

Применение Технологий Big Data в Медицине

Гюляра Мурадова

Азербайджанский Технический Университет

gularamu@aztu.edu.az

Абстракт – в статье рассматриваются вопросы применения больших данных в медицине и биомедицинских исследованиях. Приведены примеры, демонстрирующие как анализ больших массивов данных может эффективно снизить проблемы в сфере здравоохранения. Показаны перспективные направления применения Big Data в медицине.

Ключевые слова – источники Big Data, анализ “больших данных”, персональная медицина, носимая электроника для диагностики, биомаркер, трехмерная цифровая медицина

I. ВВЕДЕНИЕ

Мир больших данных радикально изменяет все и применение технологий Big Data позволяет не просто трансформировать применяемые ранее методы, но и найти новые модели принятия решений, которые раньше были попросту невозможны. Устаревшие методы анализа и сбора данных не соответствуют современным требованиям, поэтому в настоящее время наблюдается повышенный интерес к ИТ-решениям в Big Data. Развитие технологий больших данных и адаптация под новые задачи является целью изучения Big Data. Постоянное накопление и анализ больших массивов данных позволяет создавать модели, с которыми можно проводить эксперименты, т.е. исследователи, извлекая выгоду из данных, получают результаты, представляющие научную ценность. Анализируя большие массивы данных, установив закономерности, можно создать модель объекта, которую можно использовать в самых разных целях. С помощью модели можно изучать свойства уже описанного реального объекта, а также делать вероятностные прогнозы. Эффективность использования больших объемов данных поможет улучшению системы здравоохранения, в частности при выборе соответствующих путей лечения, диагностики и т.д.[1]. Для того чтобы справиться с огромным объемом данных, необходимы новые решения. Анализ “больших данных” может улучшить планирование медицинского обслуживания. Big Data требует пересмотра существующего подхода с учетом новых технологий. Количество данных в таких областях, как биология, медицина и генетика увеличивается с возрастающей скоростью. Медицина является одной из отраслей, которая становится потребителем технологий больших данных. Источники «больших данных» могут быть разные данные. С медицинской точки зрения источниками данных являются:

1. Данные соц. медиа: Facebook, Twitter, LinkedIn, медицинские веб-сайты, приложения и т.д.
2. Данные отслеживающих устройств: с удаленных датчиков, томографов, счетчиков и др.
3. Данные транзакций: данные страховой медицины, биллинговые операции и др.
4. Биометрические данные: отпечатки пальцев, почерк, сетчатка глаза, рентгеновские и другие медицинские изображения, артериальное давление, пульс др.
5. Данные, генерируемые человеком: заметки врачей, бумажные документы, рецепты врачей.

Электронная форма работы в медицинских учреждениях сделала возможным отслеживание взаимосвязи между принимаемыми препаратами, их эффективностью, а также другие факторы, которые столь важны для фармакологических компаний. Главной задачей использования больших данных является сокращения времени и средств при разработке новых препаратов и вывод их на рынок за счет многофакторного анализа их эффективности. Теперь врач в реальном времени изучая данные, полученные от множества специалистов, занимающихся лечением пациентов с единичными заболеваниями технологий может позволить себе полагаться не только на собственные знания. А это позволит намного лучше анализировать эффективность применяемых методов лечения. Примеры использования технологий больших данных это создание регистров пациентов с наследственными заболеваниями для счета объемов и потребностей в препаратах, для упрощения диагностики, для поиска скрытых зависимостей. Существующие множество ограничений на распространение данных о пациентах, накладывает определенные ограничения на возможность такого анализа [2].

Для успешного принятия решения о внедрении решений Big Data необходимо четкое планирование первоначальных действий. Культурный уровень сводится к осознанию возможностей, которые открываются при извлечении новых знаний из накопленных данных, и, как следствие, увеличению когнитивной способности человечества. Научный уровень образован математическими методами анализа данных, позволяющими совершать открытия в химии, биологии, медицине и других дисциплинах. Технологический уровень предполагает переориентацию ИТ со счетных

задач как основного приложения на сбор, хранение, обработку больших массивов данных с поддержкой тех или иных аналитических методов[3].

