

Михнев Илья Павлович

канд. техн. наук, доцент,

Заслуженный работник науки и образования

Челнокова Анастасия Дмитриевна

студентка

Реут Александра Дмитриевна

студентка

Волгоградский институт управления (филиал)

ФГБОУ ВО «Российская академия народного хозяйства

и государственной службы при Президенте РФ»

г. Волгоград, Волгоградская область

DOI 10.21661/r-470090

ТЕХНОЛОГИИ BIG DATA И ИХ ПРИМЕНЕНИЕ В СФЕРЕ СОВРЕМЕННОГО ВЫСШЕГО ОБРАЗОВАНИЯ

Аннотация: в статье рассмотрена значимость технологии Big Data на современном этапе развития, проанализированы перспективы применения в сфере высшего образования. Показаны результаты исследований применения технологий Big Data. Представлены современное состояние, тенденции развития технологий Big Data и примеры компаний, внедривших технологии больших данных в свою деятельность.

Ключевые слова: большие данные, технологии Big Data, прогнозирование, информация, анализ данных, модели больших данных, высшее образование.

Технологии «Большие данные» (Big Data) способны превратить современное высшее образование в продукт с отличным предсказуемым результатом, а преподавателя ВУЗа – в суперпрофессора. Анализ Big Data позволяет ускорить решение различных исследовательских, научных и педагогических недочетов. При изучении статистики, можно работать и с индивидуальными векторами, и с образовательными системами глобального уровня. Технологии Big Data помогают значительно улучшить так называемый

«дизайн педагога» – в тех местах, где студент «засыпает от скуки», такую систему образования лучше заменить. Эта технология автоматизирует поведение образовательной системы и дает фактические рекомендации и подсказки, если студенты застревают на одном месте. И если эти рекомендации и подсказки не помогают, – технология оповещает преподавателей и родителей, что таким студентам нужна дополнительная помощь [1; 5–8].

Технология Big Data снимает со студента различные данные и анализирует, как конкретный студент учится. Фиксирует, где студент ошибается, что решает медленно, что решает быстро, когда отвлекается и составляет полный детальный портрет обучаемого: сколько времени и на какие действия потратил, правильно решил или нет, сколько перемещал мышкой по экрану, сколько раз возвращался к решению одной и той же задачи. Big Data помогает обработать опыт сотен тысяч преподавателей и студентов, и на основе анализа получить эффективную образовательную методику. Такая образовательная методика становится продуктом массового опыта. С помощью технологии Big Data можно персонализировать контент под потребности каждого студента. Например, Big Data анализирует сотни тысяч текстовой информации в глобальной сети и подбирает тот текст, который содержит необходимое количество нужных к изучению новых слов и фраз [2; 3–5].

Используемые методы в анализе Big Data, родом из распознавания образов, компьютерного обучения, статистики и психометрии. В современном высшем образовании большие данные пока не такие уж и большие. Существующие сегодня специальные репозитории «DataShops», позволяют собирать данные и прямо там их анализировать. На сегодняшний день в самом популярном репозитории «PSLC DataShop», собрана и хранится информация, обработанная более чем за 260 000 часов, проведенных студентами в образовательных программах, – это примерно 55 миллионов действий, ответов и различных результатов [3–6; 8].

Одна из самых интересных моделей работы с Big Data – это прогнозирование, где комбинация известных данных позволяет обеспечить

прогноз искомого неизвестного. Множественные данные собираются из записей интернет-сервисов, студенческих систем, опросов, социальных сетей и различных наблюдений во время экспериментов. Сбор и обработка подобных данных, это огромное дело, так как нужно знать, на какие моменты смотреть, и уметь выявлять нужную полезную информацию. Модель может работать для прогнозирования настоящего, используя статистику за прошедший час, узнавая, интересно ли сейчас студенту смотреть онлайн-курс, или прогнозирования будущего (используя предыдущие оценки), сможет ли студент решить следующую задачу и с каким результатом. Современные алгоритмы технологии Big Data принимают в расчет цену ошибки и эффективность правильного использования образовательной системы. Например, если за одну минуту студент усваивает 0,05% изучаемого курса, то неправильный прогноз «стоит» ему одну лишнюю минуту обучения, а правильный добавляет 0,03% [4; 6–8].

