

Отзыв официального оппонента

кандидата технических наук, доцента Ямашкина Станислава Анатольевича, доцента ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва» на диссертационную работу Зоткиной Алены Александровны «Методы и алгоритмы формирования психологического портрета пользователя социальной сети для эффективного подбора кадров», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.8. – «Информатика и информационные процессы» (отрасль науки – технические) в объединенный диссертационный совет 99.2.113.02 при ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В.Ф. Уткина», ФГБОУ ВО «Пензенский государственный технологический университет».

1 Актуальность темы

Согласно указу Президента Российской Федерации от 18 июня 2024 г. № 529 "Об утверждении приоритетных направлений научно-технологического развития и перечня важнейших наукоемких технологий" определен перечень важнейших наукоемких технологий, в рамках которого определена целесообразность развития сквозной технологии искусственного интеллекта в отраслях экономики, социальной сферы (включая сферу общественной безопасности) и в органах публичной власти. В данном контексте, решение важной задачи эффективного подбора кадров для различных отраслей народного хозяйства в настоящее время целесообразно проводить с вовлечением методов, алгоритмов и средств разведки по открытым источникам (OSINT). Автоматизированный интеллектуальный анализ больших данных при этом вовлекается в процессы поиска, выбора, сбора, анализа и систематизации информации из общедоступных источников о человеке.

Кандидатская диссертация Зоткиной Алены Александровны направлена на совершенствование методов формирования психологического портрета пользователя социальной сети, основанных на анализе информации, размещаемой ими в открытом доступе, для прогнозирования их профессионального поведения и разработки эффективных стратегий развития сотрудников и повышения уровня их удовлетворенности с целью эффективного подбора кадров.

2 Соответствие диссертации и автореферата паспорту научной специальности

Диссертационная работа и автореферат соответствуют паспорту специальности 2.3.8. – «Информатика и информационные процессы», охватывает следующие направления:

п. 1 – разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией пользователями и возможностей используемого программно-аппаратного обеспечения;

п. 7 – разработка методов обработки, группировки и аннотирования информации, в том числе, извлеченной из сети интернет, для систем поддержки принятия решений, интеллектуального поиска, анализа.

3 Структура и содержание диссертации

Диссертационная работа состоит из введения, 4 глав, заключения, библиографического списка из 133 наименований и 2 приложений. Текст работы изложен на 140 (132 без приложений) страницах, содержит 29 рисунков и 9 таблиц.

Во введении показана актуальность работы, сформулированы цель и основные задачи исследования, научная новизна и практическая ценность работы, приведены основные положения, выносимые на защиту. Представлены сведения об апробации и внедрении результатов.

Материалы, изложенные в диссертации, позволяют сформировать достаточно полное представление о проведенных исследованиях и полученных результатах.

В главе "1 Обзор современного состояния больших данных и методов интеллектуального анализа текстовых данных в социальных сетях" представлены результаты анализа международного опыта работы с большим данными в контексте решаемых автором задач, проанализированы современные системы, используемые для обработки и анализа больших данных. Проведен анализ современного состояния моделей и методов интеллектуального анализа текстовых данных в социальных сетях. Для составления портрета предложено использовать методологию

МВТІ, т.к. она включает профессиональную ориентацию и трудоустройство, чтобы помочь работодателям лучше понять личностные характеристики кандидатов и их потенциальное поведение в рабочей среде.

В главе "**2 Методы и алгоритмы формирования психологического портрета пользователя социальной сети**" рассмотрены ключевые проблемы анализа тональности текстов и психологического профилирования пользователей социальных сетей, охарактеризованы разработанные алгоритмы, методы и подходы к анализу данных из социальных сетей:

- метод оценки сходств признаков выражения, текстовых объектов, записей множества пользователей социальных сетей и реализующий ее алгоритм работы поиска аккаунтов пользователя, которые в отличие от существующих, учитывают разнообразные аспекты активности пользователей (публикации, участие в сообществах, комментарии, лайки к комментариям и публикациям);

- метод интеграции информации, публикуемой пользователем на разных платформах социальных сетей, позволяющий восстанавливать данные о пользователях, проявивших активность хотя бы на одной из этих платформ, который отличается тем, что учитываются полные данные об активности пользователя на протяжении длительного периода времени;

- методика кросс-доменного аспектно-ориентированного анализа тональности текста и алгоритм на ее основе, которая в отличие от существующих, фокусируется на выделении аспектов и анализе тональности отношения к ним в тексте;

- методика и реализующий ее алгоритм, позволяющие осуществлять определение психологических характеристик пользователя социальной сети, с использованием типологии МВТІ и нейронных сетей.

