

Задачи и методы интеллектуального анализа образовательных данных для поддержки принятия решений

Горлушкина Наталия Николаевна
кандидат технических наук, доцент, заведующая кафедрой интеллектуальных технологий в гуманитарной сфере,
Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики,
Кронверкский пр., д.49, г. Санкт-Петербург, 197101; тел. (812)2328645;
nagor.spb@mail.ru

Коцюба Игорь Юрьевич
аспирант кафедры интеллектуальных технологий в гуманитарной сфере,
Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики,
Кронверкский пр., д.49, г. Санкт-Петербург, 197101, тел. (812)2328645;
igor.kotciuba@gmail.com

Хлопотов Максим Валерьевич
старший преподаватель кафедры интеллектуальных технологий в гуманитарной сфере,
Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики,
Кронверкский пр., д.49, г. Санкт-Петербург, 197101; тел. (812)2328645;
hlopotov@list.ru

Аннотация

В статье обосновывается актуальность применения методов интеллектуального анализа данных, возникающих в рамках образовательного процесса, для поддержки принятия решений по управлению образовательным процессом. Подробно изложены основные задачи и методы интеллектуального анализа данных в сфере образования, а также дан обзор основных работ, посвященные рассматриваемому направлению. Предложено программное приложение, позволяющее проводить анализ данных с использованием сложных алгоритмов классификации. Дано описание работы приложения и возможных способах его применения. Приведенные методы могут найти широкое применение в системах поддержки принятия решений на всех уровнях управления образовательным процессом.

This article explains how to use methods of educational data mining for decision support during educational process' management. The article also shows main purposes and methods of educational data mining and main articles devoted to this scientific direction. A software being able to analyze data with complex classification algorithms is offered. Working process of the software and possible ways of its application are described. These methods can be used in systems of support decision making in different levels of educational process' management.

Ключевые слова

интеллектуальный анализ данных, образовательный процесс, поддержка принятия решений, алгоритмы классификации
data mining, educational process, decision support, classification algorithms

Введение

Всегда, а особенно в современном мире, образование рассматривается как важнейший компонент экономического роста и развития государства. Его роль для человека и общества постоянно возрастает. Современный человек живет и действует в условиях, требующих высокого профессионализма и значительных интеллектуальных усилий для принятия правильных решений в различных жизненных и профессиональных ситуациях. Увеличение информационных потоков, их разносторонность, явный недостаток времени на их осмысление, возросшие конкурентность.

Принятие решений в сфере образования – сложный, многоаспектный процесс, в который вовлечен большой круг заинтересованных лиц. Немаловажным критерием принятия эффективного решения является анализ информации, поступающей от участников образовательного процесса на его различных этапах.

В течение длительного периода в информационных системах образовательных учреждений накапливалась информация о студентах и их успеваемости, о преподавателях и их научно-образовательной работе. Создавались дистанционные курсы обучения, компьютерные средства обучения, образовательные форумы, системы тестирования и анкетирования студентов, и многое другое. Таким образом, за последние годы была накоплена масса данных, представленных в различных форматах, в той или иной степени, имеющих отношение к тем или иным аспектам образовательного процесса.

Также, как указано в работе [1], к компьютерным средствам обучения предъявляются требования обеспечения двухстороннего семантического диалога с обучающимися и в результате этого диалога должно создаваться адресное содержание обучения. Следовательно, содержание обучения должно полностью соответствовать дидактической ситуации, сложившейся здесь и сейчас в отношении конкретного ученика, а в этом как раз и может помочь интеллектуальный анализ образовательной информации.

Теоретические основы и особенности интеллектуального анализа образовательных данных

Данные, накопленные о разных аспектах работы образовательного учреждения, могут быть использованы для принятия решений по управлению образовательным процессом. Исследования [2,3] показывают, что большое влияние на процесс принятия решений в сфере образования оказывают экспертные оценки ответственных лиц (в том числе, заведующих кафедрами), а также опыт и интуиция преподавателей, основывающихся на субъективных мнениях о месте, роли учебных дисциплин в процессе подготовки выпускника вуза, участвующих в формировании учебных планов и так далее. Таким образом, лицо, принимающее решение, несет большую ответственность за его последствия. Чтобы сделать процесс принятия решения более объективным, необходимо пользоваться развитым специализированным научным аппаратом, реализованным в системах поддержки принятия решений.

