

НАУЧНЫЙ ОБЗОР

УДК 004.6:330

АНАЛИЗ, ПРОБЛЕМЫ И ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ В ГОРОДСКОМ ХОЗЯЙСТВЕ

Массеров Д.Д., Массеров Д.А.

ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет им. Н.П. Огарева», Саранск, e-mail: resurs2003@bk.ru

Аналитика больших данных может быть использована во многих областях экономики, в том числе для анализа больших объемов данных в городском хозяйстве. В данной работе мы анализируем применение больших данных в таких областях, как производство, энергетика, транспорт и даже здравоохранение, везде, где интеллектуальные машины вовлечены в бизнес-процесс. Назрела необходимость в согласовании аппаратных технологий (т.е. машин и датчиков) с программными технологиями (т.е. представление данных, связь, хранение, анализ и контроль из машинного оборудования). Авторы рассмотрели будущие разработки встраиваемых систем, которые превращаются в «киберфизические системы». Перед наукой и специалистами стоят задачи по синхронизации совместной разработки аппаратного обеспечения (вычислений, датчиков и сетей) и программного обеспечения (форматов данных, операционных систем и систем анализа и управления). С одной стороны, это добавляет новую зависимость от использования больших данных, а именно зависимость от аппаратных систем, их разработки и ограничений. С другой стороны, это открывает новые возможности для решения более интегрированных систем с приложениями для использования больших объемов данных в качестве основы поддержки бизнес-решений в городском хозяйстве.

Ключевые слова: интеллектуальная сеть, хранение данных, анализ данных, обработка данных, сбор данных, безопасность данных

ANALYSIS, CHALLENGES AND OPPORTUNITIES FOR THE USE OF BIG DATA IN URBAN MANAGEMENT

Masserov D.D., Masserov D.A.

Mordovia State University, Saransk, e-mail: resurs2003@bk.ru

Big data analytics can be used in many areas of the economy, including big data analytics in the urban economy. In this paper, we analyze the application of big data in areas such as manufacturing, energy, transportation, and even healthcare, wherever smart machines are involved in the business process. The need to align hardware technology (i.e., machines and sensors) with software technology (i.e., data presentation, communication, storage, analysis, and control from the machinery) is urgent. The authors reviewed future developments in embedded systems that are evolving into “cyber-physical systems.” Science and specialists face the challenge of synchronizing the joint development of hardware (computing, sensors, and networks) and software (data formats, operating systems, and analysis and control systems). On the one hand, this adds a new dependency on the use of big data, namely the dependency on hardware systems, their development and limitations. On the other hand, it opens up new opportunities to solve more integrated systems with applications to use big data as the basis for supporting business decisions in urban areas.

Keywords: phase measurement devices, smart grid, data storage, data analysis, data processing, data collection, data security

Огромный потенциал для инноваций имеет применение аналитики больших данных в государственном секторе [1], банковском деле [2], здравоохранении [3], сфере услуг [4], агропромышленном комплексе [5], Интернете вещей (IoT) [6, 7], связи [8], умных городах [9, 10] и транспорте [11]. Соответственно существует большой потенциал для применения больших данных в электросетях городского хозяйства, как в настоящее время, так и в будущем [12, 13]. Для эффективной работы электроэнергетических систем, обеспечивающих доступную, надежную, устойчивую и качественную энергию для конечных потребителей, в электросетях используются всевозможные инновации в области управления, связи, измерений и информатики. Для сбора общесистемных электрических измерений

и развертывания массивной инфраструктуры в электрических сетях начал широко использоваться расширенный учет (АМ) в умных счетчиках и фазоизмерительных устройствах (PMU) [14]. Эффективное использование больших данных повышает наблюдаемость за электросетями. Для надежной и экономичной работы электросетей в условиях города необходим мониторинг общесистемных условий сети, поведения конечных потребителей и доступность возобновляемых источников энергии.

Масштаб увеличившихся измерительных устройств, а также данных на основе симуляторов и данных из неэлектрических источников приводит к колоссальному объему широко варьирующихся данных в электросетях городов, измеряемых тысячами терабайт информации каждый год [15]. Эти данные

поступают из различных источников, включая, программируемые термостаты, умные штекеры, умные приборы и счетчики, PMU, географическую информационную систему (ГИС), информацию о погоде, систему диспетчерского контроля и сбора данных (SCADA), информацию о трафике и сетей коммуникаций, как показано на рисунке [16].

