сколь долго продолжалась она, а также кто они, этот всадник, и его лошадь? Солнце, пересекая Небесный экватор, который подразделяет Мир на две полусферы – Южную и Северную, в день *весеннего равноденствия* начинает смещение в просторах последней (*весенний астрономический сезон*; 94 дня) до трёх дней летнего солнцестояния (начало последней декады июня, наибольшего сближения светила с астрономическим Севером). А далее следуют сутки солнцеворота, *начала движения Солнца на Юг*, в сторону Небесного экватора, к точке *осеннего равноденствия* (*летний астрономический сезон*; 92 дня). По-видимому, солнцеворот следует принять за момент поднятия на рога дневного светила, вызванный стремлением вызванное стремлением чудовища переместить его в Южную сферу, т.е. в Нижний Мир, обитель душ умерших, где в то время наблюдается Луна. Тогда-то и начинается погоня, вызванная желанием всадника предотвратить скатывание Солнца в Южную сферу, а заодно и возвратить в Северную сферу Луну, летнее светило Мира мёртвых (зимой она наблюдается на Севере). Поскольку в руках всадника аркан, то погоня предполагает не убийство, а захват, и затем изменение направления движения чудовища в противоположную сторону, на Север, чтобы наступили там, на Земле, тёплые дни весны и лета.

Но, видно, удача не сопутствует всаднику до конца суток летнего сезона, в день *осеннего равноденствия*, а также в последующие три месяца *осеннего астрономического сезона* (89 дней), когда Солнце начинает смещаться в южной сфере Мироздания, *крае мёртвых*, до *начала последней декады декабря*. Остановка дневного светила в ближайшей к Югу точке горизонта на трое суток – это момент, который мог интерпретироваться творцом мифа как момент меткого броска аркана на рога уставшего чудовища (Солнце в декабре смещается по горизонту всё медленнее и медленнее) и, наконец, обретает неподвижность в начале *последней декады декабря*, *всходя и заходя в одних и тех же точках горизонта на Востоке и на Западе*. А далее следуют сутки *зимнего солнцеворота*, вслед за которым начиналось всё более ускоряющееся смещение Солнца на Север, к Небесному экватору и *дню весеннего равноденствия*. Возможно, всадник сначала волок за собой светило на аркане, постепенно убыстряя ход коня, а после пересечения Небесного экватора, побеждённое чудовище более не сопротивлялось и покорно следовало за всадником со знаменем в течение всего *весеннего астрономического сезона* (см. рис. 3 *а*). Что это было то же самое гибридное существо подтверждает знаковая «запись» 19-летия, которую составляют 18 отростков двух ветвей оленьих рогов и 1 несущая их голова лося, лишённая ипаговидного рога (см. рис. 2 *г*; позиционирование чисел 7 и 12 сохраняется; *серьга остаётся вне учёта*). Если так, то композиция на плоскости 60 второго яруса скалы XIX представляет собой второй акт космической драмы, вслед за которой вновь последует первый акт – погоня всадника с арканом. Такая ритмика событий будет продолжаться «до скончания века» – времени существования Вселенной.

Сказать о всаднике что-либо новое затруднительно. Это, конечно же, эпический герой курыкан, прославленный в боях, но он, видимо, и предок клана вождей. Его могущество стало, в

конечном счете, столь беспредельным, что он оказался способным вмешиваться в ритмы бытия самого Мироздания, обеспечивая его стабильность. Подстать ему была и лошадь, видимо, подобие крылатого коня мифологии нивхов, который после удара плетью был способен вознестись в Небо [см. Штернберг, 1908].

Миф о космической погоне не ограничивается преследованием существа, которое *воплощает 19-летний лунно-солнечный цикл*. На той же плоскости 60 второго яруса скалы XIX располагается иная композиция охоты всадника на гибридное существо (рис. 3 б). Знаковую «запись» составляют 10 отростков рога, 1 ухо, 1 голова животного и 1 «серьга» (всего 13 элементов; см. рис. 2 в). Если каждый из них символизирует *лунный месяц*, то 12 знаков призваны отразить длительность *двух следующих один за другим простых лунных годов*, а 12+1 («серьга») – *третий год (эмболистмический, с дополнительным месяцем, выравнивающим лунное трёхлетие с трёхлетием солнечным)*. Иносказательная содержательность мифа, связанная с охотой на воплощение *трёхлетнего лунно-солнечного цикла*, остаётся той же, что и в случае с существом, воплощающим *19-летие*.

Краткий итог исследования. Курыканский миф о «Небесной охоте» представляет собой, по-видимому, синтез космологических сказаний степняков юга Сибири и коренных обитателей прибайкальской тайги – эвенков. Восприятие головы гибридного существа, носителя многоветвистых рогов и подшейной «серьги», как образно-знакового «текста» календарно-астрономического содержания, в очередной раз убеждает, что *древняя мифология, не есть плод отрешённых от реалий Природы «придумок» людей, бессильных понять действительность*. Это, в реальности, *мировоззренчески глубокие иносказания, за которыми скрываются положительные знания о Мире. Они, эти знания, – обретение не средневековья, а наследие эпох первобытности Северной и Центральной Азии, воспринятое, творчески переработанное и дополненное. Они – не результат заимствований из центров протоцивилизаций юга Евразии, а возвращённое в степях и тайге Гипербореи*.

ЛИТЕРАТУРА

- Анисимов А.Ф. Религия эвенков в историко-генетическом изучении и проблемы происхождения первобытных верований. – Москва–Ленинград, 1958.
- Анисимов А.Ф. Космологические представления народов Севера. – Москва–Ленинград, 1959.
- Кириллов И.И., Рижский М.И. Очерки древней истории Забайкалья. – Чита, 1973.
- Ларичев В.Е. Космографическая миниатюра из Раймонден (Солнце, Луна и Вселенская жертва в искусстве древнекаменного века) // Гуманитарные науки в Сибири. Серия: Археология и этнография. №3, 1996.
- Ларичев В.Е. Космографическая композиция из «Большого зала» Ляско // Гуманитарные науки в Сибири. Серия: Археология и этнография. №3, 1998.
- Ларичев В.Е. Устремлённая в Небо (страны Света, светила и Время в мировоззрении жречества китойской культуры Забайкалья) // Сибирь на перекрестье мировых религий. – Новосибирск, 2005.
- Ларичев В.Е. Семантика образов человека, птицы и змеи в контексте календарно-астрономических «записей» в наскальном искусстве Забайкалья (реконструкция систем счисления времени в эпоху палеометалла Бурятии) // Сибирь на перекрестье мировых религий. – Новосибирск, 2006.
- Ларичев В.Е. Время, запечатлённое в камне (семантика пластинчатой скульптуры с числовыми знаковыми «записями» из неолитического погребения Забайкалья) // Каменная скульптура народов Евразии. Труды Сибирской ассоциации исследователей первобытного искусства. Вып. 3. – Барнаул, 2007 а.
- Ларичев В.Е. Китойский календарно-астрономический жезл (к проблеме реконструкции систем счисления времени в раннем неолите Восточной Сибири) // Note Bene. Вып. 1. Случайная находка. – Новосибирск, 2007 б.
- Ларичев В.Е. Астроархеология: «Сквозь тернии – к звёздам!». Начало становления «непопулярной научной традиции». // Астроархеология – естественно-научный инструмент познания протонаук и астральных религий жречества древних культур Хакасии. – Красноярск, 2009.
- Ларичев В.Е. Наскальное искусство юга Якутии и Верхнего Приамурья: астроархеологические методы интерпретаций образно-знаковых «записей» эпох неолита и палеометалла (к проблеме реконструкций систем счисления времени в древних культурах юга Восточной Сибири) // Традиционная культура Востока Азии. Выпуск 6. Посвящается памяти археолога и этнографа А.И. Мазина (1938 – 2008). – Благовещенск, 2010.
- Ларичев В.Е. Оленные камни: опыт интерпретаций отдельных стел и рядов их в структурах святилищ и погребальных комплексов (к проблеме космичности мировоззрения жречества ранних кочевников Северной и Центральной Азии) // Информационные технологии в гуманитарных исследованиях. Вып. 16. – Новосибирск, 2011.
- Ларичев В.Е., Рижский М.И. Озёрный неолит и ранняя бронза Восточного Забайкалья (стоянки в районе озёр Иргень и Кенон) // Сибирский археологический сборник. – Новосибирск, 1966.
- Мельникова Л.В., Николаев В.С., Демьянович Н.И. Шишкинская писаница. Том I. История и методика изучения, проблемы музеефикации, описание петроглифов. – Иркутск, 2011.
- Окладников А.П. Неолит и бронзовый век Прибайкалья. Части I и II. Материалы и исследования по археологии СССР, № 8. – Москва–Ленинград, 1950.
- Окладников А.П. Шишкинские писаницы. Памятник древней культуры Прибайкалья. – Иркутск, 1959.
- Окладников А.П., Запорожская В.Д. Ленские писаницы: Наскальные рисунки у деревни Шишкино. – М.–Л., 1959.

- Окладников А.П., Кириллов И.И.** Юго-Восточное Забайкалье в эпоху каменного века и ранней бронзы. – Новосибирск, 1980.
- Пиков Г.Г., Мякин Т.Г. М.И. Рижский** – учёный-библеист // Сибирь на перекрестке мировых религий. – Новосибирск, 2002.
- Рижский М.И.** Из глубины веков. Рассказы археолога о древнем Забайкалье. – Иркутск, 1965.
- Титов Е.И.** Некоторые данные о культе медведя у нижнеангарских тунгусов кандигирского рода. Сибирская живая старина. – Иркутск, 1923.
- Штернберг Л.Я.** Материалы по изучению гилацкого языка и фольклора. Т. III. Образцы народной словесности. Часть I. – СПб, 1908.

КОММЕНТАРИЙ

В.Е. Ларичев – один из наиболее авторитетных и творческих представителей отечественной археологии и этнологии. Обсуждаемая статья продолжает серию его работ, посвященных проблеме, которая была впервые поставлена им еще в начале 1980-х гг. Речь идет о ключевой роли, которую играла астрономия в календаристике у аборигенов Сибири, начиная с эпохи палеолита и неолита. Последовательно расширяя и углубляя эту тематику, В. Е. Ларичев проделал за прошедшие десятилетия значительную работу по изучению сибирских народов, событий их истории и малоисследованных периодов жизни.

Статья посвящена объяснению символики одного из сюжетов Шишкинских писаниц – охота героя на мифологического лося с оленьими рогами. Автор представляет новую, с использованием данных астрономии, версию интерпретации наскальных изображений, найденных вдоль побережья реки Лены в 1929 г. А. П. Окладниковым, копии которых были недавно выполнены Л.С. Мельниковой и В. С. Николаевым. Точность и тщательность копирования стали для В.Е.Ларичева важным фактором при анализе нескольких композиций. Особое внимание автора привлекли в них образы всадника и лося с детально выполненными рогами оленя. А. П. Окладников неоднократно отмечал, что на Шишкинских скалах изображения лося превалируют над рисунками других зверей, и объяснил это особым статусом животного у курыкан [Окладников, 1959, с. 94]. В.Е. Ларичев привнес новый смысл в разгадку семантики лося, обратив внимание на знания курыкан в области астрономии и систем счисления времени. Такой подход к проблеме весьма неожиданный, и он, тем не менее, заслуживает внимания, поскольку автор статьи детально обосновывает свои соображения, используя методы астроархеологии.