Министерство труда Германии использует технологию Big Data в работе, связанной с анализом поступающих заявок на выдачу пособий по безработице. Проанализировав информацию, стало понятно, что 25% пособий выплачивалось незаслуженно. С помощью технологии Big Data министерство труда сократило расходы на 10 млрд евро. Big Data – обозначение данных огромных объемов и значительного многообразия, эффективно анализируемых горизонтально масштабируемыми программными инструментами. Они были созданы в конце 2000-х годов в качестве замены устоявшимся системам управления базами данных и решениям класса Business Intelligence [5; 6; 8].

Термин Big Data относится к наборам данных, габариты которых превосходит возможности классических баз данных по хранению, управлению и анализу информации. В настоящее время множество компаний следят за развитием технологий Big Data. Благодаря отчету компании IDC в 2012 г. «Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East», появились сведения о том, что в ближайшее время объемы информации будут удваиваться каждые два года. За ближайшие семь лет на каждого жителя планеты будет приходиться более чем 5200 Гигабайт данных. В качестве главных

характеристик для Big Data выделяют VVV «три V»: (*volume*, *velocity*, *variety*), то есть физический объём, скорость и многообразие. Под скоростью понимают скорость прироста, высокоскоростную обработку для получения результатов, а под многообразием – возможность одновременно обрабатывать различные типы структурированных и полуструктурированных данных. В дальнейшем возникли различные вариации этого признака. Это «четыре V» (добавлялась *veracity* – достоверность, использовалась в рекламных материалах IBM), «пять V» (в этом варианте прибавляли *viability* – жизнеспособность, и *value* – ценность), и даже «семь V» (добавляли также *variability* – переменчивость, и *visualization*).

Особенностями технологий Big Data являются:

- обработка различной информации в огромных масштабах;
- обновление информации в многообразных источниках;
- метод, отличающийся по качеству открывающей аналитики для выявления практических знаний, которые монетизируются в прибыль;
- наглядное отображение отчетов и возможности сценарного анализа («что, если...»);
- улучшение качества работы, создание новых продуктов и повышение конкурентоспособности.

Объёмы различной и быстро поступающей цифровой информации обработать традиционными инструментами невозможно. Сама обработка материала позволяет обнаружить определённые и незначительные закономерности, которые не может заметить человек. Это позволяет улучшить все сферы нашей жизни – от государственного управления до производства и телекоммуникаций. Например, несколько лет назад некоторые компании защищали своих клиентов от мошенничества, что влияло на доверие и прибыль компаний со стороны клиентов. В настоящее время предприятия создают большое количество неструктурированных данных, в которые входят видеозаписи, текстовые документы, таблицы, графики и т. д. Весь этот материал находится во множестве репозиториев, иногда даже за пределами той или иной

организации. Компании могут иметь доступ к собственным материалам и не иметь необходимых инструментов, которые могли бы установить корреляцию между этими материалами и сделать на их основе выводы. Классические методы анализа информации не успевают за масштабными объемами постоянно растущих и меняющихся данных, за что в итоге и принимаются технологии Big Data. Количество информации на предприятии растет за счет данных, полученных с измерительных устройств. Самыми перспективными устройствами считаются датчики, которые могут передавать информацию в режиме реального времени. Все устройства на предприятии с их помощью могут быть объединены в сеть, а технологии больших данных позволяют анализировать информацию, исходящую от них, и проводить необходимые процедуры в автоматическом режиме. Благодаря датчикам, предприятия могут ежеминутно получать информацию о состоянии своего оборудования и на основе этих материалов предсказывать наиболее благоприятное время для изменения и обслуживания. Слишком ранняя замена приведет к дополнительным расходам, а поздняя – к утрате прибыли вследствие простоя оборудования. В настоящее время технологии Big Data нашли свое применение во многих отраслях: банкинг, здравоохранение, сфера производства. Благодаря большим данным можно: улучшить производительность, выявить недостатки и совершиТЬ ревизию продукции, улучшить удобства использования продукта на основе поведения потребителей. Рассмотрим примеры зарубежных и отечественных компаний, внедривших технологии больших данных в свою деятельность. Зарубежные компании: «Apple». Цель внедрения Big Data: использование материалов о поведении потребителей для улучшения качества продукта. Благодаря своей популярности технические средства Apple довольно качественно собирают ценную информацию. Поэтому эта компания имеет огромное количество данных о том, как потребители используют устройства фирмы Apple. Пользователь дает компании знать, что именно требует исправления, и каким должен быть дизайн и характеристики последней версии устройства. Так же благодаря почти индивидуальному подходу с помощью специальных приложений и Apple Watch

