

¹ Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия

² Институт экономики и организации промышленного производства СО РАН
пр. Акад. Лаврентьева, 17, Новосибирск, 630090, Россия

³ Государственная публичная научно-техническая библиотека СО РАН
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия

E-mail: kostin@ieie.nsc.ru

WEB-СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА АРХЕОЛОГИЧЕСКИХ ДАННЫХ*

В соответствии с планом работ по созданию автоматизированной обучающей Web-системы для обработки археологической информации, проработаны и зафиксированы основные принципы и требования к архитектуре, определен перечень инструментальных средств. Программу можно условно разделить на 3 основных структурных компонента: база данных, блок запуска вычислительных методов и интерфейс пользователя. Для хранения данных выбрана бесплатная реляционная СУБД MySQL, удовлетворяющая требованиям по скорости доступа при ожидаемых объемах хранимой информации и достаточно простая в разработке и сопровождении. В качестве основного источника методов используется R-язык программирования для статистической обработки данных и работы с графикой, разрабатываемый интернет-сообществом под лицензией GNU GPL. Интерфейс пользователя реализован на языке PHP с применением технологии AJAX, позволяющей выводить на экран новое содержимое без повторной загрузки всей Web-страницы. Содержательная информация структурирована, как в файловой системе современных компьютеров. Общедоступные данные хранятся в корневом разделе, а также в личных архивах, открытых владельцами или администратором системы по поручению владельца. Реализован рабочий вариант пробной версии Web-системы, выставленный на сайте <http://ko.ieie.nsc.ru/archaeology/ru>. На примере решения нескольких задач проведено сравнение оригинальной системы Discovery, которая будет включена в систему, с наиболее широко используемыми методами Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services. Проведена разработка системы Visual Discovery для визуального извлечения интерпретируемой информации из данных в терминах онтологии.

Ключевые слова: археология, информационные системы, интеллектуальный анализ данных, статистические методы, стратегии решения задач, Data Mining.

Археология исследует историю человечества путем изучения вещественных остатков древних поселений. Наряду с единичными, уникальными предметами, она собирает и массовый материал, характеризующий образ жизни и технологии древних производств. Накапливаемые массивы создают почву для применения методов математической статистики в археологии. Ис-

пользование этих методов восходит к временам, предшествовавшим появлению ЭВМ и персональных компьютеров, когда начиная с 20-х гг. XX в. в России и на Западе стали применяться методы математической статистики в археологических исследованиях. С появлением компьютерных технологий и возникновением течения «новой археологии» стали активно внедряться методы

* Исследование выполнено при финансовой поддержке РГНФ (проект № 12-01-12026 «Автоматизированная обучающая Web-система для обработки археологической информации методами интеллектуального анализа данных (Data Mining)»).

математической статистики в археологию. Это работы Л. Бинфорда [Binford, 1972], Д. Кларка [Clarke, 1968], ряд сборников статей, а также учебных пособий, посвященных применению количественных методов в археологии [Doran, Hodson, 1975; Orton, 1982; 2000; Fletcher, Lock, 2005; Baxter, 1994; 2003; Drennan, 1996; Shennan, 1997].

В середине 70-х гг. началось применение методов математической статистики в СССР. Стали издаваться сборники статей по использованию статистико-комбинаторных методов в археологии. Отрядным фактом в 80-е гг. XX столетия было появление учебного пособия по математической статистике [Федоров-Давыдов, 1987], которое до сих пор является единственным отечественным изданием в данном направлении. Здесь ощущается явное несоответствие потребностям развития отечественной археологической науки. В этом отношении намного более продвинутое состояние наблюдается в исторической науке благодаря созданной профессором Л. И. Бородкиным ассоциации «История и компьютер», которая регулярно проводит конференции по квантитативной истории и издает сборники работ «Круг идей» и др. В МГУ издано учебное пособие [Белова и др., 1999].

В конце 80-х гг. на Западе был разработан оф-лайн статистический пакет для археологов *BASP*. В нашей стране подобных разработок для археологов не проводилось. Антропологи и некоторые археологи используют статистические пакеты (*Statistica*, *Statgraphic*, *SPSS* и др.). Массовому использованию статистических методов препятствует 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 статистической подготовки.

В структурах большинства исторических и археологических НИИ и вузов отсутствуют специализированные лаборатории информатики (за исключением немногочисленных лабораторий в МГУ, АГУ, ИАЭТ СО РАН с ограниченным штатом сотрудников), призванные обеспечить исследователю квалифицированный статистический анализ наблюдений, который к тому же не позволяет в полной мере дать основы системного подхода с грамотной формулировкой статисти-

ческих гипотез научным сотрудникам, аспирантам и докторантам, проводящим свои исследования в этих НИИ и вузах. А наметившаяся тенденция к поголовному сокращению подготовленных специалистов в области информатики (как непрофильных в археологии) лишь ухудшает и без того низкую статистическую культуру археологов. Кроме того, отсутствует отраслевая нормативная база (отраслевые археологические стандарты), регламентирующая статистический анализ как завершающий этап кропотливой работы многих специалистов, которая в большинстве случаев выполняется самоучками, людьми, не имеющими профессиональной подготовки в этой области.

