

Б. А. Кондратенко, А. Б. Кондратенко

ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ АНАЛИЗА БОЛЬШИХ ДАННЫХ В СОВРЕМЕННОМ ОБРАЗОВАНИИ

Раскрываются возможности и перспективы использования анализа больших данных в образовании в условиях быстрого развития технологий. Анализируются пути модернизации образования, которые станут возможными благодаря массовому внедрению технологий анализа данных и свободному обращению к неструктурированным данным. Предлагаются приемы и инструменты, которые в будущем могут помочь разрешить существующие на сегодняшний день в системе образования трудности, и рассматриваются некоторые внушающие опасения негативные последствия массового внедрения больших данных в цифровом обществе.

117

The paper describes possibilities and prospects of big data analysis in education given a rapid development of technology. The author explores new approaches to the modernization of the system of education, which are based on big data technologies, including unstructured data. The article offers techniques and tools that could help to solve the existing problems in education and describes some negative consequences of the massive use of big data in modern digital society.

Ключевые слова: современное образование, влияние на образование, большие данные, последствия массового анализа данных, перспективы.

Key words: modern education, impact on education, big data, mass data analysis effects, perspectives.

Начать изучение темы внедрения больших данных в современное образование стоит с конкретизации самого понятия. Что же такое большие данные и зачем их анализировать? К сожалению, устоявшийся понятийный аппарат в данной области еще не сформирован, но содержательные характеристики различных определений понятия «большие данные» можно условно представить следующим образом: большие данные — это совокупность подходов, инструментов и методов, предназначенных для накопления и анализа неструктурированных данных с целью извлечения из них нового знания [5].

При этом важно сделать два уточнения. Во-первых, такие подходы, инструменты и методы предназначены именно для быстрой (относительно других методов) обработки баз данных большого объема со скоростью, максимально приближенной к реальному времени [17]. Во-вторых, из-за своего объема и структуры (а точнее, ее отсутствия) большие данные практически невозможно обрабатывать при помощи традиционных программных и аппаратных средств анализа информации, а без обработки и анализа они оказываются бесполезными или практически бесполезными [13]. Два этих факта и послужили толчком к созданию для больших данных собственного инструментария и технологий хранения и извлечения знаний из них.



Большие данные — специфическая технология, которая дает возможность анализировать и извлекать новые знания из массивов неструктурированных данных. Еще буквально семь лет назад на носителях информации различных организаций накапливались невероятные объемы информации, применение которой было экономически совершенно нецелесообразным из-за отсутствия четкой структуры и высокой сложности извлечения полезных данных. Большие данные позволяют при помощи набора математических и статистических инструментов формировать универсальную структуру данных и извлекать из бездействовавших ранее массивов полезную информацию, применяя ее в самых разных сферах деятельности.

Подобные решения могут применяться повсеместно, в том числе и в рамках образовательного процесса. Одним из наиболее ярких примеров внедрения анализа больших данных в сфере образования является успех компании Skillsoft, которая на своей платформе предоставляет образовательные материалы для онлайн-обучения более чем 20 миллионам пользователей. Благодаря анализу больших данных компании Skillsoft удалось адаптировать предлагаемые образовательные ресурсы под каждого из миллионов своих клиентов с учетом уровня их квалификации, конкретных потребностей и бизнес-интересов. Это позволило компании достичь серьезных положительных результатов в области взаимодействия пользователей с образовательной средой: треть пользователей воспользовалась хотя бы одним из предложенных в рамках первой же электронной рассылки курсов и оплатила его; только 15 % получателей рассылок не заинтересовались рекомендациями, разосланными в рамках серии электронных писем; уровень удовлетворенности пользователей персонализированным образовательным контентом повысился более чем вдвое по сравнению с ранними не персонализированными версиями учебных курсов [14].