пользователь может с легкостью следить за здоровым сном, употреблением пищи, тратой калорий и т. д. Зарубежная компания «Intel». Цель внедрения Big Data: уменьшение себестоимости продукции. Фирма Intel производит компьютерные компоненты, в том числе, микропроцессоры, каждый из которых, перед выходом на рынок, должен пройти более 18000 проверок. Анализируя материал по всему производственному процессу, аналитические программы способны устанавливать, какие испытания проводить не требуется, оставляя лишь самые главные проверки. Таким образом, значительно сократилось время тестирования микропроцессоров, а также затраты на проведение тестов. Результат: экономия трех миллионов долларов на одной линейке процессоров Intel Core. Отечественные компании: «Газпром нефть». Цель внедрения Big Data: установление причин сбоя работы оборудования. «Газпром нефть» совместно с Teradata (американское предприятие специализируется на программно-аппаратных комплексах для обработки материалов) осуществили проект внедрения предиктивной аналитики в процессы регулирования электроцентробежными насосами. Задачей проекта стало обнаружение источников сбоя автоматического перезапуска насосов после аварийного разъединения с электропитанием. В процессе диагностики были использованы свыше 200 млн. записей с контроллеров систем управления на тысячи скважинах, в результате были созданы визуализированные модели цепочек событий, влияющие на процесс автозапуска насосов и карты вероятностного распределения причинно-следственных связей. Результат: получение материалов о ранее неизвестных взаимосвязях в работе насосного оборудования и устранение возникших неполадок. Цель внедрения Big Data в «Сбербанк»: управления рисками, предотвращение случаев с мошенничеством. В 2014 году была введена система использования кейса АС САФИ. Она сравнивала фотографии из базы, которые поступали туда с веб-камер для идентификации клиентов. Основа этой системы была биометрическая платформа. Результат: происшествия с мошенничеством уменьшились в десять раз [1; 4–7].

На сегодняшний день ещё не достаточно хорошо решены задачи использования Big Data. Один из примеров – задачи, связанные с образованием: автоматический анализ успеваемости, выработка индивидуальных программ и рекомендаций, прогнозирование показателей, социологические исследования коллективов студентов. В нашей стране пока ещё даже нет доступных и удобных массивов информации для решения подобных задач. Очень хочется надеяться, что у нас к Big Data будут относиться с должным вниманием. Что касается методов, которые здесь используются, то они достаточно стандартные для математика-специалиста в анализе данных. Как ни удивительно, с точки зрения статистики, результаты ответов на вопросы тестов очень похожи, например, на степень удовлетворённости просмотром фильмов. Поэтому не важно, что рекомендовать: фильмы для просмотра или темы для повторения – механизм рекомендаций одинаков.

Список литературы

1. Моррисон А. Большие Данные: как извлечь из них информацию / А. Моррисон [и др.] // Технологический прогноз. Ежеквартальный журнал. – 2010. – №3. – С. 22–29.
2. Михнев И.П. Информационная безопасность в современном экономическом образовании // Международный журнал прикладных и фундаментальных исследований. – 2013. – №4. – С. 111–113.
3. Михнев И.П. Обучение и контроль знаний студентов с помощью UniTest // Фундаментальные исследования. – 2008. – №1. – С. 94–95.
4. Банько Ю.А. Современные компьютерные угрозы: что реально угрожает бизнесу? / Ю.А. Банько, А.М. Кокорева; науч. рук. И.П. Михнев // Приоритетные направления развития образования и науки: Материалы IV Междунар. науч.-практ. конф. (Чебоксары, 24 дек. 2017 г.) / Редкол.: О.Н. Широков [и др.] – Чебоксары: ЦНС «Интерактив плюс», 2017. – С. 169–171.
5. Михнев И.П. Информационная безопасность на просторах мобильного интернета // Образовательные ресурсы и технологии. – 2015. – №4(12). – С. 66–70.

6. Черняк Л. Большие Данные – новая теория и практика // Открытые системы. СУБД. – 2011. – №10. – С. 36–41.

7. Что такое Big data: собрали всё самое важное о больших данных [Электронный ресурс]. – Режим доступа: <https://rb.ru/howto/chto-takoe-big-data/> (дата обращения: 11.03.2018).

8. Big Data: проблема, технология, рынок [Электронный ресурс]. – Режим доступа: <http://compress.ru/Article.aspx?id=22725> (дата обращения: 11.03.2018).