Вообще анализ образовательных данных – это не новое явление, это давняя практика. Однако, в связи с ростом применения информационных технологий в образовании возник интерес к новым методикам и подходам, к автоматизированному выявлению новых, порою скрытых, взаимосвязей в данных и их интерпретации в интересах лица, принимающего решение, а также всех участников образовательного процесса.

Существует множество задач, в которых методы статистики, машинного обучения и извлечения знаний полезны для всех участников образовательного процесса: обучающихся, преподавателей и административных работников сферы образования. Большинство исследований в этой области предполагает использование

учебном процессе так называемых интеллектуальных обучающих систем (Intelligent Tutoring Systems, ITS) [4, 5]. Однако, несмотря на разработки и применение ITS педагогами-исследователями [6, 7], подобные системы пока ещё не получили столь широкого распространения в образовательных учреждениях России, в отличие от информационных систем управления и систем дистанционного обучения (Learning Management System, LMS). В связи с этим наибольший интерес представляет анализ образовательных данных, полученных из этих источников, а не из интеллектуальных обучающих систем.

Вообще термин «data mining» (интеллектуальный анализ данных) подразумевает выявление новых, нетривиальных и практически полезных знаний, необходимых для принятия решений в различных областях деятельности. Educational data mining (EDM) – это направление в рамках интеллектуального анализа данных. Целью EDM является обработка и анализ данных, полученных в рамках образовательного процесса.

Одна из основных задач EDM – нахождение скрытых закономерностей в данных. Таким образом, суть EDM – это методы, инструменты и исследования, связанные с автоматическим извлечением данных, относящихся к учебной деятельности.

Учитывая сказанное выше, дадим определение. Интеллектуальный анализ образовательных данных – это совокупность методов обнаружения в данных ранее неизвестных, практически полезных и интерпретируемых знаний об образовательном процессе и его участниках с целью поддержки принятия решений.

Наиболее характерными из задач интеллектуального анализа образовательных данных являются следующие [8]:

- мониторинг сформированности профессиональных компетенций;
- проектирование учебных планов, индивидуальных учебных планов, программ учебных дисциплин;
- анализ и прогнозирование повышения конкурентоспособности студентов на рынке труда;
- прогнозирование и проектирование тех качеств выпускника, которые предполагается получить «на выходе» образовательного процесса;
- диагностика уровня качества образования для своевременной компенсации нежелательных отклонений;
- оценка реального качества образования на его соответствие стандартам.

Основные цели EDM:

- улучшить образовательный процесс путем поддержки принятия рациональных решений;
- направить студентов по целесообразной образовательной траектории;
- дать рекомендации студентам и преподавателям по корректировке образовательного процесса;
- вникнуть в самую суть учебного процесса — выявить неявные взаимосвязи данных и, как следствие, понять, каким образом человек усваивает информацию, приобретает навыки и умения.

Интеллектуальный анализ образовательных данных — сравнительно новое направление исследований. В 2008 году группа учёных впервые провела конференцию по EDM в канадском Монреале. С тех пор конференция стала ежегодной. Конференции проводились в США, Испании, Греции. В следующем году состоится уже седьмая конференция, которая на этот раз пройдёт в Institute of Education (Лондон, Великобритания) [9]. С 2010 года издаётся журнал «Educational data mining» [10]. С октября по декабрь 2013 года на сайте Coursera Райан Бейкер (Ryan Baker), профессор Колумбийского университета, являющийся одним из ведущих учёных в области EDM, провёл курсы под названием «Big Data in

Education» [11]. Основное содержание курса посвящено вопросам использования методов математической статистики и интеллектуального анализа данных в образовании.

Для активных научных исследований в области EDM в России необходимо решение следующих задач:

- проанализировать основные процедуры, методы и алгоритмы EDM (их сходство, различия, назначение);
- рассмотреть области применения, то есть возникающие в образовательном процессе задачи, которые могут быть решены при помощи EDM, с учетом российской специфики.

