

ИСПОЛЬЗОВАНИЕ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ В ЗДРАВООХРАНЕНИИ

А.И. Панов, Нижегородский государственный университет им. Н.И. Лобачевского, alexs.npc481@gmail.com.

УДК 65.011.56:614.2:004.4

Аннотация. Внедрение в сферу здравоохранения аналитики больших данных (*Big data analytics, BDA*) позволит использовать новые технологии как в лечении пациентов, так и в управлении самой отрасли. В данной статье предложены способы использования аналитики больших данных в медицинских учреждениях, а также отражены проблемы и преимущества, связанные с ее эксплуатацией.

Ключевые слова: большие данные; *BDA*; здравоохранение; аналитика; информационные технологии.

THE USE OF BIG DATA ANALYTICS IN HEALTHCARE

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

Annotation. Big Data Analytics (*BDA*) are methods and tools used to analyze and extract information from big data. Its introduction into the healthcare sector will allow the use of new technologies both in the treatment of patients and in the management of the industry itself. The article presents the means of using big data analytics, and shows the problems and advantages associated with its operation.

Keywords: big data; *BDA*; healthcare; analytics; information technology.

Введение

Большие данные – это огромное количество наборов данных, которые невозможно сохранить, обработать или проанализировать с помощью традиционных инструментов. Из-за отсутствия четко определенной схемы работа с информацией затруднена, и, следовательно, требуется специальная технология и метод извлечение ценности из них [1]. Таким образом, раскрывается потенциал аналитики больших данных. *BDA* – это методы и инструменты, используемые для анализа и извлечения информации из больших данных. Результаты анализа больших данных могут быть использованы для прогнозирования будущего. Они также помогают в отслеживании тенденций в прошлом. Когда речь идет о здравоохранении, *BDA* позволяет анализировать большие наборы данных от тысяч пациентов, выявлять кластеры и корреляцию, а также разрабатывать прогностические модели с использованием методов интеллектуального анализа данных [2].

В данной статье в качестве объекта исследования в области аналитики больших данных рассматривается здравоохранение. Здравоохранение – это сложная система с различными заинтересованными сторонами: пациентами, врачами, больницами, фармацевтическими компаниями и лицами, принимающими решения в данной области. Этот сектор также ограничен строгими правилами и нормами. Однако во всем мире можно наблюдать отход от традиционного подхода «врач-пациент». Врач становится партнером, а пациент вовлекается в терапевтический процесс [3]. Здравоохранение больше не сосредоточено исключительно на лечении пациентов. Приоритетом для лиц, принимающих решения, должно быть содействие надлежащему отношению к здоровью и профилактика заболеваний, которых можно избежать [4]. Это стало заметным и

важным, особенно во время эпидемии *Covid-19* [5]. Пандемия показала, что пациенты должны иметь доступ к информации о состоянии своего здоровья, возможность цифрового анализа этих данных и доступ к надежной медицинской поддержке в режиме онлайн. Мониторинг состояния здоровья и сотрудничество с врачами с целью профилактики заболеваний могут фактически произвести революцию в системе здравоохранения. Одним из наиболее важных аспектов изменений, необходимых в данной сфере, является постановка пациента в центр системы по учету состояния здоровья.

Технологий недостаточно для достижения этих целей. Поэтому изменения должны быть сделаны не только на технологическом уровне, но и в управлении и проектировании полных процессов здравоохранения, и, более того, они должны повлиять на бизнес-модели поставщиков услуг. Использование *BDA* становится все более распространенным на предприятиях [4, 5]. Тем не менее, медицинские учреждения по-прежнему не могут идти в ногу с информационными потребностями пациентов, врачей, администраторов и политики разработчиков. Внедрение подхода анализа больших данных позволит реализовать персонализированную медицину, основанную на личной информации, предоставляемой в режиме реального времени и адаптированной к индивидуальным пациентам.

Использование аналитики больших данных в здравоохранении

В последние годы наблюдается постоянно растущий спрос на решения, предлагающие эффективные аналитические инструменты. Эта тенденция заметна и при анализе больших объемов данных (*Big data, BD*). Организации ищут способы использовать возможности больших данных для улучшения процесса принятия решений, конкурентного преимущества или эффективности бизнеса [5]. Считается, что *BD* предлагают потенциальные решения для государственных и частных организаций [6].

