На правах рукописи

Русначенко Николай Леонидович

Модели, методы и программные средства извлечения оценочных отношений на основе фреймовой базы знаний

Специальность 05.13.11 —
«Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей (технические науки)»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Москва — 2022

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)».

Научный руководитель: **Лукашевич Наталья Валентиновна**
доктор технических наук

Официальные оппоненты: **Котельников Евгений Вячеславович**
доктор технических наук, доцент
Федеральное государственное бюджетное образовательное учреждение высшего образования «Вятский государственный университет»
профессор

Бурцев Михаил Сергеевич
кандидат физико-математических наук
Федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (национальный исследовательский университет)»
заведующий лаборатории Нейронных систем и глубокого обучения

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования “Казанский (Приволжский) федеральный университет”

Защита состоится «28» апреля 2022 г. в 15:00 на заседании диссертационного совета Д 999.216.02 при МАИ и МГТУ им. Н.Э. Баумана по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1, зал Ученого совета ГУК.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана на сайте http://bmstu.ru.

Отзывы на автореферат в двух экземплярах, заверенные печатью учреждения, просьба направлять по адресу: 105005, г. Москва, 2-я Бауманская ул., д. 5, стр. 1, ученому секретарю диссертационного совета Д 999.216.02.

Автореферат разослан «___» _____ 2022 года.

Ученый секретарь
диссертационного совета
Д 999.216.02,
доктор технических наук, доцент А.Н. Алфимцев

Общая характеристика работы

Актуальность темы. Анализ тональности является одной из наиболее востребованных задач в автоматической обработке текстов, которая состоит в определении отношения (позитивного или негативного) некоторого лица относительно содержания текста или каких-то его аспектов. На практике анализ тональности подразделяется на множество различных подзадач, таких как определение общей тональности текста или предложения, тональность автора по отношению к упомянутым сущностям и другие.

Одной из мало исследованных подзадач анализа тональности является извлечение тональности отношений между сущностями, упомянутыми в тексте (оценочные отношения). В новостных и аналитических текстах тональность оценочных отношений сложным образом коррелирует с другими тональностями, например, с тональностью отношения автора текста к обсуждаемой тематике. Таким образом, извлечение оценочных отношений является как подвидом задачи анализа тональности, так и задачи извлечения отношений. Актуальными на настоящий момент методами в решении таких задач являются модели на основе различных методов машинного обучения, включая классические методы машинного обучения, нейронные сети сверточного и рекуррентного типов, а также нейронные сети с вниманием, в том числе языковые модели типа BERT. Основными ограничениями в организации процесса обучения таких методов являются: общий недостаток разметки и сложность ее ручного выполнения для составления обучающего корпуса.

Среди отечественных и зарубежных ученых, занимающихся исследованием задачи анализа тональности и применением методов машинного обучения в такой области, наиболее известными являются: Е. Котельников, О. Кольцова, P. Turney, D. Zeng, Y. Choi, J. Devlin и др.

Актуальность исследования заключается в том, что на настоящий момент нет универсальных методик автоматической разметки оценочных отношений, которые бы позволили увеличить объем обучающих данных. Предложенный подход по автоматической разметке данных и проведения *опосредованного обучения* (от англ. Distant Supervision) на их основе, позволяет повысить эффективность моделей нейронных сетей.

Объектом исследования являются комбинированные подходы, включающие базу знаний и нейросетевую модель для извлечения оценочных отношений из текстов.

Предметом исследования является структура и состав базы знаний для анализа тональности текстов на русском языке.

Целью диссертационного исследования является разработка методов извлечения оценочных отношений между именованными сущностями из текстов средств массовой информации с использованием русскоязычной базы знаний.

Для достижения поставленной цели были решены следующие **задачи**:

1. Разработать базу знаний для описания структуры тональностей слов-предикатов;
2. Реализовать методы машинного обучения для извлечения оценочных отношений между именованными сущностями из текстов новостных и аналитических статей;
3. Реализовать модель и методы порождения автоматически размеченных оценочных отношений на основе лексико-семантических ресурсов;
4. Реализовать методы извлечения оценочных отношений на основе подхода опосредованного обучения (от англ. Distant Supervision) и комбинированной обучающей выборки, включающей как ручную, так и автоматическую разметку;
5. Создать программные средства для обработки новостных и аналитических текстов, которые на основе текста статьи порождают список оценочных отношений между упомянутыми именованными сущностями.