Основные методы интеллектуального анализа образовательных данных базируются на статистике, машинном обучении и теории баз данных [8]. К этим методам относятся:

- поиск аномалий (Anomaly detection);
- поиск ассоциаций (Association rule learning);
- классификация (Classification);
- кластерный анализ (Clustering);
- регрессионный анализ (Regression Analysis);
- факторный анализ (Factor Analysis);
- нейронные сети (Artificial neural network) и др.

В рамках перечисленных методов реализовано большое количество подходов, базирующихся на различных алгоритмах. Эффективность работы по анализу образовательных данных во многом зависит от представления анализируемых объектов в виде системы показателей. К ведущим принципам методологии педагогических измерений, являющейся ядром такого системного подхода, в том числе, относят [8]:

- сочетание количественного и качественного уровня измерения;
- привлечение психодиагностических методик для измерения изменений качества образования;
- использование факторного, дисперсионного и метаанализа для изучения данных о качестве подготовки студентов и т.д.

Отметим, что одной из отличительных особенностей интеллектуального анализа данных образовательного процесса является то, что, помимо базовых методов и алгоритмов анализа данных, в EDM используются и некоторые специальные методы. В частности, для описания особенностей и характеристик среды обучения применяются психометрические методы. Таким образом, интеллектуальный анализ образовательных данных – это область, тесно связанная с педагогической психологией. Помимо этого, здесь применимы научные методы, такие, как когнитивные, объединяющие подходы из психологии и математический аппарат, позволяющие формализовать мнения экспертов в виде формальной и субъективной модели, что позволит проанализировать их с помощью математических методов.

Как показывает анализ [12], в современной психолого-педагогической литературе большое внимание уделяется выявлению факторов, влияющих на успеваемость студентов. Для получения численных характеристик влияния изучаемых параметров на успеваемость применяются методы статистического анализа. Выводы, получаемые в результате статистического анализа, нередко содержат проблемы, связанные с тем, что зачастую результаты, полученные при тестировании определённой выборки студентов, оказываются недействительными даже для отдельных студентов той же выборки. Это приводит к необходимости выявить возможные факторы, влияющие на ограничение адекватности исследований,

а также обращать внимание на специфику анализируемых данных и размер выборок из генеральной совокупности, на которых он будет проводиться.

Интеллектуальный анализ может быть проведен для различных данных, имеющих отношение к образовательному процессу. Данные, необходимые для анализа, могут быть взяты из баз данных университетов, в которых хранится информация о студентах, преподавателях, успеваемости и т. д. К тому же, могут быть собраны данные от работодателей о рынке труда, данные о личностных качествах студентов, данные из приёмной комиссии и т.д.

Данные, получаемые из образовательного процесса (например, уровень школьников, абитуриентов, студентов, учителей и преподавателей, учебного заведения, образовательной системы региона и страны в целом) должны быть правильным образом охарактеризованы и зафиксированы в определенных шкалах измерения. Информация в таком случае обычно бывает очень разобщенной и зачастую недостаточной для поддержки принятия решений.

Источниками образовательных данных становятся такие информационные системы, как компьютерные образовательные программы (в том числе интеллектуальные), информационная система вуза, социальные сети, работы обучающихся и т.д. – например, логи, результаты тестирования и анкетирования, тексты студенческих работ, учебные планы, рабочие программы дисциплин и т.д. Также данные для анализа получают из информационных образовательных систем и ресурсов, которые связаны с учебным процессом. К таким системам относят, кроме упомянутых LMS, массовые онлайн-курсы (Massive Open Online Courses, MOOC). В качестве источника данных в таких системах чаще всего используют файлы с логами, содержащими информацию о действиях пользователя, например, просмотры страниц, клики, переходы, оценки. Для масштабной аудитории курсов, например, на таких ресурсах как Coursera или edX, применение алгоритмов интеллектуального анализа образовательных данных становится особо важным.

Следовательно, одна из основных задач EDM – это *предсказание успеваемости обучающихся*. В частности, работа [13] посвящена прогнозированию успеваемости с использованием методов классификации. В статье [14] прогнозирование успеваемости осуществляется на основе логов из LMS Moodle. В [15] решается частная задача предсказания успеваемости – предсказание неудовлетворительной оценки по учебной дисциплине с помощью генетических алгоритмов.