Здравоохранение всегда генерировало огромные объемы данных, и в настоящее время внедрение электронных медицинских карт, а также огромное количество данных, отправляемых различными типами датчиков или генерируемых пациентами в социальных сетях, приводит к постоянному росту потоков информации. Кроме того, медицинская промышленность включает различные протоколы, медицинские изображения, геномные данные и прочее. Проблема связана не только с обработкой большого количества клинических данных, но и со сложностью их обработки.

BD можно рассматривать как массивные и постоянно генерируемые цифровые наборы данных, которые создаются посредством взаимодействия с онлайн-технологиями [6]. Описывая большие данные, нельзя упускать из виду, что этот термин относится скорее к явлению, чем к конкретной технологии. Все больше авторов описывают *BD*, давая им характеристики, включающие набор, связанный с их природой [5-7]:

- Объем (относится к объему данных и является одной из самых больших проблем в аналитике больших данных).
- Скорость генерации – скорость, с которой генерируются новые данные (задача состоит в том, чтобы иметь возможность эффективно управлять данными в режиме реального времени).
- Разнообразие – множество различных типов медицинских данных (задача состоит в том, чтобы получить понимание, рассматривая все доступные разнородные данные целостным образом).

- Вариативность – несогласованность данных (задача состоит в том, чтобы исправить интерпретацию данных, которая может значительно варьироваться в зависимости от контекста).
- Достоверность (насколько достоверны данные, их качество).
- Интерпретируемость (способность интерпретировать данные и результирующие идеи).
- Ценность – цель *BDA* (обнаружить скрытые знания из огромных массивов данных).

Организации должны по-другому подходить к этому типу неструктурированной информации. Прежде всего, организации должны начать рассматривать данные как потоки, а не запасы, что влечет за собой необходимость внедрения так называемой потоковой аналитики [7]. Указанные особенности обуславливают необходимость использования новых ИТ-инструментов, позволяющих наиболее полно использовать имеющиеся массивы данных [8].

Концепция *BD* постоянно развивается, и в настоящее время она фокусируется не на огромных объемах данных, а скорее на извлечении из них ценной информации [8]. Большие данные собираются из различных источников, которые имеют разные свойства и обрабатываются разными организационными подразделениями, в результате чего создается целая цепочка таких данных [8]. В сфере здравоохранения потоки *BD* состоят из различных типов данных, а именно [8, 9]:

- Клинические данные, т.е. данные, полученные из электронных медицинских карт, больничных информационных систем, имидж-центров, лабораторий, аптек и других организаций, предоставляющих врачебные услуги, данные о состоянии здоровья, сгенерированные пациентом, заметки врача в свободном тексте, геномные данные, данные физиологического мониторинга [9].
- Биометрические данные, предоставляемые с различных типов устройств, которые контролируют вес, давление, уровень глюкозы и т.д.
- Финансовые данные, составляющие полный отчет о хозяйственных операциях, отражающий осуществляемую деятельность.
- Данные научно-исследовательской деятельности, т.е. результаты исследований, разработки медицинских изделий и новых методов лечения.
- Данные, предоставляемые пациентами, включая описание предпочтений, уровня удовлетворенности, информацию из систем самоконтроля их активности: упражнения, сон, рацион и т.д.
- Данные из социальных сетей.

Информация, которая была сгенерирована до сих пор в секторе здравоохранения, хранится как в бумажной, так и в цифровой форме. Таким образом, сущность и специфика процесса анализа больших данных означает, что организациям необходимо столкнуться с новыми технологическими и организационными вызовами [10].

Поэтому потенциал виден в анализе больших данных, особенно в аспекте повышения качества медицинской помощи, спасения жизней или снижения организационных и операционных затрат [10]. Извлечение заданных правил, моделей и тенденций позволит поставщикам медицинских услуг и другим заинтересованным сторонам в секторе здравоохранения предлагать более точные и более исчерпывающие диагнозы пациентов, персонализированное лечение, наблюдение за пациентами, профилактическую медицину, поддержку медицинских исследований и здоровья населения, а также лучшее качество

медицинских услуг и ухода за пациентами при возможном снижении издержек, как показано на рис. 1.



Рисунок 1

Основная проблема, связанная с *BD*, заключается в том, как обрабатывать такой объем информации и использовать ее для принятия решений на основе данных во многих областях. Еще одна серьезная задача заключается в настройке хранения больших данных, представления результатов анализа и выводов на их основе в клинических условиях. Системы анализа данных, внедряемые в здравоохранении, предназначены для описания, интеграции и представления сложных данных соответствующим образом, чтобы повысить их интерпретируемость, как показано на рис. 2. Это повысит эффективность сбора, хранения, анализа и визуализации больших данных в сфере здравоохранения.