Рассматривая историю вопроса, заметим, что В.Е.Ларичев ранее проводил анализ изображения палеолитического оленя с необычными рогами из «Большого зала» Ляско (Франция) и, используя числовой анализ, пришел к выводу, что отростки рогов оленя заключали в себе определенный код. Расшифровка рисунка сложных по своей структуре рогов оленя древнекаменного века показала, что в них заключена информация о календарной системе – ключевых моментах лунного и солнечного годовых циклов [Ларичев (а), 1998, с. 44-51]. В последующих статьях В.Е.Ларичев обосновывал свою гипотезу о том, что творцы древних сибирских культур разработали свои системы счисления времени и могли даже, возможно, предсказывать даты солнечных и лунных затмений. Такое знание могло быть получено лишь после длительного наблюдения за этими пугающими человека космическими явлениями, когда, в конечном счете, удалось заметить ритмику их повтора. В.Е.Ларичеву выявляет астрономичность и календарность в деталях, которые остаются вне оценок других интерпретаторов. Так, например, в статье, посвященной изображению танцев в деревне Шишкино, он на основании количества и определенных позиций фигур обнаруживает факты фиксации курыканами астрономических наблюдений [Ларичев (б), 2012, с. 163-173].

В.Е.Ларичев отмечает, что фиксация астрономических сведений с помощью отростков рогов небесного лося, вероятнее всего, связана с тем, что именно это животное было космическим символом у народов Сибири. Это подтверждается открытием рисунков лося с выбитым на его туловище солнцем [Эвенкийские, 1990, с. 24] и ассоциированием образа животного в мифах народов Севера с Большой Медведицей и другими созвездиями [Мазин, 1984, с. 40]. Гибридный облик странного существа, у которого туловище и голова с подшейной серьгой на подбородке – лося, а рога – оленя, наглядно демонстрирует идею – для курыкан это животное было мифическим, космическим по сути своей.

Автор предлагает оригинальную интерпретацию эвенкийской легенды о лосе Хэглэне, обитателе Верхнего мира, похитителе Солнца и виновнике погружения Среднего мира людей во тьму. Мифический герой отправляется охотиться на лося, чтобы вернуть Солнце людям. А.Ф.Анисимов высказал мнение, что легенда эвенков объясняет смену дня и ночи [Анисимов, 1959, с.12-13]. Сходный мотив возвращения героем Солнца, украденного лосем, встречается и в других мифосказаниях народов Севера [Мазин, 1984, с. 9]. В. Е. Ларичев впервые предположил, что на Шишкинских писаницах изображен тот же самый космический лось, о котором повествуют легенды, но анализ количества отростков на рогах мифического зверя позволил ему предложить иное их прочтение. Миф о небесном лосе объясняет, по его мнению, укорочение и удлинение дня после равноденствий и смену сезонов. Возможно, неслучайно мифический герой эвенков отправляется на поимку небесного зверя именно осенью, в период наступления нового дня у северян, когда день начинал укорачиваться [Мазин, 1984, с. 9].

Связь мифического рогатого существа с календаристикой курыкан подтверждает якутская мифология. А.И.Гоголев на основании лингвистических данных убедительно доказал, что мифологический олень связан с древним солнечным рогатым божеством якутов Сах. Он отметил, что слово «сахсагар» – «густо сросшиеся

и нависшие ветки», которые обнаруживаются на рогах оленя, в якутском языке близко по смыслу словам «Солнце» и «Время»: «В быту дореволюционные якуты пользовались выражением «Сахпыт тһһлһһтһ?» «Сколько времени?». В данном случае под словом «сах» они имели в виду солнце» [Гоголев, 2002, с. 15]. Поэтому резонно предположить, что в мифе курыкан рога-ветки небесного лося имеют прямое отношение к пониманию ими времени.

По версии В.Е. Ларичева, небесный лось курыкан олицетворял собой целый годовой цикл и все его сезонные перемены. Курыканский всадник Шишкино держит в руках не оружие убийства, лук, а аркан. Автор справедливо считает, что, вероятнее всего, целью небесного героя было не убийство животного, а разворот его движения в обратную сторону, что знаменует завершение осеннего астрономического сезона, начало зимнего, а с ним и смещение солнца на север ко дню весеннего равноденствия. Композицию «Всадник с флагом и лось с рогами оленя,двигающийся в обратную сторону» можно считать документом, который подтверждает интерпретацию В.Е.Ларичева. Флаг (а не аркан) в руках этого всадника определяет его социальный статус. Он определенно – эпический герой курыкан или предок клана их вождей. Действительно, у азиатских народов знамя обычно окружалось религиозным ореолом и пользовалось культовым почитанием. Так, Д.Банзаров писал, что у бурят и монголов царское знамя являлось материальным воплощением «сулдэ», божественного могущества [Банзаров, 1997, с. 46]. Курыканский всадник был, вероятно, не просто вождем, а объектом поклонения. На скалах святилищ, скорее всего, изображались наиболее почитаемые предки и родоначальники народа.

Подводя итог комментариям к статье В.Е.Ларичева, хотелось бы особо подчеркнуть методическую новизну интерпретаций. Ему удалось раскрыть неизвестные сведения о культуре одного из сибирских народов. Ключевой тезис, который подробно и последовательно обоснован в статье, заключается в том, что курыкане владели обширными познаниями астрономии. Результаты исследования В.Е.Ларичева подтверждают важный факт: верования современных якутов, бурят и эвенков взаимосвязаны с мифологией курыкан. Следовало бы только шире использовать информацию по этнографии, фольклору и лингвистике якутов, эвенков и северных бурят, в таком случае вполне возможно привлечь больше аргументов к уникальной интерпретации сюжета Шишкинских писаниц. Тот факт, что при решении проблемы В.Е.Ларичев использует междисциплинарный подход, указывает на актуальность и комплексность проблематики и широкую масштабность исследования, так как узость одной дисциплины в этом случае не позволяет объективно решить задачи. Выявление роли астрономии в культурах народов Сибири является актуальной тематикой для молодых ученых.

Отметим, что подход В.Е.Ларичева к наскальным рисункам как к иллюстративному научному открытию курыкан оправдан тем, что для них эти сюжеты являлись отражением их мировоззрения. Фольклор и мифы каждого народа, воспроизведение их мотивов на скалах не могут быть истолкованы однозначно как фантастические вымыслы, которые придумывались нашими предками для неких развлечений. Современный человек часто воспринимает мифы как чистейший вымысел. Однако для людей прошлого мифическое пространство было реальностью, в которой он жил. А.Ф.Лосев и Ф.Х.Кессиди неоднократно подчеркивали, что миф является, в первую очередь, своего рода иносказательным объяснением загадочных для народа явлений природы, результатом долговременного поиска причин природных изменений [Лосев, 1994; Кессиди, 1972].

ЛИТЕРАТУРА

- Анисимов А.Ф.** Космологические представления народов Севера. – М.-Л.: Изд-во Академии наук СССР, 1959. – 108 с.
- Банзаров Д.** Черная вера или шаманство у монголов // Банзаров Д. Собрание сочинений. – Улан-Удэ: Изд-во БНЦ СО РАН, 1997. – С. 29-56.
- Гоголев А.И.** Истоки мифологии и традиционный календарь якутов. – Якутск: Изд-во Якутского ун-та, 2002. – 104 с.
- Кессиди Ф.Х.** От мифа к логосу. – М.: Мысль, 1972. – 315 с.
- Ларичев В.Е.** Космографическая композиция из «Большого зала» Ляско // Гуманитарные науки в Сибири. – № 3, 1998. – С. 44-51.
- Ларичев В.Е.** Курыканские пляски (календарно-астрономические аспекты правздничных торжеств кочевников средневековья юга Восточной Сибири в дни летнего солнцестояния) // Дальневосточно-Сибирские древности. – Новосибирск: Изд-во Института археологии и этнографии СО РАН, 2012. – С. 163-173.
- Лосев А.Ф.** Миф. Число. Сущность. – М.: Мысль, 1994. – 919 с.
- Мазин А.И.** Традиционные верования и обряды эвенков-орочонов (конец XIX – начало XX в). – Новосибирск: Наука, 1984. – 200 с.
- Окладников А.П., Запорожская В.Д.** Ленские писаницы. Наскальные рисунки у деревни Шишкино. – М.-Л.: Изд-во Академии наук СССР, 1959. – 201 с.
- Эвенкийские** героические сказания. – Новосибирск: Наука, 1990. – 392 с.

Лыгденова В.В., научный сотрудник Института археологии и этнографии СО РАН, кандидат философских наук.

Зольников И.Д.,
Постнов А.В.
Лямина В.А.,
Новикова О.И.,
Глушкова Н.В.,
Славинский В.С.,
Чупина Д.А.,
Кузьмин Я.В.,
Бондаренко А.,
Деменьтев В.Н.,
Селяницкая Н.А.

ГИС-моделирование условий обитания древнего человека юга Западной Сибири для определения наиболее благоприятных зон его обитания¹

***Abstract.** Examples of application of GIS technology to analyze the conditions habitats of ancient man is revised. At the core of GIS modeling paleoenvironment conditions are ideas about Geological and morphological paleolandscapes framework that in Western Siberia remained almost unchanged the last 50,000 years. We used the DTM SRTM and geological map of scale 1:200 000. In the Altai Mountains was studied paleolith. Here the basic parameters, who controlled the localization of sites were chosen: the presence of flat areas suitable for settlement, the area illumination, the proximity to sources of materials for stone tools, water supply (density of rivers and their confluence nodes). Verification of the the known Paleolithic monuments showed the adequacy of the GIS model. Vengerovsky area on South of the West Siberian Plain is selected for the study of spatial distribution of monuments of the Bronze and Iron Ages. It was found that the monuments of different eras and cultures associated to different paleolandscapes (lakes, streamside, thecrests, the valleys and interfluves, etc.). This is due to the change of paleoclimatic conditions and changes in lifestyles of ancient people (gathering, fishing and hunting, sedentary pastoralism, nomadic herding). Thus, on examples of different geomorphological conditions and chronological range shown the effectiveness of GIS-modeling to optimize the search for new sites, as well as to determine the adaptation of an ancient human interactions with the paleolandscapes.*

Геоинформационные технологии достаточно активно применяются в археологических исследованиях для решения задач, связанных с анализом закономерностей пространственного распределения археологических памятников и их локализации в зависимости от благоприятности условий палеосреды. Один из пионерных ГИС-проектов для Сибири включал создание банка геоданных (с доступом через интернет) по датированным радиоуглеродным методом памятникам палеолита Сибири и местонахождениям крупных млекопитающих (в частности - мамонтов). ГИС-анализ ареалов проживания палеолитического человека и крупных млекопитающих на разные временные срезы [Кузьмин и др., 2004; Орлова и др., 2000] в частности позволил обосновать археологическими и палеонтологическими данными вывод об отсутствии сартанского (23-11 тысяч лет назад) озера-моря на юге Западно-Сибирской равнины. На протяжении ряда лет сотрудниками ИАЭТ СО РАН и ГИС-центра ИГМ СО РАН проводятся совместные исследования, нацеленные на выяснение закономерностей приуроченности археологических памятников Западной Сибири к различным палеогеографическим обстановкам палеосреды для определенных хроностратиграфических интервалов.

В предлагаемой работе приводятся результаты геоинформационного моделирования обстановок палеосреды обитания древнего человека изучаемой территории по двум районам. В одном из них изучался палеолитический этап, а в другом – этап палеометалла (бронзы и железа). В качестве района для ГИС-моделирования обстановок жизнеобитания человека времени палеометалла был выбран Венгеровский район Новосибирской области [Зольников и др., 2008], а для ГИС-моделирования обстановок благоприятных для поселения палеолитического человека была выбрана территория, охватывающая зону проектирования газопровода на Горном Алтае [Славинский и др., 2011]. Два обсуждаемых примера кардинально различаются по геолого-

¹ Исследования выполнены при поддержке гранта РФФИ 12-06-00045-а

геоморфологическим условиям (что обуславливает и палеогеографическую специфику), а также по историко-возрастному охвату моделируемых палеообстановок жизнеобитания древнего человека. Это позволяет наглядно проиллюстрировать разные аспекты разрабатываемого подхода. Поскольку для древнего человека характерен адаптационный механизм взаимодействия с окружающей средой, то рельеф местности в качестве структурного каркаса ландшафтов и палеоландшафтов, является ключевым при оценке благоприятности обстановок жизнеобитания. Для геоморфологического анализа показательных районов использовалась цифровая модель рельефа SRTM открытого доступа через интернет с исходным размером ячейки для данных широт 90X60 м, которая была приведена к пространственному разрешению 60x60 м.