Web-интерфейс системы предполагается реализовать с помощью одной из имеющихся бесплатных систем управления содержанием сайтов (CMS – content management system).

Для реализации дополнительных функций серверной части портала предполагается использовать язык PHP, выбор которого связан с тем, что на этом языке написаны почти все CMS. Для реализации дополнительных функций интерактивного поведения интерфейсных элементов клиентской части можно применить язык JavaScript и библиотеки jQuery и jQuery UI, которые позволяют разрабатывать динамичный и удобный пользовательский интерфейс, использующий технологию AJAX для реализации динамически изменяющихся Web-страниц без их перезагрузки.

В качестве базы данных выбрана бесплатная реляционная СУБД MySQL, удовлетворяющая требованиям по скорости доступа при ожидаемых объемах хранимой информации и достаточно простая в разработке и сопровождении.

Для предоставления возможностей проведения расчетов методами статистики и интеллектуального анализа данных, а также свободного конструирования стратегий анализа для специалистов, предполагается подключить бесплатный Open Source пакет анализа «R-язык», развиваемый и регулярно обновляемый интернет-сообществом.

Систему можно условно разделить на 3 основных структурных компонента: базу данных, блок запуска вычислительных методов и интерфейс пользователя. Рассмотрим кратко основные принципы и требования к архитектуре каждого из этих компонентов.

База данных обеспечивает хранение содержательной и служебной информации. К содержательной информации относятся данные, которые обрабатывают пользователи. К служебной – данные о паролях и правах отдельных пользователей и групп пользователей, а также системные настройки, влияющие на интерфейс. Содержательная информация может быть персонализированной (полный доступ только у владельца) либо же общей. Для поддержки групповой работы необходимо вводить права доступа для групп пользователей. Права доступа определяют допуск к информации:

R – read, чтение (просмотр средствами интерфейса);

C – copy, копирование (и экспорт);

A – append, добавление (новых данных);

W – write, запись (поверх имеющихся);

E – edit, редактирование (отдельные изменения внутри данных);

D – delete, удаление (архива или фрагментов данных).

Содержательная информация распределена по архивам, структурированным в виде дерева, подобно файловой системе современных операционных систем. Общедоступные данные хранятся в системном архиве, расположенном в корневом разделе, а также в открытых архивах пользователей, которые могут быть открыты владельцем или администратором системы по поручению владельца. За системный архив отвечает администратор системы. Кроме того, для незарегистрированных пользователей создаются персональные временные архивы на одну сессию, сохраняемые в базе данных определенное время после последнего обращения к серверу, и добавляемые к архиву пользователя, если он регистрируется в системе или войдет в нее под своим именем.

Данные для расчетов хранятся в архивах в виде набора таблиц типа объект-свойство с информацией о памятниках и экспедициях, где был получен этот эмпирический материал, а также ссылки на публикации с анализом, если таковые имеются. Не вызывает сомнений, что наиболее полным источником информации является полевой журнал с подробными описаниями всех находок, их локализацией на памятнике и стратиграфией. В этом случае предварительный искусственный отбор материала при формировании выборочной совокупно-

сти не накладывает дополнительных ограничений на данные, что делает анализ максимально объективным.

Блок запуска вычислительных методов предоставляет инструменты анализа данных. В настоящее время существует целый ряд коммерческих и свободно распространяемых пакетов и сред для анализа данных. Многие из них позволяют производить вычисления в пакетном режиме – по предварительно записанной программе на командном языке пакета. Кроме того, в каждом из них есть возможность импортировать данные из разных форматов, в том числе текстовых. Это создает принципиальную возможность реализовывать стратегии решения археологических задач на многих инструментах, каждый из которых наиболее удобен той или иной группе пользователей.

Проектируемая Web-система должна обеспечивать возможность решения небольших задач пользователя на сервере, а больших (по затратам памяти и времени) на стороне клиента. Порог масштаба допустимых задач должен рассчитываться динамически, исходя из текущей загруженности сервера.

Разумным выбором для решения мини-задач на стороне сервера нам представляется R-язык, который разрабатывается интернет-сообществом по лицензии, предусматривающей свободное распространение и использование. Положительным аргументом в его пользу является широта представления методов в его наборе, охватывающих практически все необходимое для целей анализа данных. Отрицательным моментом является явно недостаточное быстродействие современной реализации R-языка, которое вынуждает в случае трудоемких задач или большого числа пользователей, одновременно запускающих счетные задачи, искать обходные архитектурные решения, в частности – поддержку альтернативных пакетов анализа данных для запуска методов на стороне клиента. С другой стороны, такое вынужденное решение делает систему более универсальной и удобной для пользователя, так как увеличивает его свободу в выборе инструмента решения задач.

