

Использование технологий Data Mining для обучения младших школьников поиску информации

Симдянкина Елена Константиновна

*Приамурский государственный университет им. Шолом-Алейхема
студент*

Коняхин Александр Олегович

*Приамурский государственный университет им. Шолом-Алейхема
студент*

Лучанинов Дмитрий Васильевич

*Приамурский государственный университет им. Шолом-Алейхема
старший преподаватель кафедры информационных систем, математики и
методик обучения*

Аннотация

В статье раскрыта сущность технологии Data Mining, представлено описание способов использования данной технологии для обучения младших школьников поиску информации, рассмотрены направления Data Mining, применяемые к анализу различных видов информации, описаны этапы интеллектуального анализа данных.

Ключевые слова: информационные технологии, технология Data Mining, учебный процесс, интеллектуальный анализ данных.

Use of Data Mining technologies for teaching younger students information retrieval

Simdyankina Elena Konstantinovna

*Sholom-Aleichem Priamursky State University
student*

Konyakhin Aleksandr Olegovich

*Sholom-Aleichem Priamursky State University
student*

Luchaninov Dmitry Vasilyevich

*Sholom-Aleichem Priamursky State University
Senior lecturer of the Department of Information Systems, Mathematics and
teaching methods*

Abstract

The article describes the essence of Data Mining technology, describes the ways in which this technology can be used to teach junior students information retrieval, examines the directions of Data Mining, applied to the analysis of various types of information, describes the stages of data mining.

Keywords: Information technology, Data Mining technology, educational process, data mining.

Вследствие развития информационных технологий, объем информации, накопленной человечеством в электронном виде, очень быстро растет. Эта информация существует вокруг нас в различных видах: тексты, аудио, изображения, видео, реляционные базы данных, гипертекстовые документы и т.д. В результате постоянного использования людьми сети Интернет, облегчающей доступ к информации появилось огромное количество различных данных. Но, несмотря на это большее количество предоставляемой информации не несет какой-либо пользы для конкретного человека. Появляется проблема извлечения полезной для человека информации из ее общего количества.

На современном этапе развития элементы искусственного интеллекта являются активно внедряющимися в нашу практическую деятельность. Отличительной чертой технологии интеллектуального поиска и анализа данных Data Mining от традиционных систем искусственного интеллекта, является не попытка моделирования естественного интеллекта, а усиление его возможностей мощностью современных вычислительных серверов, хранилищ данных и поисковых систем.

Data Mining это процесс поддержки принятия решений, который основан на поиске данных в скрытых закономерностях (шаблонах информации) [5].

Достаточно точное определение технологии Data Mining дает один из основателей этого направления Григорий Пиатецкий-Шапиро. Он определяет Data Mining в качестве процесса нахождения в сырых данных ранее нетривиальных, неизвестных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах деятельности человека.

В настоящее время методы «Data Mining» используются в различных сферах человеческой деятельности. Исследованиями в этой области занимаются такие ученые, как А.А. Барсегян, М.С. Куприянов, Г. Пятецкий-Шапиро, Х. Ромесбург, Дж. Хан [3, 7]. Проблемы анализа данных в образовательном процессе рассматривали в своих работах таких ученых, как Р. Бакер, Л.И. Григорьев [4] и другие.

Современные изменения в сфере образования требуют применения и новых методов анализа для подготовки организационных и управленческих решений, соответствующих современным задачам, поставленным перед учителями. Исходя из этого использование информационно-аналитического

обеспечения выходит на первый план в решении проблемы модернизации качества образования.

Одной из причин развития электронного обучения считается внедрение интерактивных систем обучения. Взаимодействие таких систем с обучаемыми вызывает огромное количество данных, которое может быть использовано для изменения процесса обучения. Способом решения этой задачи могут стать методы Educational Data Mining, занимающиеся разработкой способов для изучения данных, которые возникают в контексте образования. Educational Data Mining использует различные методы, к которым относятся как стандартные методы Data Mining, такие как классификация, кластеризация, регрессия, визуализации, корреляция и др., так и ряд специфических методов, например, из области психометрики.

Работа Educational Data Mining направлена на конкретизацию области источника крупных данных для их обработки и её целью является поиск паттернов (шаблонов, образцов, закономерных регулярностей, схем), которые встречаются в процессе образования. Б. Браун [1], утверждает, что такие данные позволяют влиять на успеваемость учащихся и способствуют лучшей организации учебного процесса. Ф. Кастро [2] высказывает мнение о том, что применение технологии Data Mining способно разнообразить методики обучения и делает учебный процесс более интересным и продуктивным.