Другой важнейшей задачей EDM является индивидуализация образовательной траектории. Например, в статье [16] рассматриваются методы анализа образовательных данных для обеспечения персонализированной траектории обучения творческим специальностям. Отметим, что для решения этой задачи применяется алгоритм классификации, основанный на деревьях решений. Авторы [17] показывают связь рекомендательных систем и траекторий обучения, и предлагают модель, основанную на теории графов для построения такой траектории.

Немаловажную роль играют методы интеллектуального анализа образовательных данных в обеспечении обратной связи для преподавателей. В работе [18] описан образовательный портал, на котором реализованы методы интеллектуального анализа данных, для учеников начальной школы. В статье [19] показан способ организации обратной связи с помощью алгоритмов генетического программирования. Авторами реализован поиск ассоциативных правил в данных, полученных из результатов тестирования.

Одним из перспективных направлений интеллектуального анализа образовательных данных является решение задачи конструирования учебного курса. Как указано в работе [1], компьютерные средства обучения должны выступать не только в виде источника, но и приемника, и переработчика семантической информации. В работе [20] указывается значимость интеллектуализации разнообразных содержательных интерпретаций в образовательном процессе,

необходимость формализации для дальнейшего улучшения объективизации получаемых данных. Также следует отметить значимость интеллектуального анализа данных деятельности обучающихся в социальных сетях. Важность этих вопросов и их разноплановость отмечают многие педагоги и исследователи, например, в работе [21] приводятся такие вопросы, которые требуют решения: как образовательные учреждения собираются реагировать на изменения, и имеет ли социальное программное обеспечение место в новом сценарии? как можно помочь педагогам и поддержать обновление учебных программ, чтобы принять практики обучения и жизни, которые возникают в информационную эпоху? как можно соединить возможности С-обучения (обучения с помощью социальных сетей) с информированным профессионализмом? как можно признать и подтвердить знания, которые некоторые молодые люди уже имеют за счет использования социального ПО? что можно сделать для стимулирования появления нового инновационного социального программного обеспечения, специально предназначенного для поддержки образовательной деятельности?

На эти вопросы дать ответы может помочь интеллектуальный анализ образовательных данных. А решать эти вопросы необходимо, как отмечается далее в [21], эти вопросы остаются открытыми, и можно только констатировать, что потенциал социального программного обеспечения существенно превышает те возможности, которые используются в образовании в настоящее время. В [21] показано, что активная деятельность молодежи в социальных сетях способствует формированию компетенций, необходимых в профессиональном образовании, технология вики позволяет организовывать различные виды учебной деятельности в процессе профессиональной подготовки и формировать динамичную образовательную среду для подготовки современных специалистов.

К другим важным задачам относятся: анализ и визуализация данных, моделирование поведения студентов, выявление нежелательного поведения студентов (прогулы, плагиат), распределение студентов по группам и подгруппам, анализ социальных связей студентов, разработка концептуальных карт.

Программное приложение для проведения анализ данных с использованием сложных алгоритмов классификации

Кафедра интеллектуальных технологий в гуманитарной сфере Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики (Университет ИТМО) активно занимается проблемами интеллектуального анализа данных в сфере образования.

В течение последних лет научных исследований, а также в рамках выпускных квалификационных работ, с использованием методов интеллектуального анализа данных были разработаны следующие приложения:

– Компонент мониторинга активности аспирантов для информационной системы кафедры.

Компонент позволяет частично автоматизировать сбор и обработку информации о деятельности аспирантов. Созданы инструменты для генерирования отчетов. Реализована возможность самостоятельного добавления аспирантом отметок о своей активности в информационной среде кафедры и подтверждения их научным руководителем. Разработанный компонент позволяет руководству кафедры получать более полное представление о деятельности аспирантов, видеть сравнительную характеристику этой деятельности и динамику ее изменения.

– Компонент сравнения результатов анкетирования в процессе лонгитюдного исследования.

