

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, доцента Кашириной Ирины Леонидовны на диссертационную работу Гришиной Любови Сергеевны на тему: *«Методы и алгоритмы интеллектуальной поддержки принятия решений в медицинской практике на основе обработки естественных языков»*, представленную на соискание ученой степени кандидата технических наук по специальности 2.3.1 – «Системный анализ, управление и обработка информации, статистика»

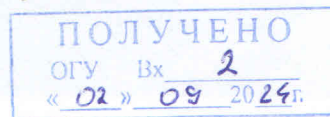
Актуальность темы исследования

В условиях современного здравоохранения наблюдается стремительный рост объемов данных, генерируемых в процессе оказания медицинских услуг. Большинство информации в медицинских информационных системах хранится в виде разношаблонных текстовых документов, таких как выписки, истории болезней и другие клинические записи. Данный подход приводит к тому, что полезная информация может быть представлена в разных форматах, что существенно затрудняет их обработку и анализ. В результате, слабоструктурированные данные могут оставаться невостребованными, что негативно сказывается на качестве принятия клинических решений.

Развитие цифровой медицины предполагает организацию медицинской помощи, основанной на использовании больших объемов данных в цифровом формате. Эффективное использование результатов обработки и анализа этих данных может значительно повысить качество диагностики и лечения, а также оптимизировать процессы управления в здравоохранении. Однако для реализации этого потенциала необходимо преодолеть существующие барьеры, связанные с анализом слабоструктурированной информации.

Методы обработки естественных языков предоставляют мощные инструменты для извлечения полезной информации из текстовых данных. Они позволяют автоматизировать процесс анализа, выявлять ключевые паттерны и закономерности, а также поддерживать врачей в принятии более обоснованных решений. Применение методов обработки естественных языков в медицинских информационных системах может способствовать созданию систем поддержки принятия решений, которые будут учитывать индивидуальные особенности пациентов и предлагать оптимальные варианты лечения на основе анализа больших объемов данных.

Таким образом, разработка интеллектуальных подходов к анализу слабоструктурированной текстовой информации электронных-медицинских карт на основе методов обработки естественных языков для поддержки принятия решений в медицинской практике могут способствовать улучшению качества



медицинской помощи, повышению удовлетворенности пациентов и эффективному использованию ресурсов здравоохранения. Вышеизложенное подчеркивает актуальность выбранной темы диссертационного исследования.

Степень обоснованности научных положений, выводов и рекомендаций

Содержание диссертационного исследования полностью соответствует сформулированным автором цели и задачам, а также определенным им предмету и объекту исследования.

В процессе проведения исследований применены известные методы системного анализа, обработки информации, методы машинного обучения, нейросетевые технологии, методы обработки естественного языка, а также методы глубокого обучения.

В своем исследовании автор опирается на классическую и современную теоретико-методическую базу, которая включает в себя работы ведущих зарубежных и отечественных ученых. Кроме того, автор использует обширную информационную базу, включающую в себя официальные статистические данные, деперсонализированные данные электронных медицинских карт медицинских организаций Оренбургской области, различные законодательные и нормативные акты. Данный подход обеспечивает обоснованность и достоверность полученных научных результатов.

Текст диссертации дополнен необходимыми рисунками и таблицами, структурирующими и иллюстрирующими представленный материал. Стиль изложения соответствует требованиям, предъявляемым к академическим текстам. Работа обладает внутренним единством и научной логикой, имеет понятную структуру, соответствующую последовательности решаемых задач. Содержание работы полностью соответствует общепринятой схеме диссертации по техническим специальностям, и состоит из введения, пяти разделов, заключения, списка литературы и приложений. В конце каждой главы последовательно описаны частные выводы, все поставленные задачи решены, цель диссертационного исследования достигнута.

Анализ текста диссертации, автореферата, а также публикаций автора позволяет сделать вывод о достаточной степени обоснованности результатов, полученных в данной работе. Кроме того, достоверность результатов исследования подтверждается апробацией основных положений работы на конференциях различного уровня.

Достоверность и новизна полученных результатов

Достоверность полученных результатов диссертационной работы подтверждается проведенными теоретическими и экспериментальными исследованиями, апробацией разработанных алгоритмов и программно-аппаратного комплекса в практике медицинских исследований ГАУЗ

«Оренбургской областной клинической больницы имени В.И. Войнова» и ГАУЗ «Бузулукской больницы скорой медицинской помощи им. академика Н.А. Семашко». Также результаты работы используются в учебном процессе Оренбургского государственного медицинского университета. Достоверность подтверждается также публикациями в научных изданиях из списка ВАК, публикациями в изданиях, индексируемых в базе Scopus и Web of Science, а также участием автора исследования в международных и всероссийских научных конференциях.

