

ИССЛЕДОВАНИЕ НАУЧНОГО ИНТЕРЕСА К БОЛЬШИМ ДАННЫМ В ОБРАЗОВАНИИ

Оспанова У.А.

uljansbox@mail.ru

Докторант, специальность «Регионоведение»
ЕНУ им. Л.Н.Гумилева, Нур-Султан, Казахстан

Научный руководитель – Ж. Е. Нурбаев

Цифровизация всех сфер жизнедеятельности человека как глобальный тренд набирает все большие обороты: в информационный век генерируется рекордное количество данных. Так согласно докладу Дж. Дежардана представленному на Мировом экономическом форуме 2019, на начало 2020 г. примерный объем всех сгенерированных данных в мире составил 44 зеттабайта [1]. В связи с этим особую актуальность приобретает проблема недостаточной реализации потенциала этих данных, извлечения практической пользы из имеющихся огромных информационных ресурсов: отмечалось, что в мире анализируется лишь 0,5% данных [2].

Аналитика больших данных, продемонстрировавшая огромную эффективность в области маркетинга, бизнеса, банковского дела и медицины, в настоящее время трансформирует методологию науки, в том числе и социальных наук. Знаменитое эссе К. Андерсона «Конец теории. Поток данных делает научный метод устаревшим», в котором отмечалось, что модели, объясняющие причинно-следственные связи между переменными, как и традиционный подход к научному исследованию, подразумевающий логическую цепочку «выдвижение гипотезы – моделирование – тестирование гипотезы» стали устаревшими, дало толчок многочисленным дебатам в научном сообществе [3]. И хотя такое заявление, отрицающее важность причинно-следственных связей, справедливо подвергается критике в академических кругах, трансформация логики исследования, при которой первоначально происходит анализ данных и поиск закономерностей, и лишь за тем следует поиск объяснений и выявление причинно-следственных связей, все же проникает в науку из бизнес-среды.

В сфере образования интерес к большим данным также растет. Исследовательской областью, являющейся центральной для настоящей статьи, является динамика научного интереса к большим данным в области образования. Ответы на исследовательские вопросы об изменении научного интереса к большим данным в области образования в последние 10 лет, и о трендах, прослеживающихся в 100 наиболее цитируемых научных статей Web of Science Core Collection, в дисциплинарной области «Образование и исследования в области образования». Источником данных для исследования стала база данных Web of Science Core Collection компании Clarivate Analytics, являющаяся одной из самых авторитетных баз данных научных исследований в мире [4]. Мы проанализировали данные базы научных статей по информационному запросу «большие данные» + «образование». Для более качественного объяснения трендов, выявленных в ходе анализа базы научных статей, дополнительно был проведен экспертный опрос 40 казахстанских ученых, специализирующихся на образовании (лица, имеющие степень не ниже магистерской по специальностям педагогика и менеджмент в образовании). Респондентам в формате интервью были заданы вопросы, способные объяснить выявленные в ходе анализа базы данных Web of Science Core Collection тренды и их причины.

За период с 2010-2020 гг. во всех изданиях, входящих в базу данных Web of Science Core Collection было опубликовано 3 227 статей, в теме которых были указаны слова «большие данные» + «образование».

Динамика объемов научных работ, посвященных большим данным в области образования, как видно из Рис. 1, показывает резкий рост с 2014 г. Данные за 2020 г. актуальны на 15 августа, что может объяснить спад, пришедшийся на этот период. Необходимо также отметить, что респонденты опроса (34 из 40) на просьбу высказать предположения о причинах такого спада, помимо неполных данных за год (самый популярный ответ: 39 из 40) отметили, замедление процесса публикации или отсутствие возможности проводить исследования под влиянием коронавирусной пандемии. Среднегодовой темп прироста количества публикаций составил 30,51% без учета 2020 г., и 22,02% с учетом 2020 г.



Рисунок 1. Количество статей на тему больших данных в образовании в период с 2010-2020 гг.

Динамика количества цитирований, представленная на Рис.2, демонстрирует схожие тенденции. Однако снижение количества цитирований в 2020 г. не такое резкое, как снижение количества опубликованных работ.



Рисунок 2. Количество цитирований статей на тему больших данных в образовании в период с 2010-2020 гг.

