

ТЕХНОЛОГИЯ DATA MINING*БНТУ, Минск**Научный руководитель: Дробыш А.А.*

В результате развития информационных технологий, количество данных, накопленных человечеством в электронном виде, растет быстрыми темпами. Эти данные существуют вокруг нас в различных видах: тексты, изображения, аудио, видео, гипертекстовые документы, реляционные базы данных и т.д. Огромное количество данных появилось в результате повсеместного использования сети Интернет, которая значительно облегчила доступ к информации из географически удаленных точек Земли. Однако подавляющая часть доступной информации не несет для конкретного человека какой-либо пользы. Человек не в состоянии переработать такое количество сведений. Возникает проблема извлечения полезной для пользователя информации из большого объема «сырых» данных. Данное руководство посвящено Технологии Интеллектуального Анализа Данных (Data Mining) – одной из активно развивающихся областей информационных технологий, предназначенной для выявления полезных знаний из баз данных различной природы.

Технология интеллектуального анализа данных (Data mining) может рассматриваться как результат естественной эволюции информационных технологий. По пути такой эволюции индустрия баз данных занималась разработкой следующих функциональностей: накопление данных, управление данными, а также анализ данных. Технология баз данных начиная с середины 1980-х характеризовалась популяризацией, широким внедрением, и концентрацией исследовательских усилий на новые, все более мощные СУБД. Появились

гетерогенные системы баз данных, а также глобальные информационные системы, такие как Всемирная Паутина – World Wide Web (WWW), которые играют ключевую роль в индустрии информационных технологий. Это способствовало всплеску индустрии информационных технологий и сделало огромное количество баз данных и репозитории информации доступными для управления транзакциями, извлечения данных и анализа данных. Одна из недавно появившихся архитектур – это хранилища данных, репозиторий множества разнородных источников данных, организованных в рамках единой схемы в одном месте, предназначенный для принятия управленческих решений.

Технология хранилищ данных включает очистку данных, интеграцию данных, а также онлайн-аналитическую обработку (On-Line Analytical Processing – OLAP), то есть технологию анализа с такими функциональностями, как консолидация, агрегация, подведение итогов, просмотр информации «под разными углами». Быстро растущие объемы накопленных данных быстро превысили способности человека в их обработке. В результате большие базы данных стали «могилами» данных – архивами, которые редко посещаются. Как следствие, важные решения принимаются не на основе информационно-насыщенных баз данных, а на основе интуиции человека, принимающего решения, так как он не имеет подходящих инструментов для извлечения полезных знаний из огромных объемов данных.

Технология Интеллектуального Анализа Данных позволяет извлечь полезные знания, важные паттерны, способствуя совершенствованию бизнес-стратегий, баз знаний, научных и медицинских исследований. Интеллектуальным анализом данных мы будем называть процесс определения новых, корректных и потенциально полезных знаний на основе больших массивов данных.

В англоязычной литературе вместо термина «интеллектуальный анализ данных» обычно используется термин Data Mining (дословный перевод – «добыча данных»), а также близкий термин Knowledge Discovery in Databases (KDD) – «Обнаружение знаний в больших базах данных». Извлеченное знание в результате интеллектуального анализа данных мы будем называть термином паттерн. Паттерном может быть, например, некоторое нетривиальное утверждение о структуре данных, об имеющихся закономерностях, о зависимости между атрибутами и т.д.

Таким образом, задачей интеллектуального анализа данных является эффективное извлечение осмысленных паттернов из имеющегося массива данных большого размера. Для отсева большого количества возможных малополезных паттернов может вводиться функция полезности. В реальности оценка полезности знания имеет субъективный характер, то есть зависит от конкретного пользователя.

Интересные знания, закономерности, высокоуровневая информация, полученные в результате анализа данных, могут быть использованы для принятия решений, контроля за процессами, управления информацией и обработки запросов. Поэтому технология интеллектуального анализа данных рассматривается как одна из самых важных и многообещающих тем для исследований и применения в отрасли информационных технологий.

Традиционно выделяются следующие этапы в процессе интеллектуального анализа данных:

1. Изучение предметной области, в результате которого формулируются основные цели анализа.
2. Сбор данных.
3. Предварительная обработка данных.
4. Анализ данных.
5. Интерпретация найденных паттернов.

Обычно в системах интеллектуального анализа данных выделяются следующие главные компоненты:

1. База данных, хранилище данных или другой репозиторий информации. Это может быть одна или несколько баз данных, хранилище данных, электронные таблицы, другие виды репозитория, над которыми могут быть выполнены очистка и интеграция. Виды баз данных:

- Реляционные базы данных;
- Хранилища данных;
- Транзакционные базы данных;
- Объектно-ориентированные базы данных;
- Объектно-реляционные базы данных;
- Пространственные базы данных (Spatial databases);
- Временные базы данных (Temporal databases);
- Текстовые базы данных;
- Мультимедийные базы данных;
- Разнородные базы данных;
- Всемирная Паутина.
- Сервер базы данных или хранилища данных.
- База знаний.
- Служба добычи знаний.
- Модуль оценки паттернов.
- Графический пользовательский интерфейс.

Многие компании пытаются обрабатывать данные, сгенерированные при выполнении ежедневных операций. Вооружившись технологиями машинного обучения и визуализации можно среди такой в общем-то беспорядочной информации обнаружить довольно ценные, хорошо интерпретируемые взаимосвязи. Приложения Data Mining, построенные на этих технологиях, успешно применяются в различных областях, в том числе в розничной торговле и маркетинге, позволяя

компаниям добывать информацию, дающую конкурентные преимущества.

УДК 321

Прокопеня А.

ИНТЕРАКТИВНЫЙ МЕТОД ОБУЧЕНИЯ

БНТУ, Минск

Научный руководитель: Зуенок А.Ю.

Интерактивный метод. Интерактивный («Inter» – взаимный, «act» – действие) – процесс взаимодействия между преподавателем и обучающимся, в результате которого происходит передача и усвоение знаний, умений и навыков, предусмотренных содержанием обучения. В отличие от активных методов, интерактивные ориентированы на более широкое взаимодействие обучающихся не только с преподавателем, но и друг с другом и на доминирование активности обучающихся в процессе обучения. Место преподавателя в интерактивных занятиях сводится к направлению деятельности обучающихся на достижение целей Занятия. Преподаватель также разрабатывает план занятия (обычно, это интерактивные упражнения и задания, в ходе выполнения которых ученик изучает материал).

Методы интерактивного обучения: интерактивные игры; «Мозговой штурм»; «Общая дискуссия»; проекты; семинары; «Общий галдѣж»;

Классификация игр:

– по области деятельности: интеллектуальные, социальные, психологические, физические, трудовые;

– по игровой среде: компьютерные, технические, настольные, телевизионные;

– по игровой методике: предметные, сюжетные, ролевые, деловые, имитационные, драматизация;