Научная новизна исследования согласуется с поставленными в диссертационной работе задачами и обусловлена необходимостью разработки и совершенствования методов извлечения и структурирования знаний из электронных медицинских карт для поддержки принятия решений при диагностике и лечении заболеваний.

Соискателем вынесены на защиту следующие результаты, имеющие научную и практическую значимость:

1. Разработанная концептуальная модель анализа клинических данных и поддержки принятия решений для автоматизации процессов заполнения электронных медицинских карт позволяет формализовать этапы структурирования текстовых данных и построения интеллектуальных моделей генерации рекомендаций к лечению диагностированных заболеваний.

2. Разработанная иерархическая модель данных амбулаторных карт пациентов для сбора информации позволяет обрабатывать разношаблонные XML-документы медицинских информационных систем на основе рекурсивного подхода для обеспечения семантической интероперабельности.

3. Разработанные метод и алгоритм прогнозирования укрупненной группы заболеваний пациентов на основе методов обработки естественных языков и модели логистической регрессии использует уникальный узкоспециализированный корпус текстов, построенный на основе слабоструктурированных текстовых данных электронных медицинских карт.

4. Разработанные метод и алгоритм автоматической генерации индивидуальных шаблонов листа назначений и рекомендаций к лечению на основе алгоритмов обработки естественных языков и модели глубокого обучения GPT-3 позволяет автоматизировать процесс заполнения документов. В отличие от действующего функционала систем поддержки принятия решений в медицинской практике, предложенный подход использует современные предобученные нейросетевые модели трансформеров для анализа слабоструктурированной информации и генерации текстовых данных.

5. Построенный прототип автоматизированного программного комплекса интеллектуальной поддержки принятия решений в медицинской практике

позволяет применять разработанные модули искусственного интеллекта для диагностирования заболеваний и формирования рекомендаций к лечению на основе методов обработки естественных языков.

Теоретическая и практическая значимость полученных результатов

Теоретическая значимость работы заключается в разработке алгоритмов обработки слабоструктурированной текстовой информации разношаблонных документов информационных систем, а также построении узкоспециализированных языковых моделей, построенных на основе методов обработки естественного языка.

Практическая значимость полученных результатов заключается в разработке программного комплекса, позволяющего производить автоматизированный анализ состояния пациента и генерацию индивидуального листа назначений и рекомендаций к лечению на основе методов глубокого обучения. Разработанные алгоритмы прошли апробацию на множестве реальных деперсонализированных данных электронных медицинских карт, полученных из базы данных медицинских организаций Оренбургской области.

Оценка структуры и содержания диссертации

Во введении обосновывается актуальность темы диссертации, формулируется цель и задачи исследования, раскрывается научная новизна работы и ее теоретическая и практическая значимость, представлены основные положения, выносимые на защиту.

В первой главе проведен аналитический обзор в области поддержки принятия решений при диагностике и лечении заболеваний на основе методов машинного обучения. Произведен обзор существующих медицинских информационных систем с поддержкой моделей искусственного интеллекта. Определены проблемы обработки слабоструктурированных медицинских данных при решении задач поддержки принятия решений. Сделан вывод об актуальности исследования подходов к анализу слабоструктурированной текстовой информации электронных-медицинских карт и целесообразности применения методов обработки естественных языков для повышения эффективности принятия решений в медицинской практике.

Во второй главе представлено описание концептуальной модели анализа клинических данных и поддержки принятия решений для автоматизации процессов заполнения электронных медицинских карт. Описана иерархическая модель структурирования данных амбулаторных карт пациентов для сбора информации из XML-документов. Предложены алгоритмы автоматической выгрузки данных электронных медицинских карт и извлечения информации из разнородных XML-документов.

В третьей главе подробно рассматриваются подходы к прогнозированию укрупненных групп заболеваний на основе слабоструктурированных текстовых данных жалоб пациентов и данных объективного осмотра из электронных медицинских карт, приведены результаты реализации моделей прогнозирования на основе методов машинного обучения.

В четвертой главе рассмотрен подход к автоматической генерации индивидуальных листов назначений и рекомендаций к лечению. Представлены алгоритмы токенизации и языковые модели на базе архитектуры трансформер в рамках решения задачи языкового моделирования для автоматической генерации текстовой информации. Приведены результаты обучения языковых моделей генерации медицинского текста для заполнения индивидуальных листов назначений и рекомендаций к лечению.

В пятой главе представлен прототип автоматизированного программного комплекса интеллектуальной поддержки принятия решений в медицинской практике с использованием веб-фреймворка для доступа к построенным алгоритмам и моделям искусственного интеллекта. Кроме того, проведена оценка эффективности программного комплекса на практике в медицинских организациях с учетом затрат рабочего времени.

Содержание автореферата кратко передает основное содержание диссертации, и позволяет получить полное представление об исследовании. Сформулированные научные задачи успешно решены, что позволяет сделать вывод о завершенности работы.

