

АНАЛИЗ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ В ПРОМЫШЛЕННОЙ СФЕРЕ

А.А. Абрамова, Нижегородский государственный университет им. Н.И. Лобачевского, nastya-abramova-2017@bk.ru.

УДК 65.011.56:614.2:004.4

Аннотация. В соответствии с тенденцией экономической глобализации интеллектуальное производство привлекло большое внимание со стороны науки и промышленности. Связанные с этим технологии делают обрабатывающую отрасль более прогрессивной. В статье представлена теоретическая основа анализа технологий, базирующихся на больших данных, для принятия решений в интеллектуальном производстве, которая включает проблемы и будущие исследования в данной области.

Ключевые слова: производство; искусственный интеллект; ИИ; большие данные; принятие решений.

ANALYSIS OF BIG DATA USAGE FOR DECISION MAKING IN THE INDUSTRIAL SECTOR

A.A. Abramova, Nizhny Novgorod State University N.I. Lobachevsky.

Annotation. In accordance with the trend of economic globalization, intellectual production has attracted a lot of attention from science and industry. Related technologies make the manufacturing industry more progressive. This article presents a theoretical framework for analyzing big databased technologies for decision-making in intelligent manufacturing that includes problems and future research in this area.

Keywords: manufacturing; artificial intelligence; AI; big data; decision-making.

Введение

В эпоху больших данных огромное количество генерируемой информации имеет многочисленные характеристики [1]. Использование данных огромной размерности их потенциальной ценности и разработка модели потока, подходящей для новой производственной среды, является сложной задачей [2]. В настоящее время анализ, основанный на больших данных (*Big Data, BD*), может принести много преимуществ промышленному сектору при взаимной поддержке соответствующих инновационных технологий на фоне новой промышленной революции.

Решения, построенные на базе анализа *BD* и умных вычислениях, постепенно используются для снижения сложности обработки значительных объемов информации. Компании все чаще применяют сильную стратегию, основанную на анализе данных, для повышения своей конкурентоспособности [2]. В последние годы анализ больших данных стал основной движущей силой для предприятий, делая промышленный бизнес более прогрессивным.

Анализ больших данных – это революционный скачок в традиционной обработке информации. Характеристики *BD* обобщены и представлены на рис. 1 [2-4]: объем (большой объем данных), вариативность (данные, генерируемые различными источниками, появляются в разных формах), скорость генерации (данные генерируются и обновляются с высокой скоростью), ценность (огромная потенциальная ценность, скрытая в данных) и точность. По сравнению с традиционной системой анализа, данные, управляемые платформой *BD*, поступают как от физических объектов, так и из виртуального цифрового окружения.



Рисунок 1

Интеллектуальное производство охватывает многие аспекты промышленной сферы, не только технологии, но и интеграцию, направленную на преобразование данных, полученных на протяжении всего жизненного цикла продукта, в производственную информационную базу. Анализ *BD* может предлагать решения в режиме реального времени для решения разнообразных проблем. Методы, основанные на больших данных, окажут влияние на управление качеством, а анализ, связанный с качеством продукции, сможет обеспечить поддержку принятия решений для контроля и гарантии в производственной системе. Интеллектуальное производство направлено на создание высокоинтегрированной промышленной экосистемы, которая может реагировать на динамично меняющиеся требования и условия окружающей среды во всей цепочке создания стоимости в режиме реального времени.

В сегодняшних конкурентных условиях компании заинтересованы не только в понимании технических аспектов аналитики больших данных, но и все чаще, в том, чтобы научиться использовать информацию и потенциал *BD* для генерации перспективных идей и, чтобы эффективно использовать их в своих стратегических и оперативных процессах принятия решений и инновациях.

Архитектура и компонентные механизмы платформы, управляемой большими данными

На данном этапе в промышленных процессах применяются системы сбора данных через интернет вещей (*Internet of things, IoT*) и киберфизическая концепция. Как правило, большая часть данных процесса обнаруживается и собирается в среде с информационным шумом. В промышленном процессе входными данными модели прогнозирования часто являются данные процесса, полученные путем быстрой выборки, а выходные обычно являются критическим фактором качества. Однако на самом деле только несколько наборов данных помечены, а большинство из них часто не имеют маркировки. Эти немаркированные данные могут также содержать большое количество информации о процессе, которая внесет существенный вклад в моделирование. Чтобы решить эту проблему, некоторые организации внедряют полуконтролируемые методы моделирования в анализ данных промышленных процессов, максимизируя ценность обрабатываемой информации, включая помеченные и немаркированные данные. Однако ряд методов моделирования, таких как полуконтролируемая вероятностная регрессия

главных компонент, ограничены малыми наборами данных и не подходят для больших. Интеллектуальное производство – это совокупность технологий, использующая взаимосвязанные машины и инструменты для оптимизации производительности [4].