Прежде всего, рассмотрим палеолитический пример, для которого геоинформационная модель создавалась на территорию республики Горный Алтай вдоль предполагаемой трассы газопровода. Длина изученного участка трассы проектируемого газопровода составляет 677 км; ширина буферной зоны, в которой проводились исследования – 50 км; общая площадь исследованной зоны – 32690 км². В основном изучались долины рек Песчаная, Урсул, Катунь, Чуя и их притоков. В качестве основных факторов, контролирующих распределение мест стоянок древнего человека, с учетом местной специфики, были приняты следующие: 1) наличие ровных пологих площадок, пригодных для поселения; 2) тепло-светобеспеченность территории (освещенность пологих склонов и площадок); 3) близость к источникам каменного сырья для орудий; 4) водообеспеченность территории (наличие рек, а также узлов их слияния).

Поскольку стоянки древнего человека обычно расположены на пологих, хорошо освещенных склонах, то для определения таких благоприятных мест на ЦМР (цифровой модели рельефа) были выделены области с углами падения склонов не более 10°. Дополнительно, были построены две сетки освещенности площадок и склонов, отдельно для зимнего и летнего сезонов. Схемы освещенности строились следующим образом: для каждого сезона (на июль и январь месяц соответственно) было построено по 5 теневых моделей рельефа по ЦМР. Теневые модели строились на момент восхода солнца, полдень, момент захода солнца и 2 промежуточные модели между восходом и полуднем и между полуднем и закатом. Время восхода, полудня и заката солнца определялось по астрономическому калькулятору (<http://suncalc.net>). Далее по калькулятору положения солнца (<http://planetcalc.com>) рассчитывался азимут и угол положения солнца над исследуемой территорией. Поскольку территория сильно вытянута с севера на юг, то брались усредненные показатели азимута и угла положения солнца над горизонтом. Далее все пять теневых моделей складывались (процедура сложения сеток в ARCGIS), и получалась интегральная схема освещенности территории за сутки. После этого, на основе статистических характеристик распределения освещенности были выделены наиболее освещенные участки для июля и января. В итоге получилась схема освещенности территории, которая отражает участки, освещенные только в зимний период времени, участки, освещенные только в летний период времени и участки, освещенные и летом и зимой.

В качестве источников сырья на данной территории выступают кислые эффузивы, а также контактные зоны интрузивных массивов с терригенными породами, в которых часто образуются роговики, а также вторично кремнистые песчаники, алевролиты и другие породы, пригодные для изготовления каменных орудий [Кулик, Постнов, 2010]. Для создания в ГИС-проекте слоя источников сырья на данную территорию вокруг интрузивов были построены буферные зоны размером от 500 до 50 м в зависимости от размера интрузии. Далее эти буферные зоны посредством оверлейных операций с векторной геологической картой проверялись на состав вмещающих пород. В качестве выделов каменного сырья оставались только те участки буферных зон, в которые попадали терригенные толщи. Кроме буферных зон, моделирующих контактовые ореолы вокруг интрузивных массивов, слой источников сырья также включает в себя контуры выходов на земную поверхность кислых эффузивных пород, которые сами по себе пригодны для изготовления каменных орудий труда. Еще одним фактором, контролирующим размещение стоянок древнего человека, является наличие воды. Поэтому были построены буферные зоны с радиусом 400 м вокруг рек, а также буферные зоны с тем же радиусом вокруг точек слияния рек. Вышеперечисленные слои легли в основу геоинформационной модели, с использованием которой на исследуемой территории был проведен анализ распределения мест стоянок древнего человека в зависимости от контролирующих факторов.

На данной территории располагаются 35 известных стоянок палеолитического человека, для которых был создан отдельный слой археологических памятников. 27 из них открыты осенью 2011 г [Славинский и др., 2011]; остальные были известны ранее [Постнов и др., 2007]. На плоские

места попадают 100%, т.е. все 35 стоянок. Из них на места, освещенные только в январе, попадают 13 стоянок (37%), освещенные только в июле – 14 стоянок (40%) и на места, освещенные и в январе и в июле – 4 стоянки (11,4%). Суммарно на освещенные участки попадает 30 стоянок (85,7%). При этом следует учесть, что площадь освещенных пологих участков составляет не более 16,8% от всей обскитываемой территории вследствие ее гористости. Оставшиеся 5 стоянок (14,3%) расположены не просто в затененных участках, а непосредственно на днищах относительно узких долин второстепенного порядка, которые освещены хуже, чем днища широких долин и выбирались древними людьми для проживания по каким-то другим причинам (возможно как убежища). Близость к источникам каменного сырья анализировалась из того расчета, что благоприятным является место, которое расположено на расстоянии не более, чем час-полтора пешего хода до ближайшего выхода пород, потенциально пригодных для изготовления орудий. Поэтому в качестве зоны близости был выбран буфер с шириной 5 км. В буферные зоны вокруг источников каменного сырья попало 33 стоянки (94,3%). При этом буферные зоны занимают 67,3% от общей площади анализируемой территории. Оставшиеся две стоянки расположены менее, чем в 10 км от ближайшего источника сырья.

Для анализа распределения мест стоянок древнего человека в зависимости от расстояния до рек была построена сетка ближайшего расстояния до рек и всем местам стоянок присвоено значение этой сетки. Большинство стоянок (27 стоянок – 77,1%) расположены на расстоянии не более 400 м до ближайшей реки. При этом площадь, занимаемая буферными зонами с радиусом 400 м вокруг рек, составляет 32,3% от всей территории. Аналогичный анализ был проделан для точек слияния рек, которые традиционно считаются индикаторным признаком для обнаружения палеолитических памятников. Однако, обнаружилось, что стоянок, находящихся на расстоянии до 400 м от ближайшего слияния рек всего 6 (17,1%), а буферные зоны вокруг узлов слияния рек радиусом 400 м занимает 3,9% от всей площади территории. Затем, было выяснено, что радиус буферной зоны вокруг узлов слияния, в которую попадают 31 стоянка (88,5%) составляет 2 км. При построении таких буферных зон ими покрывается 48,7% изученной территории.

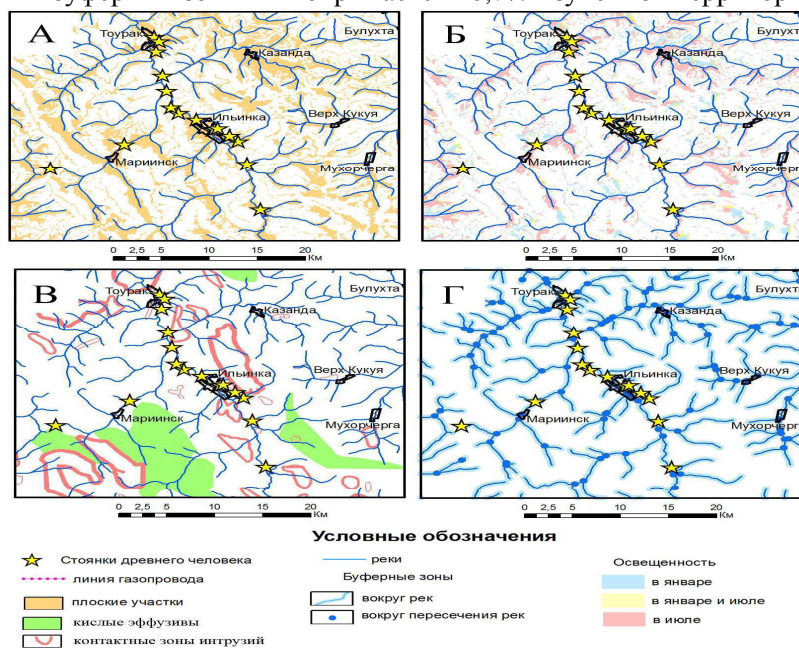


Рис. 1. Факторы, обуславливающие территориальную локализацию палеолитических стоянок в районе поселка Ильинка: А – пологие поверхности; Б – освещенность; В – источники сырья; Г – водообеспеченность. Условные обозначения: 1 – стоянки древнего человека, 2 – линия газопровода, 3 – плоские участки, 4 – кислые эффузивы, 5 – контактовые зоны интрузий, 6 – реки, 7 – буферные зоны вокруг рек, 8 – буферные зоны вокруг пересечений рек, 9 – освещенность в январе, 10 – освещенность в январе и июле, 11 – освещенность в июле.

Таким образом, проведенные исследования показали, что освещенные пологие склоны и площадки являются ключевым условием для оценки перспективности поиска археологических памятников палеолита в данном регионе. Важным признаком является также близость к источникам каменного сырья. Здесь же особо отметим, что хотя более 90% палеолитических стоянок попало в пятикилометровую зону вокруг выходов каменного сырья, тем не менее, при этом около 80% памятников попало в трехкилометровую зону. Следовательно, при необходимости, область поиска по данному параметру может быть сужена. Достаточно существенным значением для локализации палеолитической стоянки обладает фактор близости реки (менее 400 м) и

расстояние до ближайшего узла слияния рек (менее 2 км). В целом, по результатам проведенного анализа можно сделать вывод, что параметры, выбранные для создания геоинформационной модели действительно значимы для территориальной оптимизации поисков палеолитических стоянок. По сопряженному анализу вышеперечисленных параметров выделился ряд районов. В их числе район поселка Ильинка, данные по которому приводятся на рисунке 1.

На рис. 1 проиллюстрированы ключевые факторы, контролирующие локализацию археологических памятников палеолита на территории, окружающей поселок Ильинка. На этой территории найдено 19 памятников. Оверлейные операции по всем слоям, характеризующим вышеперечисленные параметры, позволили получить слой с пересечением оптимальных условий палеосреды для человека каменного века. В итоговом пересечении, составляющем 6,8% по площади от всего района, оказалось 14 памятников из 19, что составляет 73,7%. Если рассмотреть ситуацию по каждому из параметров отдельно, то на пологие склоны и площадки попало 19 памятников (100%), хотя они занимают 30% площади района Ильинки, а на освещенные участки, которые занимают 19,4% района, попало 16 памятников (84,2%). Что касается буферных зон по источникам сырья, то в них попали все 19 памятников, что и не удивительно, т.к. в районе поселка Ильинка эти зоны составляют 90% от анализируемой территории. В буферные зоны вокруг рек, занимающие 35,8% площади района, попало 27 памятников (77,1%), а в буферные зоны вокруг пересечений рек, занимающие 48,9% района, попало 17 памятников (89,5%). Это достаточно хороший результат, т.к. всего 5 из 19 памятников (24,3%) оказались выходящими по одному или двум параметрам недалеко за пределы статистически рассчитанной оптимальности. В таблице 1 приводятся данные, характеризующие условия палеосреды для данного района в радиусе 5 км, т.е. в пределах краткого перехода от места стоянки.

Таблица 1. Условия палеосреды в радиусе 5 км от памятников в районе п. Ильинка.

	Длина рек (км)	Количество пересечений рек	Площадь плоских мест	Площадь источников сырья	Площадь освещенной территории
Среднее	30,4	7,5	28,66%	15,7%	27,8%
Разброс значений	23,7-53,2	3-12	14-38,5%	5,5-31%	11,3-57,8%

Очевидно, что выявленные закономерности адекватны для региональных условий и применимы только для горных местностей. Полученные результаты позволяют создать для Горного Алтая геоинформационную модель на основе серии плотностных сеток, учитывающих распределение параметров, характеризующих условия палеосреды в определенном радиусе (в нашем случае принят радиус 5 км) вокруг каждой точки изучаемой территории. Сопряженный анализ таких сеток позволяет выделять ареалы с разной степенью оптимальности условий палеосреды для палеолитического человека, что обеспечивает возможность оптимизации территориальных поисков археологических памятников каменного века.