В главе "**3 Разработка и реализация программного обеспечения для анализа психологического портрета пользователей социальных сетей**" представлены особенности программной реализации алгоритма обработки информации, размещаемой пользователями в социальной сети, с целью формирования их психологического портрета. Алгоритмы разработаны с учетом основных принципов анализа данных, учитывают специфику социальных сетей и их влияние на поведение пользователей. Разработанные алгоритмы интегрированы в программное обеспечение, которое предоставляет возможность проводить анализ психологического портрета пользователя и оценивать вероятность отклоняющегося поведения в режиме реального времени, что позволяет учитывать данные за определенный временной интервал, обеспечивая точное и актуальное представление о состоянии пользователя.

В главе "**4 Проведение экспериментального тестирования алгоритмов**" изложены результаты экспериментов, проведенных с применением предложенных алгоритмов, реализованных в виде программного комплекса. Исследование включает анализ различных методов поиска аккаунтов одного и того же человека как в различных социальных сетях. В процессе экспериментов протестированы различные архитектуры нейронных сетей, использующие модель ІbDA-LSTM-CRF. Автором сформулированы методические рекомендации по использованию нейросетевых моделей и настройке гиперпараметров для достижения максимальной эффективности решения поставленной задачи.

4 Научная новизна исследований и полученных результатов

В качестве новых научных результатов диссертантом предложены:

1. Метод оценки сходств признаков выражения, текстовых объектов, записей множества пользователей социальных сетей и реализующий ее алгоритм работы поиска аккаунтов пользователя, которые в отличие от существующих, учитывают разнообразные аспекты активности пользователей (публикации, участие в сообществах, комментарии, лайки к комментариям и публикациям).

2. Метод интеграции информации, публикуемой пользователем на разных платформах социальных сетей, позволяющий восстанавливать данные о пользователях, проявивших активность хотя бы на одной из этих платформ, который отличается тем, что учитываются полные данные об активности пользователя на протяжении длительного периода времени, что позволяет составить более полный и подробный психологический портрет пользователя.

3. Методика кросс-доменного аспектно-ориентированного анализа тональности текста и алгоритм на ее основе, которая в отличие от существующих, фокусируется на выделении аспектов и анализе тональности отношения к ним в тексте, что позволяет получить более детальное представление о содержании и оценке текста, в отличие от других рассматриваемых методик.

4. Нейросетевая методика и алгоритм, ее реализующий, для определения психологических характеристик пользователя социальной сети, с использованием типологии MBTI, которая в отличие от других подходов, фокусируется на изолированных личностных чертах, что позволяет предоставить комплексное представление личности.

Полученные результаты обеспечивают решение задачи развития методов формирования психологического портрета пользователя социальной сети, основанных на анализе данных из открытых источников, а представленные новые научно обоснованные технические решения и разработки имеют существенное значение для решения задачи эффективного подбора кадров.

5 Достоверность результатов исследования

Обоснованность и достоверность научных положений и выводов, представленных в диссертационном исследовании, подтверждается результатами проведенных экспериментов, опытом внедрения результатов исследования в практическую и научно-исследовательскую деятельность ряда организаций, а также апробацией и обсуждением результатов диссертации на международных и всероссийских научных конференциях.

Разработанные методы и алгоритмы согласно приложенным документам внедрены в учебный процесс на кафедре «Программирование» ФГБОУ ВО ПензГТУ и используются при подготовке студентов по направлениям бакалавриата 09.03.01 «Информатика и вычислительная техника» и 09.03.04 «Программная инженерия» в рамках дисциплин «Методы машинного обучения и искусственного интеллекта», «Сбор и управление большими данными», «Технологии больших данных». Результаты НИОКР внедрены в деятельность в АО «НПП «Рубин» (г. Пенза) в процессе выполнения научно-исследовательской работы по теме «Метрика-Р», в ООО «ТД «ПЗЭМ» (г. Пенза) в рамках выполнения научно-исследовательского проекта по теме «Кадры для цифровой экономики», в Ассоциацию разработчиков программного обеспечения Пензенской области «Секон» при разработке решений для систем подбора кадров ряда организаций входящих в Ассоциацию (CodeInside, Tortuga), в АО «ИПО ЭВТ им. В.А. Ревунова» для принятия решений по подбору сотрудников.

По результатам диссертационного исследования опубликовано 38 научных работ, в том числе 6 статей в журналах, рекомендованных ВАК РФ, 2 статьи, индексируемые в международной базе данных Scopus, получено 2 свидетельства о государственной регистрации программ для ЭВМ. Количество публикаций в изданиях ВАК является достаточным.

Автореферат выдержан по форме, объему и отражает основные положения диссертационного исследования. Язык и стиль написания диссертации характеризуется ясностью и четкостью изложения материала.