Анализ дисциплинарных областей, в которых представлены научные о больших данных в образовании согласно категоризации Web of Science за период с 2010 по 2020 гг. (см. Табл. 1), показывает, что количество научных работ посвященных большим данным в образовании в областях связанных с образованием и педагогическими дисциплинами (выделение синим), уступает объему научных исследований посвященных большим данным в образовании в области компьютерных и информационных наук (выделение красным). Данная тенденция объясняется тем, что инструментарий аналитики больших данных подразумевает либо наличие навыков программирования, либо умение работать со статистическими программами для

анализа больших данных. Количество ученых в области образования, владеющих им крайне низко за счет новизны самого концепта: хотя словосочетание «большие данные» впервые было упомянуто немногим ранее, современные авторы сходятся во мнении, что сам концепт в его классическом понимании был впервые раскрыт Ф. Дайбодом в 2000 г. [5]. Многие университеты стали внедрять аналитику больших данных в рамках курсов по специальностям Data Science и специальностям, связанным с компьютерными науками, однако программы обучения для специализирующихся в области образования, либо не включают курсов по обучению аналитике больших данных, либо в редких случаях проводят сокращенные курсы факультативно. К примеру, даже программа курса повышения квалификации по специальности «Менеджмент образования» Гарвардского университета не включает такого модуля [6]. Проведенный нами мини-опрос показал, что лишь 1 человек из 40 опрошенных владеет инструментарием необходимым для анализа больших данных, и он же отметил, что им владеют его коллеги по месту работы. Предположительно, при большей выборке, процент владеющих им был бы ниже.

Таблица 1. Дисциплинарные категории научных статей, посвященных большим данным в области образования

| Дисциплинарные категории | Кол-во | % |
|---|--------|----------|
| Образование и исследования в области образования | 819 | 25,380 % |
| Методология и теория компьютерных наук | 330 | 10,226 % |
| Компьютерные науки. Информационные системы | 277 | 8.584 % |
| Электрическая инженерия и электроника | 236 | 7.313 % |
| Компьютерные науки. Междисциплинарные прикладные исследования | 214 | 6.632 % |
| Междисциплинарные социальные науки | 188 | 5.826 % |
| Образовательные научные дисциплины | 184 | 5.702 % |
| Управление | 174 | 5.392 % |
| Экономика | 148 | 4.586 % |
| Компьютерные науки. Искусственный интеллект | 139 | 4.307 % |
| Общественная гигиена охраны труда | 114 | 3.533 % |
| Бизнес | 107 | 3.316 % |
| Инженерное дело. Мультидисциплинарные исследования | 102 | 3.161 % |
| Информационные науки. Библиотечное дело | 99 | 3.068 % |
| Телекоммуникации | 90 | 2.789 % |
| Науки об окружающей среде | 76 | 2.355 % |
| Мультидисциплинарные науки | 74 | 2.293 % |
| Компьютерные науки. Программная инженерия | 64 | 1.983 % |
| Медицина и здравоохранение | 62 | 1.921 % |
| Общая медицина | 56 | 1.735 % |
| Экология. Технологии устойчивого развития | 51 | 1.580 % |
| Психология. Мультидисциплинарные исследования | 49 | 1.518 % |
| Сестринское дело | 48 | 1.487 % |

| | | |
|---|----|---------|
| Компьютерные науки. Архитектура аппаратного обеспечения | 42 | 1.302 % |
| Психология образования | 39 | 1.209 |

Далее в нашем исследовании мы сосредоточились на анализе содержания 100 наиболее цитируемых работ в дисциплинарной категории «Образование и исследования в области образования». Наблюдения, сделанные в ходе анализа, представлены ниже.

Исследовательская часть крайне малого количества работ (15 из 100) содержит либо фактический анализ больших данных, проведенный авторами, либо представляет способ совершенствования методологии анализа больших данных, основана на инструментарии анализа больших данных. Так, крайне интересная работа П. Дайвера и И. Мартинеза, посвященная аналитике данных массовых открытых онлайн курсов для изучения феномена прокрастинации, в исследовательской части опирается на классические статистические за исключением геолокации, методы [7]. И хотя аналитика больших данных опирается именно на статистику и математику, потенциал новых методов таких как машинное обучение, глубокое обучение в данной, без сомнения, выдающейся работе не раскрыт.

