

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»
(СПбГУТ)

Юридический адрес: набережная реки Мойки,
д. 61, литера А, Санкт-Петербург, 191186

Почтовый адрес: пр. Большевиков, д. 22, корп. 1,

Санкт-Петербург, 193232

Тел. (812) 3263156, Факс: (812) 3263159

<http://sut.ru>

E-mail: rector@sut.ru

ОКПО 01179934 ОГРН 1027809197635

ИНН 7808004760 КПП 784001001

ОКТМО 40909000

11.07.2022 № 1656/54
на № _____ от _____

УТВЕРЖДАЮ

И.о. проректора по научной работе,
д-р техн. наук, с.н.с.



Шестаков
Александр Викторович

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

на диссертацию Андреева Ильи Алексеевича на тему «Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность)

Актуальность научной работы

В настоящее время мир вошел в «информационную эру». Каждый человек, живущий в эту эру, оставляет какой-то цифровой след в сети Интернет. Большое количество информации при этом хранится в социальных сетях, примерами такой информации являются сообщения пользователей, комментарии к новостям, реакции на события (в виде так называемых «лайков» и «дизлайков»). Одной из особенностей этой информации является ее слабая структурированность, а часто – отсутствие какой-либо структуры, что усложняет процесс ее обработки. Количество информации, генерируемой пользователями социальных сетей, год от года только растет, поэтому существует необходимость поиска новых методов по извлечению, обработке и структурированию информации из социальных сетей.

Одной из задач, решаемых путем анализа социальных сетей, является построение социального портрета пользователя, т.е. набора характеристик

человека, которые он представляет на открытых ресурсах. Обзоры работ для русского языка показывают, что в настоящий момент существует недостаточное количество методов анализа текстов социальных сетей с учетом их особенностей, а эффективность существующих не достаточна для повсеместного использования.

Поэтому тематика данной диссертации, посвященная исследованию методов и алгоритмов формирования социального портрета пользователя при помощи обработки текстовой информации социальных сетей, является актуальной задачей для исследования.

Новизна исследования, полученных результатов, выводов и рекомендаций, сформулированных в диссертации

Научная новизна результатов исследования заключается в следующем:

- Разработан алгоритм формирования обучающей выборки для обучения моделей классификации в задачах сентимент-анализа текстовых данных, отличающийся совместным использованием словарей авторских символов выражения эмоций и ключевых фраз, позволяет эффективно решать задачу обучения нейронной сети в процессе сентимент-анализа русскоязычных текстов социальных сетей.
- Предложен подход к сопоставлению профилей пользователей в разных социальных сетях, отличающийся гибридизацией подходов анализа графической информации, структурированных данных анкет, текстовых данных, а также социальных графов профилей, реализован в программном комплексе и автоматизирует процесс поиска профилей пользователя в задаче построения социального портрета.
- Разработан метод определения психологических характеристик пользователя социальных сетей, отличающийся гибридизацией алгоритмов обработки естественного языка текстовых данных, машинного обучения и метода «Большой пятерки», позволяет классифицировать пользователя по пяти основным факторам данной модели.
- Предложен алгоритм анализа эмоциональной окраски русскоязычных текстовых данных социальных сетей, отличающийся интеграцией семантических подходов и методов машинного обучения, позволяет повысить точность классификации текстов социальных сетей по классам тональности.

Обоснованность и достоверность научных положений и выводов

Достоверность разработанных научных положений и выводов подтверждена экспериментами и успешными внедрениями, а также широкой апробацией на следующих конференциях, семинарах, симпозиумах, конгрессах и других мероприятиях, приравненных к ним: XXIX Международная конференция «Computational Science and Its Applications»-ICCSA-2019 (г. Санкт-Петербург, 2019), Международная научно-техническая конференция «Автоматизация» – RusAutoConf-2020 (г. Сочи, 2020), XII Международная конференция Developments in eSystems Engineering – DESE-2019 (г. Казань, 2019), V Международная научно-техническая конференция «Открытые семантические технологии проектирования интеллектуальных систем» (г. Минск, 2015); Международная научно-практическая конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (г. Коломна, 2015, 2020), I Международная научная конференция «Интеллектуальные информационные технологии в технике и на производстве» (г. Сочи, 2016), Международная конференция и молодежная школа «Информационные технологии и нанотехнологии» (г. Самара, 2017, 2019–2021), VIII Международная конференция «Системный анализ и информационные технологии» САИТ – 2019 (г. Иркутск, 2019), I Международная Поспеловская летняя школа-семинар для студентов, магистрантов и аспирантов «Методы и технологии гибридного и синергетического искусственного интеллекта» (г. Светлогорск, 2014), V Всероссийская Поспеловская конференция с международным участием «Гибридные и синергетические интеллектуальные системы» (г. Светлогорск, 2020), XVII национальная конференция по искусственному интеллекту с международным участием «КИИ-2019» (г. Ульяновск, 2019), IV Всероссийская научно-практическая мультikonференция с международным участием «Прикладные информационные системы» – ПИС-2017 (г. Ульяновск, 2017), 6-я Всероссийская научно-техническая конференция аспирантов, студентов и молодых ученых ИВТ-2014 (г. Ульяновск, 2014).

Значимость для науки и практики результатов, полученных автором диссертации

Теоретическая значимость диссертационной работы заключается в разработке новых алгоритмов, подходов и методов обработки текстовой информации социальных сетей, применяемых для составления социального портрета пользователя. Практическая значимость диссертационной работы заключается в разработанном программном комплексе, который позволяет автоматизировать построение социального портрета пользователя путем объединения и анализа профилей одного человека в различных социальных сетях.