Еще один фактор, который, на наш взгляд, должен способствовать массовому внедрению анализа больших данных в области образования, — это значительное снижение себестоимости хранения единицы информации. Уровень развития технологий хранения информации привел к снижению стоимости хранения одного гигабайта информации на 46 % за последние 9 лет, а за последние два десятилетия — более чем на 250 % [7]. Эти изменения столь значительны, что сегодня существенная часть информации уже представлена только в цифровом виде — традиционные бумажные носители не применяются или почти не применяются (по сравнению с прошлым) во множестве отраслей. Повсеместно, в том числе и в образовательных организациях как России, так и зарубежных стран, все больше субъектов переходят к электронному документообороту, а также к сохранению максимально возможного объема информации о взаимодействии со своими партнерами и обучающимися. Появляющиеся и упрощающиеся (с точки зрения применения) новые программные и аппаратные средства анализа данных позволяют получать новые знания из массивов информации, еще пять лет назад считавшихся «информационным балластом», от которого раньше не задумываясь избавились бы. На сегодняшний день анализ



данных об обучающихся — один из основных ресурсов образовательных платформ, который позволяет выстраивать более продуктивные образовательные отношения [4].

Это подтверждается не только тенденциями в практике образовательной деятельности. Так, ряд ученых, в частности Д. Боллиер и Ч. Файрстоун, отмечает, что базы данных больше не являются просто способом хранения данных. Они превращаются в современный инструмент получения новых знаний, благодаря которому процессы адаптации и коллаборации становятся более эффективными, а принимаемые в реальном времени решения — более персонализированными, учитывающими специфику конкретных участников взаимоотношений [4].

Накопленные ранее колоссальные объемы данных, измеряемые в сотнях петабайт (если такой объем информации разместить в печатных книгах, то пространство, которое будет необходимо для их хранения в стопках высотой 50 метров, по площади окажется сопоставимым с территорией Маршалловых островов), теперь имеют реальные перспективы стать источником информации, которая станет залогом не только коммерческого успеха, но и высоких стандартов образовательных услуг для всех обучающихся, вне зависимости от особенностей личности и специальных потребностей.

Не стоит сбрасывать со счетов и другое важное изменение в технологиях обработки данных, которое обеспечивает их быструю передачу, — сеть Интернет. По данным Internet World Stats, количество людей, пользующихся ресурсами глобальной сети на постоянной основе, с 2000 по 2017 г. возросло почти десятикратно [16]. Количество пользователей сети Интернет неразрывно связано с количеством устройств, которые передают информацию по вычислительным сетям: в 2017 г. их число достигло 5 устройств на одного пользователя (всего более 20 миллиардов устройств). За одну минуту к сети подключается в среднем около пяти тысяч новых устройств и пропорционально вырастает объем передаваемой и хранимой информации. Исследователи, занимающиеся проблемами развития информационных технологий, прогнозируют увеличение этих параметров в 30 раз до 2025 г. [1].

Согласно отчету, опубликованному компанией IBM, объем цифрового контента в глобальной сети в настоящее время составляет около 8 зеттабайт (напечатанная на бумаге информация такого объема покрывает всю площадь поверхности Земли слоем толщиной в 8,5 метров). Ожидается, что это значение удвоится уже к концу 2018 г. [4].

К тому же скорость генерации новых данных также демонстрирует сопоставимые значения. В 2012 г., по приблизительным оценкам специалистов в области коммуникации, в сети Интернет ежедневно создавалось около 2,5 эксабайт информации (1 эксабайт равен 1 миллиарду гигабайт), причем предполагалось, что это значение будет удваиваться каждые 40 месяцев. Тем не менее уже в 2016 г. пользователи сети Интернет ежедневно создавали приблизительно 44 эксабайта новых данных, а к 2025 г. прогнозируется десятикратный прирост, то есть по сравнению с ожидавшимся увеличением рост составил более 550 %.



Чтобы сделать наглядными произошедшие изменения, достаточно отметить, что за одну секунду в Интернете передается больше данных, чем составлял его полный объем 15 лет назад.