II. НАПРАВЛЕНИЯ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ BIG DATA В БИОМЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ

Современная медицина превращается в науку, где большие объемы информации о пациенте позволяют специалистам вести персонализированное лечение. Это новая модель организации медпомощи, основанная на выборе диагностических, лечебных и профилактических средств, которые являются оптимальными для конкретного пациента, с учётом его генетических, физиологических, биохимических, поведенческих и других особенностей. Исходя из этого можно утверждать, что теперь любой процесс в организме человека можно наблюдать и измерять. Медицине нужны удобные инструменты для обработки, для хранения, для передачи массива данных. Модель организации медицинской помощи людям за счет рационального подбора лекарств актуальное направление технологий Big Data. Геномный Проект (PGP), использующий большие данные — поставила перед собой цель секвенирования или иными словами «прочесть» молекулу, 100000 добровольцев, чтобы направить исследования на персональную геномику и персональную медицину [4].

Это исследование разработано в Гарвардской Медицинской Школе. В рамках проекта планируется выполнить расшифровку персональных геномов множества добровольцев, их генетические данные вместе с записями из медицинских карт будут доступны для дальнейшего анализа. В рамках проекта расшифровка персональных геномов добровольцев, генетические данные которых вместе с записями из медицинских карт будут доступны для дальнейшего анализа [5,6]. Применение технологий Big Data в биоинформатике позволяет использовать методы компьютерного анализа в сравнительной геномике. Проект GenePool создан для поиска биомаркеров в геноме человека, а также анализа и хранения полученной в ходе исследований информации[7]. С помощью биомаркеров возможно оценить общее состояние организма, выявить отклонения и уничтожить конкретный вредоносный ген. Для определения наличия или отсутствия раковых клеток вводится в организм специальное вещество, в состав которого входит онкомаркер, затем оно выявляет или не выявляет зараженные клетки. Биомаркер может определять и степень поражения тканей, и область распространения метастазов вокруг них. В результате эти вещества позволяют точно скорректировать зону лечения и избежать удаления здоровых клеток.

Проект GenePool использует для поиска биомаркеров последовательности ДНК молекулы, объясняющей ее строение. Один человеческий геном содержит в себе около трех миллиардов символов GenePool систематизирует весь огромный поток полученных

данных. [8]. Применение технологий Big Data, расшифровывающей этот массив данных и выявляет закономерности в геномном мире. С помощью технологий Big Data учёные проанализировали геномы многих тысяч человек, больных и здоровых. Основываясь на анализе огромного количества информации генетического кода, система подбирает лечение. Так, например, болезнь Альцгеймера зависит от генов, участвующих в развитии рассеянного склероза и паркинсонизма[9,10,11].

В определении наиболее эффективных в клиническом аспекте и рентабельных применяемых методов лечения может помочь анализ больших данных. Для фармацевтических и медицинских секторов обнаружение заболевания на ранних стадиях дает возможность моделировать будущий спрос на свою продукцию. Это является основой для принятия решений об инвестициях в научные исследования и разработки. Важнейшей задачей повышения уровня здравоохранения является анализ глобальных закономерностей заболеваемости с целью выявления тенденций закономерностей заболеваемости на ранней стадии [12].

