

АНАЛИЗ ДАННЫХ В МЕДИЦИНСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ DATA MINING

DATA ANALYSIS IN HOSPITAL INFORMATION SYSTEM BY USING DATA MINING TECHNOLOGY

**N. Markina
S. Kasyuk
T. Shamaeva**

Summary. The article describes an overview of Data Mining technology for analyzing patient's records in hospital information system (HIS). The operation of HIS based on electronic medical record within the framework of functioning unified state healthcare system is described. Main directions and objectives of Data Mining technology for analysis of medical data are described. Architecture of medical data mining is outlined.

Keywords: hospital information system, medical data, electronic medical record, data analysis, Data Mining.

Маркина Нэлли Васильевна

*К.т.н., доцент, ФГБОУ ВО «Южно-Уральский
государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(г. Челябинск)*

markina.nel@yandex.ru

Касюк Сергей Тимурович

*К.т.н., доцент, ФГБОУ ВО «Южно-Уральский
государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(г. Челябинск)*

sergey.kasyuk@gmail.com

Шамаева Татьяна Николаевна

*К.п.н., доцент, ФГБОУ ВО «Южно-Уральский
государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(г. Челябинск)*

shamtan@rambler.ru

Аннотация. В статье представлен обзор использования технологии Data Mining для анализа данных пациентов в медицинских информационных системах (МИС). Дано описание работы МИС на базе интегрированной электронной медицинской карты (ИЭМК) в рамках функционирования Единой государственной системы в сфере здравоохранения (ЕГИСЗ). Приведены основные направления применения и задачи технологии Data Mining для анализа медицинской информации. Приведены сведения об архитектуре медицинских систем Data Mining.

Ключевые слова: медицинская информационная система, медицинская информация, электронная медицинская карта, анализ данных, Data Mining.

Работа медицинских организаций (МО) сопровождается накоплением больших объемов данных о состоянии здоровья пациентов. В связи с развитием информационных технологий рукописные заметки врача в карте пациента или истории болезни начинают заменять электронными записями, и все больше медицинских данных начинают храниться в цифровом виде. Это позволяет, привлекая технологию интеллектуального анализа данных Data Mining, решать задачи постановки диагноза, назначения лечения, прогнозирования развития заболевания и повышения эффективности здравоохранения в целом.

Технология Data Mining используется для обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации закономерностей [1].

Особенности медицинской информации

К особенностям медицинской информации следует отнести:

1. *Неоднородность.* Медицинская информация является неоднородной по своей природе, собира-

емой из различных источников: флюорография, компьютерная томография, ультразвуковые исследования, лабораторные исследования, жалобы пациентов, анамнезы и т.д. [2].

2. *Большие объемы.* Медицинская информация представляется в виде больших массивов данных, требующих соответствующего места для хранения и инструментария для анализа.
3. *Сложность.* Медицинские документы являются сложными по своей природе для структурирования и математического описания.
4. *Исключительная важность.* Медицинская информация является исключительно важной, поскольку она служит для диагностики и лечения заболеваний.
5. *Чувствительность.* Медицинская информация является чувствительной к мнению и оценкам врача, медицинским осмотрам и клиническим анализам. Большинство диагнозов и анализов содержат некоторый процент ошибок [2].
6. *Правовые, этические и социальные аспекты.* Медицинская информация — это в основном данные пациентов. В связи с этим возникают вопросы обеспечения конфиденциальности персональных медицинских данных в соответствии с действующим законодательством.

Медицинские информационные системы

На сегодняшний день медицинская информация доступна из тысяч баз данных МО, являющихся составной частью *медицинских информационных систем* (МИС) МО. В ходе работы МИС МО могут генерироваться гигабайты информации, нуждающиеся в анализе.

Создание МИС МО на территории РФ регламентируется федеральным законодательством, гостами и нормативными документами Министерства здравоохранения.

Согласно Федерального закона от 21 ноября 2011 г. № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» МО имеют право «...создавать медицинские информационные системы, содержащие данные о пациентах, об оказываемой им медицинской помощи, о медицинской деятельности медицинских организаций с соблюдением требований, установленных законодательством Российской Федерации в области персональных данных, и соблюдением врачебной тайны» [4].

А.В. Гусев, Т.В. Зарубина формулируют основную задачу МИС МО на текущем этапе как «...переход на преимущественное ведение *электронных медицинских карт пациентов* (ЭМК), накопление первичной медицинской информации в электронном виде и автоматиза-

ция базовых задач, таких как сокращение непрофильных затрат медицинского персонала на ведение документации, автоматическое формирование медицинской статистики, сдача реестров по ОМС, автоматизация учета лекарственных средств и т.д.» [5].