Рассматривался компонент информационной системы анкетирования студентов, который позволяет автоматизировать рутинные операции, возникающие

при проведении лонгитюдных исследований, и предоставляет возможность для проведения длительных повторных исследований по интересующим вопросам.

– Программное приложение для многокритериального оценивания результатов продуктивной деятельности учащихся.

Оценивание результатов работ учащихся является неотъемлемой частью образовательного процесса. В рамках настоящей работы предложен подход, который, во-первых, позволяет сделать более прозрачным и понятным процесс оценивания, а, во-вторых, снижает нагрузку на преподавателя и делает оценку более объективной.

– Компонент диагностики процесса формирования профессиональной компетенции для информационной системы кафедры.

– Модуль ведения электронного портфолио студента для информационной системы кафедры.

– Модуль создания тестов с использованием встроенной системы тегов в информационно-образовательной среде LMS Moodle.

– Модуль автоматизированной проверки знаний глоссария для LMS Moodle.

– Система генерации математических задач и ее интеграция в LMS Moodle .

– Веб-сервис для визуализации учебного плана.

– Визуализатор результатов выбора траектории обучения.

– Информационный образовательный ресурс для реализации интегрированного образовательного маршрута в начальной школе.

– Программное приложение для нечёткой кластеризации данных в неевклидовых метриках с применением генетических алгоритмов.

– Программное приложение для распределения и учета учебной и внеучебной нагрузки преподавателя.

Одна из недавних разработок сотрудников кафедры – это программное приложение, позволяющее проводить анализ данных с использованием сложных алгоритмов классификации. Достоинство приложения заключается в том, что им могут пользоваться даже люди далёкие от программирования и математических методов. Приложение представляет графический интерфейс пользователя, выполненный в виде стандартной архитектуры «мастера». С помощью приложения можно осуществить все основные этапы интеллектуального анализа данных.

На первом шаге происходит загрузка файла с данными. Обычно это табличные данные, представленные в формате CSV. Это универсальный формат, который поддерживают многие информационные образовательные системы.

Загруженные в приложение данные проходят предварительную обработку. В процессе предобработки данных для дальнейшего анализа также необходимо учитывать дополнительные проблемы, возникающие при работе с данными, и принимать во внимание такие процедуры, как очистка и предобработка данных, фильтрация данных, сортировка данных, перекодирование данных (в том числе, создание категориальных переменных из количественных), пропуски в данных и методы их решения и выявление аномальных данных. После того как завершён первый шаг, исследователь может приступить к выбору алгоритма классификации.

В общем виде задачу классификации можно представить следующим образом. Имеется множество объектов, которые разделены на классы по некоторым признакам. Например, успевающие студенты и отстающие студенты; абитуриенты, поступившие по результатам ЕГЭ, и абитуриенты, поступившие по результатам предметных олимпиад. В обучающей выборке задано конечное множество объектов и их признаков. Например, перечень всех студентов учебного заведения и все оценки по прошедшим и текущим дисциплинам. Для каждого из объектов обучающей выборки известно, к каким классам они относятся. Принадлежность же остальных объектов к классам неизвестна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества, то есть указать

наименование (или номер) класса, к которому объект отнесён в результате применения алгоритма классификации.

Математически это можно записать следующим образом. Пусть X – это множество описаний объекта, Y – это множество классов. Существует неизвестная целевая функция $f: X \rightarrow Y$, значения которой известны только на обучающей выборке $X_m = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)\}$, где $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ – векторы признаков объектов, а y_1, y_2, \dots, y_m – наименование классов, к которым принадлежат соответствующие объекты. Требуется построить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

В настоящий момент приложение поддерживает только бинарную классификацию с помощью восьми разных алгоритмов. Бинарная классификация предполагает наличие всего двух классов, т.е. $Y = \{-1; +1\}$.

Как показано в статье [22] почти половина всех исследований в области EDM используют различные алгоритмы классификации. К наиболее часто используемым алгоритмам относятся деревья решений, метод k ближайших соседей и наивный байесовский классификатор. В программном приложении реализованы наиболее распространённые в области EDM алгоритмы классификации. Кроме названных трёх это: SVM – метод опорных векторов, Random Forest – ансамбль деревьев решений, Ada Boost – алгоритм усиления классификаторов, градиентный бустинг и логистическая регрессия. Алгоритмы классификации описаны в [23]

После выбора алгоритма на третьем шаге происходит его настройка. Пример формы для настройки алгоритма приведён на рисунке 1.

