Научная статья
Original article
УДК 004.8

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

INTELLIGENT DATA ANALYSIS

Бохан Петр Андреевич, магистрант, Южно-Российский институт управления — филиал федерального государственного бюджетного образовательного учреждения высшего образования «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации» (344002 Россия, Ростовская обл., г. Ростов-на-Дону, ул. Пушкинская, д. 70), тел. 8 (928) 186-68-81, v.petr100@mail.ru

Petr A. Bokhan, Master's student, South Russian Institute of Management - Branch of Federal State Educational Institution of Higher Education "Russian Presidential Academy of National Economy and Public Administration" (70 Pushkinskaya St., Rostov-on-Don, 344002 Russia, Rostov-on-Don region), tel. 8 (928) 186-68-81, v.petr100@mail.ru

Аннотация. В данной статье раскрывается понятие технологии интеллектуального анализа данных (Data Mining) как одного из трендов развития систем анализа данных. Определены основные задачи, которые способен выполнить метод Data Mining. Выявлена актуальность данной технологии и ее особенности. Раскрыты проблемы и ограничения в использования интеллектуального анализа данных в современных информационных системах. Проанализирована современная ситуация и

тенденции использования технологии интеллектуального анализа данных в различных компаниях. Предложены возможные варианты развития сфер применения метода Data Mining в связи с закономерностями трансформации современного технологического процесса.

Abstract. This article reveals the concept of intelligent data analysis technology (Data Mining) as one of the trends in the development of data analysis systems. The main tasks that can perform the method of Data Mining are defined. The relevance of this technology and its features are revealed. Revealed the problems and limitations of using data mining in modern information systems. Analyzed the current situation and trends in the use of data mining technology in various companies. Possible options for the development of the areas of application of the Data Mining method in connection with the patterns of transformation of the modern technological process are proposed.

Ключевые слова: интеллектуальный анализ данных, метод Data Mining, выявление неявных и скрытых связей и закономерностей, развитие сферы применения интеллектуального анализа данных, проблемы и ограничения использования интеллектуального анализа данных, обнаружение новых знаний.

Keywords: data mining, Data Mining, identifying implicit and hidden relationships and patterns, development of the scope of data mining, problems and limitations of using data mining, the discovery of new knowledge.

В настоящее время тенденции бурного развития всех сфер общественной жизни как никогда заметны, особенно в технологической сфере. Это связанно с огромным ростом объемов информации, которые требуют развития технологий автоматизированной и интеллектуальной обработки данных, способов нахождения необходимой информации и выявления неявных закономерностей и связей в ретроспективных данных. Новые полученные знания необходимы для принятия качественных управленческих решений, чья важность также повышается из-за усложнения общественных отношений и

систем. В связи с этим повышение надежности систем и их гибкости реагирования на неочевидные изменения в настоящее время крайне актуальны. Однако среди всего массива данных человек способен вычленить только небольшую часть новых и полезных данных. Поэтому активное использование метода интеллектуального анализа данных или Data Mining способно увеличить объем получаемых полезных знаний и выявления большего количества неявных закономерностей.

Наиболее точно раскрыл сущность понятия «Data Mining» один из основателей данного направления Григорий Пиатецкий-Шапиро. Он определил Data Mining как процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. При этом сутью и целью данного метода является выявление в большом массиве данных неявных, объективных и полезных закономерностей [1].

Основой технологии «Data Mining» является концепция шаблонов, которая представляет из себя закономерности, свойственные подвыборкам данных, которые можно выразить в понятной для человека форме. Если дословно переводить слово «Mining», то в русском языке оно будет эквивалентно словосочетанию «добыча полезных ископаемых». В свою очередь поиск связей и закономерностей в массиве информации является аналогичным процессом.

Таким образом, понятие интеллектуальный анализ данных является не конкретным методом или технологией, а самими процессом поиска закономерностей, отклонений, взаимосвязей и тенденций на основе математических и статистических алгоритмов. Данный процесс осуществляется с целью представления данных, описывающих динамику явления, и построения модели, которая способна оценить и спрогнозировать процессы, влияющие на явление [2].