Фармацевтические компании, используя потенциал «больших данных» сокращают время и средства при разработке новых препаратов на этапе их вывода на рынок и затраты на проверку заведомо неперспективных компонентов при разработке лекарств. [13]. Благодаря технологиям «Больших данных» Германии, например, по анализу крови пациентов и доноров выявляются онкологические заболевания либо предрасположенность к ним. В результате своевременной диагностики существенно повышается эффективность лечения и снижаются затраты государства и самих людей. Российские разработчики «Элайд Инновейшенс» занимаются развитием проекта интеллектуальной компьютерной системы для прогноза течения онкологических заболеваний на базе технологий обработки больших данных. Помогают врачам выбирать наиболее эффективную методику лечения в каждом конкретном случае[14]. Глобализация в мире привело к тому, что население становится мобильнее, а географические границы — более размытыми и для медицинских работников всё более важно иметь возможность получить доступ к данным приезжих пациентов. Анализ и обработка “больших данных”, относящихся к определённым регионам и группам населения, улучшит планирование медицинского обслуживания. Раньше медицинские учреждения в основном имели дело с отдельными компаниями, которые разрабатывали и поставляли им аппаратуру, программное обеспечение к ней, создавали прикладные решения «местного значения». Сейчас в эту сферу стали активно вторгаться государственные организации и им требуется доступ к данным о пациентах в любой точке и в любое время. Например, в Бостоне в местном госпитале хранится история болезни. Но если в Нью-Йорке попасть в аварию, то врачу скорой помощи будет очень важна информация о хронических болезнях или аллергии на лекарства.

Быстрый доступ к таким данным поможет оказать квалифицированную помощь. В США есть программа по созданию региональной инфраструктуры здравоохранения. InterSystems участвовала в реализации этой программы в Нью-Йорке. В созданной системе содержатся записи по 25-30 млн пациентов обо всех случаях взаимодействия пациента с врачами.

При этом также важным является распознавать речь (неструктурированные данные) и превращать ее в текст (структурированные данные). Объемы данных получаются гигантскими, так как человек говорит примерно в три-четыре раза больше, чем пишет. InterSystems предлагает технологии, позволяющие фильтровать сырые данные и извлекать из них информацию, имеющую ценность[15]. Эта технология применима и для психически неполноценных людей. Так, например, можно обнаружить депрессию, так как депрессивные лица, как правило, говорят более медленно. Люди все больше занимается самолечением. Развиваются открытые информационные системы для самодиагностики. Все больше и больше людей приобретает носимую электронику, а различные мобильные девайсы постоянно развиваются и наращивают функционал (уровень сахара, частота сердечных сокращений, давление, активность, диагностические анализы и прочее). В здоровых людей могут встраиваться USB-порты для диагностики. Уже сейчас в странах Запада анализы делаются дома самостоятельно, клиническая картина сохраняется в облаке, а общение с врачом происходит дистанционно – через интернет. Больные диабетом, могут использовать различные устройства с подключением к сети, которые помогли бы людям периодически отправлять такие данные, как уровень сахара в крови. Таким образом, посредством внутренних датчиков человек в режиме реального времени отслеживает свое состояние и программа контролирует уровень сахара в крови пациента. Агрегировать (объединять полученные данные) данные об уровне сахара в крови, данные о питании и состоянии здоровья, полученные из различных источников - это возможность получения медицинской документации, необходимой врачам и диабетологам. Комбинируя эти данные с возможностями аналитических инструментов, предоставляется возможность миллионам больных диабетом получить более качественного медицинского обслуживания[16].

Внедрение и развитие подобных датчиков порождает огромное количество данных, целый поток, который необходимо оперативно не просто обрабатывать, но и хранить для дальнейшего анализа.

Другой пример обработки больших потоков данных в реальном времени. Цель медицинского центра Колумбийского университета состоит в том, чтобы обеспечить своевременной информацией медицинских работников. Если в режиме реального времени отслеживать свое состояние больных, то передовая

аналитика диагностирует серьезные осложнения о повреждении аневризмы головного мозга на целых 48 часов раньше[13].