Второй пример, иллюстрирующий геоинформационный анализ палеообстановок обитания древнего человека, приводится для Венгеровского района Новосибирской области, где имеется достаточно представительная группа археологических памятников [Молодин, Новиков, 1998]. В работе использовалась база геоданных, включающая в себя 26 памятников бронзового века (из них 13 – жилые комплексы и 13 – могильники) и 35 памятников железного века (из них 18 – жилые комплексы и 17 – могильники). Данная территория относится к Барабинской слабоволнистой гривно-озерной равнине [Земцов и др., 1988]. Высота грив 3-12 м, ширина 0,3-1 км, длина 2-8 км. Озерные поверхности представлены днищами современных и древних котловин, размеры которых в поперечнике варьируют от нескольких сот метров до нескольких километров, в некоторых случаях существенно превышая 10 км. Относительно узкие долины рек Омь, Тартас, Кама разделены на низкую пойму и первую надпойменную террасу, которая на современном этапе геологического развития территории представляет собой высокую пойму, раз в несколько лет затапливаемую аномально многоводными паводками. Площадки первой надпойменной террасы и поймы участками трудноразличимы в пределах озеровидных расширений аллювиальных долин, сформировавшихся, по всей вероятности, при спуске озер за счет их подрезания местными реками. За вычетом аллювиального, гривного и озерного рельефа остается слаборасчлененный водораздел с пологими изометричными холмами и суффозионно-просадочными западинами, сложенный лессовидными суглинками и супесями. Нередко гривы продолжают на днищах котловин, а древние озерные поверхности в свою очередь «проникают» в межгривные пространства. Гривы, а нередко и удлинённые озера, локализуясь в межгривных понижениях, а также некоторые отрезки речных русел вытянуты с юго-запада на северо-восток по доминирующей розе ветров.

Такие особенности рельефа земной поверхности Барабинской равнины диктуют иные приемы

обработки цифровых моделей рельефа, нежели в горах Алтая, поскольку очевидно, что хорошо освещенные пологие склоны и площадки охватывают фактически всю площадь исследуемой территории, а следовательно не могут служить индикаторами благоприятных обстановок проживания древнего человека. Источники каменного сырья для производства орудий труда здесь также не являются индикаторными признаками, т.к. Барабинская равнина покрыта сплошным чехлом рыхлых кайнозойских отложений, в результате чего выходы докайнозойских пород на дневную поверхность отсутствуют. В качестве местного каменного сырья для орудий труда здесь используются кварцитовые гальки из речных отложений [Молодин, Новиков, 1998]. Вместе с тем, и на равнине рельеф являлся каркасом палеоландшафтов, а, следовательно, представляется ключевым для геоинформационного моделирования палеообстановки среды обитания. На рисунке 2 отражены наиболее характерные геоморфологические особенности изученной территории.

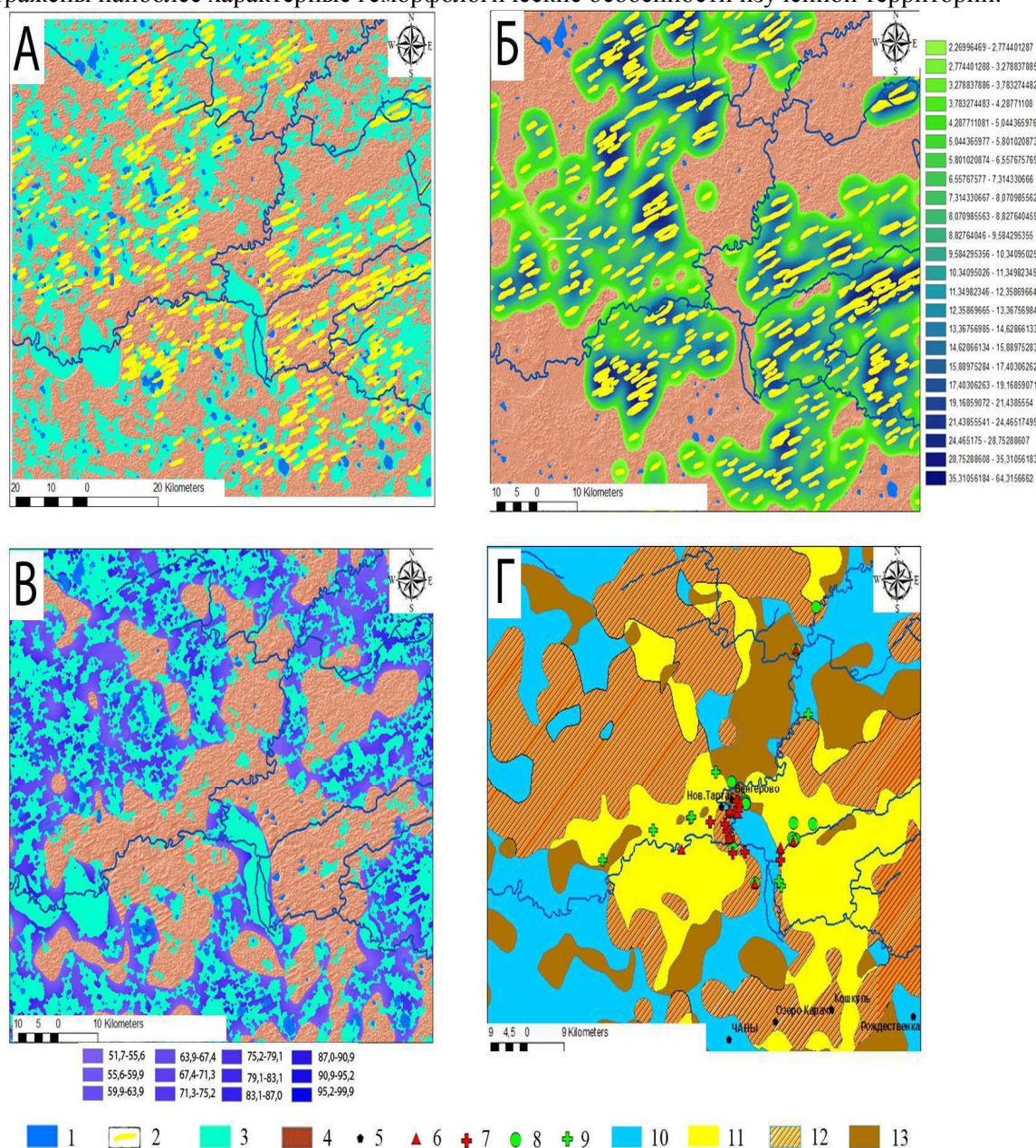


Рис. 2. Формы и типы рельефа, образующие каркас рельефа Венгеровского района Барабинской равнины: А – гряды и днища древних озерных котловин; Б – плотность распространения гряд; В – плотность распространения озерных поверхностей; Г – типы рельефа, выделенные по площадному соотношению доминирующих форм земной поверхности. Условные обозначения: 1 – гидросеть; 2 – гряды; 3 – днища древних озерных котловин; 4 – слаборасчлененные водораздельные поверхности с мелкими пологими изометричными холмами и суффозионно-просадочными западинами; 5 – населенные пункты; 6 – жилые комплексы эпохи бронзы; 7 – могильники эпохи бронзы; 8 – жилые комплексы эпохи железа; 9 – могильники эпохи железа; типы рельефа: 10 – озерный; 11 – грядный; 12 – грядно-озерный; 13 – слабо расчлененный.

Формы земной поверхности, показанные на рисунке 2А, были выделены на основе

классификации морфометрических параметров, построенных в скользящем окне по цифровой модели SRTM. Для выделения днищ древних озерных котловин использовались: диапазон углов склонов и диапазон высот, а для выделения грив: среднее аспектов (экспозиции склонов) и диапазон высот. Построение плотностных сеток для ключевых форм земной поверхности (гривы – рисунок 2Б и днища палеозер – рисунок 2В) позволило автоматически оконтурить области распространения типов рельефа, различающихся по их сочетаниям (рисунок 2Г). Реализованное на основе ГИС-технологий геоморфологическое районирование играет важную роль для реконструкции обстановок обитания древнего человека. Ландшафт современной Барабинской лесостепи представляет интразональная растительность межгривных понижений с болотно-солончаковыми, солонцово-солончаковыми и остепненными лугами, обсыхающими травяными болотами, зарастающими и заболачивающимися озерами, березовыми колками в западинах. Эта территория, чутко реагирует на климатические флуктуации. Изменения количества выпадающих осадков приводят к изменению площади озер. Предельные разливы озер показаны на рисунке 2А. Современные озера составляют незначительную часть от днищ древних озерных котловин. Этапы увлажнения (развитие озер и болот) чередовались с этапами аридизации, когда широкое распространение получали эоловые процессы (дефляция и аккумуляция грив). Здесь же отметим, что как современные, так древние болота фактически никогда не выходят за пределы озерного и гривно-озерного рельефа, а современная лесная растительность, отмеченная на топокартах и фиксируемая на космоснимках среднего пространственного разрешения, фактически не попадает в область озерного рельефа.

Именно солонцово-солончаковые луга составляют основу природных кормовых угодий Барабы, используемых современным человеком для скотоводства. Человек времени палеометалла имел достаточно ограниченный набор видов хозяйственной деятельности: охота, рыбалка, собирательство, оседлое скотоводство с зачатками земледелия, отгонное, полукочевое и кочевое скотоводство. В эпохи увлажнения обширные пространства, которые показаны на рисунке 2Г полями распространения озерного и озерно-гривного рельефа становились непригодными для скотоводства, поскольку котловины заполнялись водой, а межгривные понижения заболачивались. В свою очередь лесная растительность, которая при современных относительно аридных условиях локализуется в межгривных понижениях, в эпохи увлажнения перебиралась на гривы. Таким образом, гривный тип рельефа, наряду с озерным и озерно-гридным становился малопригодным для выпаса скота. С другой стороны при климате, существенно более влажном, чем современный, увеличивались территории, пригодные для собирательства, охоты и рыбалки.

Оверлейные операции позволили проанализировать приуроченность памятников разных эпох к формам и типам рельефа, а также к определенным интервалам морфометрических характеристик. Прежде всего, выяснилось, что к озерному и озерно-гривному рельефу приурочено 14 памятников бронзового века из 26, что составляет 53,8%. Что касается памятников железного века, то их к озерному и озерно-гривному типам рельефа приурочено 10 из 35, что составляет 28,6%. Следует тут же подчеркнуть, что это лишь общая приуроченность к районам рельефа, а не к определенным формам или даже их элементам. Однако, даже этот приближенный результат свидетельствует о предпочтительности субэдрально-водораздельного рельефа для выбора мест обитания людьми эпохи железа по сравнению с людьми эпохи бронзы (см. рисунок 2Г). Анализ приуроченности памятников непосредственно к днищам древних озерных котловин показал для бронзового века 17 попаданий (65,4%), а для железного 8 (22,9%). Анализ зависимости расположения памятников от абсолютных высот подтвердил выявленную закономерность. По характеру полимодального распределения на гистограмме статистически обособляются 4 гипсометрических интервала, к которым приурочены памятники: 93 – 98 м; 99 – 104 м; 105 – 110 м; 111 – 116 м. Памятники эпохи бронзы распределены по этим интервалам от нижнего к верхнему соответственно: 11 (42,3%); 9 (34,6%); 5 (19,2%), 1 (3,8%), а памятники железного века: 5 (14,3%) 8 (22,9%) 12 (34,3%) 10 (28,6%). Отчетливо проявлена обратная зависимость от высоты для памятников бронзового века. В свою очередь, памятники железного века наоборот, находятся в прямой зависимости от высоты (за исключением последнего интервала, к которому приурочено меньше памятников, чем к предыдущему). При этом, для жилых комплексов закономерность проявлена еще контрастней. Для бронзового века в соответствующие интервалы попадают снизу вверх: 6; 4; 2; 1 памятников, а для железного: 1; 4; 5; 8. Полученные закономерности могут быть проинтерпретированы в двух аспектах. С одной стороны, человек бронзового века вел преимущественно оседлый образ жизни, осваивая приречные ландшафты (в частности – пойменные), наиболее богатые природными ресурсами. В свою очередь железный век

характеризуется повышением роли полукочевого и кочевого скотоводства, что приводило человека к необходимости осваивать междуречные водораздельные пространства. С другой стороны, нельзя скидывать со счетов палеоклиматический фактор, который также мог влиять на локализацию мест обитания.

