

Применение средств Education Data Mining в задачах управления процессами обучения

Е. Е. Котова

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
eekotova@gmail.com

Аннотация. Обучение все больше распределяется по пространству, времени и средствам массовой информации, ввиду чего генерируется большой объем данных, называемых большими данными, об учащихся и процессе обучения. Эти данные представляют собой «следы», которые «оставляют» учащиеся, когда они взаимодействуют со все более сложной и быстро меняющейся средой обучения. Обилие данных привело к растущему интересу в применении методов аналитики образовательных данных (Education Data Mining, EDM) среди исследователей и практиков в сфере академического образования. Студенты сталкиваются в процессе учебной деятельности с переизбытком информации. Большие объемы информации в формализованном, формализовано-алгоритмизированном виде приводят к неконтролируемому увеличению когнитивной нагрузки в учебном процессе. С целью управления процессом обучения необходимо в короткие сроки получить информацию об индивидуальных особенностях познавательно-мыслительной деятельности обучающихся, чтобы перенастроить контент, адаптировать процесс под конкретного студента для снижения отрицательного эффекта. Перспективы технологии EDM в системах обучения обсуждаются в докладе.

Ключевые слова – процесс обучения; методы анализа данных; индивидуальные особенности; интегрированная среда; когнитивный потенциал; набор данных DataSet

I. ВВЕДЕНИЕ

Ряд исследовательских проектов за последние годы посвящены изучению стратегий обучения, когнитивных навыков, психологических конструкций с использованием различных типов данных. Предлагаются т.н. «mechanistic models of learning, computational models of learning», «вычислительные модели» моделирования обучения студентов в интеллектуальных системах обучения (intelligent tutoring systems, ITS) (Koedinger K. R. et al., Carnegie Mellon University). Последние десять лет разработка интеллектуальных сред для поддержки учебного процесса ведется в крупнейших университетах мира. Примерами могут служить такие программы как: EComma (<https://ecomma.coerll.utexas.edu/about-ecomma/>), EMargin (<http://emargin.bcu.ac.uk/>), Ponder (<https://www.ponder.co/welcome/>), Classroom Salon (<http://srv01.pragma.cs.cmu.edu/>) и др. Перечисленные примеры систем позволяют работать с учебными ресурсами и видеоматериалами онлайн, обсуждать

проблемные вопросы в специально подготовленных чатах, пользоваться инструментами визуализации данных, автоматически анализировать действия студентами для модификации процесса обучения.

Сфера образовательных услуг сталкивается с определенными трудностями повышения производительности процесса обучения, однако подход с использованием методов аналитики (подход Learning Analytics) [1–3, и др.] и больших данных может расширить представление о процессах обучения через анализ действий студентов и повысить его результативность [4 и др.].

II. МЕТОДЫ АНАЛИТИКИ ДАННЫХ В ОБРАЗОВАТЕЛЬНОЙ СФЕРЕ

Вопросы применения методов аналитики данных в области образования обсуждаются в профессиональных научных обществах.

Международной рабочей группой в 2011 году основано Международное образовательное сообщество интеллектуального анализа данных (International Educational Data Mining Society, <http://www.educationaldatamining.org>).

Начиная с 2011 года проводится ежегодная Международная конференция по обучающей аналитике (International Conference on Learning Analytics and Knowledge).

В 2011 году было создано профессиональное общество исследований в области аналитики в системах обучения и развития (Society for Learning Analytics Research, SoLAR, <http://www.solaresearch.org>).

На мировом уровне известны организации, развивающие основные направления исследований интеллектуальных обучающих систем в сфере образования.

В институте Human-Computer Interaction Institute в Питтсбурге (as part of Carnegie Mellon University's Simon Initiative) проводятся исследования интеллектуальных обучающих систем, познавательных и когнитивных процессов, факторов приобретения знаний, социальных и коммуникативных факторов в обучении. Пример системы, где анализируются кривые обучения по образовательным

данным, DataShop (Pittsburgh Science of Learning Center, <http://www.learnlab.org/>).

В Национальном университете Кордовы (University of Córdoba) (Department of Computer Sciences and Numerical Analysis) разработаны модели «предсказания» конечных результатов обучения с целью повышения эффективности адаптивных систем обучения [5].

В университете Гронингена (University of Groningen) исследуются динамические системы в применении к моделированию образовательных траекторий на основе динамического подхода к процессам развития [6, 7].