Наличие и доступ к данным будут основой любой городской системы, ориентированной на данные. Здоровая городская система (назовем ее «экосистемой») данных будет состоять из широкого спектра различных типов данных: структурированных, неструктурированных, многоязычных, генерируемых машинами и датчиками, статических данных и данных в реальном времени. Полученный объем достоверной информации позволяет использовать новые алгоритмы управления городским хозяйством, которые могут привести к революционным изменениям в способах планирования и эксплуатации городских сетей.

Следовательно, аналитика больших данных будет играть все возрастающую роль не только для эффективной работы будущих электрических сетей, но и для разработки эффективных бизнес-моделей в городском хозяйстве [17].

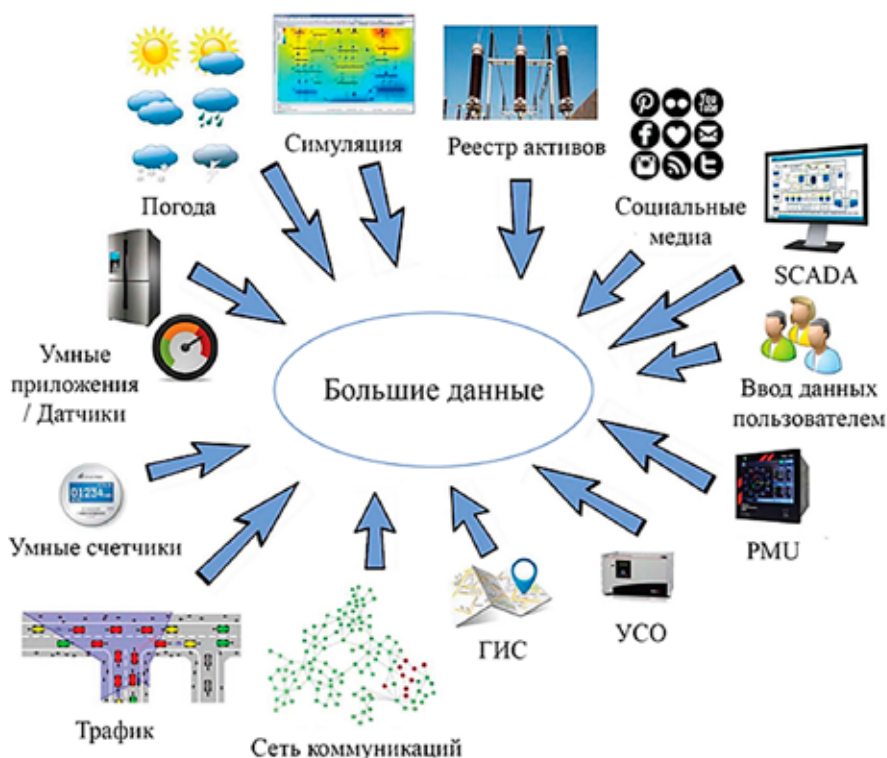
Цель исследования – дать анализ использования больших данных для текущего

применения в управлении городским хозяйством, а также определить проблемы и потенциал их применения.

Основные проблемы, связанные с развертыванием аналитики больших данных в будущих электросетях городов, заключаются в том, что, когда энергия поступает в распределительную сеть, она должна использоваться в это время. Поставщики энергии экспериментируют с устройствами хранения данных, чтобы помочь решить эту проблему, но они находятся на стадии становления и очень дороги. Поэтому проблема решается с помощью интеллектуальных измерительных приборов.

В настоящее время и в будущем объем данных (хранение данных, обработка данных, запрос данных и индексация данных) как у крупных, так и у мелких потребителей растет с экспоненциальной скоростью. Необходимы новые инновационные решения, заключающиеся в создании распределенной и масштабируемой вычислительной архитектуры [18].

Недостоверные данные как характеристика представляются отличительной особенностью достоверных данных умной сети. Данные не могут быть получены со 100% правдоподобностью, поскольку реальные данные восприимчивы к ошибкам из-за недостаточной приходящей информации [19].



Источники больших массивов данных в умных сетях города

Большие данные уже оказали влияние на многие предприятия и потенциально могут повлиять на все секторы бизнеса. Несмотря на наличие ряда технических проблем, влияние на управление и принятие решений и даже корпоративную культуру будет не менее значительным.

Однако все еще существует несколько границ. Именно проблемы конфиденциальности и безопасности должны решаться с помощью этих систем и технологий. Многие системы уже генерируют и собирают большие объемы данных, но лишь небольшой фрагмент активно используется в бизнес-процессах. Кроме того, многие из этих систем не имеют требований к работе в режиме реального времени [20]. Однако управление данными – это нечто большее, чем просто технические задачи по обработке данных.