Для обеспечения совместимости результатов при использовании различных инструментов можно придерживаться соглашения, что перенос стратегии на другой инструмент считается завершенным только после того, как на контрольных примерах

минимального объема будут достигнуты результаты, полностью идентичные эталону.

Однако не все методы из предполагаемых к использованию в системе представлены в R-языке и универсальных пакетах. В частности, такой мощный метод логико-вероятностного предсказания и естественной классификации, как система Discovery [Фирсов, Витяев, 2012], будет встроен в проектируемую систему на правах оригинального метода анализа данных наряду с некоторыми другими авторскими методами – кластерный анализ с поиском оптимального числа кластеров [Жданов, Костин, 2002], обобщенная классификация [Костин, Корнюхин, 2003] и сравнение классификаций [Костин, 2003].

Интерфейс организует рабочее место пользователя, предоставляя ему доступ к данным и методам анализа. Основные требования к интерфейсу вытекают из сверхзадачи – создание пространства общения археологов и математиков в процессе математизации археологии. Среди этих требований следует особо отметить необходимость в виртуальном конструкторском бюро, в котором археологи приходят со своими проблемами, специалисты по анализу данных предлагают свои решения в виде цепочки методов, а математики и программисты – программы для эффективных вычислений. Результатом их совместных усилий будут новые стратегии анализа данных, доведенные до полной формализации. При этом каждая такая созревающая стратегия должна выделяться в отдельный проект, степень готовности которого будет легко отслеживаться и в общем, и в деталях.

Необходимо также соблюдать и общие принципы разработки интерфейса.

Интерфейс не должен создавать технических препятствий для пользователя. Невнятность логики управления и любые лишние действия отвлекают пользователя от решения содержательных задач. Никаких лишних сложностей – максимальная простота.

Интерфейс не должен скрывать информации. Максимальная прозрачность и доступность.

Если есть несколько маршрутов движения к результату, то интерфейс не должен навязывать пользователю какой-либо один из них. Максимальная свобода действий. Инициативой владеет пользователь, а ин-

терфейс обслуживает его, создавая активную среду для разработки и использования стратегий анализа данных.

Более детальные соображения по архитектуре системы вытекают из опыта разработки пробного варианта нашей системы на базе CMS Drupal 6.0. В результате этой предварительной разработки была продемонстрирована возможность в заявленные сроки реализовать всю необходимую функциональность системы. Но были также выявлены и недостатки архитектуры и интерфейса, которые должны быть учтены при реализации рабочей версии системы.

В состав пробного варианта Web-системы входят:

Главная страница <http://ko.ieie.nsc.ru/archaeology/ru> поможет пользователю понять, будет ли для него полезна и актуальна работа с системой. В этом разделе надо дополнительно описать, какова целевая аудитория системы (археологи и специалисты в области статистики и анализа данных), что такое стратегия решения задач и как устроена система (рис. 1).

Анализ данных <http://ko.ieie.nsc.ru/archaeology/ru/analysis> – основной раздел системы, отвечающий за работу с данными и методами. От удобства пользования этим разделом в основном зависит будущая популярность системы у потенциальных пользователей. Любой просчет здесь может стать критическим (рис. 2).

В пробной версии для навигации по данным и методам был использован один и тот же специально разработанный элемент управления, сохраняющий в отведенной ему области окна все последовательно пройденные уровни навигации с возможностью прокрутки.

Хотя на разработку этого элемента управления было потрачено достаточно много времени, в процессе опытной эксплуатации выяснилось, что пользоваться такой навигацией неудобно, поскольку значительная часть видимой области окна применяется нерационально – для хранения второстепенной информации с верхних уровней дерева данных или методов, а для самой важной информации текущего уровня остается совсем немного места. Отсюда вытекают рекомендации для реализации интерфейса в проектируемой системе: освободить для актуальной информации максимально возможное пространство, свора-

Методы статистики и интеллектуального анализа данных в археологии

Археология исследует историю человечества путем изучения вещественных остатков жизнедеятельности древних людей. Наряду с единичными, уникальными предметами, она собирает и массовый материал, характеризующий образ жизни и технологии древних производств. Накапливаемые массивы массового материала создают почву для применения методов математической статистики в археологии.