Подобные изменения в области появления новой информации и возможностей ее хранения позволяют таким компаниям, как, например, самая крупная в мире сеть розничной торговли Walmart, собирать по 4 тысячи гигабайт данных о своих клиентах, их предпочтениях и сделанных покупках каждые 60 минут [6]. Совершенно справедливо предположить в такой ситуации, что и в области образования внедрение подобных прорывных технологий уже в ближайшее время позволит в режиме реального времени не только определять востребованность того или иного образовательного контента и его необходимость для каждого конкретного индивида, но и на основе анализа данных о взаимодействии последнего с образовательными и социальными платформами адаптировать под него траектории изучения материала [6].

Такие перспективы не могут не ставить амбициозную задачу по созданию новых инструментов и приемов, которые обеспечат эффективное получение информации об обучающихся, ее хранение и анализ [11].

Уже сегодня компании по всему миру получают множество разнообразных данных, связанных со своей деятельностью и с деятельностью их коммерческих и социальных партнеров. При этом (как уже демонстрировалось выше) мощность этого информационного потока столь велика, что не остается никаких иных вариантов кроме применения новых способов обработки и анализа информации, которые могут быть с успехом использованы во многих сферах: культуре, сфере услуг, розничной торговле, производстве, финансовом секторе, научных исследованиях, медицине и образовании [2]. Например, в области образования одним из наиболее часто применяемых способов принятия решений (как организационных и управленческих, так и научных и методических) был метод экспертных оценок, который при текущей скорости появления новых и обновления существующих знаний становится все менее оправданным — в этих новых условиях практически единственным эффективным инструментом принятия решений в режиме реального времени будет анализ данных, который в известной мере может быть подвергнут автоматизации.

В отчете международной консалтинговой компании McKinsey & Company, которая специализируется на решении задач, связанных со стратегическим управлением, говорится о том, что наиболее важным фактором, обеспечивающим успешность внедрения решений на базе больших данных, является их свободное обращение — так называемая концепция открытых данных [8]. Под открытыми данными принято понимать концепцию хранения данных, которая гарантирует свободный доступ к ним в машиночитаемом формате, а также отсутствие ограничений на их использование и публикацию. При этом стоит отметить, что основной объем открытых данных поступает от государственных структур и предприятий, для которых раскрытие информации обяза-



тельно в соответствии с законом. Образовательные организации различных уровней, рассматриваемые нами в рамках статьи, в значительной доле находятся в собственности государства [8].

Рассмотрим перспективные возможности применения анализа больших данных в сфере образования. В будущем анализ больших данных, собираемых организациями, а также открытых данных поможет реализовать механизм, обеспечивающий эффективное взаимодействие между педагогами и обучающимися в режиме реального времени, что даст возможность провести глубокое всестороннее изучение моделей обучения, реализуемых образовательными организациями, и произвести их оптимизацию с учетом новых знаний, доступных благодаря анализу больших данных [15]. Эти мероприятия позволят сбалансировать трудоемкость учебного процесса путем прогнозирования, кому из обучающихся необходимо больше содействия и поддержки в процессе освоения образовательных программ, чтобы избежать экстремально низкого уровня подготовки и, как следствие, отказа от обучения по причине его неэффективности [12; 15]. По нашему мнению, все это должно привести к активизации поиска новых подходов в теории и практике образования, которые будут востребованы обучающимися с особыми потребностями [15].

Рассматривая предстоящее изменение системы образования под влиянием новых технологий, важно помнить о стремительном росте роли электронных и дистанционных образовательных форм, благодаря которым люди со всего мира получают доступ к лучшим образовательным ресурсам от ведущих мировых университетов, не выходя из дома [21]. Внедрение все новых открытых образовательных ресурсов даст мировому образованию возможность аккумулировать значительные объемы данных, анализ которых, наряду с открытыми данными, поможет улучшить не только качество образования за счет превентивного удовлетворения потребностей в дополнительных учебных материалах, но и учебный опыт обучающихся [12].

На рисунке ниже продемонстрирована доля хозяйствующих субъектов в различных отраслях, которые активно внедряют решения на основе анализа больших данных или уже применяют подобные подходы в своей деятельности [3].