Интеллектуальный анализ данных используется для решение большого количества общих задач. К ним можно отнести:

- Распознавание (классификация, диагностика) ситуаций, явлений, объектов или процессов с обоснованием решений;
- Прогнозирование ситуаций, явлений, процессов или состояний по выборкам динамических данных;
- Кластерный анализ и исследование структуры данных;
- Выявление существенных признаков и нахождение простейших описаний;
- Нахождение эмпирических закономерностей различного вида;
- Построение аналитических описаний множеств (классов) объектов;
- Нахождение нестандартных или критических случаев;
- Формирование эталонных описаний образов [3].

В свою очередь, Data Mining наиболее актуален в настоящее время в сфере управления, так как данная технология позволяет провести наиболее полный и глубокий анализ данных, что позволит выбрать наиболее эффективное и правильное управленческое решение.

Технология Data Mining, целью которой является выявление неочевидных взаимосвязей и закономерностей, представляется как более эффективный метод анализа данных по сравнению с традиционными, которые направлены больше на проверку заранее сформулированных гипотез и «грубый» анализ, являющийся основой оперативной аналитической обработки данных или «OLAP». То есть интеллектуальный анализ данных способен самостоятельно находить тенденции и закономерности, а также строить модели и гипотезы на основании выявленных взаимосвязей. Таким образом, выработка гипотез на основании полученных закономерностей и взаимосвязей является очевидным преимуществом технологии Data Mining по сравнению с традиционными методами анализа данных, которые не способны на решение подобной сложной задачи. То есть традиционные методы анализа

данных используется в основном для обработки ретроспективных данных, а метод Data Mining использует эти данные для выявления тенденций и закономерностей в изучаемом процессе [4].

Однако для эффективного использования технологии интеллектуального анализа данных необходимо прежде всего понять, какие существуют проблемы, ограничения и особенности в использовании данной технологии.

Во-первых, метод Data Mining не подменяет или заменяет аналитика. Данная технология является лишь мощным инструментом анализа данных в руках аналитика, что значительно облегчает ему работу, однако данный метод не способен самостоятельно дать ответы на те вопросы, которые не были заданы.

Во-вторых, сложность разработки и эксплуатации приложений, использующих технологию Data Mining. Интеллектуальный анализ данных является междисциплинарной областью, которая основана на прикладной статистике, искусственном интеллекте и теории баз данных, поэтому разработка программ или приложений, включающих метод Data Mining требует усилий большого множества специалистов из разных научных областей, которым необходимо качественное взаимодействие для обеспечения оптимального функционирования данной технологии. Также для эффективного использования технологии Data Mining пользователь должен обладать определенным уровнем квалификации, так как различные инструменты интеллектуального анализа данных имеют разную степень сложности интерфейса.

Еще одно ограничение использования технологии Data Mining связано с необходимостью четкого понимания сути рассматриваемых данных, так как без этого невозможно получение полезной информации. То есть специалисту необходимо тщательно осуществлять выбор модели и возможную интерпретацию найденных закономерностей или связей. Выработанная модель должна быть грамотно интегрирована в динамику развития

рассматриваемого явления с целью осуществления оценки полученной модели.

Большое количество ложных или недостоверных результатов также является явной проблемной зоной технологии Data Mining. Интеллектуальный анализ данных позволяет получить действительно ценную информацию, на основе которой можно получить наиболее эффективное решение. Однако специалисты, использующую данную технологию, отмечают также наличие большого процента ложной или недостоверной информации. Поэтому полученные данные, закономерности, тенденции и выработанные на основе них модели необходимо проверять на адекватность с помощью тестовых ситуаций [4].