Использование этих методов восходит к временам, предшествующих появлению ЭВМ и персональных компьютеров, когда, начиная с 20-х гг. XX в. В России и на Западе стали применяться методы математической статистики в археологических исследованиях. С появлением компьютерных технологий и возникновением течения «новой археологии» стали активно внедряться методы математической статистики в археологию. Это работы Л.Бинфорда, Д.Кларка, ряд сборников статей, посвященных применению количественных методов в археологии. С этого времени началось активное издание учебных пособий по их применению в археологии. Это книги: J.E.Doran & F.R.Hodson. *Mathematic & Computer in Archaeology* (1975); C.Orton. *Mathematics in Archaeology* (1982); M.Fletcher, G.R.Lock. *Digging for archaeologists: elementary statistics for archaeologists* (1991); M.J.Baxter. *Exploratory multivariate analysis in archaeology* (1994); R.D.Drennan. *Statistics for archaeologists* (1996); S.Shennan. *Quantifying Archaeology* (1997); C.Orton. *Sampling in archaeology* (2000), M.J.Baxter. *Statistics in archaeology* (2003) и др.

В середине 70-х гг. началось применение методов математической статистики в СССР. Стали издаваться сборники статей по применению статистико-комбинаторных методов в археологии. Отрядным фактом в 80-е годы XX столетия было появление учебного пособия по математической статистике (Г.А.Федоров-Давыдов «Статистические методы в археологии» – М. 1987), которое до сих пор является единственным отечественным изданием в данном направлении. Здесь ощущается явное несоответствие потребностям

Последние новости

IBM купила разработчика ПО для анализа данных Butterfly Software

Digit.ru – информационно-аналитический IT-портал

Butterfly разрабатывает корпоративные решения для планирования операций,...

Школа анализа данных Яндекса

Школа анализа данных открыта в 2007 году. Инициатором создания выступила компания «Яндекс». Она же взяла на себя финансирование Школы,...

Лаборатория Касперского: спам как глобальная угроза, границы стираются

Новостная служба Feega

Рис. 1. Главная страница портала «Автоматизированная обработка археологической информации»

Диски

- Публичные архивы пользователей
- Общедоступный архив
- Временный архив

Публичные ...

- belkosyak
- admin
- manager
- anaefthen
- sach23omm

belkosyak

- Раскопки
- Черепка

Черепка

- Объем
- v3
- v4
- v5

пол

- Показать столбец данных
- Шкала: Номинальная
- Количество валидных значений: 35

Категории

- Задачи археологии
- Анализ данных
- Статистика

Анализ дан...

- Нейронные сети
- Метод главных компонент
- Линейная регрессия
- Множественная логит-регрессия
- k-means++
- Линейный дискриминантный анализ

Линейный дискриминантный анализ

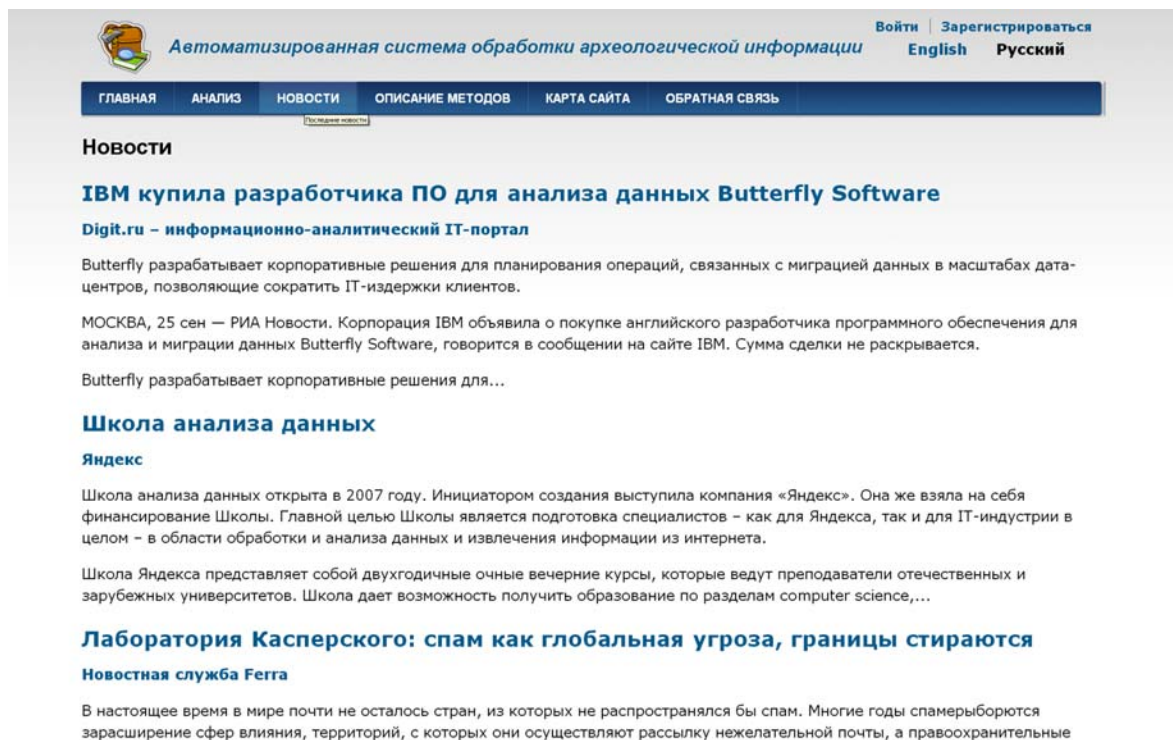
Входные признаки

Зависимая переменная:
пол [id:6842].
Признак объекта, который означает принадлежность наблюдения к определенному классу.