В соответствии с *Методическими рекомендациями* [6] базовые функциональные возможности МИС МО должны обеспечивать: персонифицированный учет оказанной медицинской помощи на основе ведения базы данных отчетных форм, взаиморасчеты со сторонними организациями, взаимодействие с реестром нормативно-справочной информации, взаимодействие с интеграционным шлюзом для передачи и получения данных, построение медико-статистических отчетов, ведение электронной медицинской карты пациента, обмен данными внутри МО, управление потоками пациентов, ведение расписаний работы.

Единая государственная информационная система в сфере здравоохранения

В РФ эксплуатация МИС МО рассматривается в качестве составной части и первичного звена *Единой государственной информационной системы в сфере здравоохранения (ЕГИСЗ)*, формирующей единое *информационное пространство* в сфере здравоохранения [5].

ЕГИСЗ представляет собой совокупность информационно-технологических и технических средств, обеспечивающих информационную поддержку методического и организационного обеспечения деятельности участников системы здравоохранения. ЕГИСЗ обеспечивает информационное взаимодействие организаций системы здравоохранения, участников лечебно-диагностического процесса в среде единого информационного пространства [6, 7].

Согласно *Концепции* [7] целью создания ЕГИСЗ в РФ является *обеспечение эффективной информационной поддержки процесса управления системой медицинской помощи, а также процесса оказания медицинской помощи.*

Проблема заключается в том, что в настоящее время МИС конкретных МО проектируются и разрабатываются децентрализованно в условиях отсутствия единой методологии, что создает *трудности с интеграцией данных* различных МИС и *анализом деятельности* системы здравоохранения в целом [7].

Следует отметить, что разработка аналитических систем в области медицины является общемировой тен-

денцией. Например, к политическим решениям ЕС в области Big Data для здравоохранения и телемедицины относится разработка новых *аналитических инструментов*, таких как Data Mining, обеспечивающих мощные предсказывающие способности и интегрирующих имеющиеся клинические данные (например, биомедицинские или геномные) с контекстными реальными данными [8].

Согласно рабочего документа Еврокомиссии о *возможности цифровой трансформации здравоохранения в единый цифровой рынок* [9] в ЕС в настоящее время разрабатываются подходы к *организации высокопроизводительных вычислений, анализу данных и искусственному интеллекту* для создания и тестирования продуктов в области здравоохранения, обеспечения быстрой диагностики и повышения эффективности лечения.

Интегрированная электронная медицинская карта

Особое место в структуре ЕГИСЗ занимает *Интегрированная электронная медицинская карта (ИЭМК)*, позволяющая перейти на единый формат *структурированных электронных медицинских документов* в МИС МО и являющаяся *совокупностью электронных персональных медицинских записей, относящихся к одному человеку, собираемых и используемых несколькими МО* [10].

Особенность ведения ИЭМК заключается в том, в ней аккумулируется медицинская информация, полученная из различных МО, для предоставления её уполномоченным медицинским работникам других МО, связанных с диагностикой, лечением или профилактикой субъекта ИЭМК [11].

Основные задачи ИЭМК представлены на сайте ЕГИСЗ¹:

1. получение, проверка, обработка и хранение структурированных обезличенных сведений, указанных в статье 94 Федерального закона № 323-ФЗ [4], о лицах, которым оказывается медицинская помощь, а также о лицах, в отношении которых проводятся медицинские экспертизы, медицинские осмотры и медицинские освидетельствования;
2. формирование баз данных обезличенной информации по отдельным нозологиям и профилям оказания медицинской помощи, позволяющие систематизировать информацию для изучения

¹ ЕГИСЗ [Электронный ресурс]. — Режим доступа: <https://egisz.rosminzdrav.ru> (дата обращения 12.04.2019).

течения и исхода заболеваний, изучения клинической и экономической эффективности методов профилактики, диагностики, лечения и реабилитации при отдельных заболеваниях, состояниях в соответствии с *Международной статистической классификацией болезней и проблем, связанных со здоровьем, и номенклатурой медицинских услуг*, утверждаемой уполномоченным федеральным органом исполнительной власти.

На *Портале оперативного взаимодействия участников ЕГИСЗ* размещены проекты национальных стандартов «Электронная медицинская карта. Интегрированная электронная медицинская карта» [11] и «Электронная медицинская карта. Электронная медицинская карта, используемая в медицинской организации» [12].