С позиции инклюзивности модель производства и потребления можно интерпретировать как перераспределение, которое показывает перестройку функциональной системы между предприятиями и заинтересованными лицами. «Перераспределение» означает, что потребители демонстрируют более активное участие в процессе проектирования и производства. В источнике [5] выдвинуты семь требований к процессу обработки больших данных в промышленной сфере:

- Предоставить расширяемый метод для распределения и настройки сенсорного оборудования на протяжении всего производственного процесса, а также для хранения данных в описательных процессах и моделях.
- Предоставить методы обнаружения и устранения сложных производственных событий в развернутом потоке данных датчика.
- Предоставить методы хранения данных в реальном времени и выполнять корреляционный анализ больших наборов и потоков данных в соответствии с конкретными параметрами.
- Предоставить методы для постоянной корректировки и адаптации к сложным моделям прогнозирования событий.
- Установить метод создания сигнала тревоги в качестве ожидаемого отклонения от запланированного производственного задания на основе расчетного анализа.
- Предоставить предложения и методы автоматического принятия решений для уменьшения производственных ошибок.
- Предоставить метод активной адаптации к производственному процессу на основе расчета и анализа.

В связи с глобализацией и растущим спросом на мелкосерийную продукцию по индивидуальному заказу производственная деятельность становится все более и более рассредоточенной. Чтобы повысить эффективность делового сотрудничества производственных компаний и сократить цикл разработки продукта, предлагается общая системная архитектура облачного технологического оборудования в рамках концепции «киберфизическое производство» и анализа *BD*. Эта архитектура соединяет оборудование с облаком для предоставления услуг по требованию, которые легко настраивают и оптимизируют производительность в соответствии с целевыми потребностями. Была разработана концептуальная основа анализа больших данных, позволяющая обрабатывать влияние их на перераспределение производства [5], извлекать более конкретную информацию и делать стоимость анализа более инклюзивной. В эпоху *BD* принятие решений по управлению цепочками поставок все больше зависит от данных, а не от традиционного опыта. Была применена децентрализованная архитектура распределения данных [6], которая использовала парадигму туманных вычислений для улучшения потенциальной доступности информации. Сквозное проектирование всего жизненного цикла продукта относится к интеграции и оцифровке данных, что включает в себя процесс принятия решений от производства до поставки и эксплуатации.

Источники промышленных больших данных

Промышленные *BD* делятся на системные данные (данные, генерируемые различными корпоративными информационными системами) и данные *IoT*

(данные, собираемые датчиками, такими как считыватели радиочастотной идентификации и считыватели штрих-кодов). Первичные источники промышленных больших данных показаны на рис. 2. Встраивание датчиков в интеллектуальные устройства позволяет собирать большое количество информации о производственных процессах, оборудовании цеха и состоянии продукции. Обработывающая промышленность часто представляет собой производственную и операционную среду, которая непрерывно генерирует все более сложные структуры данных. Тем не менее, руководство использует только небольшую часть и не извлекает из данных максимальную выгоду. Существует множество источников *BD*, таких как приложения электронной коммерции, записи о покупках, банковские транзакции, социальные сети, сетевые устройства и т.д. Во многих передовых производственных средах проанализированные данные передаются лицам, принимающим решения, практически в режиме онлайн и предоставляют больше возможностей для более эффективного управления [6]. Источники данных в производстве обычно делятся на следующие категории:

- Данные о производственных ресурсах. Данные о производительности интеллектуальных устройств в режиме реального времени, собранные с помощью технологии промышленного интернета вещей.
- Данные производственных систем. Включают в себя проектирование продукта, конфигурацию заказа, распределение материалов, планирование, управление бизнесом и т.д.
- Интернет-данные. Открытые веб-сайты, платформы электронной коммерции и социальных сетей.



Рисунок 2

Сбор и передача больших данных

С непрерывным ростом и совершенствованием современных предприятий многие данные, описывающие производственные процессы, собираются и сохраняются большим количеством датчиков. Обычно они анализируются и оцениваются с помощью передовой технологии мониторинга. Тем не менее, лучший технологический процесс отслеживания должен сортировать данные и группировать их в классы перед их анализом, но для этого требуется много ресурсов [7]. В противном случае в результате причина отслеживаемой неисправности, например, может быть неизвестна.