Независимые переменные:
v3 [id:6843], v4 [id:6844].

Запустить

Рис. 2. Основной раздел системы – анализ данных



Автоматизированная система обработки археологической информации [Войти](#) | [Зарегистрироваться](#)
English Русский

[ГЛАВНАЯ](#) [АНАЛИЗ](#) [НОВОСТИ](#) [ОПИСАНИЕ МЕТОДОВ](#) [КАРТА САЙТА](#) [ОБРАТНАЯ СВЯЗЬ](#)

Новости

IBM купила разработчика ПО для анализа данных Butterfly Software

Digit.ru – информационно-аналитический IT-портал

Butterfly разрабатывает корпоративные решения для планирования операций, связанных с миграцией данных в масштабах дата-центров, позволяющие сократить IT-издержки клиентов.

МОСКВА, 25 сен — РИА Новости. Корпорация IBM объявила о покупке английского разработчика программного обеспечения для анализа и миграции данных Butterfly Software, говорится в сообщении на сайте IBM. Сумма сделки не раскрывается.

Butterfly разрабатывает корпоративные решения для...

Школа анализа данных

Яндекс

Школа анализа данных открыта в 2007 году. Инициатором создания выступила компания «Яндекс». Она же взяла на себя финансирование Школы. Главной целью Школы является подготовка специалистов – как для Яндекса, так и для IT-индустрии в целом – в области обработки и анализа данных и извлечения информации из интернета.

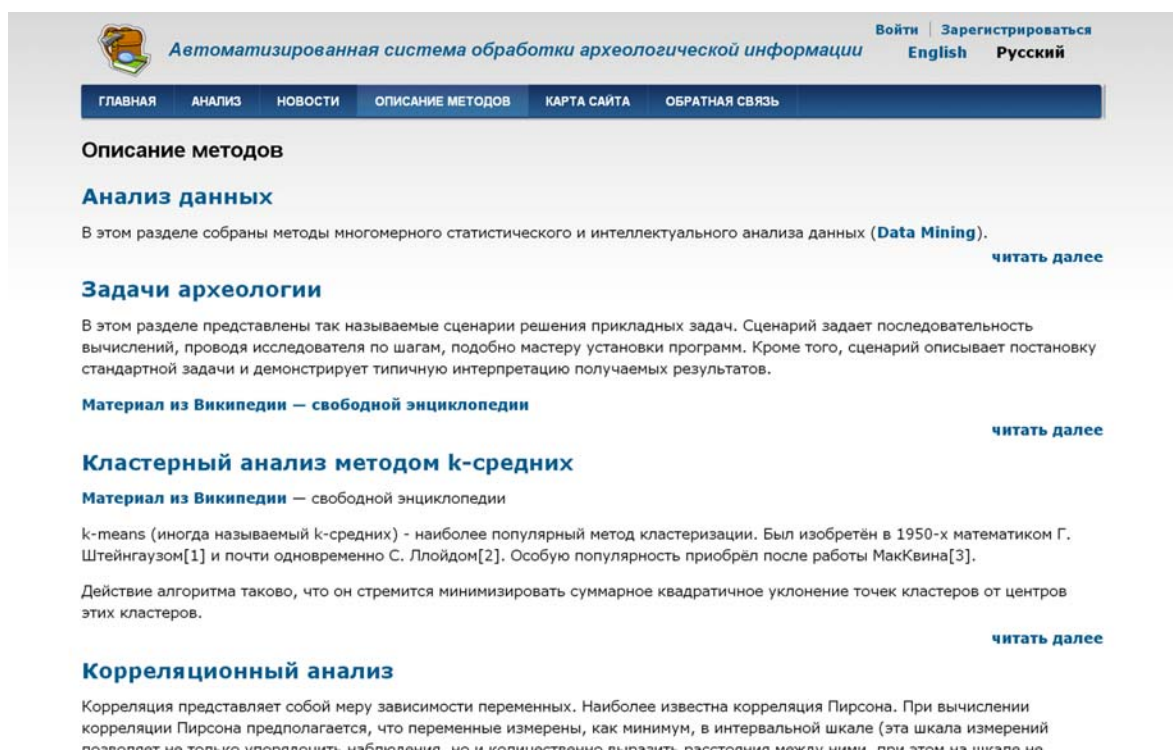
Школа Яндекса представляет собой двухгодичные очные вечерние курсы, которые ведут преподаватели отечественных и зарубежных университетов. Школа дает возможность получить образование по разделам computer science,...