Система ведения ИЭМК, аналитические системы, а также системы автоматизированного контроля и поддержки принятия управленческих решений на основе анализа первичных данных относятся к федеральным управленческим системам. Анализ данных ИЭМК должен осуществляться централизованно в *Федеральном центре обработки данных Министерства здравоохранения и социального развития РФ* [7].

На сегодняшний день в РФ решаются вопросы по технологии взаимодействия МИС МО с системой ведения ИЭМК, юридические вопросы, связанные с передачей медицинской информации пациента, вопросы электронной цифровой подписи при передаче данных в ЕГИСЗ.

Основные аспекты разработки, планы разработки и внедрения базовой ИЭМК в РФ описаны в статье Т.В. Зарубиной, С.Л. Швырева, В.Г. Соловьева и др. «Интегрированная электронная медицинская карта: состояние дел и перспективы» [13].

Основные задачи Data Mining

Основные задачи, решаемые методами Data Mining [1]:

1. *Кластеризация* — это группировка объектов на основе данных, описывающих их свойства, при этом число кластеров может быть заранее неизвестно (например, кластеризация пациентов по симптомам заболеваний).
2. *Регрессия* — установление функциональной зависимости между факторами или показателями (например, установление зависимости между показателями клинического анализа крови у пациентов).
3. *Классификация* — это определение класса объекта на основе его признаков, при этом множе-

ство классов, к которым можно отнести объект, заранее известно (например, постановка медицинского диагноза по данным жалоб пациента, его анамнеза и клинических исследований).

4. *Ассоциация* — выявление закономерностей между связанными объектами или событиями (например, выявление групп лекарственных препаратов, которые врачи выписывают пациентам совместно при определенных заболеваниях).
5. *Последовательные шаблоны* — выявление закономерностей между связанными во времени событиями (например, если пациент начал принимать лекарственный препарат, то с некоторой вероятностью через определенное время состояние этого пациента улучшится).
6. *Анализ временных рядов* — это процесс прогнозирования временных рядов, базирующийся на выявлении закономерностей, объясняющих динамику процесса в прошлом, и использовании этих закономерностей для описания развития в будущем (например, анализ динамики развития заболевания или динамики внутренней картины болезни в процессе лечения).

Анализ медицинской информации с использованием технологии Data Mining

Построение систем Data Mining для анализа медицинской информации возможно как на уровне федеральных сервисов в рамках ЕГИСЗ, так и на уровне сервисов конкретной МО в рамках МИС.

Можно выделить следующие основные направления использования технологии Data Mining в области анализа медицинских данных:

1. *Диагностика и прогнозирование заболеваний.* Применение технологии Data Mining позволяет диагностировать заболевания пациента, его стадии, характер течения болезни, а также осуществлять прогнозирование изменения клинического состояния пациента при применении различных вмешательств [3]. Пример системы прогнозирования кардиологических заболеваний приведен в работе [15].
2. *Выявление лучших методов лечения и оценка эффективности лечения.* С помощью технологии Data Mining можно выявить лучшие методы лечения из доступных, сравнивая такие факторы, как причины заболеваний, симптомы, время и стоимость лечения, а также побочные эффекты [2].
3. *Оказание пациентам качественной медицинской помощи.* Анализ больших информационных массивов медицинских данных может выявить новые закономерности (например, о ранее неизвест-

ных аспектах методов лечения или применения лекарственных препаратов), с помощью которых можно повысить качество оказания медицинской помощи. Кроме того, технология Data Mining может оценить качество работы МО на основании заданных критериев, проанализировать потребности конкретного пациента с целью улучшения качества оказываемых медицинских услуг [2].