Статьи зачастую являются обзорными, систематизируют результаты других работ либо описывают кейсы работы с большими данными. К примеру, наиболее цитируемая статья Б. Дэниела является классическим обзором, основывается на описании результатов аналитических отчетов или работ других авторов, описывает отдельные кейсы для подкрепления положений [8].

В большой части статей данной категории большие данные и их анализ лишь заявлены в качестве ключевых слов, в статье анализ больших данных упоминается опосредовано, либо статьи не имеют отношения к большому анализу данных никакого отношения.

При этом зачастую те немногие статьи, предметом исследования которых фактически являются большие данные в образовании, посвящены обсуждению перспектив и вызовов. Еще одной популярно под-тематикой является кейсовые исследования или инсайты, одним из которых может послужить статья М. Худа и соавторов, описывающая инсайты инновационной среды для обучающих онлайн-ресурсов [9].

Анализ 100 наиболее цитируемых работ во всех категориях совокупно показал, что в основном статьи посвященные совершенствованию методологии анализа больших данных, имеющие высокую научную ценность (новизна и высокий уровень цитируемости), а также статьи, исследовательская часть которых опирается на фактический анализ или работу с большими данными публикуются в дисциплинарных категориях, связанных с компьютерными науками. Более того, цитируемость статей в других научных областях превышает цитируемость статей в дисциплинарной категории «Образование и исследования в области образования» в десятки раз: 1577 цитирований против 128 цитирований топовой статьи.

Из представляющих интерес результатов проведенного нами экспертного опроса, можно отметить проблему, отмеченную всеми без исключения участниками, заключающуюся в отсутствии или ограниченности доступа к данным. Также необходимо отметить, что лишь 6 из 40 действительно в общих чертах понимают концепт больших данных и методы их анализа. Для выявления этого участникам опроса было предложено выбрать правильный вариант описания

концепта и методов анализа больших данных. 6 участников опроса отметили, что ничего не знают об аналитике больших данных.

Из вышесказанного следует, что потенциал научных исследований в области аналитики больших данных с сильно ограничен за счет того, что исследователи специализирующиеся в области образования не владеют инструментарием, необходимым для проведения эмпирических исследований в данной области, а также отсутствием доступа к базам данных. Исследования, представляющие наибольшую научную ценность, могут быть реализованы за счет объединения усилий представителей науки, чей исследовательский интерес относится к таким сферам как компьютерные науки, математика и статистика, образование. Тем не менее потенциал и возможности больших данных в образовании широко признается научным сообществом и интерес к проведению научных исследований в данной области будет продолжать расти.

Список использованных источников

1. Desjardins, J. How much data is generated each day? // Доклад. Международный экономический форум, 2019. <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/> (дата обращения: 10.08.2020)
2. Regalado A. The Data Made Me Do It. // Big Data Gets Personal MIT TECHNOLOGY REVIEW, Business report. P1-2 <http://datascienceassn.org/sites/default/files/Big%20Data%20Gets%20Personal.pdf>. (дата обращения: 10.08.2020)
3. Anderson, C. The end of theory: The data deluge makes the scientific method obsolete. Эссе. 2008. <https://www.wired.com/2008/06/pb-theory/> (дата обращения: 10.08.2020)
4. База данных научных статей Web of Science Core Collection http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=F2g66kdwqPmxfDNWCMu&preferencesSaved= (дата обращения: 08.08.2020)
5. Diebold F.X. Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting. // Discussion Read to the Eighth World Congress of the Econometric Society, Seattle, August. 2000
6. Программа курса повышения квалификации по специальности «Менеджмент образования». Гарвардский университет <https://canvas.harvard.edu/courses/8304/assignments/syllabus> (дата обращения: 11.08.2020)
7. Diver, P., & Martinez, I. (). MOOCs as a massive research laboratory: opportunities and challenges. // Distance Education. 2015, Vol. 36, No. 1, P.5–25
8. Daniel, B. (). Big Data and analytics in higher education: Opportunities and challenges. // British Journal of Educational Technology. 2014. Vol 46(5), P.904–920
9. Huda M., Maseleno A., Atmotiyoso P., Siregar M., Ahmad R., Jasmi K., Muhamad N. Big Data Emerging Technology: Insights into Innovative Environment for Online Learning Resources // International Journal of Emerging Technology in Learning, 2018. Vol. 13, No 1, P.23-36