В Колумбийском университете (Columbia University) разработаны методы, алгоритмы интеллектуального анализа данных в области образования, моделирования кривых индивидуального обучения на основе вероятностных моделей с использованием методов интеллектуального анализа данных, приводятся примеры разработок [8, 9].

Отдельные примеры исследований подтверждают значимость развития методов и средств EDM для образовательных организаций.

Например, представлены кривые производительности обучения (learning curves), где производительность обучения оценивается временем выполнения задач и вероятностью допущенных ошибок [10].

Исследуется анализ динамики роста знаний студентов при изучении текстовых документов. На основе анализа терминов формируются модели компонентов знаний, учитывается число неизученных компонентов знаний. Применяются методы классификации, основанные на логистической регрессии, марковских моделях [11].

Рассматриваются модели обучения студентов на основе Байесовской трассировки знаний (Bayesian Knowledge Tracing) и когнитивных моделей [12].

Приведем примеры отечественных разработок.

Для оценки навыков/умений обучаемого используется Байесовский подход. Математическая модель адаптивного процесса управления обучением реализуется как многошаговый процесс принятия решений [13].

Рассматривается мультиагентная имитационная модель учебного процесса, отражающая взаимодействия обучаемого и преподавателя с учетом психофизиологического, эмоционального и когнитивного состояний, динамику передачи и накопления знаний с анализом и прогнозированием качества образования [14].

Модели компьютерного и имитационного моделирования, использующие методы распределения учебной информации, и определения уровня знаний [15].

Данные исследования фокусируются на методах анализа данных, статистического анализа, дисперсионного, факторного анализа, Байесовском подходе, вероятностных моделях, методах классификации, марковских моделях, необходимых для вычислительной поддержки и принятия решений в образовательных системах.

III. ВОПРОСЫ МОДЕЛИРОВАНИЯ СТУДЕНТА. МОДЕЛЬ КОГНИТИВНО-СТИЛЕВОГО ПОТЕНЦИАЛА

Моделирование студентов играет решающую роль в развитии учебно-методических технологий. Модели учащихся являются критически важными для точной оценки навыков обучения, получения адаптивных инструкций и улучшения процесса [16]. Отмечается, что до сих пор актуален вопрос, как люди управляют обучением внутри и между контекстами обучения [17].

Система управления учебным процессом организована по принципу обратной связи, разделения дидактических ресурсов и адаптации к индивидуальным характеристикам обучаемого [18].

В соответствии с динамическим подходом к исследованию учебного процесса, скорость роста уровня знаний/когнитивного роста зависит от индивидуальных показателей и стратегий учащихся, дидактических ресурсов (когнитивной нагрузки), усвоенных знаний, способностей к обучению и др. Скорость роста уровня знаний представим в виде дифференциального уравнения первого порядка

$$\frac{dy}{dt} = \Phi(u, y, f, \mathbf{q}), y_0,$$

где: $dy(t)/dt$ – скорость роста уровня знаний, %/ед. вр., $y(t)$ – уровень усвоенных знаний, %; y_0 – начальный уровень знаний; $u(t)$ – интенсивность дидактических ресурсов, усл. ед./ед. вр.; f – внешние возмущения, \mathbf{q} – вектор параметров модели студента. Результат обучения, уровень знаний, «накопленный» за период обучения T представляется в виде:

$$y(T) = \int_0^T \Phi(u, y, f, \mathbf{q}) dt + y_0.$$

В работе приняты следующие допущения.

1. Траектории обучения фиксируются в течение 12 недель активного периода обучения за один стандартный семестр.

2. Дидактические ресурсы приняты в условных единицах и могут быть представлены в виде учебных модулей соответствующих дисциплин (лекций, практических занятий, иных форм – электронных модулей, индивидуальных заданий, лабораторных работ и др.).

3. Промежуточные значения индивидуальных траекторий на интервалах между управляющими воздействиями являются ненаблюдаемыми. Наблюдаются и измеряются только точки траекторий обучения в дискретные моменты процесса.

4. Оценка каждого занятия (результаты, контрольные баллы) осуществляется независимо от предыдущего.

Важное значение придается исследованиям индивидуальных различий, когнитивных стилей, стилей обучения, подходов к обучению и моделям обучения [17].