Для того, чтобы оценить приуроченность мест обитания к палеоландшафтным обстановкам нами был проведен анализ тех памятников, которые были диагностированы с точностью до культурной принадлежности в контексте палеоклиматических изменений. Вторая половина атлантического периода голоцена (от 6,5 до 5 тысяч лет назад) была временем климатического оптимума, когда доминировали теплые и умеренно влажные условия. В суббореальном периоде голоцена (от 5 до 2,5 тысяч лет назад) происходило постепенное иссушение климата и к его середине уже господствовали теплые и сухие условия. Памятники ранней бронзы (байрыкская, одинцовская культуры) расположены на берегах рек и ныне высохших стариц в пределах озерных расширений. По всей видимости, озерные расширения в то время не заливались паводками, а представляли собой поверхность первой надпойменной террасы, где существовала субэральная обстановка в условиях сухого теплого климата. Такая обстановка палеосреды по-видимому была оптимальной для охотников, рыболовов и собирателей ранней бронзы. Памятники кротовской культуры, относящейся к развитой бронзе, также находятся в озерном расширении реки Тартас вблизи ныне отмерших палеомеандр. Вероятно, природные условия на поверхности озерного расширения способствовали появившемуся у кротовцев отгонному скотоводству. Андроновские памятники развитой бронзы локализованы в пределах Омь-Тартасского озерного расширения около рек. Преобладание у андроновцев среди скота лошадей и овец позволяет говорить о кочевом образе жизни [Молодин, Новиков, 1998]. По всей видимости, количество андроновцев и палеоландшафтные условия позволяли им осуществлять кочевое или полукочевое скотоводство в пределах озерных расширений. Памятники ирменской культуры (оседлое скотоводство), которая относится уже к поздней бронзе, расположены на берегах Оми, Камы и Тартаса в пределах гривного рельефа, из них два в пределах озерного расширения. Таким образом, памятники бронзового века локализованы преимущественно в озерных расширениях и на берегах рек. Это обусловлено с одной стороны доминированием оседлого образа жизни у немногочисленного населения, а с другой стороны наличием богатых кормовых ресурсов (как для человека, так и для скота) на площадках 1 террасы и на поверхности озеровидных приречных расширений.

Памятники железного века обладают совершенно другой спецификой локализации относительно форм и типов рельефа. Начало железного века совпадает с финалом суббореального периода голоцена (от 5 до 2,5 тысяч лет назад), который считается влажным и холодным. Вероятнее всего, на рассматриваемой территории повышенное количество осадков вызвало подтопление площадок 1 надпойменной террасы и поверхности озерных расширений, что вывело эти участки из субэральных условий в режим высокой поймы. Так, из 8 местонахождений саргатской культуры 7 расположены на гривном и слаборасчлененном междуречном рельефе, из них 3 у границы озерных расширений Тартаса, Оми и Камы и 2 непосредственно на берегу рек. Таким образом, с одной стороны удовлетворяется потребность в воде, а с другой памятники располагаются вне зоны потенциального подтопления. Здесь же отметим, что памятники саргатской культуры локализованы на территориях современного распространения лесной растительности. Именно спецификой палеоландшафтов, на наш взгляд может быть объяснена ориентация саргатцев на разведение не мелкого, а крупного рогатого скота. Если проанализировать рисунок 2Г, то становится очевидным, что при избыточном увлажнении из пастбищных ресурсов выпадает не только приречное пространство, но и значительные территории озерного и озерно-гривного рельефа, занятого озерами, болотами и лесом. Всего лишь треть района занято гривным и слаборасчлененным междуречным рельефом, но и здесь пригодность территории определяется близостью постоянных источников воды.

Памятники сперановской и горносталевской культур (охотники и рыболовы) расположены на берегах рек и палеопротоков в пределах гривного или слаборасчлененного междуречного рельефа в районах, где была велика вероятность произрастания лесов. Горносталевцы практиковали также отгонное скотоводство. Таким образом, железный век, ознаменовавшийся в более южных степных районах развитием кочевого скотоводства, в Венгеровском районе Барабинской равнины характеризуется другими видами хозяйственной деятельности, что обусловлено спецификой палеоландшафтов рассматриваемой территории, которая не позволяла в тех палеоклиматических условиях массово содержать курдючных овец. При увлажнении первая надпойменная терраса и поверхность озерных расширений подтапливалась, что делало малоприспособленным для активной

эксплуатации аллювиальные ландшафты. Что касается междуречных пространств, то там происходило заполнение водой древних озерных котловин, заболачивание межгрядных понижений и зарастание гряд лесом. Поэтому, площадь потенциальных пастбищных угодий на водоразделах тоже сокращалась. Что же касается древнетюркских кочевников, которые появились на территории около 1,5 тысяч лет назад, то памятники этого времени встречаются как на приречных, так и на междуречных территориях. Вероятно, это связано с неоднократной сменой палеоландшафтных условий в ходе короткопериодических циклов увлажнения и аридизации на протяжении средневековья.

Проведенные исследования показали, что палеоландшафтное районирование на основе обработки цифровых моделей рельефа SRTM, даже в пределах таких слабо расчлененных территорий как Барабинская полого-волнистая равнина, обеспечивает возможность выявления закономерностей пространственной локализации памятников палеометалла в зависимости от хозяйства и условий палеосреды. Поскольку около половины археологических памятников палеометалла Венгеровского района не имеет возрастной и культурной привязки, полученные закономерности можно использовать для предварительной диагностики памятников по условиям палеоландшафтной локализации. По существу удалось создать геоинформационную модель, определяющую палеоландшафтные предпочтения людей эпохи палеометалла в зависимости от палеоклиматических обстановок увлажнения и аридизации. Геоморфологическое районирование по SRTM в пределах Горного Алтая позволило создать геоинформационную модель предпочтений обстановок палеосреды для палеолитического человека. Полученные результаты могут быть использованы для оптимизации поисков археологических памятников определенного возраста и культурно-хозяйственной специализации, а также для предварительного прогноза археологических исследований при изыскательских работах.

ЛИТЕРАТУРА

- Земцов А.А., Мизеров Б.В., Николаев В.А.** и др. Рельеф Западно-Сибирской равнины. - Новосибирск: Наука. Сиб.отд-ние, 1988. 192 с.
- Зольников И.Д., Кузьмин Я.В., Чемякина М.А., Новикова О.И.** Геоархеологические наблюдения в центральной части Барабинской равнины летом 2008 года // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий: Материалы Годовой сессии Ин-та археологии и этнографии СО РАН 2008 г. - Новосибирск: Изд-во Ин-та археологии и этнографии СО РАН, 2008. - Т. XIV. - С. 160-163
- Кузьмин Я.В., Зольников И.Д., Орлова Л.А., Зенин В.Н.** Палеогеография Западно-Сибирской равнины во время максимума сарганского оледенения (в связи с находками мамонтов и палеолитических памятников) // Доклады Академии Наук" (РАН), 2004, т. 398, № 4. С. 542-544.
- Кулик Н.А., Постнов А.В.** Геология, петрография и минералогия в археологических исследованиях // Методы наук о Земле и Человеке в археологических исследованиях. Новосибирск: изд-во НГУ, 2010. С.39-99.
- Молодин В.И., Новиков А.В.** Археологические памятники Венгеровского района Новосибирской области. Новосибирск, 1998. 139 с.
- Постнов А.В., Зольников И.Д., Гуськов С.А., Челваков Л.М.** К вопросу о стратиграфическом положении палеолитических памятников вдоль чуйского тракта в долинах Чуи и Катунь // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий. Том XIII. Материалы годовой сессии ИАЭ СО РАН. 2007, с.149-155.
- Славинский В.С., Постнов А.В., Марковский Г.И., Басова Н.В., Зольников И.Д., Рыбин Е.П.** Результаты разведки памятников палеолита в Алтайском крае и Республики Алтай в 2011 году // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий. – Новосибирск: Изд-во ИАЭТ СО РАН, 2011. – Т. XVII – С. 469–472.
- Орлова Л.А., Кузьмин Я.В., Зольников И.Д.** Пространственно-временные аспекты истории популяции мамонта (*mammuthus primigenius blum.*) и древний человек в Сибири // Археология, этнография и антропология Евразии. 2000. Т. 3. № 3. С. 31.

Кулемзин В.М.

Сибиреведение и полевая этнография

В данной статье Сибиреведение рассматривается как нераздельная часть этнографической науки вообще. Выводы как общего, так и частного характера всегда опирались на результаты, полученные учеными-сибироведами. В статье проводится мысль о необходимости совершенствования методики полевых исследований, включающей знание других научных дисциплин: археологии, фольклористики, культурологи. Основной акцент сделан на продуктивности полевых материалов, которые составляют ценную часть кандидатских и докторских диссертаций, монографий, статей современных исследователей.

Ключевые слова: сибиреведение, полевая этнография, политическая антропология, педагогическая антропология, аграрная антропология, юридическая антропология, обычаи.

Уже целый ряд десятилетий в Сибири идет процесс интенсивного индустриального освоения. Вследствие значительного притока населения из других регионов страны и зарубежья меняется в целом облик Сибири: экологическая и демографическая ситуация, формы собственности и хозяйственной деятельности, а так же жизнь и мировоззрение аборигенов (если под мировоззрением понимать суммарное представление о мироздании).

Как известно, контакты с пришлым населением (в основном русским) начались уже давно и насчитывают не одно столетие. Тогда же появились первые сведения о коренных обитателях Сибири. Для науки представляет интерес сам поступательно-линейный процесс исследования: первые сообщения – это то, что видно совершенно невооруженным глазом. За ними следуют более детальные описания различных сторон быта, деятельности, мировоззрения, которые требуют некоторого проникновения, сопоставления по простейшей схеме «мы – не мы». Совершенствование методики этнографической науки, стационарные экспедиции в Советский период истории, несмотря на известные сложности, дали не только обильный и прекрасный материал, но и ключ к пониманию этого материала. Особняком стоит период, связанный с деятельностью КОМСОДА.

Еще большие успехи сделаны в последние десятилетия, когда в сферу науки включились носители самих традиционных культур. Являясь одинаково знакомыми с той и другой культурами, владея тем и другим языком, они своими работами убедительно доказали, что полевая этнография, в противоположность распространяемому кабинетными учеными мнению, отнюдь не клонится к закату. На ежегодных сессиях проводившихся Институтом этнографии (г. Москва), посвященных полевым исследованиям, этнографы из всех регионов Сибири делали сообщения, из которых следовало, что теории этноса есть не самостоятельное и независимое направление, а тесно связанное с полевой этнографией. Именно полевая этнография вкупе с междисциплинарными исследованиями, использование достижений зарубежной этнологии сулят сегодня ещё большие перспективы.

В данной статье Я акцентирую внимание именно на тех направлениях и вопросах, которые невозможно решить и исследовать, не собрав необходимый полевой материал. Кроме того, я хотел бы обратить внимание этнографов и культурологов на то большое значение, которое имеют полевые материалы для решения актуальных задач, поставленных и решенных в диссертационных исследованиях.