Сообщества Educational Data Mining и Learning Analytics, преследуют примерно одинаковые цели, но используют различные подходы к их достижению и показывают неодинаковое понимание образовательного процесса. Группа Learning Analytics смотрит на учебный процесс целостно и системно, а Educational Data Mining – более упрощенно и ориентируется в основном на анализ больших данных.

Процесс обработки данных состоит из следующих этапов:

1. Постановка цели и задач.
2. Приведение данных к форме, которая пригодна для использования конкретных методов Educational Data Mining и Learning Analytics.
3. Непосредственно применение методов Educational Data Mining и Learning Analytics.
4. Проверка созданных моделей.
5. Интерпретация готовых моделей человеком с целью их применения для принятия решений.

В Educational Data Mining выделяют две основные цели: прикладная – улучшение процесса обучения и определение направления для обучаемых и фундаментальная – более глубокое понимание процесса познания. Эти параметры сложно измерить количественно и они требуют собственного специального набора метрик

Технология Data Mining подразделяется на несколько направлений, которые, в свою очередь работают с информацией определенных видов: Web Mining предназначено для исследования и извлечения информации из Web-документов и сервисов. Для анализа текстовой информации используются

методы Text Mining. A Multimedia Mining анализирует изображения, видео и аудиоинформацию.

Технология Text Mining относится к методам анализа неструктурированного текста. Примерами документов, содержащих такой текст, являются электронная почта, Web-страницы, нормативные документы и т.п. Также эти документы могут содержать не только текст, но и графическую информацию.

Процесс анализа текстовых документов состоит из следующих шагов:

1. Поиск информации. Здесь является необходимым определить, какие документы должны быть подвергнуты анализу, и обеспечить их доступность. При большом количестве документов необходимо применять варианты автоматического отбора по заданным параметрам.

2. Преобразование документов для представления их в виде, с которым работают методы Text Mining. Происходит удаление лишних слов и придание тексту более строгой формы.

3. Извлечение информации из выбранных документов, предполагает нахождение в них ключевых понятий, которые в дальнейшем будут подвергнуты анализу.

4. Применение методов Text Mining. На данном шаге извлекаются шаблоны и отношения, которые имеются в текстах. Этот шаг является основным в процессе анализа текстов.

5. Интерпретация результатов. Как правило, это заключается или в представлении результатов на естественном языке, или в их графическом виде [6].

Визуализация также может быть использована как средство анализа текста. Для этого извлекаются ключевые понятия, которые и представляются в графическом виде. Такой подход помогает пользователю быстро идентифицировать главные темы и понятия, а также определить их важность.

Технология Web Mining способствует облегчению поиска информации в глобальной сети Интернет.

Данная технология автоматически подстраивается под конкретного пользователя путем анализа его запросов и того, как он воспринимает полученную информацию, исходя из этого, система выдает информацию необходимую только данному пользователю.

К этапам применения Web Mining относятся:

1. Поиск ресурсов неизвестных документов и сервисов в Web-среде.

2. Автоматическое извлечение определенной информации из найденных Web-ресурсов.

3. Выявление общих шаблонов в отдельных и пересекающихся множествах сайтов.

4. Интерпретация найденных шаблонов пользователю[6].

Таким образом, интеллектуальные технологии Data Mining представляют собой удобный способ использования знаний, неявно содержащихся в огромных массивах накопленной человечеством информации. Обучать младших школьников поиску информации с помощью

технологии Data Mining необходимо для того, чтобы они могли с легкостью ориентироваться в огромных потоках информации и выбирать нужную и актуальную на данный момент.

Библиографический список

1. Brown B., Manyika J., Chui M. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, 2011.
2. Castro F. Applying Data Mining Techniques to e-Learning Problems / F. Castro, A. Vellido, A. Nebot, F. Mugica // Studies in Computational Intelligence. 2007. Т. 62. P. 183-221.
3. Барсегян А.А. Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. СПб.: БХВ-Петербург, 2009. С. 512.
4. Григорьев Л.И. Научно-методические и технологические основы информационной системы управления качеством учебного процесса. М.: Нефть и газ, РГУ нефти и газа им. И.М. Губкина, 2008. 132 с.
5. Константиновский Д.Л., Агранович М.Л., Дымарская О.Я. От сбора статистических данных – к информационному обеспечению принятия решений. М.: Логос, 2006. 160 с.
6. Маннинг К. Введение в информационный поиск / К. Маннинг, П. Рагхаван, Х. Шютце. М.: Вильямс, 2011. 528 с.
7. Пятецкий-Шапиро Г. Data Mining и перегрузка информацией: Вступительная статья // Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров и др.. СПб.: БХВ-Петербург, 2009. С. 13–14.