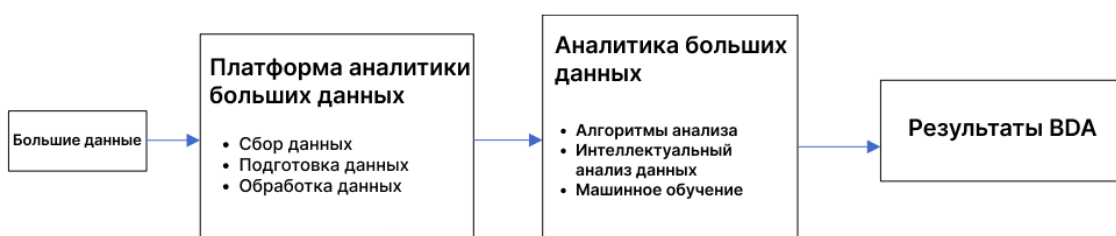


Рисунок 2

Под *BDA* в медицине и здравоохранении понимается интеграция и анализ большого количества сложных разнородных данных, таких как различные омики (геномика, эпигеномика, транскриптомика, протеомика, метаболомика, интерактомика, фармакогенетика, диасомика), биомедицинские данные, телемедицина (датчики, данные медицинского оборудования) и информация электронных медицинских карт [10]. Анализируя феномен *BD* в сфере здравоохранения, следует отметить, что его можно рассматривать с точки зрения трех направлений: эпидемиологической, клинической и деловой.

1. С клинической точки зрения анализ больших данных направлен на улучшение здоровья и состояния пациентов, позволяет делать долгосрочные прогнозы о состоянии их здоровья и проводить соответствующие терапевтические процедуры. В конечном счете, использование анализа данных в медицине заключается в том, чтобы позволить адаптировать терапию к конкретному пациенту, т.е. персонализированную медицину.

2. С эпидемиологической точки зрения желательно получить точный прогноз заболеваемости, чтобы заблаговременно реализовать профилактические программы.
3. В бизнес-контексте анализ больших данных может позволить предложить персонализированные пакеты коммерческих услуг или определить вероятность возникновения отдельных заболеваний и инфекций. Стоит отметить, что *BD* означает не только сбор и обработку информации, но, прежде всего, вывод и визуализацию данных, необходимых для получения конкретных преимуществ для бизнеса.

Для внедрения новых методов управления и новых решений с точки зрения эффективности и прозрачности возникает необходимость сделать данные более доступными, цифровыми, а также анализируемыми и визуализируемыми.

Ожидается, что в сфере здравоохранения *BDA* улучшит качество жизни и снизит операционные расходы [11]. Также она позволяет организациям улучшать и углублять понимание информации, содержащейся в данных.

Аналитику больших данных в здравоохранении можно разделить на следующие категории [11-13]:

- Описательная аналитика в здравоохранении используется для понимания прошлых и текущих решений в данной области, преобразования данных в полезную информацию для понимания и анализа, результатов и качества медицины. Ее можно использовать для создания отчетов (например, о госпитализациях пациентов, производительности врачей, управлении использованием), визуализации, настраиваемых отчетов, детализированных таблиц или выполнения запросов на основе исторических данных.
- Прогнозная аналитика работает с прошлыми показателями, чтобы предсказать будущее, изучая исторические или обобщенные данные о состоянии здоровья, обнаруживая закономерности взаимосвязей, а затем экстраполируя эти отношения для прогнозирования. Она может быть использована для исследования реакции различных групп пациентов на различные лекарства (дозировки) или воздействия (клинические испытания), прогнозирования риска и нахождения взаимосвязей данных о состоянии здоровья и выявления скрытых закономерностей. Таким образом, можно прогнозировать распространение эпидемии, предвидеть спрос на обслуживание и планировать ресурсы здравоохранения.
- Предписывающая аналитика возникает, когда проблемы со здоровьем связаны со слишком большим количеством вариантов или альтернатив. Этот вид аналитики использует здоровье и медицинские знания в дополнение к данным или информации. Предписывающая аналитика используется во многих областях здравоохранения, включая рецепты на лекарства и альтернативы лечения. Персонализированная медицина и доказательная медицина поддерживаются предписывающей аналитикой.

Успех анализа больших данных и его точность во многом зависят от инструментов и методов, используемых для анализа, способности предоставлять надежную, актуальную и ценную информацию различным заинтересованным сторонам. Считается, что внедрение *BDA* организациями здравоохранения может принести много пользы в ближайшие годы, включая снижение затрат, лучшую диагностику и предсказание заболеваний и их распространения, улучшение ухода за пациентами и разработку протоколов предотвращения повторной госпитализации, оптимизацию персонала и оборудования, прогнозирование

потребности в больничных койках, операционных, лечение и совершенствование цепочки поставок лекарств [14].