Для создания релевантных прогнозов функционирования производственных систем необходимы сбор информации для различных уровней детализации, контекстно-зависимый анализ и оценка на основе исторических и полученных в режиме реального времени данных [8]. Выходные данные прогноза запускают

механизм самокоординации системы, чтобы выполнять функции распределения и самонастройки для снижения частоты сбоев производства. Хранение и сбор информации имеют жизненно важное значение для предприятий. Однако способы хранения данных зачастую недостаточно безопасны. На безопасность влияют многие факторы, такие как права доступа, программные и аппаратные сбои и нестабильность системы, что приводит к коллапсу и финансовым потерям. Вопросы, связанные с конфиденциальностью и безопасностью, привлекают все больше внимания со стороны соответствующего персонала. Избыточная система остаточных чисел может использоваться для разделения файлов пользовательских данных на несколько реплицируемых блоков и их параллельного распределения среди гибридных мультиоблачных поставщиков хранилищ [8]. Она гарантирует конфиденциальность данных с помощью стратегий обфускации и шифрования. В процессе прогностического моделирования датчиков из-за шумности могут возникать такие проблемы, как неполная передача данных, погрешность измерений и так далее. Метод обнаружения аномалий обычно используется для мониторинга данных, которые отклоняются от ожидаемой тенденции. Полуконтролируемый способ используется для работы с потерянными и непомяченными данными. Однако эти методы не подходят для обнаружения аномалий в режиме онлайн и не могут решить проблемы отсутствующих данных и шума в режиме реального времени. Способ, основанный на дистанционном опознавании и динамическом обучении, может реализовать поиск отклонений данных, прогнозировать значение целевых переменных и временного диапазона, а также определять положение источника информации. Умные узлы собирают *BD* и передают их в облако по беспроводной сети. Была предложена [9] структура сбора информации на основе компрессионного зондирования для снижения энергопотребления, которая могла бы свести к минимуму объем сбора при обеспечении качества собираемых данных.

Хранение и анализ больших данных

Интеллектуальность производственной системы зависит от способности накапливать и анализировать большие данные. Обработка *BD* может улучшить обслуживание клиентов, повысить качество продукции и иметь большую ценность для предприятий.

Выявление реальных взаимосвязей и закономерностей, скрытых в данных, может сыграть существенную роль [9]. Инструменты интеллектуального анализа данных являются эффективным методом обработки для всех видов данных для поддержки принятия решений. Для работы с *BD* была применена модель фабрики, управляемая обобщенными данными. Эта модель охватывает процессы сбора, передачи, интеграции, хранения и анализа. Ее основной принцип заключается в содействии обмену данными между поставщиками и пользователями путем пакетной обработки и приверженности функциональной совместимости и стандартизации. Промышленные *BD*, организованные каким-то образом с помощью аналитики, могут быть ценным ресурсом для предприятий. В среде с интенсивным использованием данных решение руководителей и внешних организаций о создании сети для совместной работы оказывает положительное влияние на корпоративную эффективность. Анализ *BD* также является нематериальным ресурсом. Таким образом, стратегическая цель эффективного использования инструментов больших данных создаст устойчивые конкурентные преимущества для компаний [10].

Интеллектуальный анализ и интерпретация *BD*

Одной из основных задач аналитики больших данных является корректная интерпретация результатов для поддержки принятия решений на всех уровнях

обрабатывающей промышленности. Способность предприятий адаптироваться к экономическим изменениям определяет развитие рыночной экономики. Динамика и гибкость обрабатывающей промышленности повышают ее конкурентоспособность, а постоянная эволюция структуры производственной сети также является серьезной проблемой.

В настоящее время цифровое интеллектуальное оборудование и автоматизированные линии широко внедрены на предприятиях по производству сложной продукции. Из-за высокой стоимости сбора данных методы динамического накопления информации в реальном времени в некоторых изолированных промышленных средах частично ограничены технологией радиочастотной идентификации, которая не может эффективно оценивать процесс онлайн. В промышленности концепция цифрового двойника используется для того, чтобы система производственного субъекта и цифро-аналоговая концепция были соединены эквивалентно, сокращали временную задержку передачи информации, обеспечивали согласованность и единство процесса и позволяли контролировать изготовление продукции в режиме реального времени для повышения визуальной прозрачности информации для полной поддержки стратегического и операционного плана предприятия [11].

