

Об одном подходе к персонификации обучения в рамках компьютерных обучающих систем

А. П. Рыжов, А. Д. Журавлев (МГУ им. М.В. Ломоносова)
А. Н. Вахов, В. В. Кривцов (ООО «Учи.ру»)

Рассмотрены основные проблемы развития компьютерных обучающих систем. Показано, что ключевой задачей и точкой роста для таких систем является персонификация процесса обучения. Формулируются необходимые условия персонификации обучения. Приводятся результаты экспериментов с данными, накапливаемыми в процессе обучения. Основные положения иллюстрируются на примере платформы Учи.ру.

Ключевые слова: компьютерные обучающие системы, персонификация, кластеризация, ранжирование.

1. Введение

Понятие компьютерных обучающих систем возникло практически одновременно с понятием искусственного интеллекта и претерпело значительную эволюцию одновременно с развитием компьютерных технологий. В докладе проводится такой исторический экскурс, близкий к обзору [1]; отдельно обсуждаются изменения, происшедшие в последнее время и происходящие сейчас [2, 3].

Сложившийся к настоящему времени ландшафт направления компьютерных обучающих систем (EdTech в англоязычной терминологии) наиболее полно представлен в [4]. В докладе обсуждаются основные типы таких систем, приводится их классификация. Уникальными или специфическими именно для этого рынка являются системы, помогающие эффективно усваивать необходимый учебный материал. Именно такие системы рассматриваются ниже.

2. Необходимые условия персонификации процесса обучения

Несоответствие запроса современной экономики на массовое обучение новым специальностям, навыкам, знаниям, вызванное быстрой сменой технологий, с одной стороны и предложения образовательных институтов — с другой, является главным источником инноваций и роста в этой области. Инструменты обучения практически не менялись со времени начала массового использования книг, то есть на протяжении нескольких веков. Массовая доступность компьютеров и развитие ИТ технологий создают новую инфраструктуру, сравнимую по своему влиянию на повседневную жизнь с началом массового использования книг. Поэтому естественным является вопрос: что является необходимыми условиями для использования особенностей новой технологической среды в обучении?

В докладе процесс обучения рассматривается как управление в рамках стандартной кибернетической системы. С этой точки зрения мы должны: (1) иметь измеримую цель, (2) понимать что происходит с объектом управления (уметь измерять нужные нам параметры) и (3) иметь варианты воздействия на объект. Информация (1) и (2) поступает в модель, которая выбирает/ генерирует управляющее воздействие (3). В зависимости от доступных измерений и изученности предметной области, могут быть разные модели. Так, для технических систем это обычно уравнения, описывающие процесс. В нашем случае нет такого рода моделей, но есть специалисты, систематически решающие эту задачу — педагоги. В такой ситуации наиболее предпочтительными (часто — единственно возможными) является модели управления на основе нечетких правил [7]. Именно они и рассматриваются ниже.

3. Возможности персонификации процесса обучения

Нечеткое управление базируется на правилах типа «Если..., то...». Примером может быть правило «Если ученик слабый и занятия идут долго, то не давать ему сложное задание». Для применения таких правил мы должны уметь ранжировать задания (сложнее — проще) и классифицировать учеников (слабый — сильный). Если мы не сможем это

делать с приемлемым качеством, персонификация обучения становится невозможной.

Ниже приводятся результаты ранжирования заданий и кластеризации учеников на базе данных компании «Учи.ру» (www.uchi.ru). Система Учи.ру собирает набор необходимых измерений (время решения задачи, количество ошибок, количество правильных ответов); к настоящему времени собрана статистика по 25 миллионам решений, что более чем достаточно для тестирования и настройки алгоритмов. Контент системы Учи.ру составляет 1076 карточек, полностью разработанных в компании.

3.1. Ранжирование заданий

Критерием для ранжирования является сложность задачи. Для этого используем широко применяемую Item Response Theory (IRT) [8, 9]. Реализация IRT имеется в библиотеке `ltm` языка R [10]. Входные данные — результаты решения всех задач всеми учениками в формате, представленном в табл. 1 : где X_{ij} — величина, принимающая значения из множества $0,1$. Ноль — в случае если ученик j допустил менее 6 ошибок при решении задачи i , и 1 — в противном случае. Результат моделирования может быть представлен в виде множества характеристических кривых (Item Characteristic Curve / ICC) для всех задач (каждая кривая — отдельная задача — рис. 1). Результатом моделирования является эвристическая оценка сложности задания, то есть искомый параметр, на основе которого можно провести кластеризацию задач, и использовать это разбиение для правых частей «Если ... то ...» правил (раздел 3).

ID ученика	Задача 1	Задача 2	Задача 3	...	Задача N
1	X11	X21	X31	...	X1N
...
M	X1m	X2m	X3m	...	XNm

Таблица 1. Формат представления данных для IRT.