4. *Прогнозирование потребности в медицинской помощи* на основании данных об уровне и характере заболеваемости по различным нозологическим формам, группам населения, территориальным и другим признакам [7].
5. *Контроль за распространением внутрибольничных инфекционных заболеваний.* Инфекционный контроль с помощью технологии Data Mining является более чувствительным, чем традиционные наблюдения за распространением инфекции [3]. Примером может служить система инфекционного контроля, разработанная в университете Алабамы [14].
6. *Выявление пациентов высокого риска.* Используя методы прогнозирования Data Mining, можно оценивать риски возникновения тяжелых заболеваний у пациентов, оперативно выявлять лиц, подверженных риску резкого ухудшения состояния здоровья, а также заболевших лиц [7]. Примером является система предсказания генетически-обусловленных заболеваний, описанная в работе [17].
7. *Построение рейтинга МО.* Технологию Data Mining можно использовать для анализа качества оказания медицинской помощи населению в различных МО, оценки оснащенности МО необходимыми материально-техническими ресурсами, анализа соответствия материально-технической обеспеченности МО стандартам оснащения и построения рейтинга этих МО. Такое ранжирование используется для выявления МО, способных лечить пациентов высокого риска [2].
8. *Управление ресурсами и лечебно-диагностическим процессом в МО.* Технология Data Mining позволяет создавать модели для управления ресурсами и совершенствования лечебно-диагностического процесса в МО [2].
9. *Контроль и надзор в сфере здравоохранения.* Используя технологию Data Mining можно осуществлять контроль за соблюдением гарантированного объема и качества оказания медицинской помощи с учетом медицинских и медико-экономических стандартов, мониторинг отпускных и закупочных цен на лекарственные средства, анализ соответствия предоставленных лекарственных средств гарантированным перечням лекарственного обеспечения для категорий граждан [7]. Технология Data Mining позволяет

осуществлять мониторинг и экспертизу действий медицинского персонала, в ходе которых оцениваются: адекватность вмешательства диагнозу, корректность дозировок и длительности приема лекарственных средств, соответствие организационных действий состоянию больного, поиск похожих клинических случаев [3].

10. *Выявление злоупотреблений и случаев страхового мошенничества.* Технология Data Mining позволяет выявить злоупотребления и случаи мошенничества: выставление счетов за фактически не проведенные процедуры, списание медикаментов, назначение ненужных дорогостоящих препаратов, выставление завышенных счетов за медицинские услуги и т.д. Примером является система прогнозирования стоимости лечения по данным страховых компаний, описанная в работе [16].

Анализ клинических текстов

Большинство медицинской документации является текстами на естественном языке, содержащими сведения, нуждающиеся в извлечении и структурировании. Анализ клинических текстов является одним из наиболее перспективных направлений использования технологии Data Mining в медицине. Так, А.В. Гусев в своей работе по искусственному интеллекту в здравоохранении пишет, что в этой области возможны «...автоматический анализ врачебных записей и медицинских документов и последующая постановка предварительного диагноза с выводом рекомендаций по тактике обследования и лечения» [18].

В исследовании, выполненном при поддержке РФФИ (проект № 13–04–12062), показано, что «...большинство существующих методов и систем анализа медицинских текстов работают только с английским языком, системы анализа медицинских текстов на русском языке отсутствуют». Авторы этого исследования отмечают, что «...большинство существующих решений ориентированы на обработку текстов общего характера, например, таких как новостные сообщения. Однако стилистика клинических текстов сильно отличается от стилистики обычных текстов, поэтому требуется как значительная доработка существующих методов и инструментов по анализу текстов на естественном языке, так и создание новых специфичных подходов» [3].

В настоящее время ведутся работы по разработке подходов, методов, программных продуктов для анали-

за клинических текстов, однако разрабатываемые системы в основном ориентированы на анализ англоязычных текстов и основаны на известных платформах обработки текстов на естественном языке общего назначения, таких как GATE и Apache UIMA [3].

Архитектура медицинских систем Data Mining

Основные сведения об архитектуре медицинских систем Data Mining приведены в исследовании [3]: многопроцессорные вычислительные системы, вычислительные кластеры, системы распределенных вычислений в Интернете, грид-системы (географически распределенные вычислительные платформы). Авторы исследования исходя из требований по стоимости, распределенности ресурсов, защищенности каналов, объемам данных и масштабируемости и делают вывод о целесообразности использования систем Data Mining, архитектуры которых основаны на грид. В таких системах часто применяется архитектура Data Mining Grid Architecture (DMGA), а качестве программных платформ используются Globus Toolkit и UNICORE.

Заключение

На сегодняшний день в РФ происходит массовое внедрение информационных технологий в работу МО, создание МИС МО, разработка ЭМК и ИЭМК, построение ЕГИСЗ, что приводит к накоплению больших объемов медицинских данных. Аналитические технологии Data Mining внедряются как федеральные сервисы на уровне ЕГИСЗ, так и сервисы на уровне МИС МО. К основным вопросам внедрения технологии Data Mining следует отнести: разработку форматов структурированных электронных медицинских документов для ИЭМК и ЭМК, отладку технологии взаимодействия МИС МО с системой ведения ИЭМК, совместимость форматов медицинских документов МИС в различных МО, анализ клинических текстов на русском языке, анализ больших объемов медицинских данных различными методами Data Mining, обеспечение конфиденциальности медицинских данных, выбор архитектуры и программных платформ для систем Data Mining. Решение перечисленных вопросов позволят использовать технологию Data Mining как для повышения качества оказания медицинской помощи, так и для контроля и надзора в сфере здравоохранения.