Базовые основы исследований когнитивных процессов, познавательно-мыслительной деятельности обучающихся на индивидуальном уровне заложены в области когнитивной науки, психологии и педагогики [19–24]. Сегодня этот вопрос актуален более, чем когда-либо ввиду необходимости высокого уровня педагогического понимания индивидуальных потребностей студентов в новых средах обучения с применением электронных средств и технологий OCW (open course ware), OER (open educational resource), MOOC (massive open online courses), средств E-Learning, blended learnig (B-Learning).

Когнитивные конструкторы стиля играют важную роль при изучении эффективности обучения студентов в условиях повышенной или возрастающей сложности, требующей производительности высокого уровня [25].

Ввиду направленности исследований на значимость когнитивно-мыслительной сферы обучающихся, нами предложена модель когнитивного потенциала студента, в составе которой показатели когнитивных стилей и когнитивных контролей, являющиеся определяющими факторами продуктивности учебно-познавательной деятельности в современной образовательной среде. Показатели когнитивно-стилевого потенциала (КСП) фиксируются в системе экспресс-диагностики с помощью компьютерной версии методик диагностики параметров когнитивных стилей [26–28]. На верхнем уровне иерархии модели КСП фиксируются параметры когнитивного контроля, на следующем уровне параметры когнитивных стилей [26]. На основе индивидуальных моделей ставятся следующие задачи. 1 – задача идентификации параметров моделей студентов различающихся групп по продуктивности процесса обучения. 2 – задача прогнозирования успешности обучения. Принимается концепция т.н. «среднего студента» (average-profile student). Уровень знаний «среднего студента» является относительным и субъективным, зависит от контингента обучающихся, от дидактических ресурсов/когнитивной нагрузки учебных дисциплин и от стратегий преподавания. Задачу прогнозирования успешности обучения имеет смысл решать на ранних стадиях обучения в вузе, на первом курсе. В публикациях отмечается, что изучение стратегий студентов во время начального периода обучения в вузе, характер обучения по отношению к требованиям конкретных дисциплин имеет важное значение [29].

Реализация поставленных задач потребовала разработки структурированного набора данных AnalyticSYSTEMS-hDS (StudCSP-DataSet).

IV. ФОРМИРОВАНИЕ НАБОРА ДАННЫХ DATASET КОГНИТИВНОГО ПОТЕНЦИАЛА ОБУЧАЮЩИХСЯ

Одна из основных задач Learning Analytics – распознавание потребностей отдельных учащихся с целью управления и адаптации индивидуальной поддержки [30]. Спектр анализа, представлен на макро-, мезо- и микроуровнях [29]. Макроуровень включает сбор данных на общегосударственном (региональном и др.) уровне. На мезоуровне анализируется информация от образовательных учреждений. Микроуровень

поддерживает данные отдельных процессов: активности учащихся (или групп), мотивации, интересов, причин и сложностей обучения, рекомендаций по улучшению учебной деятельности. Данные на микроуровне могут быть получены максимально быстро [29].

В настоящей работе представлен набор данных (НД) StudCSP-DataSet, который включает параметры когнитивно-стилевого потенциала, данные траекторий обучения, значения факторов моделей обучающихся и др. Параметры КСП формируются на основе результатов экспресс-диагностирования с использованием методик по решению когнитивных задач в реальном масштабе времени. По данным диагностирования определены признаки внутренней дифференциации обучающихся на условные подгруппы типовых профилей: High-profile (H), Average-profile (A), Special-profile (S), Exeptional-profile (E) – HASE-profile. H-группа – группа студентов низкого уровня риска, или высокого уровня продуктивности. A-группа среднего уровня риска, среднего (минимально допустимого) уровня продуктивности. S-группа – высокого уровня риска, или недостаточного уровня продуктивности учебной деятельности (ниже среднего). E-группа – группа студентов низкого уровня продуктивности. Обучающиеся S и E-групп требуют особого внимания в организации индивидуального процесса обучения.

Разнородность контингента обучающихся является отличительной особенностью учебного процесса. Модель КСП позволяет быстро получить необходимые данные для уточнения будущих студентов на этапе поступления в вуз. На рис. 1 в качестве примера изображена сравнительная диаграмма данных модели КСП обучающихся для шести групп студентов первого курса СПбГЭТУ «ЛЭТИ».