На графике, иллюстрирующем результаты опроса, проведенного Tech Pro Research, самое широкое применение анализ больших данных нашел в области телекоммуникаций, в проектировании производства, в государственных органах, сфере финансов, информационных технологий и логистике; в то же время в здравоохранении и образовании — областях, по нашему мнению, благодаря большому объему накопленной информации в значительной степени готовых к внедрению технологии раньше прочих, — большие данные остаются серьезно недооцененными. Количество образовательных организаций, активно внедряющих технологии работы с большими данными, составляет всего 15 % [3].

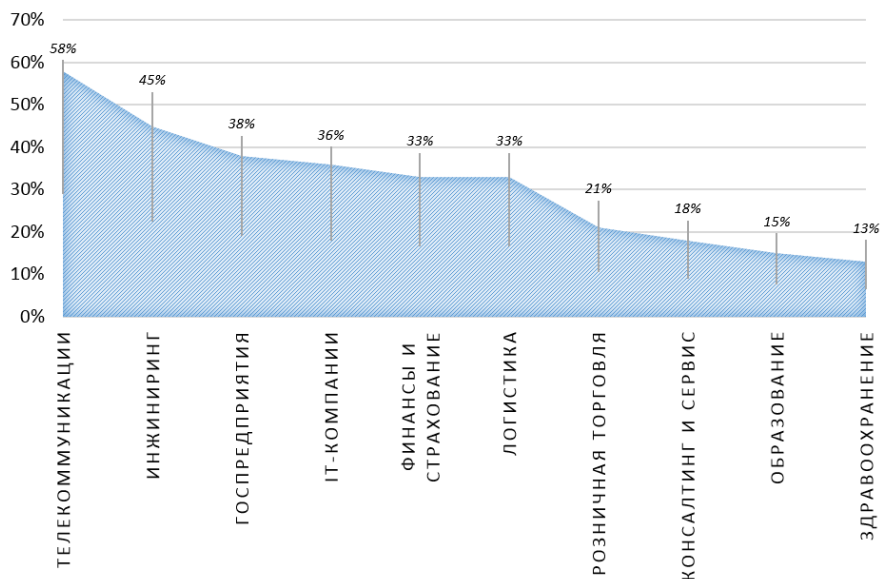


Рис. Доля компаний, применяющих аналитику больших данных

Как уже отмечалось ранее, большие данные помогают находить скрытые взаимосвязи между объектами, предметами и явлениями. Не это ли тот самый «Священный Грааль» педагогики, который позволит сделать образование более личным и повысить эффективность освоения учебного материала, поможет исследователям найти скрытые ранее связи между теми или иными личностными особенностями обучающихся и их успехами как в учебе, так и в профессиональной и общественной деятельности [10]?

Итак, анализ больших данных открывает перспективы по созданию нового положительного учебного опыта – все участники образовательного процесса смогут беспрепятственно обмениваться друг с другом информацией, взаимно и высокоэффективно расширяя свою компетентность на протяжении всей жизни [10].

Конкретизируем такие перспективы:

1. *Ранняя профессиональная ориентация.* Большие данные способны помочь обучающимся в выборе тех или иных образовательных продуктов и их наборов среди в наибольшей степени соответствующих свойствам их личности (особенностям поведения) и социальным запросам. А это потенциально позволит не только более осмысленно выбирать образовательные организации, но и обеспечит обучение и развитие способностей (обучение) в соответствии с врожденными задатками [18; 20].

2. *Новые адаптивные образовательные траектории.* Анализ больших данных, опирающийся на опыт взаимодействия человека с образова-



тельными и социальными платформами, позволит на раннем этапе определять способности и задатки обучающегося, на основе которых могут быть созданы образовательные траектории, наилучшим образом способствующие развитию необходимых компетенций конкретных обучающихся с учетом их способностей, мотивации и потребностей как общества, так и самих обучающихся [19].

3. *Контроль профессиональной траектории.* Аналитика больших данных может позволить образовательной организации отслеживать профессиональные успехи своих выпускников, ориентируясь по тысячам различных источников, начиная с их профилей и фотографий в социальных сетях и заканчивая публикациями в средствах массовой информации и на сайтах организаций. Это позволит поддерживать актуальность информации о статусе и квалификации выпускников образовательных организаций всех уровней, осуществлять поддержку сообщества выпускников при помощи наставничества и предоставления дополнительных образовательных услуг, а также выведет портфолио обучающихся на новый уровень значимости [19].