Лаборатория Касперского: спам как глобальная угроза, границы стираются

Новостная служба Ferrat

В настоящее время в мире почти не осталось стран, из которых не распространялся бы спам. Многие годы спамеры борются за расширение сфер влияния, территорий, с которых они осуществляют рассылку нежелательной почты, а правоохранительные

Рис. 3. Страница новостей



Автоматизированная система обработки археологической информации [Войти](#) | [Зарегистрироваться](#)
English Русский

[ГЛАВНАЯ](#) [АНАЛИЗ](#) [НОВОСТИ](#) [ОПИСАНИЕ МЕТОДОВ](#) [КАРТА САЙТА](#) [ОБРАТНАЯ СВЯЗЬ](#)

Описание методов

Анализ данных

В этом разделе собраны методы многомерного статистического и интеллектуального анализа данных (**Data Mining**). [читать далее](#)

Задачи археологии

В этом разделе представлены так называемые сценарии решения прикладных задач. Сценарий задает последовательность вычислений, проводя исследователя по шагам, подобно мастеру установки программ. Кроме того, сценарий описывает постановку стандартной задачи и демонстрирует типичную интерпретацию получаемых результатов.

Материал из Википедии — свободной энциклопедии [читать далее](#)

Кластерный анализ методом k-средних

Материал из Википедии — свободной энциклопедии

k-means (иногда называемый k-средних) - наиболее популярный метод кластеризации. Был изобретён в 1950-х математиком Г. Штейнгаузом[1] и почти одновременно С. Ллойдом[2]. Особую популярность приобрёл после работы МакКвина[3].

Действие алгоритма таково, что он стремится минимизировать суммарное квадратичное отклонение точек кластеров от центров этих кластеров. [читать далее](#)

Корреляционный анализ

Корреляция представляет собой меру зависимости переменных. Наиболее известна корреляция Пирсона. При вычислении корреляции Пирсона предполагается, что переменные измерены, как минимум, в интервальной шкале (эта шкала измерений позволяет не только измерять, но и количественно выразить расстояние между ними. При этом из шкалы не