ЛИТЕРАТУРА

1. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям. — СПб.: Питер, 2009. — 624 с.
2. Pandey S. C. Data Mining techniques for medical data: a review // International conference on Signal Processing, Communication, Power and Embedded System. — 2016. — November. — P. 972–982.

3. Методы и средства комплексного интеллектуального анализа медицинских данных / А. А. Баранов, Л. С. Намазова-Баранова, И. В. Смирнов [и др.] // Труды ИСА РАН. — 2015. — Т. 65. — № 2. — С. 81–93.
4. Федеральный закон от 21 ноября 2011 г. N323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». [Электронный ресурс]. — Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_121895 (дата обращения: 12.04.2019).
5. Гусев А. В., Зарубина Т. В. Поддержка принятия врачебных решений в медицинских информационных системах медицинской организации // Врач и информационные технологии. 2017. — № 2. — С. 60–72.
6. «Методические рекомендации по обеспечению функциональных возможностей медицинских информационных систем медицинских организаций», утвержденные Министром здравоохранения Российской Федерации В. И. Скворцов 1 февраля 2016 г. [Электронный ресурс]. — Режим доступа: <https://portal.egisz.rosminzdrav.ru/materials> (дата обращения: 12.04.2019).
7. «Концепция создания единой государственной информационной системы в сфере здравоохранения», утвержденная Приказом Минздравсоцразвития России от 12.04.2012 № 348 [Электронный ресурс]. — Режим доступа: <http://portal.egisz.rosminzdrav.ru/materials/99> (дата обращения: 12.04.2019).
8. Study on Big Data in Public Health, Telemedicine and Healthcare: Final Report of the European Commission, December 2016. [Электронный ресурс]. — Режим доступа: https://ec.europa.eu/health/sites/health/files/ehealth/docs/bigdata_report_en.pdf (дата обращения 12.04.2019).
9. Commission staff working document Accompanying the document communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society (SWD/2018/126 final) [Электронный ресурс]. — Режим доступа: <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52018SC0126> (дата обращения 12.04.2019).
10. Зингерман Б. В., Шкловский-Корди Н. Е., Карп В. П., Воробьев А. И. Интегрированная электронная медицинская карта: задачи и проблемы // Врач и информационные технологии. — 2015. — № 1. — С. 24–34.
11. Проект национального стандарта «Электронная медицинская карта. Интегрированная электронная медицинская карта». [Электронный ресурс]. — Режим доступа: <http://portal.egisz.rosminzdrav.ru/materials/311> (дата обращения 12.04.2019).
12. Проект национального стандарта «Электронная медицинская карта. Электронная медицинская карта, используемая в медицинской организации». [Электронный ресурс]. — Режим доступа: <http://portal.egisz.rosminzdrav.ru/materials/309> (дата обращения 12.04.2019).
13. Интегрированная электронная медицинская карта: состояние дел и перспективы / Т. В. Зарубина, С. Л. Швырев, В. Г. Соловьев [и др.] // Врач и информационные технологии. — 2016. — № 2. — С. 35–44.
14. Brosette S. E., Sprague A. P., Jones W. T., Moser S. A. A data mining system for infection control surveillance // Methods of Information in Medicine — 2000. — Dec; 39 (4–5). — P. 303–310.
15. Soni J., Ansari U., Sharma D. Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction // International Journal of Computer Applications (0975–8887). — 2011. — Volume 17. — № 8, March. — P. 43–48.
16. Big data in health care: using analytics to identify and manage high-risk and high-cost patients / David W. Bates, Suchi Saria, Lucila Ohno-Machado et al. // Health Affairs. — 2014. — V. 33. — № 7. — P. 1123–1131.
17. Tassy O., Pourquie O. Manteia, a predictive data mining system for vertebrate genes and its applications to human genetic diseases // Nucleic acids research. — 2014. — Jan; 42. — P. 882–891.
18. Гусев А. В. Перспективы нейронных сетей и глубокого машинного обучения в создании решений для здравоохранения // Врач и информационные технологии. — 2017. — № 3. — С. 92–105.

© Маркина Нэлли Васильевна (markina.nel@yandex.ru),

Касюк Сергей Тимурович (sergey.kasyk@gmail.com), Шамаева Татьяна Николаевна (shamtan@rambler.ru).

Журнал «Современная наука: актуальные проблемы теории и практики»