

**ВОРОНЕЖСКИЙ ИНСТИТУТ ЭКОНОМИКИ  
И СОЦИАЛЬНОГО УПРАВЛЕНИЯ**

**ИНТЕЛЛЕКТУАЛЬНЫЕ  
ТЕХНОЛОГИИ В  
ОБРАЗОВАНИИ, ЭКОНОМИКЕ И  
УПРАВЛЕНИИ**

**(Материалы II Международной  
научно-практической  
конференции)**

**ВОРОНЕЖ – РОССИЯ**

**2005**

## **ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

*Институт информационных технологий*

*Национальной Академии Наук  
Азербайджана, г. Баку, Республика  
Азербайджан*

**Юсифов Ф. Ф.**

Две важные и активные области текущего исследования – анализ данных (Data mining) и Всемирная паутина (WWW). Термин Data mining переводится как “интеллектуальный анализ данных” или “раскопка данных”. Нередко рядом с Data mining встречаются слова “обнаружение знаний в базах данных” (knowledge discovery in database). Возникновение всех указанных терминов связано с новым витком в развитии средств и методов обработки данных [1,2].

До начала 90-х годов казалось, не было особой нужды переосмысливать ситуацию в этой области. Все шло своим чередом в рамках направления, так называемой прикладной статистики. Вместе с тем, практики всегда знали, что попытки применить теоретические экзерсисы для решения реальных задач в большинстве случаев оказываются бесплодными. В связи с совершенствованием технологий записи и хранения данных на людей обрушились колоссальные потоки информации в самых различных областях. Традиционная математического статистика, долгое время претендовавшая на роль основного инструмента анализа данных, откровенно спасовала перед лицом возникших проблем. Методы математической статистики оказались полезным, можно сказать, только для проверки заранее сформулированных гипотез и для разведочного анализа, составляющего основу оперативной аналитической обработки данных (online analytical processing).

В основу современной технологии Data mining положена концепция шаблонов (паттернов), отражающих фрагменты многоаспектных взаимоотношений в данных. Эти шаблоны представляют собой закономерности, свойственные подвыборкам данных, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределения значений анализируемых показателей [1].

## СЕКЦИЯ IX. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБРАЗОВАНИИ, ЭКОНОМИКЕ И СОЦИАЛЬНОМ УПРАВЛЕНИИ

Можно отметить, что сфера применения Data mining ничем не ограничена – она везде (розничная торговля, банковское дело, телекоммуникации, страхование, медицина и др.), где имеются какие-либо данные. Но, в первую очередь, методы Data mining сегодня стали все более и более необходимыми для пользователей WWW, использования методов интеллектуального анализа данных, чтобы найти, извлечь, фильтровать и оценить желаемую информацию.

В настоящее время естественная комбинация этих двух областей – интеллектуальный анализ данных с WWW, называемых Web mining, была центром нескольких недавних научно-исследовательских работ. Термин Web mining использовался в основном в двух направлениях [4,5]. Первое, которое называется Web content mining, описывает процесс обнаружения информации или ресурса из миллионов источников из WWW. Второе, называемое Web usage mining, является процессом анализа журналов доступа Web или другой пользовательской информации и доступа на одной или нескольких окрестных сетях. Web mining может быть широко определен как открытие и анализ полезной информации из WWW. Это широкое определение с одной стороны, описывает автоматический поиск, поиск информации и средств, доступных от миллионов сайтов и взаимосвязанных баз данных с другой стороны, обнаружение и анализ пользовательских образов доступа от одних или нескольких серверов Web [4,6].

Data mining является мультидисциплинарной областью, возникшей и развивающейся на базе достижения прикладной статистики, распознавания образов, методов искусственного интеллекта (нейронные сети, эволюционное программирование, генетические алгоритмы) и др. Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data mining. Можно сказать, что многие из таких систем интегрируют в себе сразу несколько подходов. Тем не менее, как правило, в каждой системе имеется какой-то ключевой компонент, на которой делается основное предпочтение.

Выделяют пять стандартных типов закономерностей, которые позволяют выявлять методы Data mining:

- ассоциация;
- последовательность;
- классификация;
- кластеризация;
- прогнозирование.

Ассоциация имеет место в том случае, если несколько событий связаны друг с другом. Если существует цепочка связанных во времени событий, то говорят о последовательности.