Соответствие диссертации критериям, установленным Положением о присуждении ученых степеней

Диссертация Л.С. Гришиной соответствует всем требованиям, установленным Положением о присуждении ученых степеней, которое утверждено постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, а именно:

– диссертация подготовлена соискателем самостоятельно, обладает внутренним единством; научные положения, методы и решения разработаны лично автором, обладают значительным потенциалом практического использования, что удовлетворяет требованиям п. 10 Положения о присуждении ученых степеней;

– основные результаты исследования полностью опубликованы в научно-технических изданиях, как требует п. 11 Положения о присуждении ученых степеней;

– в диссертационной работе выполнено требование о размещении ссылок на источники заимствования материалов других авторов и на научные работы Гришиной Л.С., что удовлетворяет требованиям п. 14. Положения.

Соответствие диссертации и автореферата обязательным требованиям

Диссертация и автореферат Гришиной Любови Сергеевны написаны стилистически грамотно, соответствуют требованиям ГОСТ Р 7.0.11-2011. Система стандартов по информации, библиотечному и издательскому делу. Диссертация и автореферат диссертации. Структура и правила оформления.

Автореферат достаточно полно отражает основное содержание диссертации, характеризует результаты проведенных исследований.

По результатам проведенных исследований опубликовано 8 научных работ, в том числе 3 статьи в изданиях, рекомендованных ВАК РФ и отечественных изданиях, которые входят в международные базы данных и системы цитирования, а также 2 работы в изданиях, индексируемых Scopus и Web of Science, получено 1 свидетельство о государственной регистрации программы для ЭВМ.

Замечания по диссертационной работе

1. В главе 1 указывается, что обработка больших данных медицинских информационных систем и построение языковых моделей генерации текста требует использования большого объема вычислительных ресурсов GPU для проведения обучения моделей. Однако, в диссертационной работе не представлено подробное описание использованного оборудования и его характеристик, достаточных для реализации разрабатываемых моделей ИИ.

2. Автор утверждает, что система ЕГИСЗ имеет сложную техническую инфраструктуру и среди прочих проблем анализа данных этой системы выделяет наличие в документах опечаток, некорректных сокращений и многовариантные аббревиатуры (раздел 1.4). Однако, количественная оценка исходных данных на наличие выделенных проблем не приводится и не указываются методы и алгоритмы их обработки.

3. В рамках реализации алгоритма выгрузки файлов протоколов из медицинской информационной системы автор указывает, что реализует подход с выделением нескольким скриптам непересекающегося пула протоколов случаев. Однако не до конца остается ясным, на каком этапе реализуется распараллеливание в алгоритме (табл. 2.6).

4. В главе 3 указывается, что для построения модели прогнозирования заболеваний по МКБ-10 на основе слабоструктурированных данных за основу выбирается укрупненная группа заболеваний. Однако, для эффективной генерации рекомендации к лечению не менее важно знать уточненный шифр конкретного заболевания, который автором не учитывается.

5. В работе имеется ряд опечаток и стилистических неточностей (стр. 9, 13, 35, 48, 51, 74 и т.д.).

Указанные замечания не снижают значимости полученных результатов и не влияют на общую положительную оценку диссертационного исследования Гришиной Любови Сергеевны.

Заключение

Диссертация Гришиной Любови Сергеевны на соискание ученой степени кандидата технических наук является научно-квалификационной работой, в которой содержится решение актуальной задачи анализа слабоструктурированной текстовой информации электронных-медицинских карт и применения методов обработки естественных языков для повышения эффективности принятия решений в медицинской практике.

Основные положения диссертации в достаточной мере нашли отражение в публикациях автора. Автореферат выполнен с соблюдением установленных требований, полно и точно отражает содержание диссертации.

На основании вышеизложенного считаю, что диссертационная работа является законченным исследованием, соответствует требованиям, предъявляемым ВАК РФ к диссертациям на соискание ученой степени кандидата технических наук по специальности 2.3.1. – «Системный анализ, управление и обработка информации, статистика», а ее автор, Гришина Любовь Сергеевна, заслуживает присуждения ученой степени кандидата технических наук.

Официальный оппонент:

Профессор кафедры математических методов исследования операций федерального государственного бюджетного образовательного учреждения высшего образования «Воронежский государственный университет»,

доктор технических наук, доцент

И.Л. Каширина

13.08.2024

Ирина Леонидовна Каширина

Научная специальность докторской диссертации 05.13.18 - Математическое моделирование, численные методы и комплексы программ

Почтовый адрес: 394018, Россия, г. Воронеж, Университетская площадь, 1

Тел.: +7 (473) 220-82-66

e-mail: dean@amm.vsu.ru

Сайт: www.vsu.ru

