

Гусейнова Арзу

д-р экон. наук, заместитель директора
Института исследований экономических реформ
Министерства экономики Азербайджана
г. Баку, Республика Азербайджан
arzu.huseynova@economy.gov.az

ПРИМЕНЕНИЕ В ЗАДАЧАХ ПО ПРИНЯТИЮ РЕШЕНИЙ И УПРАВЛЕНИЮ ДАННЫМИ МОДЕЛЕЙ DATA MINING И FAZZY

Система поддержки принятия решений представляет собой совокупность взаимосвязанных по целям, параметрам и условиям задач, методов (методик), программных средств и технических систем, позволяющих формировать в автоматизированном режиме набор отчетных форм, содержащих информацию для принятия управленческих решений и/или варианты таких решений. Сложность решения этой проблемы заключается в необходимости использования интеллектуальных алгоритмов обработки информации, которые могли бы дать наглядные и понятные результаты для принятия решений. Традиционные методы анализа информации не позволяют решать задачи такого класса. Известные статистические методы покрывают лишь часть нужд по обработке данных, и для их использования необходимо иметь четкое представление об искомым закономерностях. Эти задачи способствуют улучшению понимания анализируемых данных. В такой ситуации методы интеллектуального анализа данных приобретают особую актуальность. Интеллектуальная обработка производится методами DM (Data Mining). DM – поиск скрытых от аналитика функциональных и логических закономерностей, тенденций в накопленной информации, построение моделей и правил, которые объясняют найденные аномалии или прогнозируют развитие процессов, о которых он не имеет представления или имеет, но очень смутное. В отличие от оперативной аналитической обработки данных (online analytical processing, OLAP), в Data Mining бремя формулировки гипотез и выявления необычных (unexpected) шаблонов переложено с человека на компьютер. Существует целый ряд методов интеллектуального анализа данных, многие из которых специально предназначены для статистического анализа и попыток извлечь возможную новую информацию из существующих данных. Сегодня растущее число методологий интеллектуального анализа данных способно также обрабатывать неопределенности (fuzzy) в существующих данных, и эти методологии предполагают и учитывают неоднозначные и расплывчатые ответы в выходных результатах аналитических запросов.

Ключевые слова: data mining, fazz.

APPLICATION IN THE PROBLEMS OF DECISION MAKING AND DATA MANAGEMENT MODELS DATA MINING AND FAZZY

In general, the decision support system is a set of methods (methods), software tools and technical systems interconnected by the goals, parameters and conditions that allow to create in an automated mode a set of reporting forms containing information for making managerial decisions and / or options for such decisions solutions. The complexity of solving this problem lies in the need to use intelligent algorithms for processing information that could provide clear and understandable results for decision making. Traditional methods of analyzing information do not allow solving problems of this class. Known statistical methods cover only a part of the data processing needs, and for their use it is necessary to have a clear idea of the required patterns. In such a situation, methods of data mining become particularly relevant. Intellectual processing is performed by DM (Data Mining) methods. In contrast to online analytical processing (OLAP) in Data Mining, the burden of formulating hypotheses and identifying unusual (unexpected) templates has been shifted from person to computer. There are a number of methods for data mining, many of which are specifically designed for statistical analysis and attempts to extract possible new information from existing data. Today, a growing number of data mining methodologies can also handle uncertainties (fuzzy) in existing data, and these methodologies assume and take into account ambiguous and vague responses in the output of analytical queries.

Key words: data mining, fazyy.

Понятие Data Mining «добыча данных».

Data Mining (DM) (KDD – knowledge discovery in databases) представляет собой новейшее направление в области информационных технологий, ориентированное на решение задач поддержки принятия решений на основе количественных и качественных исследований сверхбольших массивов разнородных ретроспективных данных.

Интеллектуальная обработка производится методами *DM (Data Mining)*. *DM* – поиск скрытых от аналитика функциональных и логических закономерностей, тенденций в накопленной информации, построение моделей и правил, которые объясняют найденные аномалии или прогнозируют развитие процессов, о которых он не имеет представления или имеет, но очень смутное. Что такое Data Mining? StatSoft определяет понятие «добыча данных» как процесс аналитического исследования больших массивов информации (обычно экономического характера)

с целью выявления определенных закономерностей и систематических взаимосвязей между переменными, которые затем можно применить к новым совокупностям данных.

Методы анализа Big Data можно условно разбить на следующие категории:

- Краудсорсинг – категоризация и обогащение данных силами широкого, неопределённого круга лиц, привлечённых на основании публичной оферты, без вступления в трудовые отношения;
- Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы;
- Искусственные нейронные сети (ИНС), в том числе генетические алгоритмы;
- Прогнозная аналитика;
- Имитационное моделирование;
- Статистический анализ, к примеру методы A/B-тестирования или анализ временных рядов;
- Методы класса Data Mining: обучение ассоциативным правилам (англ. *association rule learning*), методы классификации, регрессионный анализ.