Проблемы и потенциальные преимущества использования аналитики больших данных в здравоохранении

Современная аналитика дает возможность не только получить представление об исторических данных, но и обрести информацию, необходимую для понимания того, что может произойти в будущем, даже когда речь идет о прогнозировании действий, основанных на фактах. Многие плательщики и поставщики проводят анализ данных для снижения риска, выявления мошенничества, повышения эффективности и спасения жизней. Таким образом, некоторые области, в которых расширенный анализ данных и аналитика могут дать наибольшие результаты, включают различные заинтересованные стороны в области здравоохранения.

Организации здравоохранения видят возможность роста за счет инвестиций в *BDA*. Стоит отметить, что в структуре сферы медицины происходит множество изменений. Цифровизация и эффективное использование больших данных в здравоохранении могут принести пользу каждой заинтересованной стороне в этом секторе. Отдельные специалисты выиграют так же, как и вся система здравоохранения. Потенциальные возможности достижения преимуществ и эффектов от *BD* в медицине можно разделить на четыре группы [15]:

Повышение качества медицинских услуг

- Оценка диагнозов, поставленных врачами, и способа лечения указанных ими заболеваний на основе системы поддержки принятия решений, работающей на сборах *BD*.
- Выявление более эффективных, с медицинской точки зрения, и более экономичных способов диагностики и лечения пациентов.
- Анализ больших объемов данных для получения практической информации, полезной для выявления потребностей, внедрения новых услуг здравоохранения, предотвращения и преодоления кризисных ситуаций.
- Прогнозирование заболеваемости.
- Выявление тенденций, которые приводят к улучшению здоровья и образа жизни общества.
- Анализ генома человека для внедрения персонализированного лечения.

Поддержка работы медицинского персонала

- Сравнение врачами текущих медицинских случаев со случаями из прошлого для лучшей диагностики и корректировки лечения.
- Выявление заболеваний на более ранних стадиях, когда их легче и быстрее вылечить.
- Выявление эпидемиологических рисков.
- Выявление пациентов, которые, по прогнозам, имеют наибольший риск специфических, угрожающих жизни заболеваний, путем сопоставления данных об истории наиболее распространенных заболеваний.
- Управление здоровьем каждого пациента в отдельности (персонализированная медицина) и управление здоровьем всего общества.
- Сбор и анализ больших объемов данных из больниц в режиме реального времени, устройства мониторинга жизни для безопасности и прогнозирования нежелательных явлений.

- Анализ профилей пациентов для выявления людей, для которых следует применять профилактику, изменить образ жизни или подход к профилактической помощи.
- Способность прогнозировать возникновение конкретных заболеваний или ухудшение результатов прохождения лечения пациентами.
- Прогнозирование прогрессирования заболевания, оценка риска осложнений.
- Выявление лекарственных взаимодействий и их побочных эффектов.

Поддержка научно-исследовательской деятельности

- Поддержка работы над новыми препаратами и клиническими испытаниями, благодаря возможности анализа множества потенциальных образцов вместо выбора одного исследуемого образца.
- Возможность выявлять пациентов со специфическими биологическими особенностями, которые будут принимать участие в специализированных клинических исследованиях.
- Подбор группы пациентов, для которых исследуемый препарат, скорее всего, окажет желаемый эффект и не будет иметь побочных эффектов.
- Использование моделирования и прогностического анализа для разработки более совершенных лекарств и устройств.

Бизнес и менеджмент

- Сокращение затрат и противодействие злоупотреблениям в практике консультирования.
- Более быстрое и эффективное выявление некорректных или несанкционированных финансовых операций с целью предотвращения злоупотреблений полномочиями и устранения ошибок.
- Повышение рентабельности за счет выявления пациентов, генерирующих высокие затраты, или выявления врачей, чья работа, процедуры и методы лечения стоят больше всего, и предложения им решений, которые уменьшают количество затрачиваемых средств.
- Выявление необязательных медицинских процедур, например, дубликатов анализов.

Приведенная выше спецификация не является полным перечнем потенциальных областей использования анализа больших данных в здравоохранении, поскольку возможности эксплуатации практически не ограничены. Кроме того, передовые аналитические инструменты позволяют обрабатывать данные из всех возможных источников и проводить перекрестный анализ для лучшего понимания [16].

