

КОНЦЕПЦИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Однобокова А.С.

*Однобокова Анастасия Сергеевна – студент,
факультет математики, информатики и физики,
Волгоградский государственный социально-педагогический университет, г. Волгоград*

Аннотация: в данной статье раскрываются концепции и задачи систем интеллектуального анализа данных, описываются основные инструменты реализации данного метода анализа и представляются направления их развития.

Ключевые слова: интеллектуальные системы, data mining.

С развитием информационных технологий увеличивается количество, объем и сложность хранилищ и баз данных. Объем хранимой в них информации может достигать большого количества записей. Поэтому возникает необходимость в разработке программных средств для автоматического анализа больших объемов данных с целью извлечения из них значимой информации. Для этого используются системы интеллектуального анализа данных, целью которых является обнаружение неявных закономерностей в наборах данных [3].

Существующие технологии могут помочь быстро найти нужную информацию, например, в базах данных, но во многих случаях этого недостаточно. Необходимо найти взаимосвязь между отдельными событиями в большом количестве данных, для этого требуются методы из многих областей, таких как математическая статистика, теория баз данных, теория искусственного интеллекта и др. [4]

Интеллектуальный анализ данных — это изучение данных с использованием методов искусственного интеллекта с упором на атрибуты, которые придают системам искусственный интеллект [3].

Интеллектуальный анализ данных — изучение и обнаружение ранее неизвестных, важных, легко интерпретируемых, практически полезных скрытых данных знаний «машинами» (алгоритмами, искусственным интеллектом) [1].

Интеллектуальный анализ данных может состоять из двух или трех этапов:

1. Выявление закономерностей (свободный поиск).

На этапе свободного поиска наборы данных рассматриваются с целью поиска скрытых закономерностей.

2. Использование выявленных закономерностей для предсказания неизвестных значений (прогностическое моделирование).

Прогнозирующее моделирование включает в себя следующие операции

- прогнозирование неизвестных величин (outcome prediction);
- прогнозирование развития процессов (forecasting).

3. Анализ исключений - стадия предназначена для выявления и объяснения аномалий, найденных в закономерностях.

Действие, выполняемое на этой стадии, - обнаружение отклонений (deviation detection). Для выявления отклонений необходимо определить критерии, которые будут вычисляться на этапе свободного поиска.

Итак, процесс Data Mining может быть представлен рядом таких последовательных стадий: Свободный поиск (в том числе валидация), прогностическое моделирование, анализ исключений [6].

Учитывая разнообразие представлений данных, используемых алгоритмов и областей применения, интеллектуальный анализ данных может осуществляться с помощью следующих классов программных продуктов [4]:

- специализированных «коробочных» программных продуктов для интеллектуального анализа;
- математических пакетов;
- электронных таблиц (и различного рода надстроек над ними);
- средств интегрированных в системы управления базами данных;
- других программных продуктов

Модели ИАД создаются путем применения алгоритмов к данным. Однако это не просто алгоритм или контейнер метаданных. Это набор данных, статистики и шаблонов, которые могут быть применены к новым данным для формирования прогноза и выводимых взаимосвязей.

Структура интеллектуального анализа данных может быть представлена как совокупность исходных данных и описания способов их обработки. Структура содержит модели, которые используются для анализа ее данных. В частности, одна структура может поддерживать несколько моделей. В структуре интеллектуального анализа данных можно выделить обучающий и проверочный набор данных, задав процентное отношение или объем данных [3].

Модель ИАД - это комбинация самих данных, алгоритма добычи данных и набора значений

параметров и фильтров, которые управляют использованием и обработкой данных. Модели ИАД определяются с помощью языка расширения Data Mining или с помощью мастера Data Mining в среде BI DevStudio [5].

Модели ИАД содержат информацию, полученную в результате статистической обработки данных, например, закономерности, обнаруженные в результате анализа.

Каждая модель добычи данных имеет свойства, которые определяют модель и ее метаданные. Эти свойства включают имя, описание, дату последней обработки модели, разрешения модели и фильтры для данных, используемых для обучения.

Каждая модель ИАД также содержит свойства, унаследованные от структуры ИАД, которые описывают колонки данных, используемые в модели. Если какие-либо столбцы, используемые в модели, являются вложенными таблицами, к ним также можно применять отдельные фильтры [5].