Проектирование и внедрение системы управления большими данными в интеллектуальном производстве

Системы больших данных, как правило, состоят из шести необходимых подсистем: генерация, сбор, передача, обработка, хранение и методы анализа [11]. Исследован новый интеллектуальный фабричный фреймворк с функцией независимого агента и интеграцией обратной связи и координации анализа *BD* [12]. Программная платформа органично сочетает в себе промышленную сеть, облако и терминалы контроля и управления с интеллектуальными объектами цехов (машинами, конвейерами и продуктами и т.д.) и делит интеллектуальное моделирование объектов на четыре типа, процесс эффективно работает за счет автономного принятия решений и распределяет сотрудничество между агентами, определяет центрального координатора в хранилище для обратной связи, а также предлагает интеллектуальный механизм переговоров для облегчения межведомственного взаимодействия и сотрудничества.

Для более эффективной реализации умного производства базовая архитектура платформы, управляемая большими данными, показана на рис. 3, которая в целом охватывает следующие три характеристики [13]:

- Горизонтальная интеграция через сеть создания стоимости. Содействие сотрудничеству между предприятиями и формирование эффективной операционной системы посредством плавной интеграции и взаимной корреляции.
- Вертикальная интеграция и сетевые производственные системы. Это означает очень гибкое, реконфигурируемое, интеллектуальное предприятие. Динамическая конфигурация управления производством реализуется с помощью исполнительных механизмов и датчиков. Для повышения прозрачности производственного процесса собирается и обрабатывается большой объем данных.
- Сквозная цифровая интеграция всей цепочки создания стоимости. Сквозная интеграция означает участие пользователей в настройке изготавливаемой продукции.

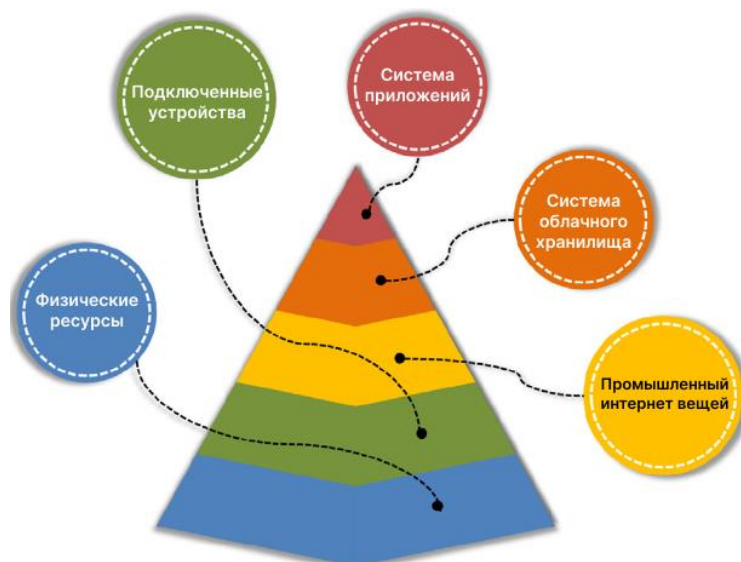


Рисунок 3

Проблемы промышленных больших данных в интеллектуальном производстве

Развитие технологий, основанных на *BD*, также сталкивается со многими серьезными проблемами, обусловленными временем. Они заключаются в следующем:

1. Проблема управления качеством данных. Высококачественные *BD* играют важную роль в интеллектуальном производстве. Масштаб набора данных выходит за пределы допустимого диапазона обычных программных средств, что приводит к неудовлетворительной обработке. Большое количество полезной информации практически не используется, чтобы помочь менеджменту принимать точные решения, что приводит к пустой трате ресурсов [14].

2. Проблема безопасности данных и защиты конфиденциальности. Разработка и применение инструментов анализа *BD* требует больших инвестиций и дополнительной работы. Существующая технология управления большими данными является дорогостоящей. В настоящее время большинство технологий не могут удовлетворить текущие требования к инфраструктуре и нет стратегии обмена информацией между организациями. Различные типы данных из разных источников могут привести к усложнению интеграции. При предоставлении пользователям достаточной информации посредством интеллектуального анализа, возникает серьезная угроза конфиденциальности и безопасности данных.

3. Проблема общности концептуальной основы в реальном производстве. Решение не может описывать практический метод представления результатов анализа и фактических условий в реальной производственной среде и не может дать полную и правильную информацию для пользователей и лиц, принимающих решения. Большинство концептуальных основ и методов анализа полезны только для определенного типа производства, но не универсальны в промышленной деятельности. Они недостаточно подробно описывают методы процесса обработки и не охватывают критические элементы, необходимые для разработки конкретных схем анализа данных [15].