Сбор данных связан со сбором данных из множества разнородных источников. Особенностью их сбора является конфиденциальность и безопасность при получении и передаче данных. Для этого используются подходы шифрования-дешифрования данных и агрегации-деагрегации [21].

Хранение данных представляет собой управление данными в специальных хранилищах. При сборе данных с интеллектуальных измерительных устройств первой проблемой является хранение большого объема данных. Например, если предположить, что 1 миллион устройств сбора данных извлекает 5 Кб данных за один сбор, потенциальный рост объема данных за год может составить до 2920 ТБ [22].

Анализ данных предназначен для извлечения ценной информации из набора данных. Вытекающие из этого задачи заключаются в анализе этого огромного объема данных, сопоставлении этих данных с информацией о клиентах, распределении сети и пропускной способности по сегментам, информацией о местной погоде и данными о стоимости энергии на спотовом рынке [23].

Использование этих данных позволит коммунальным службам лучше понять структуру затрат и стратегические варианты в рамках своей сети.

Одним из таких подходов от компании Lavastorm является проект, который исследует проблемы аналитики с такими инновационными компаниями, как FalbygdensEnergi AB (FEAB) и Sweco. Чтобы ответить на ключевые вопросы, используется аналитическая платформа Lavastorm. Аналитический движок Lavastorm – это решение для бизнес-аналитики с самообслуживанием, которое позволяет аналитикам быстро

получать, преобразовывать, анализировать и визуализировать данные, а также делиться ключевыми идеями и надежными ответами на бизнес-вопросы с нетехническими менеджерами и руководителями [24].

Поступающие из нескольких источников большие данные в умных сетях приносят ценную информацию для всех заинтересованных сторон, планирования и принятия оперативных решений.

Технологии хранения больших данных являются ключевым фактором для продвинутой аналитики, которая потенциально может трансформировать общество и способ принятия ключевых бизнес-решений. Это имеет особое значение в традиционно не связанных с ИТ секторах, таких как энергетика. В то время как эти секторы сталкиваются с нетехническими проблемами, такими как нехватка квалифицированных экспертов по большим данным и регуляторные барьеры, новые технологии хранения данных обладают потенциалом для создания новых аналитических возможностей, генерирующих ценность, в различных отраслях промышленности [25].

В дополнение к преимуществам, описанным выше, существуют также угрозы технологиям хранения больших объемов данных, которые необходимо устранить, чтобы избежать любого негативного воздействия. Это относится, например, к задаче защиты данных отдельных лиц и снижения энергопотребления центров обработки данных. Киберфизическая уязвимость является критической инфраструктурой в умной сети, которая может привести к широкомасштабным последствиям. Поэтому использование аналитики больших данных в кибербезопасности позволяет своевременно выявлять кибератаки и предотвращать огромный ущерб городской системе [26].

На сегодняшних сложных рынках электроэнергии или интермодальной мобильности практически не существует сценария, в котором все необходимые данные для ответа на бизнес- или инженерный вопрос поступали бы из баз данных одного департамента. Тем не менее большинство установленных в настоящее время передовых измерительных инфраструктур обеспечивают привязку полученных данных об энергопотреблении к платежным системам коммунальных предприятий. Блокировка затрудняет использование энергетических данных для другой ценной аналитики. Кроме того, объем данных, подлежащих обмену в прошлом, был намного меньше, так что интерфейсы, протоколы и процессы для обмена данными были довольно рудиментарными.

Открытыми остаются следующие вопросы: как представлять конкретные данные об энергии и мобильности, возможно, в нескольких измерениях – и как разработать алгоритмы, которые изучают ответы на конкретные вопросы в областях энергетики и мобильности лучше, чем это могут сделать операторы – люди, и делают это проверяемым образом. Основными вопросами машинного обучения являются экономически эффективное хранение и вычисления для огромных объемов данных с высокой выборкой, разработка новых эффективных структур данных и таких алгоритмов, как тензорное моделирование и сверточные нейронные сети. Встроенная аналитика и анализ распределенных данных, облегчающие внутрисетевую и полевую аналитику (иногда называемую периферийными вычислениями) в сочетании с аналитикой, проводимой на уровне предприятия, станут катализатором инноваций в энергетике и транспорте.

Энергетический и транспортный секторы с точки зрения инфраструктуры, а также эффективности использования ресурсов, глобальной конкурентоспособности и качества жизни очень важны для России.

Анализ доступных источников данных в энергетике, а также примеров их использования в различных категориях ценности больших данных, операционной эффективности, клиентского опыта и новых бизнес-моделей поможет определить промышленные потребности и требования к технологиям больших данных. При изучении этих требований становится ясно, что простого использования существующих технологий больших данных будет недостаточно. Необходима адаптация к конкретным областям и устройствам для использования в киберфизических энергетических и транспортных системах. Инновации, касающиеся управления и анализа данных с сохранением конфиденциальности и секретности, являются главной заботой заинтересованных сторон энергетического и транспортного секторов.