Рекомендации по использованию результатов и выводов диссертации

В ходе диссертационного исследования были получены результаты, которые можно рекомендовать к использованию на практике для решения задачи построения социального портрета пользователя социальных сетей.

Алгоритмы и методы могут использоваться совокупности или по отдельности. Алгоритм сентимент-анализа может применять для оценки эмоциональной окраски любого короткого текста, а не только для оценки эмоциональной окраски текстов социальной сети. Поиск всех аккаунтов одного пользователя может применяться в различных задачах, несвязанных с построением социального портрета пользователя.

Основные теоретические и практические результаты диссертационной работы были использованы в рамках проектов: «Интеллектуальная платформа формирования социального портрета соискателя на основании семантико-когнитивного анализа профилей в социальных сетях», поддержанного Фонда содействия инновациям по программе «Старт-Цифровые технологии» для компании ООО «Центр программной инженерии и аналитики «ФаззиЛаб»; УОСОО «Федерация бадминтона» в рамках проекта «Парабадминтон: все силы – для победы», поддержанного Фондом Президентских грантов для отбора волонтеров, обеспечивающих сопровождение лиц с ПОДА (проект № 18-2-009220); предложенные в рамках диссертационной работы алгоритмы и методы интеллектуального анализа неструктурированных данных социальных сетей были использованы при разработке системы интеллектуального поиска и анализа в Интернет-СМИ и социальных сетях в рамках совместного проекта с ФНПЦ АО «НПО «Марс».

Общая оценка диссертационной работы

Диссертация состоит из введения, четырех глав, заключения, списка литературы и приложений. Объем диссертации – 166 страниц. Представлено 50 рисунков и 11 таблиц. Список использованной литературы – 170 наименований.

Диссертация и автореферат написаны грамотным техническим языком, все положения изложены аргументировано.

Основные результаты диссертационной работы в полной мере опубликованы в 32 печатных работах, в том числе 4 статьи в изданиях, рекомендованных ВАК. Также автор имеет 3 свидетельства о регистрации программы для ЭВМ.

Основные научные результаты теоретических и прикладных исследований, выводы, изложенные в диссертации, получены автором самостоятельно. В работах, опубликованных в соавторстве, вклад соискателя был определяющим.

Однако, в работе имеются некоторые **замечания**, а именно:

1) В работе приведены эксперименты по оценке эффективности алгоритмов классификации текстов по классам тональности с использованием только двух языковых моделей – Word2Vec и BERT, при этом их применение не обосновано. В настоящее время существует большой набор модификаций языковых моделей (ELMo, GloVe, TinyBERT и пр.), которые могли показать лучшую эффективность на русскоязычной текстовой выборке.

2) В качестве научной новизны исследования в работе выступает методика объединения профилей пользователей из различных социальных сетей. В экспериментах показано, что текущего набора алгоритмов сравнения данных профилей, примененных последовательно, достаточно для объединения профи 204 профилей трех социальных сетей. Однако данная методика может показать иной уровень эффективности при обработке данных более крупной выборки.

3) Критерий схожести лиц на фотографиях профилей социальных сетей, составляющий методику объединения профилей пользователей, базируется на применении метода гистограмм направленных градиентов. В качестве порогового значения вычисляемого евклидова расстояния используется

величина 0,85. В работе нет обоснования, каким образом была получена данная величина.

4) Представленная в экспериментах по оценке алгоритма психолингвистического анализа текста профилей социальных сетей выборка из 204 профилей может не являться достаточной при указанных в диссертации условиях отбора профилей для анализа – анализ анкет производился только среди добровольцев, которые заведомо обладают специфическими психологическими характеристиками.

5) В диссертации встречаются ошибки в форматировании и расположении рисунков – так, например, на странице 117 оставлено много пустого места, в который стоило бы поместить рисунок 4.1, расположенный на странице 118.

6) Список используемой литературы составлен с нарушением ГОСТ – по алфавиту, однако сначала идут источники, написанные с использованием латинского алфавита, а за ними – источники, написанные с использованием кириллицы.

Отмеченные недостатки не носят принципиальный характер и не влияют на общую положительную оценку диссертационной работы.

Заключение

Диссертация «Исследование методов и алгоритмов обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя» является законченной научно-квалификационной работой, в которой содержится решение научной задачи формирования социального портрета пользователя социальной сети, имеющей значение для развития технической отрасли знаний.

Работа соответствует критериям, предъявляемым в отношении кандидатских диссертаций, которые установлены пп. 9–14 Положения о присуждении ученых степеней (утв. Постановлением Правительства РФ от 24.09.2013 № 842), а ее автор, Андреев Илья Алексеевич, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

Доклад автора по диссертации «Исследование методов и алгоритмов


обработки текстовой информации социальных сетей в задачах формирования социального портрета пользователя» заслушан и настоящий отзыв принят на заседании кафедры инфокоммуникационных систем 21 июня 2022 г., протокол № 10.

Заведующий кафедрой инфокоммуникационных систем,
к.т.н., доцент



Зарубин Антон Александрович

Профессор кафедры инфокоммуникационных систем,
д.т.н., профессор



Гольдштейн Борис Соломонович

Доцент кафедры инфокоммуникационных систем,
к.т.н., доцент



Елагин Василий Сергеевич