Рис. 4. Описание методов интеллектуального анализа данных

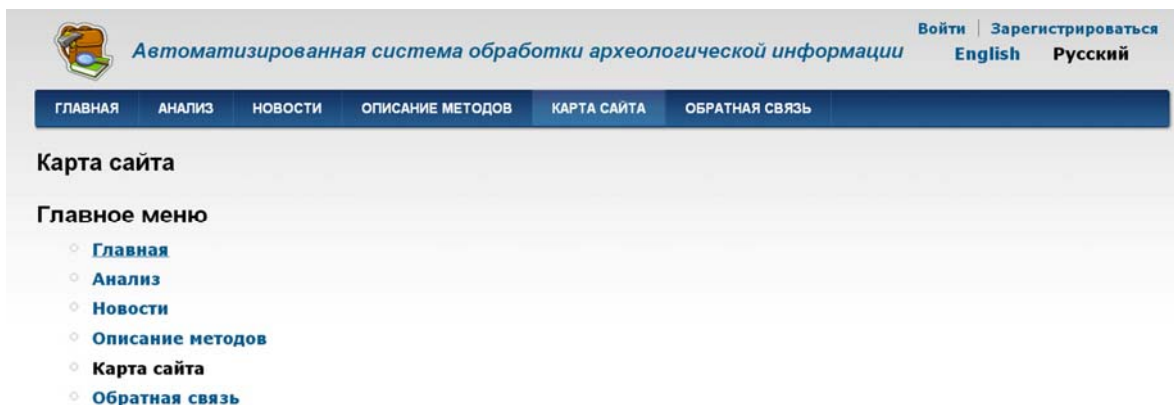


Рис. 5. Карта сайта

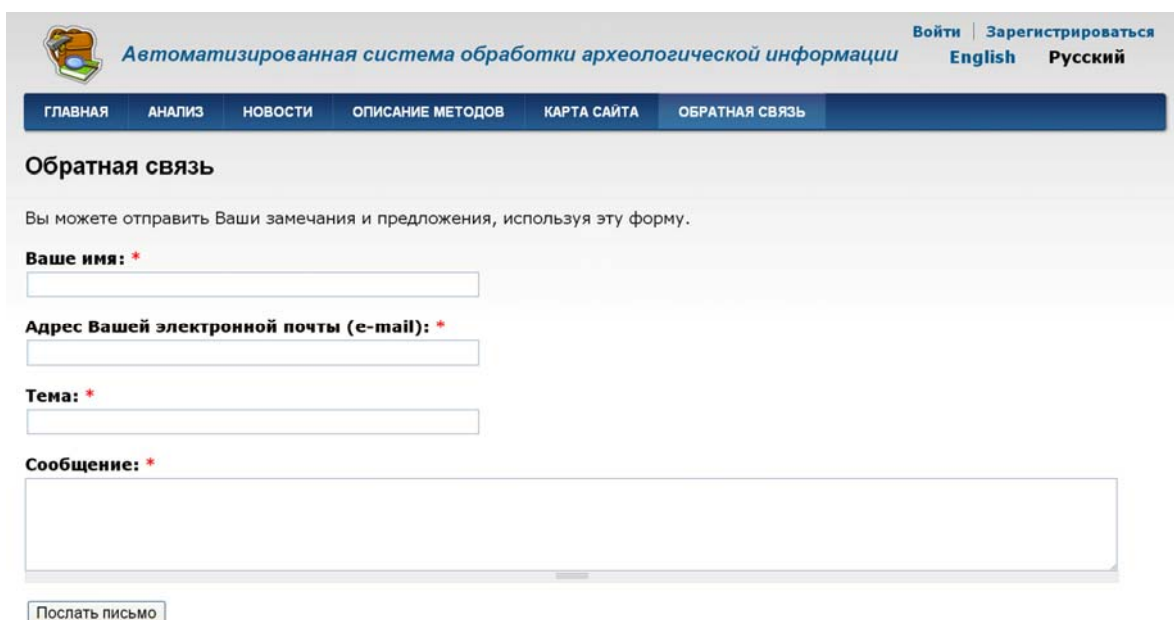



Рис. 6. Обратная связь

чивая неактуальную в данный момент информацию в как можно меньшую область экрана. Под неактуальной информацией следует понимать как верхние уровни навигации, так и всю навигацию по методам, когда мы работаем с данными, или всю навигацию по данным, когда мы работаем с методами. И только в тот момент, когда мы уже выбрали массив данных и метод, которым мы собираемся их обрабатывать, можно одновременно раскрыть обе панели – и данных, и методов, чтобы настраивать метод путем забрасывания признаков в его параметры.

Новости <http://ko.ieie.nsc.ru/archaeology/ru/news> по темам, имеющим отношение к археологии, анализу данных и прочим, помогающие пользователю расширить свои представления по теме. Заголовок выполнен в виде гиперссылки на полный текст одной этой новости в том же окне. Строка, которой указаны дата и источник новости – гиперссылка на новость в этом источнике (рис. 3).

Описание методов <http://ko.ieie.nsc.ru/archaeology/ru/methods-description> анализа данных, помогающее понять, для решения каких задач применяется каждый метод, и

какие требования к входным данным он предъявляет. Те же описания доступны из навигатора по методам – там они вызываются кнопкой  в правом верхнем углу второго и последующих уровней навигатора методов. Описания привязаны ко всему уровню навигации, а не к выделенному элементу, который при открытии уровня еще не определен. Эти описания имеет смысл дополнить примерами решения задач и замечаниями пользователей из соответствующей ветви форума (рис. 4).

Карта сайта <http://ko.ieie.nsc.ru/archaeology/ru/sitemap> для ресурса с такой простейшей структурой не имеет самостоятельного смысла, но при развитии, разветвлении структуры, может оказаться полезной. Особенно разветвленной структура станет при появлении форума для пользователей (рис. 5).

Обратная связь <http://ko.ieie.nsc.ru/archaeology/ru/contact> предоставляет пользователю возможность вносить свои замечания и предложения, посылая письма непосредственно администратору сайта. Следует заметить, что для зарегистрированных пользователей такие поля, как «Ваше имя» «Адрес Вашей электронной почты (e-mail)» должны по умолчанию заполняться значениями из регистрационных данных, чтобы по возможности не напрягать пользователя по пустякам (рис. 6).

Web-система для обработки археологической информации будет содержать оригинальные методы интеллектуального анализа данных, в частности, системы Discovery, реализующей реляционный подход для обнаружения закономерностей в данных [Витяев, 2006; 2010; Витяев, Москвитин, 1993; Kovalerchuk, Vityaev, 2000].

Для того чтобы пользователь мог в этой системе удобно и наглядно 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. Нами разработана такая система, дано описание системы и пример ее использования для решения задач из области медицины.

Проведено сравнение системы Discovery с алгоритмами Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services. Показано, что система «Discovery», во-первых, обладает

теоретическими преимуществами перед этими алгоритмами, и, во-вторых, практически лучше работает на данных, где эти преимущества проявляются явно.