6 Замечания по автореферату и диссертационной работе в целом

1) Некоторые ключевые аспекты диссертационного исследования следовало бы описать более детально. К ним относятся вопросы определения топологии нейронных сетей, способы подбора гиперпараметров и тонкой настройки моделей, алгоритмы и стратегии формирования обучающих выборок. Работу можно существенно усилить и более детальным сравнением предложенного авторского подхода с существующими способами решения схожих задач, в том числе на основе аппарата матриц ошибок и метрик, рассчитываемых на их основе. Наконец, напрашивается более детальное описание разработанного программного обеспечения. Для решения этой задачи, к примеру, хорошо подходят UML-диаграммы.

2) В перечне публикаций в научных изданиях из перечня ВАК РФ не конкретизировано, сколько научных статей опубликовано в изданиях по специальности 2.3.8. – Информатика и информационные процессы (технические науки), а сколько – в иных изданиях. Так, научных журнал «Современные наукоемкие технологии» не входит в перечень рецензируемых научных изданий по специальности, в рамках которой защищается диссертация. Это – хорошо, так как говорит о том, что результаты исследования имеют прикладное междисциплинарное значение, при этом число публикаций по специальности – более, чем достаточное. Тем не менее, в автореферате и диссертации этот факт целесообразно обозначить.

3) К структурированию некоторой информации в диссертации имеются рекомендации. На странице 19 в реестре примеров на один уровень иерархии вынесены отрасли народного хозяйства («здравоохранение»), название организации («NASA»), вид системного программного обеспечения

(«утилиты»), тип продукции («спортивные товары и фитнес»), раздел науки («кибербезопасность»). На странице 20-21 имеется схожее изложение материала: на один уровень иерархии списка выносятся «неструктурированные данные» и «данные на естественном языке» (как «... сложный подтип неструктурированных данных»). При структурировании информации необходимо придерживаться системности.

4) В диссертации встречаются определенных стилистические и орфографические ошибки, имеется ряд замечаний к иллюстративному материалу. Так, текст на некоторых рисунках (1.2, 2.1) слабо читаем. Целесообразно, чтобы размер шрифта на рисунках был сопоставим с размером шрифта основного текста. Блок-схемы (рисунки 1.5, 1.6, 3.1) построены с незначительными отклонениями от стандартов. В таблице 2.1 («Таблица символов выражений эмоций») эмодзи с кодом 😱 представлено неверно. Есть замечания к некоторым формулировкам. Так, фраза «социальная сеть представляет собой кортеж» является неточной, корректнее было бы написать: «описание социальной сети может быть представлено в виде кортежа».

Приведенные замечания не меняют общего положительного мнения о диссертации и не снижают научную и практическую значимость результатов исследования. Диссертационная работа оформлена в соответствии с требованиями ВАК РФ, предъявляемыми к кандидатским диссертациям.

7 Заключение по работе

Диссертация Зоткиной Алены Александровны является законченной научно-квалификационной работой, которая содержит новые результаты, позволяющие достичь цели совершенствования методов и алгоритмов формирования психологического портрета пользователя социальной сети, основанных на анализе информации, размещенной в открытом доступе, для эффективного подбора кадров.

Диссертация «Методы и алгоритмы формирования психологического портрета пользователя социальной сети для эффективного подбора кадров» представляет собой научно-квалификационную работу, в которой изложены новые научно обоснованные технические решения и разработки, имеющие существенное значение для развития страны, соответствует требованиям п. 9 «Положения о присуждении ученых степеней», утвержденного Постановлением Правительства Российской Федерации от 24.09.2013 г. № 842 (в редакции Постановлений Правительства Российской Федерации от 21.04.2016 г. № 335, от 02.08.2016 г. № 748, от 29.05.2017 г. № 650, от 28.08.2017 г. № 1024, от 01.10.2018 г. № 1168, от 20.03.2021 г. №426, от 11.09.2021 г. №1539, от 26.09.2022 г. №1690, от 26.01.2023 г. №101), а соискатель Зоткина Алена Александровна заслуживает присуждение ей ученой степени кандидата технических наук по специальности 2.3.8. – Информатика и информационные процессы.

Официальный оппонент

Кандидат технических наук, доцент, доцент кафедры «Автоматизированные системы обработки информации и управления» ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва». Кандидатская диссертация защищена в 2016 году по специальности 05.13.01 – «Системный анализ, управление и обработка информации».

С.А. Ямашкин

02.12.2024г

Подпись Ямашкина С.А. заверяю
Ученый секретарь Ученого совета ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва», к.филол.н., доцент

Е.А. Казеева

Организация: ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет им. Н.П. Огарёва», кафедра автоматизированных систем обработки информации и управления».

Почтовый адрес: 430005, г. Саранск, Республика Мордовия, ул. Большевистская, д. 68
Телефон: +7 (8342) 47-86-91
e-mail: yamashkinsa@mail.ru

С отзывом ознакомлена.
5.12.2024г. [Подпись]