

К. АТТО

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина), Санкт-Петербург

АДАПТАЦИЯ АЛГОРИТМОВ АНАЛИЗА ДАННЫХ ДЛЯ ИСПОЛЬЗОВАНИЯ В ИНТЕЛЛЕКТУАЛЬНОЙ СРЕДЕ ОБУЧЕНИЯ

В рамках предыдущих исследований, связанных с использованием CRM и интеллектуальных агентов в системах управления обучением, была предложена оптимальная модель коммуникации в интеллектуальной среде обучения, которая позволяет улучшить качество образовательного процесса с точки зрения своевременного реагирования и полноты ответов в соответствии с запросами учащихся на персональные рекомендации относительно развития траекторий обучения. В продолжение разработки модели предполагается формализовать набор задач, которые могут быть эффективно решены с помощью алгоритмов анализа данных, перечислить источники анализируемой информации и ожидаемые результаты.

Введение. Рост числа обучающихся и усиление конкуренции на рынке высшего образования вынуждает университеты использовать более сложные инструменты для повышения качества обучения. К сожалению, многие из них используются исключительно с целью доставки учебного материала до пользователя, но недостаточно вовлечены в сбор и обработку обратной связи от пользователя. Помимо непосредственной передачи знаний, LMS (learning management system, система управления обучением) следует использовать для автоматизации процесса оценки успеваемости обучающегося, результаты которой будут использованы в двух направлениях: выдача рекомендаций для корректировки траектории обучения, а также корректировки самих обучающих материалов. В ходе предыдущего исследования [1] по данной тематике была предложена оптимальная модель коммуникации в интеллектуальной среде обучения, также проведено ее моделирование с целью подтверждения эффективности ее использования как ассистента обучающегося. Предлагаемый доклад посвящен поиску и детализации описания алгоритмов анализа данных, которые могут быть использованы в рамках предложенной ранее модели коммуникации.

Общая задача анализа данных в электронной среде обучения заключается в том, чтобы определить связь между подходом студента к получению образовательной информации и результатами его тестирования с двумя целями [2]:

1. Оценка пригодности доступных студентам учебных материалов.
2. Выдача рекомендаций обучающимся.

На рис. 1 приведена оптимальная модель коммуникации в интеллектуальной среде. На ней показаны каналы взаимодействия между участниками образовательного процесса, системой управления обучением, а также базовые внутренние процессы самой системы. Красным цветом отмечены процессы и подсистемы, связанные с получением образовательных материалов, зеленым – непосредственно с тестированием.

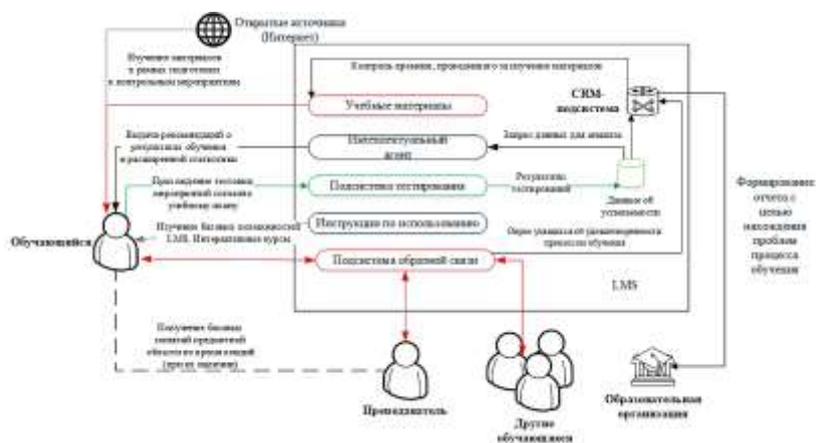


Рис 1. Оптимальная модель коммуникации в интеллектуальной среде

Перейдем к описанию процесса оценки пригодности пособий, доступных в рамках LMS. Перечислим основные каналы получения образовательного материала:

1. Материалы, доступные в LMS.
2. Общение на профильных разделах форума с преподавателями и другими студентами.
3. Самостоятельный поиск материалов в Интернете.
4. Общение с преподавателем в ходе лекций.

Очевидно, что активность через источники 3 и 4 не может быть отслежена силами LMS, поэтому она не будет учитываться в рамках непрерывного анализа, но будут участвовать в опросах учащихся по итогам семестра. Наиболее приоритетным источником является первый, поскольку, как было сказано ранее, одна из целей анализа – усовершенствование доступных материалов. Источник 2 учитывается как побочный, с целью изучения активности конкретного студента: как «общей», т. е. количества сообщений, активных тем, так и «полезной» – сообщений, отмеченных другими пользователями/модератором статусами «полезное» или «правильный ответ».

Для оценки того, насколько полезным для студента в рамках определенного курса (раздела курса) был учебный материал, введем показатель его «результативности» P

$$P(t) = \frac{t \times a}{q_{\Sigma}}, \quad (1)$$

где: t – время, потраченное на изучение материала, a – количество правильных ответов, q – количество вопросов.

Для получения более точных результатов требуется выполнить следующие условия:

1. Материалы предварительно разделены на группы по дисциплинам (может выполняться как с помощью ручного назначения категорий, так и при помощи алгоритмов классификации с использованием словарей ключевых слов).
2. Из оценок удалены очевидные промахи, для нивелирования влияния «случайного открытия» материалов вводится минимальный период сессии, во время которой материал открыт, и студент с ним взаимодействует (выполняется навигация по содержимому), который учитывается при анализе.
3. Тесты составлены так, что количество вопросов по теме соответствует количеству учебных часов, отведенных на нее в рабочей программе.