В археологии множество задач решается методами классификации в режиме самообучения. Существующие алгоритмы никак не связаны с существующими теориями психологии образа и восприятия. 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. Нами 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. Нами разработан алгоритм, реализующий данную формализацию. Проведены эксперименты, демонстрирующие ее работоспособность.

Список литературы

Белова Е. Б., Бородкин Л. И., Гарскова И. М., Измestьева Т. Ф., Лазарев В. В., Тихонов А. И. Компьютеризованный статистический анализ для историков. М.: Изд-во МГУ, 1999. 187 с.

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

Витяев Е. Е. Извлечение информации из данных // Информационные технологии в гуманитарных исследованиях: Сб. науч. ст. Новосибирск, 2010. Вып. 15. С. 9–16.

Витяев Е. Е., Москвитин А. А. Введение в теорию открытий. Программная система DISCOVERY // Вычислительные системы: Сб. науч. ст. Новосибирск, 1993. Вып. 148. С. 117–163.

Жданов А. С., Костин В. С. Значимость и устойчивость автоматической классификации в задаче поиска оптимального разбиения // Информационные технологии в гуманитарных исследованиях: Сб. науч. ст. Новосибирск, 2002. Вып. 3. С. 36–42.

Костин В. С. Статистика для сравнения классификаций // Информационные технологии в гуманитарных исследованиях: Сб. науч. ст. Новосибирск, 2003. Вып. 6. С. 57–65.

Костин В. С., Корнюхин Ю. Г. Построение обобщенной классификации // Информационные технологии в гуманитарных исследованиях: Сб. науч. ст. Новосибирск, 2003. Вып. 6. С. 65–72.

Федоров-Давыдов Г. А. Статистические методы в археологии. М.: Высш. шк., 1987. 216 с.

Фирсов Н. И., Витяев Е. Е. Сравнение системы «Discovery» с алгоритмами Microsoft Association Rules и Decision Trees, встроенными в Microsoft SQL Server Analysis Services // Информационные технологии в гуманитарных исследованиях: Сб. науч. ст. Новосибирск, 2012. Вып. 17. С. 51–63.

Baxter M. J. Exploratory Multivariate Analysis in Archaeology. Edinburgh: Edinburgh Univ. Press, 1994. 307 p.

Baxter M. J. Statistics in Archaeology. L.: Arnold, 2003. 292 p.

Binford L. R. An Archaeological Perspective. N. Y.; L.: Seminar Press, 1972. 464 p.

Clarke D. L. Analytical Archaeology. L.: Methuen, 1968. 684 p.

Doran J. E., Hodson F. R. Mathematic & Computer in Archaeology. Edinburgh: Edinburgh Univ. Press, 1975. 381 p.

Drennan R. D. Statistics for Archaeologists. N. Y.: Plenum Press, 1996. 273 p.

Fletcher M., Lock G. R. Digging Numbers: Elementary Statistics for Archaeologists. Oxford: Oxford Univ. School of Archaeology, 2005. 205 p.

Kovalerchuk B., Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid methods. Boston: Kluwer Academic Publishers, 2000. 308 p.

Orton C. Mathematics in Archaeology. Cambridge; N. Y.: Cambridge Univ. Press, 1982. 248 p.

Orton C. Sampling in Archaeology. Cambridge: Cambridge Univ. Press, 2000. 261 p.

Shennan S. Quantifying Archaeology. Iowa City: Univ. of Iowa Press, 1997. 433 p.

Материал поступил в редколлегию 07.02.2013

E. E. Vityaev, V. S. Kostin, Yu. P. Kholyushkin

WEB SYSTEM INTELLIGENT ANALYSIS OF ARCHAEOLOGICAL DATA

In accordance with the plan of automatic training Web-system development for the archaeological information processing, the following is done: main principals and demands to architecture of the system is developed and the list of instrumental devices is defined. The system contain three main components: data base, bloc of computational methods and interface with user. For data base MySQL is selected that is rather useful in elaboration and accompaniment and have sufficient speed of access for expected data. Methods will be produced by R-language for statistical processing of data and graphical representation that is developing under the GNU GPL license. Interface is developed using PHP with technology AJAX that permits to show new information without reloading the web-page. Information in the interface is structured as in the file system of contemporary computers. Public data is store in the core partition and also in personal archives that are accessible for users or administrators. Experimental version of the Web-system, that is available on the website <http://ko.ieie.nsc.ru/archaeology/ru> is developed. Some tasks is solved by the original system Discovery, that is included into the system. The comparison of the Discovery system with the Microsoft Association Rules and Decision Trees methods, from the Microsoft SQL Server Analysis Services is performed. Visual Discovery system for the visual extraction of information from data in terms of ontology is developed.

Keywords: archaeology, information systems, intelligent data analysis, statistical methods, tasks solution strategies, Data Mining.