Наибольший интерес, не только теоретический, но и практический вызывает включённость аборигенов в современное производство. Всегда, без исключения это связано с деформацией традиционной культуры как жизнеобеспечивающей саморегулирующей системы.

Интересно сопоставление: как себя идентифицируют коренной житель и приезжий (он может быть и старожилом) русский человек.

«Мы» – говорит русский человек, а далее перечисляет: семья, школа (класс), институт (или производство), общежитие, город, район, область, место службы в армии, страна. Особняком стоит слово земляки».

«Мы» – говорит коренной житель, а затем перечисляет: род (т.е. группа родственных семей), мой народ, семья, охотничья, рыболовецкая или пастушеская община. Слово «земляк» здесь отсутствует. Но вместо него указывается место нахождения семейного, родового или племенного культового места с изображением духа-божества. Государственные или административные

границы в таком сообщении всегда отсутствуют. У некоторых народов указывается предок рода (например, у обских угров – это медведь, бобёр, лось, чайка). Интересно, что здесь семья не на первом месте, а так же характерно отсутствие знаний родословной (в основном это обязанность особых лиц – исполнителей эпоса или шаманов, которые следили за соблюдением правил экзогамии. Включенность аборигена в современный производственный коллектив основанием для идентификации не является. Производственные операции, с точки зрения аборигена, обязательно должны выполняться в целях упрочения и улучшения коллективных отношений, но ни в коем случае не в ущерб им. Однако это не совместимо с другими представлениями аборигенов о совсем необязательном выполнении требований производственной дисциплины. В своей практике Я сталкивался со множеством случаев, когда аборигены, не поставив в известность администрацию, на неделю – две покидали производство и уходили в тайгу на промысел. На это же время из школы-интерната они забирали своих детей. Во всех случаях аргументация была одна: «Это чтобы ни сам, ни дети не забыли дело наших предков».

В современном обществе часто бывает так, что, что физический труд или интеллектуальная производственная деятельность является социальной необходимостью (особенно когда она является единственным источником дохода). Но со временем это становится психологической потребностью. В результате человек, достигший пенсионного возраста, остается на работе. Однако если человек занимался традиционным видом деятельности (охота, рыбная ловля, земледелие), то по достижению пенсионного возраста он обязательно возвращается к нему без всякого сожаления о производстве.

Поскольку для такого человека современное производство не является психологической потребностью, то у аборигенов не формируется чувство лидерства, конкуренции, ревности. Аборигены, которые твердо решили порвать с традиционным образом жизни, вступают на сложный путь, предполагающий новое обучение и новую социализацию. Это значит выработать и управлять новыми эмоциями, сокрыв старые. Разрыв контакта между головой и телом часто вызывает чувство растерянности.

В абсолютном большинстве случаев аборигены избегают ответственной работы, не говоря уже о создании собственного дела; они склонны выполнять не требующую квалификации работу, находиться в полной зависимости от работодателя. Такое положение дел оказывается менее сопряженным с внутренним дискомфортом.

Однако здесь имеется другая сторона. В обществе, где какая-то часть чувствует себя угнетенной, распространяются так называемые милитаристские настроения (от лат. *mille* – тысяча). Согласно им, через каждую тысячу-две тысячи лет Всемирный потоп прекращает несправедливую жизнь, но она возобновляется в традиционном виде с обилием природных ресурсов и без самолетов, бульдозеров, нефтяных вышек. Такие настроения впервые были отмечены Т.Н. Потаниным и позднее проанализированы американским этнографом П.Уорсли [Уорсли, 1963] и отечественным Н.В. Цеханской [Цеханская, 1991]. Здесь я хотел бы обратить внимание на то, что столь пессимистический взгляд на будущее отражается в современном фольклоре, где металл ассоциируется с тракторами, трубами нефтепровода, ракетами, поменял положительный знак на отрицательный. Ожидания специалистов в том, что с образованием обществ в защиту национальных культур, эти настроения исчезнут, не оправдались.

Аборигенам приходится многое преодолевать, но этого требует зависимость от привозных товаров, переход к городскому образу жизни, эволюция статуса женщины, доступ к специальному и высшему образованию. Кстати, коренные жители не согласны с термином «приниженное» положение женщины. Они полагают, что мужской мир сакрален по отношению к женскому, а женский – по отношению к мужскому. По этой причине ограничения имеются для тех и других, и эти ограничения имеют мировоззренческий характер.

В разговорный язык сейчас входят такие понятия, как доход, расход, процент, заём, аренда, долг, банк, ссуды. Все это исключает те принципы, на которых зижделись прежние отношения. Главным из них был принцип взаимного доверия.

О том, насколько не совпадает менталитет ханта-оленевода и японки, свидетельствует следующий случай. В 2004 году на международном съезде оленеводов в г. Осло познакомились хант и японка. Оформив брак, они улетели в Японию, где прожили три года. Объяснений и возражений своей жены хант и слушать не хотел. Не выдержав, он вернулся в свое стойбище. Когда этнограф из Томска Н.В. Лукина (устное сообщение 2004 г.) спросила этого ханта о причине бегства из Японии, он ответил: «Япония – страна диких нравов. Там родители своим детям дают деньги в долг через банк и требуют проценты».

Однако имеются и другие примеры. Полевые исследования А.Л. Томуран [Томуран, 2004] показали, что забайкальским старообрядцам удалось безболезненно войти в рыночные отношения. Старообрядцы совместили свои традиционные и современные экономические принципы, чему способствовала незыблемость Устава старообрядческих общин.

В этнографической науке пока нет ответа на вопрос, почему одни обычаи обнаруживают удивительную устойчивость, а другие – изменчивость и хрупкость. Представление о сакральном, содержащем в себе ритуал, поясняет многое, но не всё. Раскрытию тайны этого вопроса способствуют исследования о взаимодействии традиций и инноваций. В настоящее время новое столь быстро входит в традиционное, а традиционное столь же быстро отвергает или принимает входящее, что механизм этого сложного процесса фиксируется одним поколением. Как показывают исследования, здесь играет роль не только экономический фактор, но и многие другие: экологический, мировоззренческий, а также государственная политика. Как правило, все это находится в определенной степени совмещения с религиозными воззрениями. Последние, в свою очередь, так же составляют напластование ранних фетишистских, анимистических, политеистических, монотеистических (православно-христианских) верований. Их венчает научно-материалистическое понимание мироздания.

Иногда контакт нового и старого в бытовой, но не в сакральной сфере диктуется чисто утилитарными соображениями. К примеру, коптильню для рыбы делают из топливного бака трактора, выпрямленное алюминиевое кольцо хула-хуп используют как копье с наконечником из бивня мамонта, растопленный костный жир лося служит в качестве незамерзающей смазки замка охотничьего ружья. Такие примеры приведены в специальной статье [Кулемзин, 2009: №1, 62-69]. Сложнее обстоит дело там, где старое и новое находятся под контролем ещё действующих сверхъестественных сил, а синкретический характер культур делает задачу, казалось бы, совершенно неразрешимой.

В традиционных культурах нет отдельных видов воспитания: физического, трудового, эстетического, экологического. Все вплетено в единую ткань бытия, все находится под контролем старших, а так же местных родовых духов и бога. В тесном переплетении сосуществуют древнейшие представления о равенстве человека и им же сделанных вещей, живых и неживых объектов природы; православно-христианские воззрения о более высоком статусе человека чем вещи; научно-материалистические знания о человеке как о находящемся на самом верху иерархической лестницы. Только беседы с информаторами, порой случайно брошенные ими фразы дают ясное представление о давно ушедших отношениях человек-вещь, человек-природа, человек-общество. Этими фразами, например, могут быть такие: «Чем к гадалке идти, лучше топор подвесить и качнуть: куда отклонится, там и ищи потерянное». «Вещь свой век отживает, потому что сила, в неё вложенная, кончается. Такие негодные вещи надо нести и хоронить, как человека». «Нам вещи нужны, а значит, мы вещам так же нужны, поэтому в стародавние времена люди приносили жертву на том месте, где делали стрелы и копья». [Об отношениях человек- вещь см: Мифология..., 2000: с.45-52].

В другом случае информант-старообрядец во время воспоминаний о прежней жизни сообщает важную деталь о способе изготовления долблёной лодки. Растущий тополь – осокорь расклинивали в течение нескольких лет и только после этого валили, чтобы продолжить работу. Это единственное в науке сообщении об изготовлении большегрузных долбленных лодок [Кержак, 2011, с. 339].

Иногда информанты на предельно простом уровне объясняют сложные процессы, допустим, механизм разрушения культуры. «Вот берестяная коробка, в ней орехи. Бересту заготавливает мужчина, он раздевает березу, если она не шумит листвою. Коробку всегда делает женщина, она специалист по коже, ткани, бересте. Все делается по правилам – от начала и до конца; первую коробку – под присмотром бабушки. Последняя операция – это выскабливание узора, который означает солнечные лучи – это к сытой жизни, здоровью. Коробка готова. Есть правила изготовления, наследования, дарения, пользования, захоронения. Есть сновидения: если черная ягода в коробке – к смерти, красная – к поранению. Железная магазинная чашка не требует никаких правил и если снится в ней черная ягода – сны не сбываются. Ты её не делал и правил никаких нет – её можно выбросить, она неживая».

Кстати сказать, человек не несет наказание ни за нарушение правил обращения с покупными изделиями, ни за нарушение правил охоты, если эти нарушения вызваны государственным планом, интересами бизнеса и т.д. Очень много сведений информантов о паритетных отношениях

человек-насекомое, животные-птицы в мировоззрении народов Сибири Севера приводит В.Н. Адаев [Адаев, 2007, с. 181-191].

Иногда информанты расходятся во мнении, противоречат друг другу, но удивительно то, что о каком бы периоде ни зашла речь, они всегда, как один, скажут: «Раньше было лучше». Единодушны они и в том, что закон как регулятор отношений продуктивно «работает» только в том случае, если он является продолжением традиций. Если юридически оформленный закон является нововведением и противоречит традиционным нормам, он обязательно будет нарушаем.

«Скоро нас всех пересадят» – сказали мне кеты-охотники в присутствии главы администрации пос. Суламай Байkitского района Эвенкии В.П. Тыганова после того, как им объявили о запрете охоты на соболя.

Мне многократно приходилось беседовать с лицами, отбывшими сроки заключения. В основном они и далее идут « по кривой дорожке», редко кто возвращается к традиционным видам деятельности. Характерно, что если культура обязательно имеет этническое лицо, то контркультура (термин, как известно, введен Фридрихом Ницше) не имеет этнического лица.

Преступный мир в своих действиях поразительно одинаков и в действиях русских, алтайцев, шорцев, хантов нет абсолютно ничего этнического. Этой темы я коснулся в специальной статье [Кулемзин, 2000].

Мною была предпринята попытка, опираясь на методику французского социолога Винсента де Гольжака [Гольжак, 2003: 17], выяснить вопрос о том, что может человек требовать от общества и что общество может гарантировать человеку. Результат бесед с информантами оказался неожиданно простым. Представители традиционных культур настолько интегрированы в неё, что составляют неотъемлемую часть общества. Здесь индивидуума как бы нет. По этой причине они (информанты) не имеют ничего такого, чего бы им хотелось, а общество им отказало. Они довольны тем, что имеют и даже тем, что жен выбрали себе не они сами, а их родители в интересах рода.

Совершенно очевидно, что приведенные здесь примеры невозможно квалифицировать и оценить в рамках отдельной науки этнографии.. Синкретический характер традиционных культур делает это в принципе невозможным.

Ю.П. Холюшкин и ряд других исследователей, которых Ю.П. Холюшкин детально анализирует, полагает, что перспективным может быть выход только один: выделить из этнологии отдельные дисциплины как самостоятельные. Именно в них могут быть сформулированы законы и аксиомы. Такими дисциплинами в частности, могут быть политическая антропология, педагогическая антропология, аграрная антропология, юридическая антропология и т.д. [Холюшкин, 2011: 117-129].