В настоящее время функционирует система Watson претендующая на замещение человека при принятии решений в цифровом мире система такого класса, реально претендующая на замещение человека при принятии решений в цифровом мире. Watson является платформой IBM BIG DATA. Watson реализован на идеях массово-параллельных вычислений, адаптированных под вычислительную платформу. С 2006 года здесь работает огромный совместный коллектив медиков и программистов. Более 100 различных методов используются для анализа естественной речи, определения источников данных, генерации гипотез, поиска и ранжирования фактов, объединения и подтверждения гипотез. В медицине Watson помогает анализировать специфические симптомы, историю болезни, историю наследственных заболеваний, и выполнять синтез и обработку этих данных с доступной неструктурированной и структурированной медицинской информацией, включая медицинские книги и статьи[17]. Применяются технологии Big Data для обработки данных геопозиционирования, которые накапливаются операторами сотовой связи. Alex Pentland и David Laser из Массачусетского технологического института, в 2009 году обосновали, что анализ геоданных не менее информативен для прогноза социально значимых заболеваний, чем геном. В происхождении мультигенного заболевания значительную роль играют факторы внешней среды, которые называют экспозом. Анализируя перемещения человека с достаточной степенью детализации можно получить информацию характерных для определённого синдрома, заболевания, например, с развитием сердечно-сосудистых заболеваний или ожирения[14].

Основная задача системы трехмерной цифровой медицины — обеспечение работы с параметризованной, максимально подробной (вплоть до тканевого и клеточного уровней) 3D-моделью органов и систем человека в норме и в патологии. Это обеспечивает адресность оказания лекарственной помощи пациентам. Рост во всем мире числа раковых заболеваний — главная причина применения томографии и технологий трехмерной медицины[18]. Процесс биомеханического анализа любой биологической системы на основе разработанной физико-математической модели требует ввода данных о пространственной геометрии объекта. Для работы с полученными изображениями, современные компьютерные томографы оснащаются специальным программным обеспечением, применяемые в области лучевой терапии и онкологии[19].

III. СЛОЖНОСТИ ПРИ РАБОТЕ С BIG DATA

Множество ограничений на распространение данных о пациентах существует в медицине. Даже в США не все учреждения здравоохранения готовы к работе с большими

данными. По результатам опроса, который проводила MeriTalk среди топ-менеджеров из сферы государственного здравоохранения США менее 25% считают, что их ведомства готовы работать с большими данными, 59% уверены, что успешная работа с большими данными станет фактором повышения эффективности их учреждения в ближайшие годы, а 63% опрошенных считают, что технологии больших данных позволят более эффективно отслеживать состояние здоровья пациентов[2]. По оценкам IBM, в мире ежедневно генерируется 2,5 квинтиллиона (2,5*10¹⁸) байт данных. Большая часть из них – это Big Data. Экспоненциально увеличился и объем данных, управляемых компаниями по всему миру. При отсутствии четкой структуры и политики использования информация беспорядочно нагромождается, и, как следствие, организациям становится сложно управлять ею. «Большие данные» не гарантируют «большой смысл» них отсутствует структура корректность достоверность [20]. По определению Майкла Франклина из Университета в Беркли «Большие данные – это любые данные, работа с которыми требует значительных затрат и из которых трудно извлечь информацию»

Ряд вопросов, которые приходится решать при работе с большими данными:

1. Большую часть данных в биологии и медицине составляют сложные для анализа из-за неполноты описания явления. Необходима творческая доработка аналитической модели при помощи гипотез. Следует развивать и логику, и оптимальные стратегии работы.

2. Вопрос защиты персональной информации клиентов. Позволят ли люди обрабатывать компаниям их персональные данные и вообще всю поступающую от клиентов информацию.

3. Нужны технологии, позволяющие повысить предсказуемость затрат времени, ресурсов, а также итоговую результативность работы дата-центров. Значительное присутствие непредсказуемых и случайных факторов особенно в медицине, где многое меняется «на лету». Если клиенты генерируют большое число запросов, то это приводит к совместному использованию дата-центров, что приводит к непредсказуемости их работы[21].