В свою очередь, перед специалистами в области интеллектуального анализа данных стоит множество непростых вопрос и задач из-за многообразия данных и подходов к их анализу. При этом важным направлением деятельности специалистов в сфере Data Mining является поиск эффективных метод применения данной технологии в приложениях и программных обеспечениях. Это связано с постоянно возрастающей тенденцией необходимости использования интеллектуального анализа данных с целью решения сложных прикладных задач, поэтому расширение сферы применения метода Data Mining является объективной закономерностью. В частности, данная технология активно развивается в следующих сферах:

Интеллектуальный анализ пространственно-временных данных и движущихся объектов – использование методов Data Mining в данной сфере связанно с развитием технологий обнаружения местоположения, а также датчиков, сенсоров и других устройств, обладающих возможностью беспроводной передачи данных. Очевидно, что возникает вопрос о своевременности и качестве извлекаемых знаний с помощью подобных технологий, но бурное развитие оборудования, предназначенного для

беспроводной связи, свидетельствует о необходимости внедрения технологии Data Mining для их более эффективного использования;

Интеллектуальный анализ биологических и биомедицинских данных – актуальность данного направления связано с уникальным сочетанием сложности, важности и богатства биологических и биомедицинских данных. Метод Data Mining может использоваться как для анализа белков и ДНК с целью выявления неявных закономерностей их развития и создания моделей их жизненного цикла, так и для анализа большого объема специализированной литературы и исследований с целью выявления общих законов взаимодействия живых организмов или консолидации биологической информации;

Интеллектуальный анализ данных, связанных с разработкой программного обеспечения и проектирования систем – современное технологическое развитие обуславливает тенденцию к усложнению информационных систем, интегрирующие в себя большое количество разнородных компонентов. Данное усложнение информационных систем привело к необходимости развитие программного обеспечения в сторону усиления его надежности и отказоустойчивости. В свою очередь, процессы, которые анализируют появление ошибок в информационных системах, являются по своей сути технологией интеллектуального анализа данных. То есть во время работы программы выдает большое количество данных о собственном состоянии и происходящих процессах. А эти данные необходимо всесторонне анализировать с целью выявления неявных закономерностей, которые могут привести к неадекватному функционированию или отказу работы программы. Развитие технологии Data Mining в данной сфере позволит иметь тенденцию к созданию полностью автоматизированной программы по обнаружению ошибок в программном обеспечении, что значительно скажется на надежности и отказоустойчивости любой информационной системы;

Интеллектуальный анализ данных в реальном времени и развитие распределенного интеллектуального анализа данных – традиционные метода

анализа данных предназначены для работы с централизованном размещении. Однако данные методы показывают свою недостаточную эффективность при анализе данных в распределенных вычислительных средах, таких как беспроводные сети и облачные среды. То есть развитие технологии Data Mining с целью анализа данных в распределенных системах имеет явный потенциал, прежде всего это связано с необходимостью анализа данных и построению адекватных моделей в режиме реального времени для большинства современных приложений и программного обеспечения.

Представленные выше тенденции развития использования технологии интеллектуального анализа данных и расширение сфер его использования опирается на современный объем использования Data Mining и его эффективность для компаний, использующих его.

Например, российская компания «Билайн» в 2019 году активно внедряла технологию Data Mining для сегментации клиентов и их защиты от различного рода махинаций. Благодаря технологическому нововведению компания смогла получить 20 % дополнительной прибыли только за счет использования технологии интеллектуального анализа данных. Также «Сбербанк», используя систему биометрии с целью защиты данных потребителей, в основу которой лежит метод Data Mining, в период с 2018 по 2021 год смогли сократить объем мошенничества с документами в 10 раз [5].

Еще ярким примером использования технологии Data Mining является сеть гипермаркетов «Лента». Проведя анализ с помощью данного метода карт лояльности своих покупателей, сеть значительно смогла оптимизировать ассортимент и управлять выкладкой и ценами. В настоящее время также активно используется технология интеллектуального анализа данных в системах, которые предназначены для обработки запросов потребителей и предложения товаров и услуг, которые с большей вероятностью захотят приобрести покупатели. Данные системы построены на методе Data Mining, который анализирует данные в режиме реального времени. Так работают

голосовые помощники Alexa от Amazon, Siri от Apple и «Алиса» от «Яндекса», которые 60 % запросов обрабатывают на основе метода Data Mining.