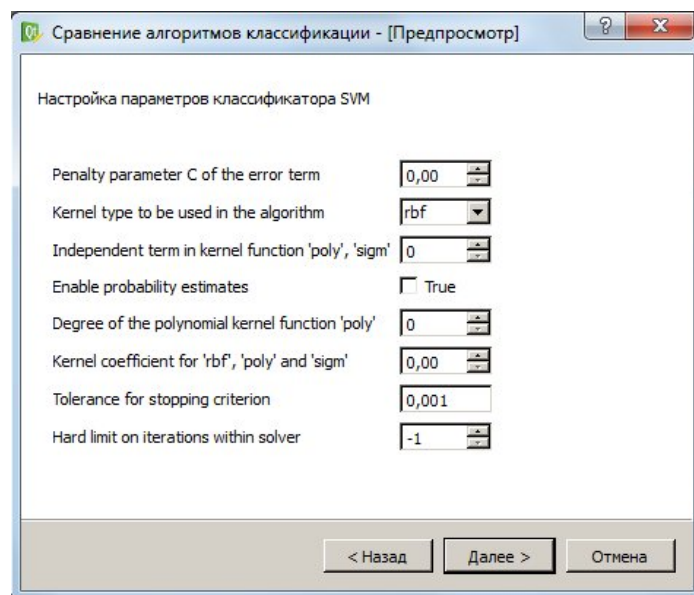


Рис. 1. Форма настройки алгоритма классификации

На рис. 1 представлена форма для настройки алгоритма SVM (support vector machine). Для настройки выводится специальная форма, в которой установлены значения по умолчанию и для каждого управляющего элемента заданы ограничения на типы данных и диапазоны. Это сделано для того, чтобы пользователь не смог ввести некорректные настройки алгоритма.

После выбора нужных настроек пользователь может перейти к визуализации работы алгоритма. Для того чтобы осуществить реализацию, программное приложение случайным образом извлекает из обучающей выборки (train set) 250 объектов. Затем опять же случайным образом из этой новой выборки выделяется 20% для формирования проверяющей выборки (test set). В проверяющей выборке

удаляются значения для всех y . С помощью этой выборки в дальнейшем будем проверяться точность алгоритма классификации.

Приложение позволяет сравнивать работу двух или нескольких алгоритмов на одинаковых данных. На рис. 2-4 представлены результаты работы приложения.

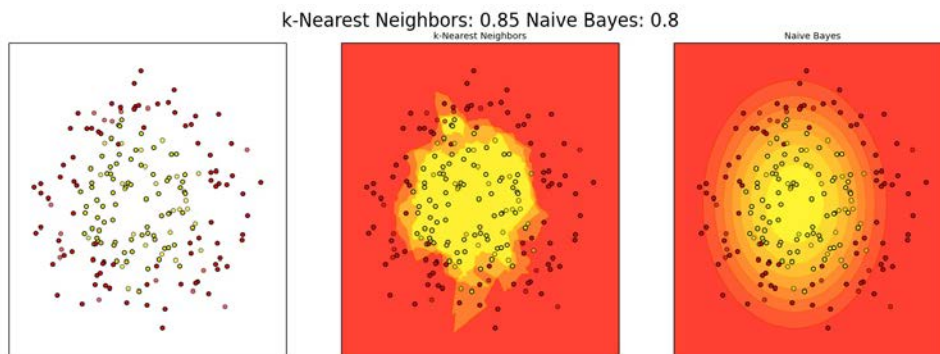


Рис. 2. Сравнение точности алгоритмов классификации (метод k ближайших соседей и наивный байесовский классификатор)