Абсолютное большинство диссертационных исследований строится на основе полевых материалов. Заметно, что Сибиреведение как наука развивается в основном благодаря диссертационным исследованиям. Их очень много и тематика самая разнообразная: практически все сферы материально и духовной культуры. Я возьму лишь несколько показательных примеров, внесших существенный вклад в Сибиреведение.

Так, например, в докторской диссертации Е.Е. Дутчак [Дутчак, 2008: 54с.], построенной исключительно на полевых материалах идет речь об актуальном по сей день «учении о побеге» старообрядческих общин в целях спасения «древнего благочестия». На огромном материале раскрыта роль старообрядческой книги, отсутствие полной изолированности «бегунов» от модернизационных процессов. Новым словом является утверждение, что «классическое» Беловодье как вольная и плодородная Земля является местом, где скрываются не «от мира», а от власти.

В диссертационном исследовании Н.С. Грибановой [Грибанова, 2011: 23] на конкретных данных доказано, что в сакральной сфере, к которой относится домотканное полотенце, в культуре сельского населения Алтая формировалось уже на месте в Сибири. Отсюда множество вариантов. Однако основа представлений была завезена из европейской части России. Автор объясняет сокращение обрядовых функций результатом распространения фабричных изделий. Н.С. Грибанова сумела увидеть за небольшими деталями большой мир сакральной сферы; показать генезис и пути эволюции этого мира.

В диссертационном исследовании М.И.Федоровой [Федорова, 2010: 24 с] анализируются функции сверхъестественных существ, управляющих вселенной и жизнедеятельностью человека в мировоззрении селькупов. Новым является объяснение обрядовой практики шаманов, а так же семейной обрядовости (рождение, смерть). Автор дает схематический анализ мировоззренческих

терминов селькупского языка, который существенно облегчает адекватно представить эволюцию мировоззрения селькупов.

В диссертации Н.В.Золотаревой [Золотарева, 2012: 22с.] предпринята удачная попытка получить от информантов ответ на вопрос: что скрывается за простыми, порой грубо изготовленными антропоморфными изображениями из камня, дерева, ткани, металла? Полученный результат автор смогла перевести на современный научный язык. Н.В.Золотарева показала, какие сложные представления о жизни, смерти, рождении, Верхнем и Нижнем мирах раскрывают незамысловатые искусственные, а иногда и естественного происхождения антропоморфные изображения. Автор сумела передать неуничтожимость этих фантастических представлений в народном сознании, их уживчивость с научными знаниями.

Совершенно новый аспект культурной традиции рассмотрен в диссертации Н.К.Карасёвой: [Карасёва, 2010 24с.]. Речь идет о кумандинцах и их традиционном хозяйственном укладе, основанном на охоте, собирательстве, рыболовстве.

По причине почти полного обрусения в течение XX в. традиционные формы хозяйства исчезли. Однако трудности, вызванные известными причинами в начале 1990-х гг., заставили реанимировать давно забытые виды деятельности, инвентарь, звероловные ловушки и т.д. Здесь показано значение тайников семейной памяти.

В ряде исследований доказано, что отдельная часть этноса, попав в иную моноэтническую или полиэтническую среду и потеряв все или почти все черты материальной культуры, сохраняется как этнос исключительно благодаря самосознанию -.

Ещё двадцать с лишним лет назад подобный вывод считался бы предельно крамольным из-за идеалистического подхода.

Новая постановка вопросов, новое решение старых задач, естественно обогащает этнографическую и вообще гуманитарную науку. Не исключено, что когда-нибудь будет создана столь общая теория этнографической науки, с позиций которой будут объяснены частности. Это было бы доказательством того, что полевая этнография сыграла положительную роль в большой биографии науки о человеке.

ЛИТЕРАТУРА

- Адаев В.Н. Традиционная экологическая культура хантов и ненцев. – Тюмень: Вектор Бук, 2007: 237 с.
- Грибанова Н.С. Полотенце в культуре русского сельского Алтая в конце XIX – начале XX в. Автореф. дисс. на соиск. уч. ст. канд. ист. наук. – Томск, 2011: 23 с.
- Дутчак Е.Е. Старообрядческие таежные монастыри: условия сохранения и воспроизводства социо-культурных традиций (вторая половина XIX – начале XXI в. Автореф. дисс. на соиск. уч. ст. докт. ист. наук. – Томск, 2011: 53 с.
- Золотарева Н.В. Явление антропоморфизации в традиционной культуре обских угров (18-20 вв.). Автореф. дисс. на соиск. уч. ст. канд. ист. наук. – Томск, 2012: 22 с.
- Гольжак Вицент. История в наследство. – М., 2003: 230 с.
- Карасева Н.К. Присваивающее хозяйство традиционной культуры кумандинцев. Автореф. дисс. на соиск. уч. ст. канд. ист. наук. – Томск, 2010: 24 с.
- Кержак А. Последний коч // Предания старых трактов – Томск: Сибирика, 2011: с. 338-343.
- Коровушкин Д.
- Кулемзин В.М. Сибирь – кладездя этнографа //Человек в истории. – Томск, 2000.
- Кулемзин В.М. Заготовка для тѣщи // Наука из первых рук – Новосибирск: ИНФОЛИО, 2009, №1: с. 62-69.
- Мифология хантов. – Helsinki- Ижевска-Budapest-Томск, 2000: 304 с.
- Уорсли П. Когда вострубит труба – М., 1963.
- Федорова М.И. Представление о душе и бессмертии в духовной культуре селькупов. Автореф. дисс. на соиск. уч. ст. канд. ист. наук. – СПб, 2010: 24 с.
- Холюшкин Ю.П. Системная классификация понятия «Основания антропологии» // Информационные технологии в гуманитарных исследованиях. Вып. 16. – Новосибирск, 2011: с.117-129.
- Цеканская Н.В. Нативистские синкретические религии Северной Америки // Локальные и синкретические культы. – М., 1991: с. 49-63.

КОММЕНТАРИЙ

К статье В.М. Кулемзина «Сибиреведение и полевая этнография»

Полевые этнографические исследования всегда были и остаются одним из основных механизмов поиска и добычи фактического материала для составления объективной картины этнической культуры. Особая потребность в такого рода исследованиях ощущается сегодня, когда на глазах меняется облик Сибири, её экология и экономика, демография и социальная стратиграфия. Все эти изменения затрагивают в том числе и аборигенные этносы Сибири. Перед современной полевой этнографией ставятся сложнейшие задачи – выработать новые методы и способы изучения культуры народов Сибири в условиях интенсивного промышленного и индустриального освоения края.

Автор статьи – В.М. Кулемзин – известнейший сибиревед, профессиональный этнограф, и при этом опытный полевик, за плечами которого два десятка полевых сезонов.

Хорошо представляя сложившуюся в этнографической науке ситуацию к началу XXI столетия, автор статьи считает, что о кризисе в этнографической науке говорить не приходится. Данный этап её развития – это начало нового витка. Автор кратко анализирует историю становления полевой этнографии, историю научных подходов, приёмов и методик изучения народов Сибири.

Как отмечает сам В.М. Кулемзин, в данной статье основное внимание было им сосредоточено на новых направлениях полевых исследований, без которых невозможно изучать современные этнокультурные процессы, а также решать любые другие актуальные задачи, выдвигаемые новым поколением этнологов в своих диссертационных исследованиях.

На современном этапе этнографических/этнологических исследований для этнографа-полевого важный теоретический и практический интерес представляет изучение механизма включённости аборигенов в современные виды деятельности. Вынужденная переориентация интересов в системе современных экономических и правовых отношений приводит к деформации традиционной культуры как жизнеобеспечивающей системы. Автор статьи настаивает на необходимости изучения этого процесса. Используя свой богатый опыт, автор на целом ряде примеров показывает особенности проявления поведения потомков аборигенов внутри новой для них сферы производственных отношений.

Важным на современном этапе (впрочем, как и всегда) является изучение вопроса взаимодействия традиций и инноваций. Современные темпы жизни, как утверждает автор, одинаково быстро заставляют, как принимать в культуру новое, перерабатывая его под свой менталитет, так и отвергать предлагаемые жизнью инновации. Однако если раньше для принятия и укоренения инновации и превращения её в традицию требовалось время трёх-четырёх поколений, то сегодня достаточно памяти одного поколения, чтобы усвоить или отбросить новое. Немаловажную роль в этом процессе играют экологический, экономический, мировоззренческо-религиозный и государственно-политический факторы. Без полевой этнографии понять и решить многие современные научно-исследовательские проблемы невозможно (в ряде случаев, их невозможно даже выявить и осознать как проблему, если по какой-либо причине отказаться от полевых исследований).

Автор удачно приводит примеры взаимодействия нового и традиционного. На современном этапе память человека является одним из основных источников по изучению сложного спектра традиционной культуры. Без полевой этнографии современному ученому трудно воссоздавать картину прошлого аборигенных народов Сибири, еще сложнее прогнозировать будущее.

На основе полевых материалов строится абсолютное большинство диссертационных исследований. Автор утверждает, что сибиреведение как наука на современном этапе также развивается в основном благодаря активному использованию именно методов полевой работы. Вывод небесспорный. Однако бум диссертаций по этнологии, написанных на материале сибирских народов, отмеченный за последнее десятилетие, подводит автора именно к такому выводу.

Тучков А.Г., кандидат исторических наук, доцент кафедры
отечественной истории и культурологи ТГПУ
15.04.2012.

Холюшкин Ю.П. Концепция Г.Чайлда о трех коренных поворотах в истории

Статья посвящена анализу категорий «производственных революций» и построению на основе концепций Г.Чайлда, Брейдвуда и Дамонда классификационного фрагмента

Ключевые слова: Присваивающее хозяйство, производящее хозяйство, орудийная революция, неолитическая (аграрная) революция, городская революция.

Категория «производственная революция» появилась в отечественной и зарубежной науке довольно давно, хотя встречается не так часто, как следовало бы. Интерес к таким рубежам усилился в 40–50-е годы XX века вместе с исследованием неолитической революции (Г. Чайлд и другие) и осознанием того, что мир вступил в НТР [Гринин, 2006].

Различие этапов присваивающего и производящего хозяйства – принципиально важный критерий периодизации. Он основан, на намеченной Морганом и развитой Энгельсом, принципе членения первобытного общества.

Присваивающее хозяйство — один из принципов производства, хозяйство с преобладающей экономической ролью охоты, собирательства и рыболовства, что соответствует самой древней стадии хозяйственно-культурной истории человечества. «Присваивающей» эта стадия называется довольно условно, так как деятельность охотников, собирателей и рыболовов не ограничивается простым присвоением, а включает ряд довольно сложных моментов как в организации труда, так и в переработке продукции, требующей разнообразных технических навыков, а также с созданием предметов искусства. С изобретением более совершенных орудий труда на смену присваивающему хозяйству приходит производящее хозяйство [Гринин, 2006].

Производящее хозяйство — это такая форма ведения хозяйственной деятельности, при которой происходит переход от присвоения готовых продуктов природы к их производству с помощью человеческой деятельности. Первобытные люди сознательно начинают приручать животных, заниматься окультуриванием злаков, осваивать простейшие ремесла. Появление таких навыков и умений позволило людям систематически добиваться устойчивых результатов и меньше зависеть от окружающей природной среды [Гринин, 2006].

Принципы производства				
Присваивающее хозяйство		Производящее хозяйство		
Орудийная революция		Аграрная революция		Городская революция
Типологическая археология				
Бродячая охота	Полубродячая охота	Раннее земледельческо-скотоводческое хозяйство	Позднее земледельческо-скотоводческое хозяйство	Ремесленное (городское) хозяйство

Рис. 1. Классификационный фрагмент «Производственные революции».