Среди заинтересованных сторон энергетического и транспортного секторов есть ощущение, что «больших данных» будет недостаточно. Растущий интеллект, встроенный в инфраструктуру, сможет анализировать данные для получения «умных данных». Это представляется необходимым, поскольку для аналитики потребуются гораздо более сложные алгоритмы, чем для других секторов. Кроме того, ставки в сценариях использования больших данных в энергетике и транспорте очень высоки, поскольку возможности оптимизации затронут критически важные инфраструктуры.

В энергетическом и транспортном секторах есть несколько примеров, когда технология сбора данных, т.е. интеллектуальное устройство, существует уже много лет, или заинтересованные стороны уже измеряют и собирают значительный объем данных. Благодаря последним достижениям появилась возможность передавать, хранить и обрабатывать данные с минимальными затратами.

Многие из современных технологий больших данных только ожидают адаптации и использования в этих традиционных секторах. Технологическая дорожная карта определяет и разрабатывает приоритетные требования и технологии, которые выведут энергетический и транспортный секторы за рамки современного уровня, чтобы они могли сосредоточиться на создании увеличения стоимости путем адаптации и применения этих технологий в своих конкретных областях применения городского хозяйства.

Такие инициативы, как «умные города», показывают, как различные секторы (например, энергетика и транспорт) могут сотрудничать, чтобы максимизировать потенциал оптимизации и отдачи от стоимости. Взаимное обогащение заинтересованных сторон и наборов данных из разных секторов является ключевым элементом для продвижения экономики больших данных. За последние годы сам объем данных, которые постоянно обрабатываются, увеличился. 90% данных в современном мире было получено за последние два года. Источник и характер этих данных разнообразны. Она варьируется от данных, собираемых датчиками, до данных, отражающих транзакции (онлайн). Все большая часть производится в социальных сетях и с помощью мобильных устройств. Тип данных (структурированные или неструктурированные) и семантика также различны. Тем не менее все эти данные должны быть агрегированы, чтобы помочь ответить на бизнес-вопросы и сформировать общую картину рынка.

Для бизнеса эта тенденция открывает ряд возможностей и проблем как для создания новых бизнес-моделей, так и для улучшения текущих операций, тем самым создавая рыночные преимущества. Например, интеллектуальные системы учета тестируются в энергетическом секторе. Кроме того, в сочетании с новыми системами выставления счетов эти системы также могут быть полезны в других секторах, таких как телекоммуникации и транспорт.

Анализ доступных источников данных в области энергетики, а также примеров их использования в различных категориях

ценности больших данных, операционной эффективности, клиентского опыта и новых бизнес-моделей помог определить промышленные потребности и требования к технологиям больших данных. При изучении этих требований становится ясно, что простого использования существующих технологий больших данных, используемых компаниями, занимающимися онлайн-обработкой данных, будет недостаточно. Необходима адаптация к конкретным предметным областям и устройствам для использования в киберфизических энергетических и транспортных системах. Инновации в области обеспечения конфиденциальности и управления данными и их анализа с сохранением конфиденциальности являются первоочередной задачей заинтересованных сторон энергетического и транспортного секторов. Без удовлетворения потребности в конфиденциальности всегда будет существовать неопределенность в регулировании и неопределенность в отношении принятия пользователями нового предложения, основанного на данных.