4. *Открытость и прозрачность образования.* Свободное и неограниченное обращение неструктурированных данных и доступность их для анализа и обработки откроет возможности для более глубокого вовлечения обучающихся в процессы, которые ранее в силу консервативности и традиционности образовательной системы считались прерогативой исключительно административно-управленческого персонала. Это позволит не только учитывать потребности рынка конкретного региона при обучении, но и совместными усилиями находить новые пути и решения для улучшения образовательного опыта всех участников [19].

Резюмируя вышесказанное, следует отметить, что анализ больших данных и их свободное обращение (при условии сохранения приватности) могут послужить основой для качественных изменений, формирования новой современной и динамично развивающейся системы образования. При этом сама педагогическая наука и практика получит инструмент, масштаб изменений и эффективность от применения которого можно сравнить разве что с внедрением вакцинации в здравоохранении.

В то же время невозможно предсказать все последствия массового внедрения технологий анализа свободно обращающейся информации во все сферы деятельности общества. Они могут оказать на само общество и его институты влияние, которое трудно спрогнозировать. Рассматривая эту «темную сторону» больших данных, важно выделить и ряд негативных последствий их повсеместного внедрения.

1. Опасность возникновения нового вида предиктивной дискриминации, основанной на результатах обработки больших данных с помощью несовершенных алгоритмов, выработка которых все еще подвержена влиянию человеческого фактора. Об этом говорит математик и исследователь Кэти О'Нил, называя возможные последствия «оружием математического поражения» (англ. «weapons of math destruction») [9, p. 288].



2. Большие данные неизбежно сократят количество рабочих мест, и вполне возможно, что масштабы перемен превзойдут последствия появления разделения труда, фабрик и конвейера. Радикальные перемены могут коснуться не только операторов различных услуг, таких как таксопарки (их уже сегодня активно теснят онлайн-платформы вроде Uber), но и врачей-диагностов и лаборантов (которых со временем могут заменить высокоразвитые нейронные сети), сотрудников маркетинговых отделов крупных и небольших предприятий, а возможно, даже части педагогов, занимающихся простой трансляцией знаний [5].

3. Возможность технократической сегрегации, то есть отделения успешных людей (обладающих возможностями и навыками работы с новыми технологиями анализа данных) от всех остальных (абсолютно невостребованных членов общества), что потенциально опасно с точки зрения не только морально-этической, но и социальной-экономической [19].

4. Незаконная торговля огромными объемами неструктурированных личных данных, которые ранее были практически бесполезны, а теперь могут служить основой для детального изучения людей и управления их поведением [9].

Важно понимать, что большие данные действительно могут послужить причиной изменения — без преувеличения — самого подхода к информации на уровне всего человечества, а не только сфер маркетинга услуг и образования, поскольку велика вероятность того, что все без исключения сферы деятельности возьмут на вооружение новые методы и инструменты.

Новая эпоха сегодня находится на этапе экспериментов, оформления и становления. Еще предстоит разрешить немало вопросов и найти способы нивелирования тех негативных последствий повсеместного внедрения анализа данных, которые были описаны нами ранее. Но нет сомнений, что образование как один из инструментов трансляции культурного и познавательного потенциала между поколениями должно найти максимально эффективные способы повсеместного применения анализа больших данных в новой образовательной парадигме.

Список литературы

1. *152,000 Smart Devices Every Minute In 2025: IDC Outlines The Future of Smart Things* // Forbes. URL: <https://www.forbes.com/sites/michaelkanellos/2016/03/03/152000-smart-devices-every-minute-in-2025-idc-outlines-the-future-of-smart-things/#476706b24b63> (дата обращения: 07.11.2017).
2. *Sagiroglu S., Sinanc D. Big data: A review* // International Conference on Collaboration Technologies and Systems (CTS 2013). San Diego, 2013. P. 42–47.
3. *Big Data* // Tech Pro Research. URL: <http://www.techproresearch.com/topic/big-data> (дата обращения: 07.11.2017).
4. *Bollier D., Firestone C.M. The promise and peril of big data* // Communications and Society Program. Washington, 2010.