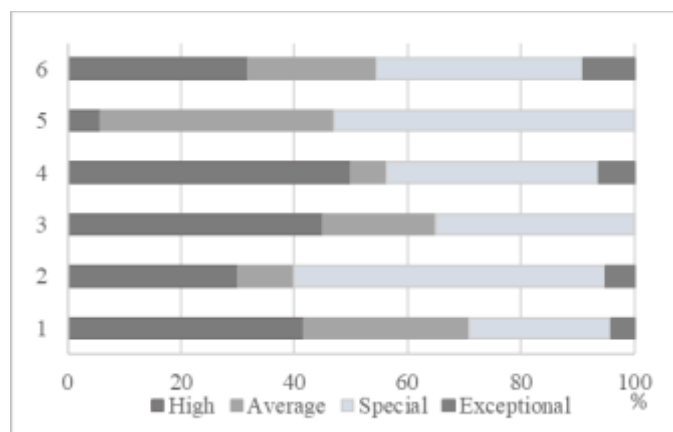


Рис. 1. Примеры распределения на подгруппы HASE-profile

Графики демонстрируют разнородность контингента обучающихся по числу студентов, входящих в каждую подгруппу HASE-profile. Данные моделей используются при прогнозировании траекторий обучения [31]. С целью прогнозирования в НД внесены параметры более 50 методов интеллектуального анализа данных.

V. ОБСУЖДЕНИЕ И ВЫВОДЫ

Сбор образовательных данных весьма ресурсоемкий процесс: данные являются сильно контекстно-зависимыми и отличаются небольшими наборами. Данные собираются из разных источников: административной информации, учебных журналов событий, онлайн-среды обучения, и др. Образовательная среда создает сложные слабо структурированные модели, которые трудно понять и интерпретировать без применения специальных методов аналитики, называемых «spaghetti models» [32].

Примеры данных в открытом доступе практически отсутствуют. Доступный для исследователей набор данных приводится в Техническом университете Эйндховена (Eindhoven University of Technology research portal, <https://research.tue.nl/en/>). Набор данных содержит временные ряды действий студентов в течение шести сессий лабораторных занятий по курсу цифровой электроники. В Питсбургском научно-образовательном центре доступны инструменты анализа данных и визуализации через веб-интерфейс: (DataShop repository at the Pittsburgh Science of Learning Center Pittsburgh Science of Learning Center, <http://learnlab.org/datashop>). Модель ученика представлена моделью компонентов знаний. В терминологии DataShop компоненты знаний (Knowledge Components, KCs) используются для представления частей знаний, концепций или навыков, которыми студенты должны владеть.

Набор данных AnalyticSYSTEMS-hDS (StudCSP-DataSet) находится в стадии государственной регистрации. DataSet является открытым хранилищем данных с соответствующими инструментами визуализации и анализа. Аналитические материалы для новых данных могут использоваться в крупномасштабной цифровой среде.