Список литературы

1. Клюева Е.Г., Ошакбай Д.М. Использование технологий «Больших данных» в государственном секторе // Информатика. 2019. № 2 (45). С. 35–38.
2. Trelewicz, J.Q. Big data and big money: the role of data in the financial sector. *IT Prof.* 2017. No. 19 (3). P. 8–10.
3. Islam M.M., Razaque M.A., Hassan M.M. et al. Mobile cloud-based big healthcare data processing in smart cities. *IEEE Access.* 2017. No. 5. P. 11887–11899.
4. Боков И.С. Анализ больших данных как эффективное средство управления клиентами // Молодой ученый. 2019. № 22 (260). С. 605–609.
5. Отмахова Ю.С., Девяткин Д.А., Усенко Н.И. Анализ цифровых технологий в агропродовольственной сфере с использованием методов обработки больших данных // Информационное общество. 2021. № 4–5. С. 334–344. DOI: 10.52605/16059921_2021_04_334.
6. Marjani M., Nasaruddin F., Gani A. et al. Big IoT data analytics: architecture, opportunities, and open research challenges, *IEEE Access*, 2017. P. 5247–5261.
7. Массеров Д.А., Массеров Д.Д. Применение искусственного интеллекта в концепции Интернета вещей // Отходы и ресурсы. 2022. Т. 9. № 2. DOI: 10.15862/051TOR222.
8. He X., Ai Q., Qiu R.C. et al. A big data architecture design for smart grids based on random matrix theory, *IEEE Trans. Smart Grid.* 2017. No. 8 (2). P. 674–686.
9. Массеров Д.А., Массеров Д.Д. Обеспечение кибербезопасности умного города в процессе цифровизации городской среды // Информационные технологии и математическое моделирование в управлении сложными системами. 2022. № 1 (13). С. 32–42. DOI: 10.26731/2658-3704.2022.1(13).32-41.
10. Массеров Д.Д. Безопасность «умного города» в процессе цифровизации городской среды // XXIV Всероссийская студенческая научно-практическая конференция Нижневартковского государственного университета: материалы конференции (Нижневартковск, 05–06 апреля 2022 г.) / Под общ. ред. Д.А. Погонишева. Нижневартковск: Нижневартковский государственный университет, 2022. С. 135–141.
11. Zhu L., Yu F.R., Wang Y., Ning B. and Tang T. Big Data Analytics in Intelligent Transportation Systems: A Survey, in *IEEE Transactions on Intelligent Transportation Systems.* 2018. Vol. 20. No. 1. P. 383–398. DOI: 10.1109/TITS.2018.2815678.
12. Hong T. Big data analytics: making the smart grid smarter [Guest Editorial], *IEEE Power Energy Mag.* 2018. No. 16 (3). P. 12–16.
13. Кузнецов А. Новая эпоха в энергетике и умные сети // Наука и инновации. 2017. № 8 (174). С. 22–27.
14. Meliopoulos A.P.S., Cokkinides G., Huang R. et al. Smart grid technologies for autonomous operation and control, *IEEE Trans. Smart Grid.* 2011. No. 2 (1). P. 1–10.
15. Михеев А.В. Анализ больших данных для обоснования решений по научно-технологическому развитию в энергетике // Информационные и математические технологии в науке и управлении. 2020. № 4 (20). С. 158–167. DOI: 10.38028/ESI.2020.20.4.014.
16. Пейсахович В.Я. «Умные сети» и их место в инженерной инфраструктуре поселений // Региональная энергетика и энергосбережение. 2017. № 2. С. 82–83.
17. Zinaman O., Miller M., Adil A. et al. Power systems of the future. *Electr. J.* 2015. No. 28 (2). P. 113–126.
18. Денисова О.Ю., Мухутдинов Э.А. Большие данные – это не только размер данных // Вестник технологического университета. 2015. Т. 8. № 4. С. 226–230.
19. Рыговский И. Анализ эффективности методов обработки больших массивов данных с использованием вычислительных систем // Проблемы информатики. 2014. № 2 (23). С. 54–58.
20. Копылов А.А. Киберфизические атаки на умные сети // Инновации. Наука. Образование. 2021. № 48. С. 1710–1720.
21. Бочков А.П., Проурзин В.А., Проурзин О.В. Анализ больших данных надежности восстанавливаемых многоканальных систем // Научное исследование в космических исследованиях Земли. 2021. Т. 13. № 4. С. 49–55. DOI: 10.36724/2409-5419-2021-13-4-49-55.
22. Аль-Раммахи А.А.Х., Громов Ю.Ю., Кошелев Е.В., Минин Ю.В. Обзор и анализ алгоритмов кластеризации больших данных // Приборы и системы. Управление, контроль, диагностика. 2021. № 9. С. 30–38. DOI: 10.25791/pribor.9.2021.1292.
23. Шайтура С.В., Галкин Д.А. Геомаркетинговый анализ больших данных // Информационные технологии. 2021. Т. 27. № 4. С. 180–187. DOI: 10.17587/it.27.180-187.
24. Олейник А.Н. Контент-анализ больших качественных данных. *International Journal of Open Information Technologies.* 2019. Т. 7. № 10. С. 36–49.
25. Касперская Н.И. Анализ больших данных в ИБ предприятий. Перспективы развития // Защита информации. Инсайд. 2019. № 3 (87). С. 34–43.
26. Березкин Д.В., Грошев В.Ю. Анализ современных средств 3D-визуализации больших данных для решения аналитических задач // Динамика сложных систем – XXI век. 2022. Т. 16. № 1. С. 22–34. DOI: 10.18127/j19997493-202201-03.