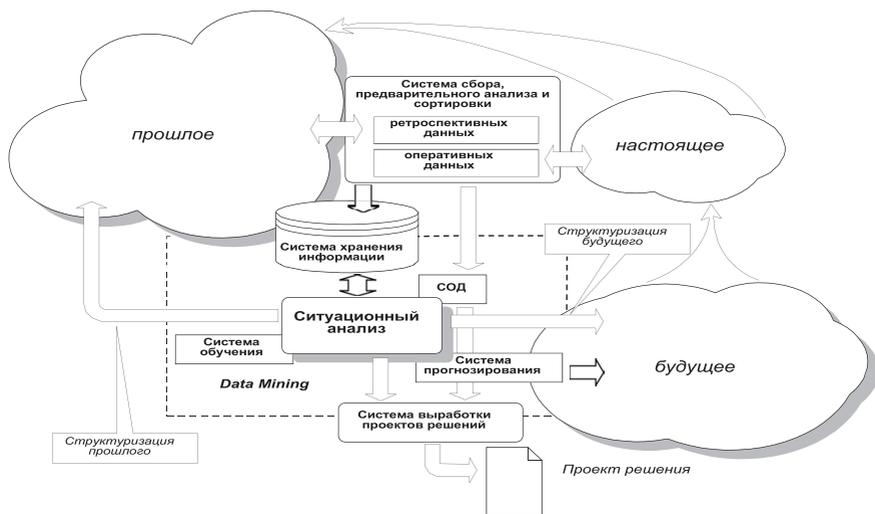


Рис. 1. Структура автоматизированного формирования управляющих решений

Обнаружение знаний в базах данных

Формирование баз знаний неразрывно связано с извлечением скрытых закономерностей, которые содержатся в базах данных. Технологии, которые позволяют это реализовать объединены под общим названием **Knowledge Discovery in Databases (KDD)** – обнаружение знаний в базах данных. **Knowledge Discovery in Databases** предназначен для поиска закономерностей в данных, который включает в себя такие этапы как: подготовка данных, выбор информативных признаков, применение методов **Data Mining** и оценки полученных результатов.

На рисунке 2 представлена типовая схема технологии обнаружения знаний в базах данных.

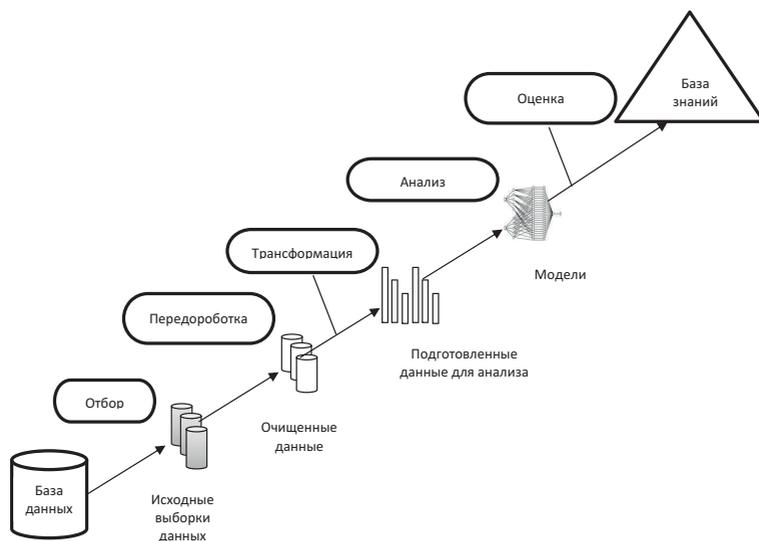


Рис. 2. Схема технологии обнаружения знаний в базах данных

Характерным признаком современных систем управления является ситуация, когда в информационных системах накапливаются большие объемы информации, которые практически не обрабатываются. Нарастание объема информации, масштабов производства, требующего обеспечения все возрастающими объемами ресурсов, приводит к значительному усложнению задач принятия решений и управления. В общем, система поддержки принятия решений представляет собой совокупность взаимосвязанных по целям, параметрам и условиям задач, методов (методик), программных средств и технических систем, позволяющих формировать в автоматизиро-