Эта пара понятий образует диадную группу.

В 1936 году Гордон Чайлд выступил с концепцией трех коренных поворотов в истории в первобытной истории, а именно орудийной, неолитической и городской революций [Childe, 1951]. Каждая из этих революций уникальна и имеет совершенно неповторимые черты. Но в то же время во всех есть и много принципиальных сходств, которые позволяют создать модель производственной революции как явления [Гринин, 2006]:

каждая революция происходит не одновременно и быстро, а представляет собой длительный по времени процесс, начинающийся в одних местах и распространяющийся на все большее количество обществ и территорий;

каждая — результат длительного накопления количественных и качественных изменений, крупный перерыв постепенности, ведущий к нарастающему усложнению общественного разделения труда и интеграции человечества. Приведем последовательно основные понятия триады.

1. Орудийная революция. Основными показателями в периодизации первобытной истории, являются успехи «в производстве средств к жизни» (указ. произв. Ф. Энгельса, 1955, с. 20), — это период преимущественно присвоения готовых продуктов природы. Переход от низшей ступени, «детства человеческого рода» (там же), к средней связан с применением огня и палеолитических орудий.

2. Аграрная революция (сельскохозяйственная, или неолитическая) начинается включением земледелия и скотоводства в постоянный производственный процесс и заканчивается созданием устойчивой (в масштабах крупного государства) системы ирригационного земледелия с переходом к использованию плуга и тягловых животных. Последнее также дало основу для завершения в более поздний период аграрной революции в неполивных районах, где без плуга и тягловых животных много земли не обработать. Концепция неолитической революции остается действенной и нынешние дни.

3. Городская революция символизирует собой процесс отделения ремесленного производства от земледелия. Вторая, урбанистическая революция, непосредственно связана с возникновением городов, цивилизации и государственных институтов как таковых. Ключевыми причинами урбанистической революции являются развитие технологий и наличие растущих излишков пищи. Чайлд предлагал 10 критериев, отличающих городскую цивилизацию от предшествовавших обществ:

увеличение размеров и плотности поселений, превращение их в города;

социальная стратификация (классовое расслоение), предусматривающая существование привилегированного правящего класса, использующего государственную машину для сохранения своего превосходства над угнетёнными;

механизмы извлечения «социальных излишков» для поддержания государственного аппарата, включая налоги или дань;

политическая организация, построенная по территориальному, а не только родственному признаку, — государство; концентрация власти;

общественное разделение труда, позволяющее выделение категорий ремесленников и специалистов в непроизводственных сферах;

интенсивная экономика, предусматривающая внешнюю торговлю;

письменность или её заменители;

возникновение зачатков точных наук;

развитое изобразительное искусство;

монументальные общественные постройки.

В пентадной группе представлены понятия, отражающие попытки некоторых археологов и этнографов детализировать приведенную в диадных и триадных группах периодизацию первобытной истории, посредством выделения в ней таких этапов, как бродячая охота полубродячая охота [Braidwood, Howe, 1962], раннее и позднее земледельческое хозяйство [Diamond, 1971].

Все эти упрощенные схемы могут быть усложнены при построении новых классификационных фрагментов.

ЛИТЕРАТУРА

Гринин Л. Е. Производительные силы и исторический процесс. Изд. 3-е. М.: КомКнига, 2006.

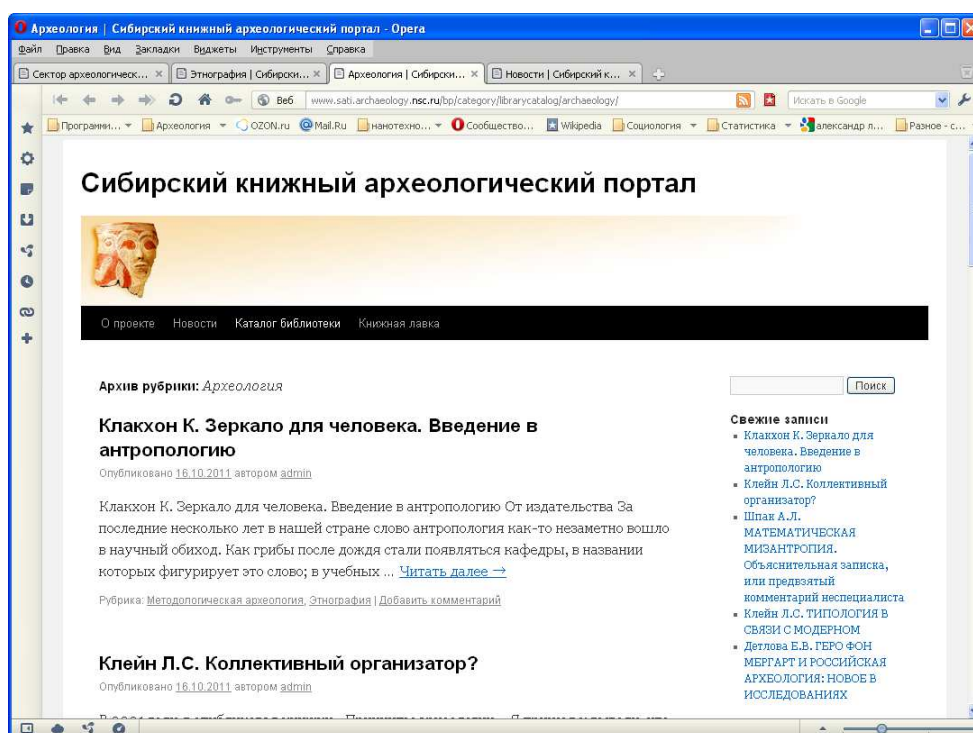
Childe V.G. Man makes himself. – N.Y., 1951.

Braidwood R.J., Howe B. Southwestern Asia beyond the lands of Mediterranean littoral. – N.Y., 1962.

Diamond A.S. Primitive law: past and present. – L., 1971.

**Холюшкин Ю.П., Новые ресурсы на сайте «Сибирика»
Рыбинцев Р.В.**

В ходе работ 2012 года завершены работы над «Сибирским книжным археологическим порталом». С точки зрения пользователя, портал является тематическим Интернет-ресурсом, обеспечивающим возможность поиска и просмотра информации в рамках заданной предметной области (археологии, этнография). Портал реализован на CMS (системе управления содержимым сайта) WordPress с открытым исходным кодом (используемой по лицензии GNU GPL), написанной на PHP и использующей в качестве базы данных MySQL. Сфера применения — от блогов до достаточно сложных новостных ресурсов и интернет-магазинов. Встроенная система «тем» и «плагинов» вместе с удачной архитектурой позволяет конструировать практически любые проекты.



За основу дизайна портала была взята тема Twenty Ten, которая была адаптирована под потребности проекта «Сибирский книжный археологический портал». Проект рассчитан на заполнение контента группой людей, которые могут добавлять в существующую структуру новые записи в соответствии со своими правами.

Поддерживается разнообразный контент: текстовые файлы, PDF, документы MS Word, большое количество форматов изображений, видео и аудио файлы.

Права по изменению структуры могут быть делегированы нескольким пользователям. Также имеется возможность модерации новых записей, в настоящий момент эта функция не используется. Имеется поддержка канала Atom семейства RSS.

RSS — семейство XML-форматов, предназначенных для описания лент новостей, анонсов статей, изменений в блогах и т. п. Информация из различных источников, представленная в формате RSS, может быть собрана, обработана и представлена пользователю в удобном для него виде специальными программами-агрегаторами.

Протокол публикации Atom (также AtomPub, от англ. Atom Publishing Protocol) основан на HTTP и позволяет создавать, изменять и удалять ресурсы, собранные в коллекции на веб-сайте (примером коллекции может служить блог). Содержимое коллекций описывается в формате Atom,

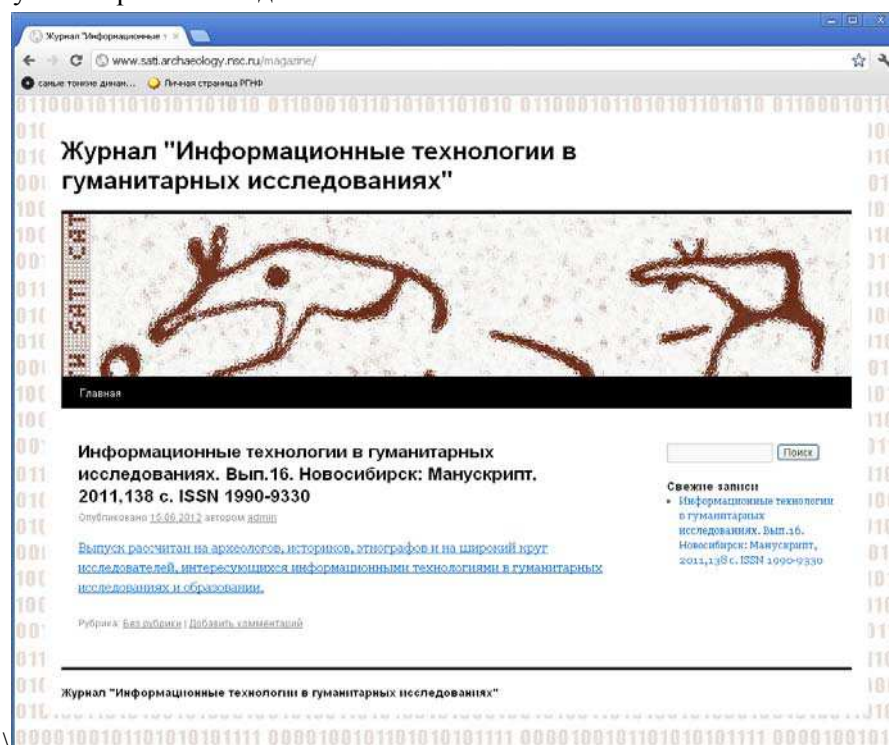
а для управления им используются стандартные методы HTTP. Предусмотрено введение системы регистрации пользователей Сибирским книжным археологическим порталом.

В ходе реализации проекта в заключительном 2012 году был так же создан сайт журнала «Информационные технологии в гуманитарных исследованиях ISSN 1990-9330» <http://www.sati.archaeology.nsc.ru/magazine/>

Появилась возможность создавать статические страницы, например «о сайте», «обратная связь», а также мощный WYSIWYG текстовый редактор;

Предусматривается разместить на сайтах: онлайн-редактор фотографий, поддерживающий функции «обрезать», «отразить», «повернуть» и «масштабировать». Не Photoshop, конечно, но вещь порою очень полезная.

В WordPress поддерживается автоматическое обновление как самого движка, так и установленных плагинов. Один клик мышкой, и движок самостоятельно скачает доступные обновления. Поиск и установка как плагинов, так и шаблонов также производится через панель управления. На этой же основе создан сайт периодического издания «Информационные технологии в гуманитарных исследованиях».



Структурно портал состоит из нескольких разделов: новости; каталог; книжная лавка.

Блок новостей содержит следующие рубрики:

1. Новинки: в настоящее время на портале размещены сообщения о вышедшей из печати археологической и этнографической литературе, начиная с 2010 года по 2012 год.

2. Разделы «конкурсы», «научные программы» и «персоналии» находятся в стадии заполнения.

3. Каталог библиотеки включает разделы: археология, естественные науки, этнография, антропология. Описание предметной области основывается на системной классификации археологической науки, предложенной Ю.П. Холюшкиным и Е.Д. Гражданниковым и развиваемой в настоящее время. Эта иерархическая системная классификация археологии включает более 500 рубрик. В настоящее время разрабатывается системная классификация этнографической науки.

Предусматривается реализация книжной продукции, как в безвозмездном виде, так и путем продажи изданий в разделе книжная лавка. Реализован поиск статей по всему portalу, так и по различным разделам. В настоящее время идет заполнение портала.