4. Проблема обработки интеграции данных в промышленных производственных системах. Промышленные *BD* зашумлены и избыточны, а при потере ценных для предприятия данных, финансовые потери слишком значительны. Объем данных огромен и разнообразен, традиционный анализ с однократной обработкой не может быстро интегрироваться для получения

разнородных знаний и целевой информации из-за различных характеристик, вопросов безопасности и других причин.

5. Проблема доступа к первичным производственным данным. Доступность данных всей производственной сети поддерживает массивные потоки информации для определения производительности, необходимой для принятия стратегических решений. Столкнувшись с неполными или недействительными данными, извлечь и преобразовать первичные данные непросто. Платформа анализа информации не может напрямую использовать данные в производственной системе для обработки. Ей необходимо открыть функцию доступа к информации, экспортировать образцы для хранения, а затем перейти на этап предварительной диагностики [16].

Заключение

В данной статье был представлен анализ использования промышленных больших данных для интеллектуального принятия решений, а также рассмотрены проблемы в этой области. В реальной производственной деятельности утечка информации, помехи и другие факторы будут оказывать косвенное влияние на принятие решений. Эта концептуальная основа новой парадигмы вводит промышленный анализ на основе больших данных в производственную систему. Однако гипотетическая модель анализа больших данных построена в идеальной среде, надежность и практичность этой концептуальной основы нуждается в дальнейшей проверке.

Литература

1. Kumar S., Mohbey K.K. A review on big data based parallel and distributed approaches of pattern mining, J. King. Saud. Univ. Comput. Inf. Sci. DOI: 10.1016/j.jksuci.2019.09.006.
2. Kuo Y.H., Kusiak A. From data to big data in production research: the past and future trends, Int. J. Prod. Res. 57, 2018. – pp. 4828-4853.
3. Tang M., Liao H. From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey, Omega? 2019. – P. 100.
4. Ghobakhloo M., Ching N.T. Adoption of digital technologies of smart manufacturing in smes, J. Ind. Inf. Integr, 2019. – P. 16.
5. Akerman M., Lundgren C., Barring M., Folkesson M., Berggren V., Stahre J., Engström U., Friis M. Challenges building a data value chain to enable data-driven decisions: A predictive maintenance case in 5G-enabled manufacturing, in: Proc. FAIM Conf., 2018. – pp. 411-418.
6. Manavalan E., Jayakrishna K. A review of internet of things (IOT) embedded sustainable supply chain for industry 4.0 requirements, Comput. Ind. Eng. 127, 2019. – pp. 925-953.
7. Hoffmann Souza M.L., C.A. da Costa, G. de Oliveira Ramos, R. da Rosa Righi. A survey on decision-making based on system reliability in the context of industry 4.0, J. Manuf. Syst. 56, 2020. – pp. 133-156.
8. Babiceanu R.F., Seker R. Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook, Comput. Ind. 81, 2016. – pp. 128-137.
9. Peres R.S., Rocha A.D., Leitao P., Barata J., Idarts - towards intelligent data analysis and real-time supervision for industry 4.0, Comput. Ind. 101, 2018. – pp. 138-146.
10. Kong L., Zhang D., He Z., Xiang Q., Wan J., Tao M. Embracing big data with compressive sensing: A green approach in industrial wireless networks, IEEE Commun. Mag. 54, 2016. – pp. 53-59.

11. Guerreiro G., Costa R., Figueiras P., Graca D., Jardim-Goncalves R. A self- adapted swarm architecture to handle big data for «factories of the future», in: Proc. IFAC Conf., 2019. – pp. 916-921.
12. Ghasemaghaei M., Calic G. Can big data improve firm decision quality? The role of data quality and data diagnosticity, *Decis. Support. Syst.* 120, 2019. – pp. 38-49.
13. Kotenko I., Saenko I., Branitskiy A. Improving the performance of manufacturing technologies for advanced material processing using a big data and machine learning framework, *Mater. Today: Proc.* 11, 2019. – pp. 380-385.
14. Kozjek D., Vrabic R., Rihtaršc B., Lavrac N., Butala P. Advancing manufacturing systems with big-data analytics: A conceptual framework, *Int. J. Comput. Integr. Manuf.* 33, 2020. – pp. 169-188.
15. Wu J.Z., Tiao P.J. A validation scheme for intelligent and effective multiple criteria decision-making, *Appl. Soft. Comput.* 68, 2018. – pp. 866-872.
16. Francis J., Bian L. Deep learning for distortion prediction in laser-based additive manufacturing using big data, *Manuf. Lett.* 20, 2019. – pp. 10-14.