ванном режиме набор отчетных форм, содержащих информацию для принятия управленческих решений и/или варианты таких решений. Сложность решения этой проблемы заключается в необходимости использования интеллектуальных алгоритмов обработки информации, которые могли бы дать наглядные и понятные результаты для принятия решений. Традиционные методы анализа информации не позволяют решать задачи такого класса. Известные статистические методы покрывают лишь часть нужд по обработке данных, и для их использования необходимо иметь четкое представление об искомым закономерностях. Специфика анализа данных в сфере ГЧП заключается в том, что большинство этих задач подразумевают построение моделей, которые раскрывают закономерности в данных. Поэтому модели интеллектуального анализа наиболее востребованными в сфере ГЧП. Эти задачи способствуют улучшению понимания анализируемых данных. В такой ситуации методы интеллектуального анализа данных (Data Mining) приобретают особую актуальность. Их основная особенность заключается в установлении наличия и характера скрытых закономерностей в данных, тогда как традиционные методы занимаются главным образом параметрической оценкой уже установленных закономерностей. Ключевой момент в таких моделях – легкость и прозрачность результатов для восприятия человеком. Возможно, обнаруженные закономерности будут специфической чертой именно конкретных исследуемых данных и больше нигде не встретятся, но это все равно может быть полезно и потому должно быть известно.



Рис. 3. Математический арсенал Data Mining

Современные технологии Data Mining (discovery-driven data mining) перелопачивают информацию с целью автоматического поиска шаблонов (паттернов), характерных для каких-либо фрагментов неоднородных многомерных данных. В отличие от оперативной аналитической обработки данных (online analytical processing, OLAP) в Data Mining бремя формулировки гипотез и выявления необычных (unexpected) шаблонов переложено с человека на компьютер. Существует целый ряд методов интеллектуального анализа данных, многие из которых специально предназначены для статистического анализа и попыток извлечь возможную новую информацию из существующих данных. Сегодня растущее число методологий интеллектуального анализа данных способно также обрабатывать неопределенности (fuzzy) в существующих данных, и эти методологии предполагают и учитывают неоднозначные и расплывчатые ответы в выходных результатах аналитических запросов.

Построение модели интеллектуального анализа данных является частью более масштабного процесса, в который входят все задачи, от формулировки вопросов относительно данных и создания модели для ответов на эти вопросы до развертывания модели в рабочей среде. Этот процесс можно представить как последовательность следующих базовых этапов:

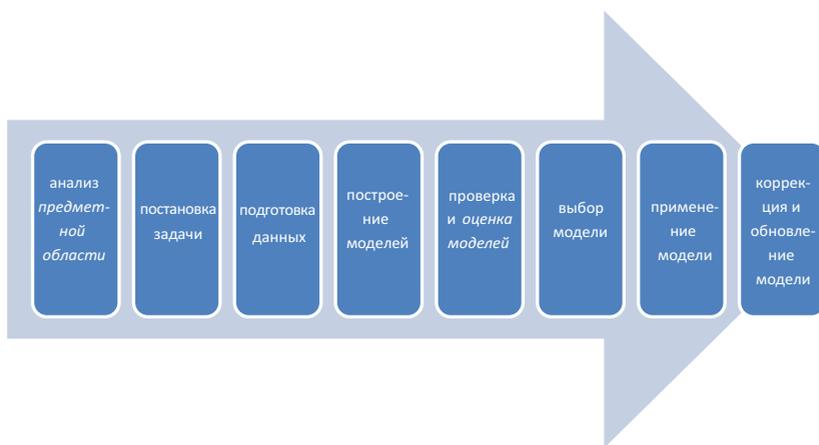


Рис. 4. Построение модели интеллектуального анализа

Нечеткая логика. Особое направление в спектре DM составляют методы, основанные на нечетких множествах (fuzzy sets).

Традиционная вероятностно-статистическая методология базируется на классической колмогоровской аксиоматике. В ее основе лежит понятие меры, определенной на множестве \square -алгебр F в пространстве элементар-

ных событий *W*. Однако в ряде практических задач, связанных, например, с лингвистическими переменными, подобную аксиоматику построить не удастся. В связи с этим в 1961 г. Л. Заде была предложена концепция нечетких множеств, позволяющая оперировать с понятием неопределенности в не метрических системах. Применение теории нечетких множеств в системе DM позволяет ранжировать данные по степени близости к желаемому результату, осуществить, так называемый, нечеткий поиск в базах данных. Однако плата за повышенную универсальность всегда была достаточно велика и проявлялась в снижении уровня достоверности и точности получаемых результатов. Поэтому число специализированных приложений данной методологии, несмотря на повышенный интерес к ней со стороны математиков-прикладников в течении последних 35 лет, весьма ограничено.

Перечисленные алгоритмы DM в настоящее время реализованы в качестве специализированных пакетов анализа данных. Среди основных программных продуктов, содержащих в себе кибернетические методы DM, следует назвать системы NeuroShell, GeneHunter, BrainMaker, OWL, PolyAnalyst, 4Thought (BI).



Рис. 5. Обобщенная функционально-структурная модель системы поддержки принятия решений