Наиболее распространенные задачи интеллектуальных методов

*II Международная конференция «ИТОЭУ - 2005», 2 декабря 2005 г., Воронеж*  
анализа – это классификация и кластеризация. Классификация позволяет выявить признаки, характеризующие однотипные группы объектов - классы для того, чтобы по известным значениям этих характеристик можно было отнести новый объект к тому или иному классу. Кластеризация обобщает идею классификации на более сложный случай, когда сами классы не predetermined. Результатом кластеризации как раз и является определение присущего исследуемым данным разбиения на классы (кластеры) [1,2].

Можно отметить, что специфика задач классификации социально-экономических объектов определяется присутствием в описании как количественных, так и качественных признаков. Границы между отдельными градациями признаков трудно установить однозначно. Как следствие - однозначная классификация и идентификация объектов по одному или нескольким признаком не всегда может быть определена достаточно строго и точно. В случае же, когда классификацию выполняют эксперты, возможны как непреднамеренные (связанные со сложной структурой данных), так и преднамеренные ошибки.

Основной для всевозможных систем прогнозирования служит историческая информация, хранящаяся в базе данных в виде временных рядов. Если удастся найти шаблоны, адекватно отражающие динамику поведения целевых показателей, есть вероятность, что с их помощью можно предсказать и поведение системы в будущем.

Использование методов интеллектуального анализа данных дает начало потребности создания серверных и клиентских интеллектуальных систем, которые могут эффективно добыть знания из WWW. Например, сделать сайт более привлекательным и полезным для пользователей или потенциальных заказчиков, может детальный анализ регистрационного журнала сайта с применением профессиональных Web-анализаторов или же с помощью сторонних организаций, специализирующихся на предоставлении таких услуг [3,6]. Однако многие администраторы, и без того перегруженные информацией, испытывают значительные трудности, интерпретируя данные регистрационных журналов Web-серверов. С этой целью для повышения эффективности работы Web-сайтов необходимо применение Web-анализаторов, способных принести компании максимум пользы. Для создания таких интеллектуальных Web-анализаторов могут быть использованы методы Data mining.

Популярность WWW продолжает увеличиваться, есть растущая потребность в разработке инструментальных средств и методик, которые помогут улучшить ее полноценность. С этой целью целесообразно использование методов интеллектуального анализа данных на WWW. Термин Web mining использован так, чтобы обратиться к раз-

## СЕКЦИЯ IX. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБРАЗОВАНИИ, ЭКОНОМИКЕ И СОЦИАЛЬНОМ УПРАВЛЕНИИ

личным видам методики, которые охватывают широкий диапазон задач.

### *Литература*

1. Дюк В.А., Самойленко А.В. *Data Mining*. Учебный курс. СПб: «Питер», 2001.
2. *Интеллектуальные модели анализа экономической информации: курс лекций*, [http://www.basegroup.ru/download/education/bi\\_lectures.pdf](http://www.basegroup.ru/download/education/bi_lectures.pdf), 2005.
3. Рабин Д. *Изучайте журналы посетителей. Сети и системы связи №1 (121)*, [http://ccc.ru/magazine/depot/05\\_01/read.html?0201.htm](http://ccc.ru/magazine/depot/05_01/read.html?0201.htm), 2005.
4. Cooley R., Mobasher B., and Srivastava J., "Web mining: Information and pattern discovery on the world wide web," *Proc. 9th IEEE Int. Conf. Tools with Artificial Intelligence*, pp. 558–567, Nov. 1997
5. Kosla R. and Blockeel H., "Web mining research a survey," *SIG KDD Explorations*, vol. 2, pp. 1–15, July 2000
6. Srivastava J., Cooley R., Deshpande M., and Tan P., *Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data*. *SIGKDD Explorations*, 1(2):12–23, 2000.

## МЕТОД ФОРМИРОВАНИЯ СИНОНИМИЧЕСКИХ РЯДОВ

ОАО «ЦентрТелеком», Калужский  
филиал, г. Калуга, Россия

**Шаров А. Г.**

Интересующая автора данной работы идея теоретической проработки и практической реализации автоматизированной обучающей системы (АОС) с использованием математических моделей искусственных нейронных сетей (ИНС) в принципе не нова. Однако сложившаяся конъюнктура современного рынка программных средств для реализации подобного рода идей делает возможным не только реализацию на уровне экспериментальных и опытных образцов, но и позволяет использовать ранее созданные базы и банки данных и накопленные базы знаний по предметным областям для конкретной реализации АОС по тому или иному курсу обучения, предмету, либо отдельной специализации (специальности).