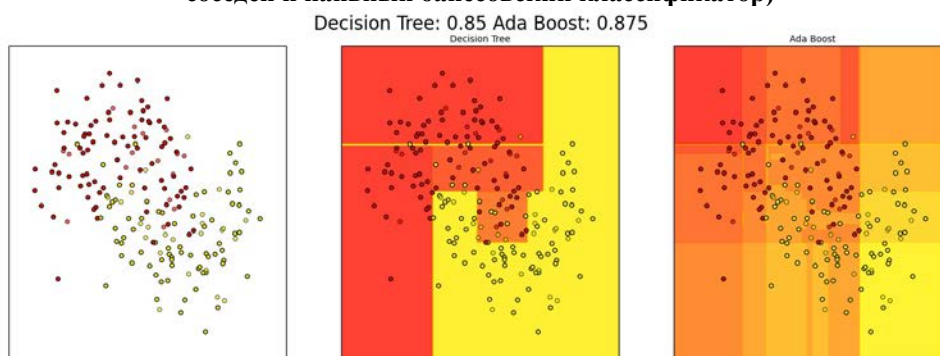


Рис. 3. Сравнение точности алгоритмов классификации (дерево решений и алгоритм усиления классификаторов)

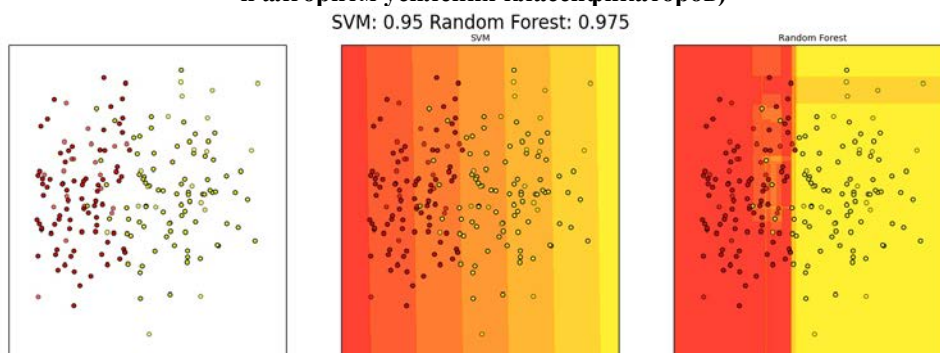


Рис. 4. Сравнение точности алгоритмов классификации (метод опорных векторов и ансамбль деревьев решений)

На первом графике отображаются исходные данные, красным и жёлтым цветом выделены объекты, относящиеся к разным классам. Второй и третий графики наглядно представляют работу алгоритмов. Точность сравниваемых алгоритмов указывается сверху, в заголовке графиков.

Приведенные на рис. 2-4 графики в явном виде свидетельствуют, что точность алгоритмов может довольно сильно отличаться в зависимости от исходных данных. Разработанное приложение позволяет протестировать алгоритмы классификации и подобрать наиболее подходящий для конкретных данных алгоритм.

Анализ и оценка разработки

Разработанное приложение позволяет решать задачу классификации образовательных данных. Примером такой задачи является в частности выявление возможности отчисления студента в первом учебном семестре, которое можно сделать по результатам анкетирования и данным, полученным из информационной системы приёмной комиссии, таким как результаты ЕГЭ, средний балл, наличие дипломов олимпиад, форма финансирования (бюджет, контракт), наличие высшего образования у родителей и т.п.

Заключение

Описаны задачи интеллектуального анализа данных образовательного процесса, подробно рассмотрены методы их решения и сферы применения. Рассмотрены перспективы развития данного научного направления, в том числе, с учетом российской специфики. Приведенные методы могут найти широкое применение в системах поддержки принятия решений на всех уровнях управления образовательным процессом.