Большие данные в свою очередь порождают множество проблем, таких как трудности в их сборе, хранении данных, анализе и визуализации [17]. Основные вызовы связаны с такими вопросами, как: структура данных (большие данные должны быть удобными для пользователя, прозрачными, но они фрагментированы, рассредоточены, редко стандартизированы и их трудно агрегировать и анализировать); безопасность – безопасность данных, конфиденциальность и конфиденциальность медицинских данных (существуют значительные опасения, связанные с конфиденциальностью); стандартизация данных (данные хранятся в форматах, совместимых не со всеми приложениями и технологиями); хранение и передача (особенно затраты, связанные с защитой, хранением и передачей неструктурированных данных); управленческие навыки, такие как управление данными, отсутствие соответствующих аналитических навыков и проблемы с

аналитикой в реальном времени (здравоохранение должно иметь возможность использовать большие данные в режиме реального времени) [6-8].

Заключение

Таким образом, можно сказать, что большие данные и аналитика больших данных представляют собой мощный инструмент для преобразования здравоохранения, повышения эффективности и улучшения результатов лечения. Использование этих технологий может привести к более точным диагнозам, персонализированному подходу и более эффективному использованию медицинских ресурсов, что способствует улучшению здоровья и благополучия пациентов. Но необходимо учитывать, что существует множество подводных камней, связанных с использованием больших данных и их аналитикой, поэтому необходимо рационально подходить к интеграции *BDA* в здравоохранении.

Литература

1. Bainbridge M. Big data challenges for clinical and precision medicine. In: Househ M., Kushniruk A., Borycki E., editors. Big data, big challenges: a healthcare perspective: background, issues, solutions and research directions. Cham: Springer, 2019. – p. 17-31.
2. Bartuś K., Batko K., Lorek P. Business intelligence systems: barriers during implementation. In: Jabłoński M., editor. Strategic performance management new concept and contemporary trends. New York: Nova Science Publishers, 2017. – p. 299-327. ISBN: 978-1-53612-681-5.
3. Williams N., Ferdinand N.P., Croft R. Project management maturity in the age of big data. *Int J Manag Proj Bus*, 2014. – № 7 (2). – P. 311-317.
4. Wang Y., Kung L., Wang W., Yu C., Cegielski C.G. An integrated big data analytics-enabled transformation model: application to healthcare. *Inf Manag*, 2018. – № 55 (1). – P. 64-79.
5. Schmarzo B. Big data: understanding how data powers big business. Indianapolis: Wiley, 2013.
6. Provost F., Fawcett T. Data science and its relationship to big data and data-driven decisionmaking. *Big Data*, 2013. – № 1 (1). – P. 51-59.
7. Olszak C.M. Toward better understanding and use of business intelligence in organizations. *Inf Syst Manag*, 2016. – № 33 (2). – P. 105-123.
8. Ohlhorst F. Big data analytics: turning big data into big money, vol. 65. Hoboken: Wiley, 2012.
9. Mohammadi M., Al-Fuqaha A., Sorour S., Guizani M. Deep learning for IoT big data and streaming analytics: a survey. *IEEE Commun Surv Tutor*, 2018. – № 20 (4). – P. 2923-2960.
10. Mikalef P., Krogstie J. Big data analytics as an enabler of process innovation capabilities: a configurational approach. In: International conference on business process management. Cham: Springer, 2018. – P. 426-441.
11. Marconi K., Dobra M., Thompson C. The use of big data in healthcare. In: Liebowitz J, editor. Big data and business analytics. Boca Raton: CRC Press, 2012. – P. 229-48.
12. Madsen L.B. Data-driven healthcare: how analytics and BI are transforming the industry. Hoboken: Wiley, 2014.
13. Lytras M.D., Papadopoulou P. Applying big data analytics in bioinformatics and medicine. IGI Global: Hershey, 2017.
14. Kruse C.S., Goswamy R., Raval Y.J., Marawi S. Challenges and opportunities of big data in healthcare: a systematic review. *JMIR Med Inform*, 2016. – № 4 (4). – P. 38.
15. Knapp M.M. Big data. *J Electron Resourc Med Libr*, 2013. – № 10 (4). – P. 215-222.
16. Janssen M., van der Voort H., Wahyudi A. Factors influencing big data decision-making quality. *J Bus Res*, 2017. – № 70. – P. 338-345.

17. Gupta V., Singh V.K., Ghose U., Mukhija P. A quantitative and text-based characterization of big data research. J Intell Fuzzy Syst, 2019. – № 36. – P. 4659-4675.