Кроме того, каждая модель ИАД имеет два особых свойства: Algorithm и Usage.

Свойство Algorithm определяет алгоритм, используемый для создания модели. Набор доступных алгоритмов зависит от используемого провайдера. Свойства алгоритма применяются к моделям добычи данных и могут быть заданы только один раз для каждой модели. Алгоритмы могут быть изменены позже, но некоторые столбцы модели добычи данных могут быть отключены, если выбранный алгоритм их не поддерживает.

Свойство Usage определяет, какие столбцы использует модель. Вы можете определить столбцы, которые используются для входных данных, прогнозов, только прогнозов или в качестве ключей. Свойство Usage применяется к отдельным столбцам в модели добычи данных и должно быть установлено отдельно для каждого столбца в модели. Если структура содержит столбцы, которые не используются в модели, устанавливаются пропущенные значения [5].

В процессе ИАД исследуются различные объекты (или варианты). В большинстве случаев, это можно представить в виде таблицы, где каждая строка соответствует одному из объектов, а столбцы содержат значения параметров, которые его характеризуют. Зависимая переменная - это параметр, его значения считаются зависимыми от независимых переменных. Эта зависимость должна быть выявлена при помощи методов интеллектуального анализа данных.

Методы Data Mining помогают решить многие задачи, с которыми сталкивается аналитик

- Задача классификации заключается в определении класса объекта в зависимости от его характеристик.

- Регрессионные задачи, такие как задачи классификации, могут быть использованы для определения значений некоторых параметров объекта по его известным свойствам. В отличие от задач классификации, значения параметров представляют собой не конечное множество классов, а набор действительных чисел.

- Целью поиска правил ассоциации является нахождение частных зависимостей (или ассоциаций) между объектами или событиями. Найденные зависимости отображаются в виде правил. Правила могут использоваться как для лучшего понимания природы анализируемых данных, так и для прогнозирования наступления событий.

- Задача кластеризации заключается в поиске независимых групп (кластеров) и их функций в анализируемом наборе данных. Решение этой проблемы помогает лучше понять данные. Кроме того, группировка однородных объектов может уменьшить количество объектов для облегчения анализа.

Рынок программного обеспечения для ИАД представлен широким спектром инструментов и является высококонкурентной борьбой за потребителей. Эта конкуренция создает новые, высококачественные решения. Все больше поставщиков стремятся включить в свои инструменты как можно больше современных методов и технологий. Инструменты интеллектуального анализа данных часто рассматриваются как неотъемлемая часть рынка бизнес-аналитики, который неуклонно растет.

В то же время некоторые эксперты отметили, что из-за сложности программной реализации некоторых новых теоретических разработок в области методов и алгоритмов ИАД, существующее программное обеспечение отстает от теоретических разработок.

В целом, рынок бизнес-аналитики, включая рынок инструментов для интеллектуального анализа данных, настолько широк и разнообразен, что любая компания может выбрать подходящий инструмент в соответствии со своими функциями и бюджетными возможностями.

Список литературы

1. Александр Бондарь. Microsoft SQL Server 2014. — СПб.: БХВ-Петербург, 2015. — 592 с.
2. Козлов А.Н. Интеллектуальные информационные системы: учебник /А.Н. Козлов; Мин-во с-х. РФ, ФГБОУ ВПО Пермская ГСХА. – Пермь: Изд-во ФГБОУ ВПО Пермская ГСХА, 2013 – 278 с.
3. Модели интеллектуального анализа данных (службы Analysis Services — интеллектуальный анализ данных) [Электронный ресурс]. Режим доступа: <https://docs.microsoft.com/ru-ru/analysis-services/data->

mining/mining-models-analysis-services-data-mining?view=asallproducts-allversions#bkmk_mdIDefine.
(Дата обращения: 05.06.22)

4. *Мусаев А.А.* Интеллектуальный анализ данных: учебное пособие. А.А.Мусаев – СПб.: СПбГТИ(ТУ), 2018. – 56 с.
5. *Нестеров С.А.* Базы данных. Интеллектуальный анализ данных: учеб.пособие / С. А. Нестеров – СПб.: Изд-во Политехн. ун-та, 2011. – 272 с.
6. *Чубукова И.А.* Data mining. Учебное пособие.