Литература

1. Печников А.Н. Е-дидактика: кому, зачем и в каком виде она нужна // Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)", 2013, том 16, №4, С. 326 - 343. URL: http://ifets.ieee.org/russian/depository/v16_i4/pdf/4.pdf (дата обращения: 19.01.2015)
2. Воробьева Н.А., Носков С.И. Программное обеспечение для автоматизации процесса разработки учебных планов. Фундаментальные исследования. 2012. № 6-3. С. 633-636.
3. Глушенко А.И. Информационная система принятия решений по формированию индивидуальных учебных планов. Управление большими системами: сборник трудов. 2006. № 15. С. 79-91.
4. Woolf В. Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-learning. – San Francisco, CA: Elsevier Inc., Morgan Kauffman, 2009. – 440 p.
5. Nkambou R., Mizoguchi R., Bourdeau J. (Eds.) Advances in Intelligent Tutoring Systems // Studies in Computational Intelligence, Vol. 308. - Springer, 2010. - 510 p.
6. Галеев И.Х. Модель управления процессом обучения в ИОС // Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)". 2010. V.13. №3. С. 285-292. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 19.01.2015)
7. Галеев И.Х. Проблемы и опыт проектирования ИОС // Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)" 2014. V.17. №4. С. 526-542. URL: <http://ifets.ieee.org/russian/periodical/journal.html> (дата обращения: 19.01.2015)
8. Хлопотов М.В., Коцюба И.Ю. Методы интеллектуального анализа данных для мониторинга и диагностики качества образования // Научный журнал "Дистанционное и виртуальное образование". - Москва, 2014. - № 5. - С. 18-25.
9. EDM2014. URL: <http://www.educationaldatamining.org/EDM2014/> (дата обращения: 19.01.2015)

10. JEDM - Journal of Educational Data Mining. URL: <http://www.educationaldatamining.org/JEDM/index.php/JEDM> (дата обращения: 19.01.2015)
11. Big Data in Education Coursera. URL: <https://www.coursera.org/course/bigdata-edu> (дата обращения: 19.01.2015)
12. Сосницкий В.Н., Потанин Н.И., Шевелева Л.В. Проблемы статистического анализа средней успеваемости студентов. // *Фундаментальные исследования*. 2013. № 10-2. С. 316-320.
13. Kabakchieva, D. Predicting student performance by using data mining methods for classification (2013) *Cybernetics and Information Technologies*, 13 (1), pp. 61-72.
14. Romero, C.a b , Espejo, P.G.b , Zafra, A.b c , Romero, J.R.b , Ventura, S.b Web usage mining for predicting final marks of students that use Moodle courses (2013) *Computer Applications in Engineering Education*, 21 (1), pp. 135-146.
15. Márquez-Vera, C.a , Cano, A.b , Romero, C.b , Ventura, S.b Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data (2013) *Applied Intelligence*, 38 (3), pp. 315-330
16. Lin, C.F.a , Yeh, Y.-C.b , Hung, Y.H.c , Chang, R.I.a Data mining for providing a personalized learning path in creativity: An application of decision trees (2013) *Computers and Education*, 68, pp. 199-210.
17. Durand, G., Belacel, N., Laplante, F. Graph theory based model for learning path recommendation (2013) *Information Sciences*, 251, pp. 10-21.
18. Levashenko, V., Zaitseva, E., Kvassay, M., Kostolny, J. Educational portal with Data Mining support for pupils of primary schools (2013) *International Conference on Digital Technologies 2013, DT 2013*, art. no. 6566281, pp. 33-38.
19. Romero, C., Zafra, A., Luna, J.M., Ventura, S. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data (2013) *Expert Systems*, 30 (2), pp. 162-172.
20. Монако А.Ф. Подход к построению формализованного описания информационных систем для образования и обучения // *Международный электронный журнал “Образовательные технологии и общество (Educational Technology & Society)”*, 2013, том 16, №1, С. 536 - 546. URL: http://ifets.ieee.org/russian/depository/v16_i1/pdf/10.pdf (дата обращения: 19.01.2015)
21. Голицына И.Н. Социальное программное обеспечение в современном профессиональном образовании // *Международный электронный журнал “Образовательные технологии и общество (Educational Technology & Society)”*, 2012, том 15, №4, С. 515 - 527. URL http://ifets.ieee.org/russian/depository/v15_i4/pdf/11.pdf (дата обращения: 19.01.2015)
22. Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works // *Expert Systems with Applications*. Vol. 41, Issue 4, Part 1. - Elsevier, 2014 - pp. 1432–1462 [in press]
23. Bishop C. M. *Pattern Recognition and Machine Learning* // *Information Science and Statistics*. - Springer, 2006 - 738 p.