Таким образом, для организации эффективного управления данными необходимо комплексное решение всех этих задач.

ЗАКЛЮЧЕНИЕ

Анализ больших потоков данных из самоотчетов и личных медицинских записей, в том числе медицинских тестов, прописанных лекарств, а также томографов, всевозможных датчиков, дает возможность резкого улучшения в сфере здравоохранения. Количество информации, требующей анализа, продолжает непрерывно расти. Доступность, непрерывность, удобство использования, масштабируемость, способность управлять на различных уровнях детализации,

конфиденциальность и безопасность решений, а также контроль качества данных имеют важное значение для эффективного управления данными. Применение технологии «больших данных» крайне востребовано, так как последние смогут решить актуальные задачи в медицине и ряде смежных областей. Учитывая экспоненциальный рост объемов данных и новые средства их расширенного анализа возможно преобразование системы здравоохранения. Поэтому применение современных компьютерных методов решения этих проблем в будущем значительно увеличит продолжительность жизни. Применение современных компьютерных методов решения этих проблем в будущем значительно увеличит продолжительность жизни.

ЛИТЕРАТУРА

- [1] Kyoungyoung Jee, Gang-Hoon Kim Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System / Healthcare Informatics Research, Korea, 2013, v.19, Issue 2
- [2] http://www.cnews.ru/reviews/new/bi_bigdata_2014/articles/chto_bolshie_dannye_mogut_dat_biznesu/
- [3] Черняк Л. Серьезно о технологиях для Больших Данных/Открытые системы <http://www.osp.ru/os/2014/01/13039646/>
- [4] http://en.wikipedia.org/wiki/Personal_Genome_Project
- [5] Васильков А. Медицина становится персонализированной под влиянием «больших данных», <http://www.computerra.ru/88609/big-data-improve-personalized-medicine/>
- [6] <http://rusangels.ru/naba/pressroom/smi-about-naba/v-kakie-startapy-rekomenduyut-vkladyvat-rossiyskie-investory.html>
- [7] <http://www.stationxinc.com/how-genepool-works>
- [8] <http://mygenome.su/articles/112/>
- [9] Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease // Nature genetics, <http://www.nature.com/ng/journal/v45/n12/full/ng.2802.html>
- [10] Swarup V.Geschwind D.H. Alzheimer's disease: From Big Data to mechanism // Nature, 2013, v500
- [11] Bettens Karolien, Slegers Kristel, Van Broeckhoven, Genetic insights in Alzheimer's disease, 2013, v12, Issue 1
- [12] "Большие данные", большое дело, большая проблема, ITU News <https://itunews.itu.int/Ru/Note.aspx?Note=4879>
- [13] Wullianallur Raghupathi, Viju Raghupathi Big Data analytics in healthcare: promise and potential / Health Information Science and Systems, 2014, v2
- [14] Трифонова О. П., Ильин В. А. «Большие данные» в биологии и медицине/Acta Naturae / Медицина и здравоохранение, ГРНТИ: Научная библиотека, 2013, № 3
- [15] Большие Данные в медицине и здравоохранении <http://www.osp.ru/medit/2014/04/13040834.html>
- [16] Новости и статьи с тегом Сколково http://leptandra58.rssing.com/chan-33366061/all_p1.html
- [17] Шмид А. Big Data <http://www.interface.ru/home.asp?artId=34991>
- [18] Турлапов В., Гаврилов Н., Сапрыкин В. Платформа цифровой медицины «Открытые системы», № 05, 2014 <http://www.osp.ru/os/2014/05/13041824/>
- [19] Skripcak T., Belka C., Bosch W. Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets Radiotherapy and Oncology, 2014, v 113, Issue 3
- [20] <http://datareview.info/article/10-zadach-kotoryie-ne-smogut-reshit-bolshie-dannye/>
- [21] Microsoft о будущем интерфейсов и Big Data <http://ria.ru/technology/20130927/966349131.html>