СПИСОК ЛИТЕРАТУРЫ

- [1] Vahdat M. et al. A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator //Design for teaching and learning in a networked world. Springer, Cham, 2015. Pp. 352-366.
- [2] Bienkowski M. et al. Enhancing teaching and learning through educational data mining and learning analytics: An issue brief //US Department of Education, Office of Educational Technology. 2012. Vol. 1. Pp. 1-57.
- [3] Del Blanco Á. et al. E-Learning standards and learning analytics. Can data collection be improved by using standard data models? //2013 IEEE Global Engineering Education Conference (EDUCON). IEEE. 2013. Pp. 1255-1261.
- [4] Manyika J. et al. Big data: The next frontier for innovation, competition, and productivity. 2011. P. 144.
- [5] Romero C. et al. Predicting students' final performance from participation in on-line discussion forums //Computers & Education. 2013. Vol. 68. Pp. 458-472.
- [6] Steenbeek H., van Geert P. The emergence of learning-teaching trajectories in education: A complex dynamic systems approach //Nonlinear dynamics, psychology, and life sciences. 2013. Vol. 17. No. 2. Pp. 233-267.
- [7] Van Geert P. Dynamic modeling for development and education: from concepts to numbers //Mind, Brain, and Education. 2014. Vol. 8. No. 2. Pp. 57-73.
- [8] Baker et al. Predicting robust learning with the visual form of the moment-by-moment learning curve //Journal of the Learning Sciences. 2013. V. 22. No. 4. Pp. 639-666.
- [9] Baker R.S., Inventado P.S. Educational data mining and learning analytics // Learning analytics. Springer New York, 2014. Pp. 61-75.
- [10] Martin et al. Evaluating and improving adaptive educational systems with learning curves //User Modeling and User-Adapted Interaction. 2011. Vol. 21. No. 3. P. 249-283.
- [11] Huang et al. A Framework for Dynamic Knowledge Modeling in Textbook-Based Learning //Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization. ACM. 2016. P. 141-150.
- [12] Pelánek R. Metrics for evaluation of student models //JEDM-Journal of Educational Data Mining. 2015. Vol. 7. No. 2. Pp. 1-19.
- [13] Галеев И.Х. Модель управления процессом обучения в ИОС // Международный электронный журнал "Образовательные технологии и общество (Educational Technology & Society)". 2010. Т.13. №3. С. 285-292.
- [14] Ивашкин Ю.А., Назойкин Е.А. Мультиагентное имитационное моделирование процесса накопления знаний //Программные продукты и системы. 2011. № 1. С. 47-52.
- [15] Майер Р.В. Основная задача математической теории обучения и ее решение методом имитационного моделирования //Международный журнал прикладных и фундаментальных исследований. 2014. № 2. С. 36-39.
- [16] Koedinger K.R., McLaughlin E.A., Stamper J.C. Automated Student Model Improvement // International Educational Data Mining Society. Proceedings of the 5th International Conference on Educational Data Mining. 2012. Pp. 17-24.
- [17] Evans C., Vermunt J.D. Styles, approaches, and patterns in student learning //British Journal of Educational Psychology. 2013. № 2. Pp. 185-195.
- [18] Имаев Д.Х., Котова Е.Е. Моделирование и имитация процессов обучения с разделением дидактических ресурсов. Динамический подход. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2014. 111 с.
- [19] Thurstone L.L. A factorial study of perception. Chicago: University of Chicago Press. 1944. 148 p.
- [20] Kagan J. Reflection-impulsivity: The generality and dynamics of conceptual tempo //Journal of abnormal psychology. 1966. V. 71. № 1. Pp. 17-24.
- [21] Stroop J.R. Studies of interference in serial verbal reactions //J. of Exper. Psychology. 1935. V. 18. Pp. 643-662.
- [22] Когнитивные стили. Тезисы научно-практического семинара. / Под ред. В. Колги. Таллинский пед. Институт им. Э. Вильде. 1986. 250 с.
- [23] Стил человека. Психологический анализ / Под ред А.В. Либина. М.: Смысл, 1998. 310 с.
- [24] Холодная М.А. Когнитивные стили: О природе индивидуального ума. Учебное пособие. М.: ПЕР СЭ. 2002. 304 с.
- [25] Evans C., Richardson J.T.E., Waring M. Field independence: Reviewing the evidence //British Journal of Educational Psychology. 2013. Vol. 83. No. 2. Pp. 210-224.
- [26] Котова Е.Е., Писарев А.С. Задача классификации учащихся с использованием методов интеллектуального анализа данных. Известия СПбГЭТУ ЛЭТИ. 2019. № 4. С. 32-42.
- [27] Имаев Д.Х., Котова Е.Е. Оценка параметров динамических моделей обучаемых по результатам экспресс диагностирования. // Известия СПбГЭТУ «ЛЭТИ». Современные технологии образования. 2015. № 1. С. 70-75.
- [28] Котова Е.Е., Падерно П.И. Экспресс-диагностика когнитивно-стилевого потенциала обучающихся в интегрированной образовательной среде. // Образовательные технологии и общество. 2015. Т. 18. № 1.
- [29] Donche V. et al. Differential use of learning strategies in first-year higher education: The impact of personality, academic motivation, and teaching strategies. //British Journal of Educational Psychology. 2013. Vol. 83. Pp. 238-251.
- [30] Shum S.B. Learning analytics policy brief //UNESCO Institute for Information Technology in Education. 2012. 12 p.
- [31] Котова Е.Е., Писарев А.С. Автоматизация прогнозирования результатов обучения студентов. Известия СПбГЭТУ ЛЭТИ. 2019. № 5. С. 31-39.
- [32] Bogarin A., Cerezo R., Romero C. A survey on educational process mining // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2018. Vol. 8. No. 1. P. 1-17.