

Д.Ю. ЕЛИСЕЕВА

М.А. ПАШКОВА

Е.А. ЛАПШИНА

ФГБОУ ВО Российский государственный социальный университет
(Москва, Россия)

ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ В НАУЧНЫХ ИССЛЕДОВАНИЯХ

Научные исследования зачастую масштабны и требуют работы со множеством различной информации, а технологии Big Data позволяют обрабатывать большие объемы неструктурированных данных систематизировать их, анализировать и выявлять закономерности, которые не способен выявить человек.

Существует множество различных определений больших данных. Возможно, самая простая характеристика - это большие наборы данных, поступающие с постоянно растущим объемом и скоростью, созданные в цифровой форме и которые, могут быть проанализированы с помощью вычислительных инструментов. Отсюда можно выявить, две особенности, наиболее часто ассоциируемые с большими данными - это объем и скорость. Объем - это размер файлов, используемых для архивирования и распространения данных. Скорость обновления характеризует быстроту, с которой данные создаются и обрабатываются. Объем цифровых данных, создаваемых исследованиями, растет с головокружительной скоростью и такими методами, которые, могут быть непонятны когнитивной системе человека и поэтому требуют технологий автоматизированного анализа. [3]

Проблема может заключаться в том, что анализ больших данных по определению не способен отличить ложные корреляции от значимых и поэтому представляет угрозу для точности научных исследований. С одной стороны, это свидетельствует о тесной зависимости между объемом и качеством данных, а с другой стороны, типом и шириной исследовательских вопросов, для решения которых данные служат доказательством. [3]

Однако деятельность человека, его взаимодействие с окружающей средой отслеживаются и регистрируются с возрастающей эффективностью, создавая огромный цифровой след. Полученные в результате “большие данные” являются настоящей находкой для исследований, и с каждым днем разрабатываются все более сложные вычислительные инструменты для извлечения знаний из таких данных. Одним из примеров является использование различных типов данных, приобретённых в ходе научных экспериментов, таких как, метеорологические исследования, при которых нужно фиксировать ежеминутно множество различных измерений за

долгосрочный период времени и в последствии подвергать их анализу. Другим примером является интеграция данных о дорожном потоке, экологических и географических условиях и поведении человека по выработке мер безопасности для беспилотных транспортных средств, с тем чтобы при столкновении с непредвиденными событиями (такими, как ребенок, внезапно выскочивший на улицу в метель или гололедицу) эти данные можно было быстро проанализировать и выявить выработку соответствующей реакции (автомобиль достаточно сильно вильнул, чтобы избежать столкновения с ребенком, а также свести к минимуму риск заноса на льду и повреждения других транспортных средств). Еще одним примером является понимание работы мозга человека в разработке искусственного интеллекта, чтобы собрать достаточное количество информации о поведенческих особенностях в различных ситуациях понадобятся множество всевозможных типов данных, которые необходимо проанализировать и структурировать. В каждом из этих случаев наличие данных и связанных с ними технологий создают новые возможности для исследований и разработки инновационных форм исследования, которые оказывают преобразующее воздействие на науку в целом.

Мир состоит из множества информации, которую человек получает ежедневно, однако, не всегда эти сведения являются достоверными и необходимыми. В случае с исследователями, которые непрерывно пополняют свой багаж знаний, изучая различные материалы, результаты экспериментов, статистические данные и многое другое, поток информации для таких людей крайне велик, обрабатывать, анализировать и искать закономерности все труднее и труднее, учитывая, что все мы видим множество данных, зачастую не нужных для нас: реклама, новостная лента, социальные сети и так далее, все это снижает эффективность работы и концентрацию.

Имея возможность связывать и сопоставлять данные из различных источников, исследователи во всех дисциплинах могут повысить точность и предсказательную силу научных открытий и эффективнее спрогнозировать будущие направления исследований, эти факторы в конечном итоге обеспечивают новую отправную точку для эмпирических исследований. В свою очередь, рост целевого финансирования, учебных программ и мест публикации доказывает, что большие данные широко рассматриваются как новый способ проведения научных исследований [2].

Большие данные обещают в корне изменить производство знаний, как в рамках науки, так и за ее пределами, предоставляя новые, высокоэффективные способы планирования, проведения, распространения и оценки исследований. За последние несколько десятилетий были созданы новые способы получения, хранения и анализа данных, кульминацией которых стало появление области науки о данных, объединяющей вычислительные, алгоритмические, статистические и математические методы для экстраполяции знаний из больших данных [1].

В эпоху больших данных каждая научная дисциплина должна найти способ решения проблем, связанных с хранением, обработкой и интерпретацией больших объемов информации. Результаты научных исследований и наблюдений накапливаются в цифровом виде, поэтому администрирование и обмен данными становятся все более важной составляющей работы ученых [4].

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. Anderson, Chris, 2008, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, Wired Magazine, 23 June 2008.
2. Boon, Mieke, 2020, “How Scientists Are Brought Back into Science—The Error of Empiricism”, in A Critical Reflection on Automated Science, Marta Bertolaso and Fabio Sterpetti (eds.), (Human Perspectives in Health Sciences and Technology 1), Cham: Springer International Publishing, 43–65. doi:10.1007/978-3-030-25001-0_4
3. Cai, Li and Yangyong Zhu, 2015, “The Challenges of Data Quality and Data Quality Assessment in the Big Data Era”, Data Science Journal, 14: 2. doi:10.5334/dsj-2015-002
4. Интернет ресурс: <https://4science.ru/articles/Tendenciya-bolshie-dannie-v-nauke>
5. Елисеева Д.Ю., Мнацаканян О.Л., Карягина Т.В., Юшков А.С. Феномен больших данных для развития общества и науки. Теория и практика проектного образования. 2019. № 1 (9). С. 47-49.