Очевидно, что улучшение списка материалов должно проводиться после составления рейтинга имеющихся, т. е. требуется выполнить их классификацию. Предлагается ввести три класса материалов относительно степени их пригодности: «не требующие доработки», «возможно требующие доработки», «точно требующие доработки». Входными данными для классификатора является признаковое описание – набор характеристик, учитываемых в показателе P . В силу того, что достаточно трудно формализовать показатели и описание «плохого» и «хорошего» учебного материала, оптимальным выглядит использование алгоритма вероятностной классификации.

В основе большинства вероятностных методов классификации лежит использование байесовского подхода [3]. Рассмотрим, как в рамках него производится задание классификатора. Процедурой классификации называется отображение $\delta(X): \Pi \rightarrow \{\omega_1, \dots, \omega_k\}$, где Π – пространство наблюдений, k – количество классов. Введем показатель C , который описывает общие потери при классификации Π .

$$C = \sum_{i=1}^k \sum_{j=1}^k c(j|i) \cdot m(j|i) \quad (2)$$

где $m(j|i)$ – число объектов из класса i , отнесенных к классу j , $c(j|i)$ – потери, которые повлечет отнесение одного объекта i -го класса к классу j ($c_{ij} = 0$ при $i = j$).

Процедура классификации называется оптимальной (байесовской), если она сопровождается минимальными потерями C среди всех других процедур классификации. Для классифицируемого объекта вычисляются функции правдоподобия каждого из классов, по ним вычисляются апостериорные вероятности классов. Объект относится к тому классу, для которого апостериорная вероятность максимальна. Для более точной классификации предлагается ввести дополнительные ограничения:

- использовать критический порог принадлежности, например, 0.7, который будет говорить об однозначной классификации элемента в выбранное множество, остальные должны

или оцениваться вручную или сравниваться с результатами работы дополнительного классификатора;

- использовать медианный подход для уменьшения количества граничных случаев: производить несколько прогонов классификатора с незначительно отличающимися целевыми показателями.

В связи с этим, в качестве первоначального алгоритма предлагается воспользоваться наивным байесовским классификатором, поскольку он обеспечивает хорошие результаты на малом числе данных и прост, значит имеет скорость выполнения, достаточную, чтобы обеспечить приемлемые временные затраты при медианном подходе.

После того, как завершена задача классификации учебных материалов по степени их полезности, можно переходить к выдаче рекомендаций студентам. Поскольку классификация результата студентов производится по одному показателю – результату тестирования, а рекомендуемые материалы сгруппированы заранее, то можно говорить о том, что данные хорошо структурированы, и воспользоваться деревом принятия решений.

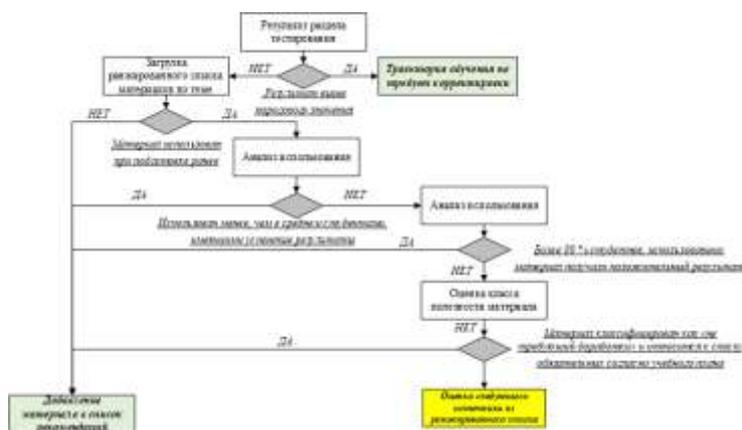


Рис 2. Дерево принятия решения формирования списка рекомендаций обучающегося

Заключение. Использование анализа данных образовательного процесса полезно для всех участников: образовательная организация повышает эффективность передачи знаний, а значит свою конкурентоспособность, студенты получают услугу более высокого качества, которая использует персонализированный подход. Препятствием, которое отсрочивает более широкое использование таких технологий на практике является высокая затратность первоначального запуска, поскольку требуются как модернизация информационных систем, так и повсеместная стандартизация образовательных процессов.

ЛИТЕРАТУРА

1. **Атто К., Котова Е.Е.** Управление коммуникативными стратегиями в электронной среде обучения // Материалы III Международной научной конференции по проблемам управления в технических системах (ПУТС-2019), Санкт-Петербург, с. 184–187. DOI: 10.1109/CTS48763.2019.8973279
2. **V. Juarez Santiago, J.M. Olivares Ramirez, J. Rodriguez-Resendiz, A. Dector, R. Garcia, J. Gonzalez-Duran Fermin, F. Sanchez,** Learning Management System-Based Evaluation to Determine Academic Efficiency Performance, Sustainability Journal, 2020, vol. 12(4256), DOI: 10.3390/su12104256
3. **Васильев В.Г., Кривенко М.П.** Методы автоматизированной обработки текстов, М: ИПИ РАН, 2008, 305 с.

K. Atto (Saint Petersburg Electrotechnical University “LETI”, Saint-Petersburg)

Adaptation of Data Analysis Algorithms for Use in the Intellectual Learning Environment

In the framework of previous studies related to the use of CRM and intelligent agents in learning management systems, an optimal communication model in an intelligent learning environment was proposed. It allows improving the quality of the educational process in terms of responses' time and completeness in accordance with students' requests for personal recommendations regarding the development of learning trajectories. In the continuation of the development of the model, it is supposed to formalize a set of tasks that can be effectively solved using data analysis algorithms, describe the sources of the analyzed information and the expected results.