

## ОТЗЫВ

официального оппонента на диссертационную работу

Алимовой Ильсеяр Салимовны

«Нейросетевой механизм кросс-внимания в задачах извлечения информации из текстов на примере биомедицинских данных», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

**Актуальность темы.** Диссертационная работа Ильсеяр Салимовны Алимовой посвящена решению задач классификации сущностей и извлечения отношений между сущностями в биомедицинских текстах. Высокая доступность в настоящее время множества текстов по биомедицинской тематике, таких как отзывы пользователей о лекарствах, твиты о здоровье, аннотации научных статей по биомедицине, электронные медицинские карты, привела к необходимости применения методов компьютерной лингвистики для автоматического анализа таких текстов. Результаты анализа используются при решении различных проблем фармакологии и персонализированной медицины, например, для поиска упоминаний о побочных действиях лекарств и о применении лекарств с нарушением инструкции.

Несмотря на большое количество исследований в данной сфере остается нерешенным ряд проблем, в частности, существующие модели тестируются как правило только в рамках одной предметной области, что ставит под вопрос их применимость на практике. Также основной пласт исследований проводится для английского языка; для русского языка разработанность тематики явно недостаточна. Таким образом, тема диссертационной работы, безусловно, является актуальной.

**Научная новизна.** Наиболее важными результатами, составляющими научную новизну диссертации, являются следующие:

– разработана нейросетевая модель классификации сущностей в текстах биомедицинской тематики, включающая механизм кросс-внимания, отличающаяся от существующих использованием множества дополнительных информативных признаков;

– разработана нейросетевая модель классификации отношений между сущностями, включающая механизм кросс-внимания, отличающаяся от существующих разделением контекста и сущностей на отдельные подсети формирования контекстно-зависимых векторных представлений.

**Практическая значимость.** В процессе выполнения диссертационного исследования разработаны два программных комплекса:

– CAFEC (Cross-Attention Feature-based Entity Classifier) – для классификации сущностей;

– CARE (Cross-Attention Relation Extraction) – для извлечения отношений между сущностями.

Оба программных комплекса могут быть использованы для автоматического анализа текстов биомедицинской тематики из разных источников, в том числе, медицинских документов, научных статей и сообщений социальных медиа.

**Достоверность и обоснованность научных результатов.** В диссертационной работе корректно использованы нейросетевые модели и методы машинного обучения; проведены экспериментальные исследования разработанных моделей на различных текстовых корпусах; полученные результаты экспериментов сопоставлены с известными.

Основные положения диссертации обсуждались на международных и всероссийских конференциях. Содержание проведенного исследования достаточно полно изложено в 14 печатных изданиях, в том числе в двух журналах из Перечня ВАК, в 10 изданиях, индексируемых базами данных Web of Science и Scopus, а также в двух сборниках тезисов докладов.

**Структура и содержание диссертации.** Диссертация включает введение, четыре главы, заключение, список литературы и одно приложение. Общий объем работы составляет 146 страниц. Список литературы содержит 182 наименования.

Во введении обоснована актуальность темы исследования, обозначены объект и предмет исследования, указаны цель и задачи, сформулирована научная новизна и практическая значимость работы, приведены основные

положения, выносимые на защиту, представлены сведения о достоверности и апробации работы, описан личный вклад автора и указаны публикации по теме диссертации.

В первой главе приведен обзор существующих методов автоматического извлечения информации из текста, методов классификации текстов биомедицинской тематики, методов извлечения отношений между сущностями в текстах.

Вторая глава посвящена описанию предложенной модели классификации сущностей на основе нейронной сети LSTM с механизмом кросс-внимания и с использованием набора извлеченных из текста признаков. Приведена формальная постановка задачи классификации сущностей. Предложенная модель сравнивается с известными моделями как при обучении на текстах одного корпуса, так и в кросс-доменной конфигурации.

В третьей главе представлена разработанная модель извлечения отношений между сущностями, использующая нейронную сеть LSTM с механизмом кросс-внимания и отдельными подсетями формирования векторных представлений контекста и сущностей. Также приводятся результаты экспериментов и сравнение с существующими моделями в рамках одного корпуса и при кросс-доменной оценке.

В четвертой главе рассматривается архитектура разработанных программных комплексов CAFEC и CARE для классификации сущностей и извлечения отношений между сущностями. Приводятся детали программной реализации и временные оценки для обоих комплексов.

Заключение содержит краткое изложение основных результатов работы и перспективы дальнейших исследований.

#### **Замечания по диссертации.**

1. В обеих предложенных нейросетевых моделях в механизме внимания применяется мультипликативный подход, использование которого в диссертации никак не обосновывается.

2. Из таблицы 2 на стр. 51 непонятно, почему применяется именно такая схема проверки значимости признаков – добавление по одному типу в определенном порядке. Более обосновано было бы использование схемы «исключение по одному из полного множества».

3. В пунктах 2.4.2 и 3.4.2 не указано, подбирались ли гиперпараметры тестируемых моделей, и если да, то каким образом.

4. В диссертации не приведены асимптотические оценки временной сложности процедур обучения и предсказания для разработанных моделей.

5. В диссертации встречаются опечатки в формулах, например, в формуле на стр. 17 для сигмоидной функции неверно указан аргумент  $x$  вместо  $z$ . Также имеется небольшое количество орфографических и пунктуационных ошибок.

Указанные замечания не влияют на общую положительную оценку диссертации.

**Заключение.** Диссертационное исследование И.С. Алимовой является самостоятельно подготовленной, завершённой научно-квалификационной работой, выполненной на высоком научном уровне, обладает научной новизной и имеет практическую ценность. Автореферат и публикации соискателя в полной мере отражают содержание диссертационной работы.

Таким образом, считаю, что диссертация Ильсеяр Салимовны Алимовой соответствует всем требованиям Положения о присуждении ученых степеней, предъявляемым к кандидатским диссертациям, а её автор заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей.

Официальный оппонент  
профессор кафедры прикладной  
математики и информатики  
ФГБОУ ВО «Вятский  
государственный университет»,  
доктор технических наук, доцент

Котельников Евгений Вячеславович