3.2. Кластеризация учеников

В системе накапливается информация о качестве и времени решения карточек учеником. Это позволяет использовать алгоритмы кластеризации для разбиения их на группы в шкале «слабый — сильный». Мы использовали реализацию алгоритма `smeans` (библиотека `e1071` R). На

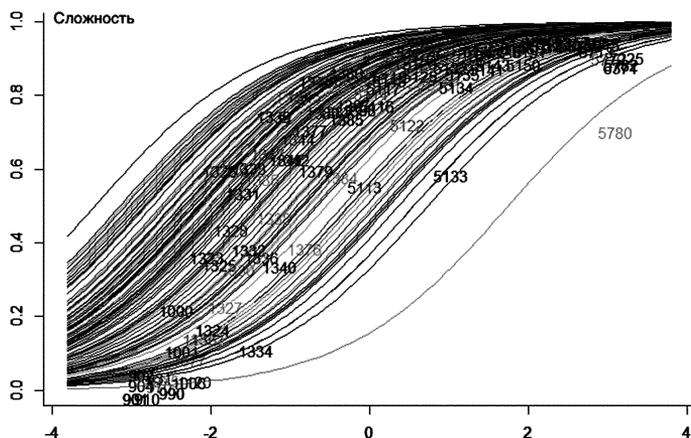


Рис. 1. Item Characteristic Curve.

вход подавались данные о количестве ошибок 31157 учеников. Решается задача выбора оптимального количества кластеров. Для этого подсчитывается степень нечеткости [12] и дисбаланс кластеров по формуле $(\max(size) - \min(size)) / \max(size)$, где $size$ — массив с размерами кластеров. Результаты собраны в табл. 2.

Количество кластеров	Степень нечеткости	Дисбаланс
3	0.268	0.813
4	0.28	0.927
5	0.28	0.887
6	0.29	0.858
7	0.28	0.906
8	0.3	0.942

Таблица 2. Характеристики качества кластеризации для разного количества кластеров.

Пример кластеризации с оптимальным количеством кластеров (3) представлен на рис. 2.

4. Выводы

Персонификация обучения на основе современных ИКТ сможет разрешить важное противоречие современной экономики — потребностью в

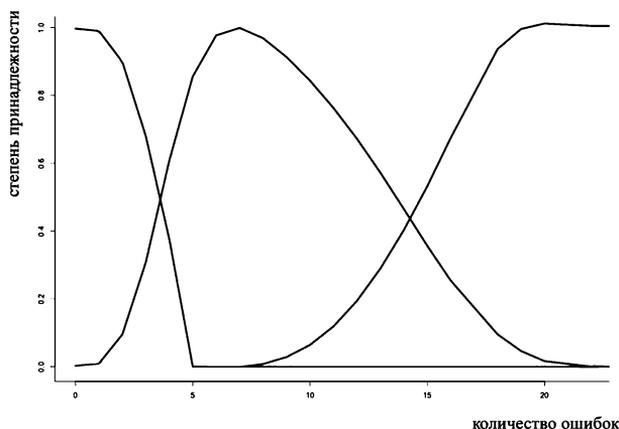


Рис. 2. Оптимальная кластеризация учащихся.

массовом переобучении миллионов людей и практическим отсутствием современных инструментов для этого. Для разработки систем компьютерного обучения, обеспечивающих его персонализацию, необходимо как минимум уметь измерять процесс усвоения материала и сложность заданий. Проведенные эксперименты на реальных данных позволяют сделать вывод о возможности разработки таких систем, имеющих потенциал качества обучения с репетитором (следствие Теоремы Кошко [11]).

Список литературы

- [1] Алисейчик П. А., Вашик К., Кнап Ж., Кудрявцев В. Б., Строгалов А. С., Шеховцов С. Г. Компьютерные обучающие системы // Интеллектуальные системы. — 2004. — Т. 8, вып. 1–4. — С. 5–44.
- [2] A gallery of disruptive technologies. — [Эл. ресурс] http://www.mckinsey.com/assets/dotcom/mgi/slideshows/disruptive_tech/index.html#
- [3] Manyika J., Chui M., Bughin J., Dobbs R., Bisson P., Marrs A. Disruptive technologies: Advances that will transform life, business, and the global economy. — McKinsey Global Institute (MGI), May 2013. — [Эл. ресурс] http://www.mckinsey.com/insights/business_technology/disruptive_technologies

- [4] Ed Tech Market Map by Flybridge Capital Partners on 11 January 2016. — [Эл. ресурс]
<https://prezi.com/xguky7u7aur6/ed-tech-market-map/>
- [5] Рыжов А. П. Некоторые задачи оптимизации и персонификации социальных сетей. — Saarbrücken: LAP, 2015.
- [6] Вахов А. Н., Зотова Е. А., Коломоец И. В., Рыжов А. П., Шварц А. Ю. Рынок компьютерных обучающих систем: состояние, перспективы, вызовы // International Journal of Open Information Technologies. — 2016. — Vol. 4, no. 1. — С. 25–30. — ISSN: 2307–8162.
- [7] Нечеткие множества в моделях управления и искусственного интеллекта / Под ред. Поспелова Д. А. — М.: Наука, 1986.
- [8] Chong Ho Yu. A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling. — 2013. — [Эл. ресурс]
<http://www.creative-wisdom.com>
- [9] de Ayala R. J. The Theory and Practice of Item Response Theory. — The Guilford Press, 2009.
- [10] [Эл. ресурс] <https://cran.r-project.org/web/packages/ltm/index.html>
- [11] Kosko B. Fuzzy Systems as Universal Approximators // IEEE Transactions on Computers. — 1994. — Vol. 43, No 11. — P. 1329–1333.
- [12] Рыжов А. П. Элементы теории нечетких множеств и измерения нечеткости. — М.: Диалог–МГУ, 1998. — [Эл. ресурс]
<http://intsys.msu.ru/staff/ryzhov/FuzzySetsTheoryApplications.htm>

