

image not found or type unknown



Термин интеллектуальный анализ данных можно понимать двояко.

В узком смысле это попытка адекватного русского перевода термина Data Mining, который ввёл в обиход Григорий Пятецкий-Шапиро в 1992 году. Согласно его определению, Data Mining — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных, доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Дословный перевод «раскопки (или добыча) данных» следует, по всей видимости, считать неудачным.

В широком смысле это современная концепция анализа данных, предполагающая, что

данные могут быть неточными, неполными (содержать пропуски), противоречивыми, разнородными, косвенными, и при этом иметь гигантские объёмы; поэтому понимание данных в конкретных приложениях требует значительных интеллектуальных усилий;

сами алгоритмы анализа данных могут обладать «элементами интеллекта», в частности, способностью обучаться по прецедентам, то есть делать общие выводы на основе частных наблюдений; разработка таких алгоритмов также требует значительных интеллектуальных усилий;

процессы переработки сырых данных в информацию, а информации в знания уже не могут быть выполнены по старинке вручную, и требуют нетривиальной автоматизации.

Необходимость интеллектуального анализа данных возникла в конце XX века в результате повсеместного распространения информационных технологий, позволяющих детально протоколировать процессы бизнеса и производства.

По составу решаемых задач Data Mining практически не отличается от стандартного набора средств, применяемых с середины XX века в области прикладной статистики, машинного обучения (machine learning), информационного поиска (information retrieval). Основное различие заключается в эффективности алгоритмов и технологичности их применения.

Подавляющее большинство классических процедур имеют время выполнения, квадратичное или даже кубическое по объёму исходных данных. При количестве объектов, превосходящем несколько десятков тысяч, они работают неприемлемо медленно даже на самых современных компьютерах. За последние десятилетия значительные усилия в области Data Mining были направлены на создание специализированных алгоритмов, способных выполнять те же задачи за линейное или даже логарифмическое время без существенной потери точности.

Интеллектуальный анализ данных — это направление информационных технологий, охватывающее всю область проблем, связанных с извлечением знаний из массивов данных.

Часто интеллектуальный анализ данных рассматривают как трансформацию термина Data Mining (раскопка, разработка данных) и считают два этих понятия синонимами. В широком смысле это вполне допустимо. Однако считать их абсолютно идентичными было бы неверно, поскольку методы Data Mining сосредотачиваются главным образом на процессах анализа данных и интеллектуальном моделировании. В то же время считается, что интеллектуальный анализ данных рассматривает весь спектр проблем, связанных с процессом извлечения знаний из баз данных.

В основе интеллектуального анализа данных, как и в Data Mining, лежит идея активного применения математических методов, таких как оптимизация, генетические алгоритмы, распознавание образов, статистика, Data Mining и т.д., а также использующих визуальное представление информации.

Однако интеллектуальный анализ данных также включает вопросы извлечения данных из различных источников, их консолидацию, профайлинг, трансформацию, предобработку, очистку и обогащение.

Модели интеллектуального анализа данных могут применяться в конкретных бизнес-сценариях, а именно:

Прогнозирование: оценка продаж, прогнозирование нагрузки сервера или времени простоя сервера;

Риск и вероятность: выбор наиболее подходящих заказчиков для целевой рассылки, определение точки равновесия для рискованных сценариев, назначение вероятностей диагнозам или другим результатам;

Рекомендации: определение продуктов, которые с высокой долей вероятности могут быть проданы вместе, создание кросс-пакетов;

Поиск последовательностей: анализ выбора заказчиков во время совершения покупок, прогнозирование следующего возможного события;

Группировка: разделение заказчиков или событий на кластеры связанных элементов, анализ и прогнозирование общих черт.

Источники информации:

http://www.machinelearning.ru/wiki/index.php?title=Интеллектуальный_анализ_данных

<https://wiki.loginom.ru/articles/data-analysis.html>

https://ru.m.wikipedia.org/wiki/Data_mining