Представленные примеры демонстрируют постоянный рост рынка Data Mining. Этому способствует деятельность крупных корпораций: SAS, IBM, Microsoft, Oracle и других. По мнению специалистов, к 2027 году тенденция использования технологий интеллектуального анализа данных вырастет на 23,1 % и достигнет $ 56,2 млдр. Данная тенденция более детализировано представлена на рис. 1.



Рис. 1. Тенденция развития технологии интеллектуального объема данных [6].

Таким образом, постоянное и непрерывное развитие технического процесса позволяет вырабатывать большое количество данных, которые необходимо анализировать с целью поиска неявных и скрытых связей и закономерностей для выявления новых полезных знаний. И в этом немалую роль играет технология интеллектуального анализа данных. Как и любой другой метод анализа данных, Data Mining, обладает своей спецификой, возможными проблемами и ограничениями в использовании. Однако несмотря на это расширение области применения данной технологии имеет положительную тенденцию. Наиболее перспективными направления развития технологии интеллектуального анализа данных является использование Data Mining в социальных и информационных сетях, биологии и медицине,

разработке программного обеспечения и анализе облачных и беспроводных сетей. Так как в этих направлениях метод Data Mining обладает наибольшим потенциалом практического использования.

Литература

1. Frawley W., Piatetsky-Shapiro G., Matheus C. Knowledge Discovery in Databases: An Overview. – AI Magazine. – 1992. – p. 220.
2. Интуит // Что такое Data Mining? [Электронный ресурс]. Режим доступа — https://intuit.ru/studies/courses/6/6/lecture/158?page=2 (Дата обращения: 25.03.2022).
3. Fetmag // Проблемы, тенденции развития и перспективные направления применения технологий Data Mining [Электронный ресурс]. Режим доступа — http://fetmag.mrsu.ru/2011-2/pdf/applications_of_data_mining.pdf (Дата обращения: 25.03.2022).
4. DocPlayer // Введение в Data Mining [Электронный ресурс]. Режим доступа — https://docplayer.com/37038386-Razdel-1-vvedenie-v-data-mining.html (Дата обращения: 25.03.2022).
5. Aiconference // Data Mining: возможности и применение [Электронный ресурс]. Режим доступа — https://aiconference.ru/ru/article/data-mining-vozmognosti-i-primenenie-86170 (Дата обращения: 25.03.2022).
6. РБК * Тренды // Банки, ретейл, медицина: кто использует Data Mining и для чего [Электронный ресурс]. Режим доступа — https://trends.rbc.ru/trends/industry/61b359739a7947c7376ef7ce (Дата обращения: 25.03.2022).

References

1. Frawley W., Piatetsky-Shapiro G., Matheus C. Knowledge Discovery in Databases: An Overview. - AI Magazine. - 1992. - p. 220.
2. Intuit // What is Data Mining? [Electronic resource]. Access mode - https://intuit.ru/studies/courses/6/6/lecture/158?page=2 (Date of access: 25.03.2022).

3. Fetmag // Problems, development trends and promising applications of Data Mining technologies [Electronic resource]. Access mode - http://fetmag.mrsu.ru/2011-2/pdf/applications_of_data_mining.pdf (Date of access: 25.03.2022).
4. DocPlayer // Introduction to Data Mining [Electronic resource]. Access mode - https://docplayer.com/37038386-Razdel-1-vvedenie-v-data-mining.html (Date of access: 25.03.2022).
5. Aiconference // Data Mining: possibilities and application [Electronic resource]. Access mode - https://aiconference.ru/ru/article/data-mining-vozmognosti-i-primenenie-86170 (Date of access: 25.03.2022).
6. RBC* Trends // Banks, retail, medicine: who uses Data Mining and what for [Electronic resource]. Access mode - https://trends.rbc.ru/trends/industry/61b359739a7947c7376ef7ce (Access date: 25.03.2022).

© Бохан П.А. 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022

Для цитирования: Бохан П.А. Интеллектуальный анализ данных // Научно-образовательный журнал для студентов и преподавателей «StudNet» №6/2022