5. Cringely R. X. Thinking about Big Data. URL: <http://www.cringely.com/2016/07/05/thinking-big-data-part-one> (дата обращения: 07.11.2017).
6. Fayyad U., Piatetsky-Shapiro G., Smyth P. The KDD process for extracting useful knowledge from volumes of data // Commun. ACM. 1996. Vol. 39, №11. P. 27–34.
7. Hard Drive Cost Per Gigabyte // BackBlaze. URL: <https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte> (дата обращения: 07.11.2017).
8. Manyika J., Chui M., Farrell D. et al. Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute, 2013. URL: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information> (дата обращения: 07.11.2017).
9. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. N. Y., 2017.
10. Podesta J., Pritzker P., Moniz E. et al. Big data: seizing opportunities, preserving values / Executive Office of the President of United States of America, May 2014. URL: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf (дата обращения: 07.11.2017).
11. Rands K. How big data is disrupting education. URL: <https://www.cio.com/article/3200685/big-data/how-big-data-is-disrupting-education.html> (дата обращения: 07.11.2017).
12. Siemens G., Gasevic D. Guest Editorial-Learning and Knowledge Analytics // Educ. Technol. Soc. 2012. Vol. 15, №3. P. 1–2.
13. Siemens G., Long P. Penetrating the fog: Analytics in learning and education // Educ. Rev. 2011. Vol. 46, №5. P. 30–32.
14. Skillsoft's Big Data Initiative Delivers Improved Learning Engagement and Outcomes // Skillsoft – eLearning: Online training. URL: http://www.skillsoft.com/about/press_room/press_releases/november_18_14_ibm.asp (дата обращения: 07.11.2017).
15. West D.M. Big Data for Education: Data Mining, Data Analytics, and Web Dashboards / Brookings Institution. Reuters, 2012.
16. World Internet Users Statistics and 2017 World Population Stats // Internet world stats. URL: <http://www.internetworldstats.com/stats.htm> (дата обращения: 07.11.2017).
17. Yan J. Big Data, Bigger Opportunities. Conference on Data Engineering Work (ICDEW). 2013.
18. Кондратенко А.Б., Кондратенко Б.А. Обоснования персонализации обучения студентов в информационном обществе // Вестник Балтийского федерального университета им. И. Канта. Сер.: Филология, педагогика, психология. 2016. №3. С. 83–89.
19. Кондратенко Б.А., Кондратенко А.Б. Социально-экономическое развитие регионов России: Реалии современности, тенденции, перспективы (посвящается 70-летию колледжа западного филиала РАНХиГС) // Возможности и перспективы учета персональных задатков и способностей обучающихся в педагогике XXI века. Калининград, 2016. С. 207–209.
20. Симлева И.Н., Клемешев А.П. К вопросу о моделях современной научно-образовательной политики в странах Балтийского региона // Вестник Балтийского федерального университета им. И. Канта. Сер.: Филология, педагогика, психология. 2014. №11. С. 7–13.
21. Системы образования зарубежных стран Балтийского региона в контексте инновационного развития экономики : монография / под ред. И.Н. Симеовой. Калининград, 2013.



Об авторах

Борис Анатольевич Кондратенко — канд. пед. наук, доц., Западный филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, Россия.

E-mail: BAKondratenko@gmail.com

Анатолий Борисович Кондратенко — д-р пед. наук, проф., Западный филиал Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации, Россия.

E-mail: Anatoliy_kondr@mail.ru

126

The authors

Dr Boris Kondratenko, Associate Professor, West Branch of Russian Presidential Academy of National Economy and Public Administration, Russia.

E-mail: BAKondratenko@gmail.com

Prof Anatoly Kondratenko, West Branch of Russian Presidential Academy of National Economy and Public Administration, Russia.

E-mail: Anatoliy_kondr@mail.ru