Научная новизна

– Предложена структура фреймовой базы знаний RuSentiFrames для описания тональностей, ассоциирующихся со словами и выражениями русского языка, включая тональность отношений между участниками ситуации, отношение автора к участникам ситуации, позитивные и негативные эффекты, связанные с ситуацией. Такая база знаний описывает значительно более сложную структуру тональностей, ассоциированных с словом, в отличие от обычных списков оценочных слов с оценками тональностей;
– Впервые для русского языка поставлена задача и выполнено исследование методов извлечения тональности отношений между именованными сущностями, упомянутыми в текстах СМИ;
– Для обучения моделей извлечения оценочных отношений предложен новый метод автоматического порождения обучающей коллекции на основе оценочных фреймов нового лексикона RuSentiFrames и использования структуры новостных текстов. Применение опосредованного обучения с использованием RuAttitudes-2.0 повысило качество языковых моделей BERT на 10-13% по метрике F1, и на 25% при сравнении с наилучшими результатами остальных моделей на основе оценочных фреймов нового лексикона RuSentiFrames и структуры новостных текстов;

Практическая значимость. Разработаны и исследованы модели извлечения оценочных отношений, а также методы автоматической обработки внешних новостных источников информации. Впервые создана и

опубликована большая база контекстов RuAttitudes-2.0 (252 тыс. примеров) с автоматической разметкой оценочных отношений, что может быть полезным для задач таргетированного анализа тональности текстов СМИ. Создан программный комплекс ARElkit для выполнения автоматической разметки коллекции новостей, а также обучения моделей на основе нейросетевых механизмов для извлечения отношений между сущностями в текстах СМИ с возможностью интерактивного человека-машинного управления.

Методология и методы исследования. В работе применяются методы обработки и анализа текстовой информации, методы классификации размеченной информации, методы объектно-ориентированного программирования для построения инструмента и проведения работы над поставленной задачей.

Основные положения, выносимые на защиту:

1. Предложена структура фреймовой базы знаний RuSentiFrames для описания тональностей, ассоциирующихся со словами и выражениями русского языка, включая тональность отношений между участниками ситуации, отношение автора к участникам ситуации, позитивные и негативные эффекты, связанные с ситуацией;
2. Предложен и реализован новый метод автоматического порождения обучающей коллекции для классификации оценочных отношений по двум и трем классам на основе словаря оценочных фреймов RuSentiFrames;
3. Программный комплекс ARElkit для создания автоматически размеченной обучающей коллекции для извлечения оценочных отношений, с программным интерфейсом для задания настроек пользователем, а также обучения методов на основе нейронных сетей;

Соответствие научной специальности. Содержание работы соответствует паспорту научной специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» (технические науки): п.4 «Системы управления базами данных и знаний», п.7 «Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения».

Апробация работы. Основные результаты работы докладывались на:

1. Международная Конференция «Диалог» (Россия, Москва, 2018);
2. 20-ая Международная Конференция Data Analytics and Management in Data Intensive Domains (Россия, Москва, 2018);
3. 21-ая Международная Конференция Text-Speech-Dialog (Чехия, Брно, 2020);

4. 12-ая Международная Конференция Recent Advances in Natural Language Processing (Болгария, Варна, 2020);
5. 25-ая Международная Конференция Natural Language & Information Systems (Германия, Саарбрюккен, 2020);
6. 10-ая Международная Конференция Web Intelligence, Mining and Semantics (Франция, Биаритц, 2020).

Личный вклад. Автором проведено исследование задачи извлечения оценочных отношений с выполнением основного объема теоретических и экспериментальных исследований, изложенных в тексте диссертационной работы. Разработана программная платформа для исследования и проведения экспериментов в предметной области на основе созданных методов. Исследование задачи извлечения оценочных отношений с применением разработанных методов рассмотрено в работах [1—8; 12].

Публикации. Основные результаты по теме диссертации изложены в 13 печатных изданиях, 2 из которых изданы в журналах, рекомендованных ВАК РФ, 8 — в изданиях, индексируемых в системах Web of Science и Scopus, 1 — в тезисах докладов. Зарегистрирован 1 патент.

Объем и структура работы. Диссертационная работа состоит из введения, четырех глав, заключения и трех приложений. Полный объём диссертации составляет 167 страниц, включая 23 рисунка и 4 таблицы. Список литературы содержит 109 наименований.

Содержание работы

Во **введении** обосновывается актуальность диссертационной работы, сформулированы цель и задачи представляемой работы, сформулирована научная новизна исследований, показана практическая значимость работы.

Первая глава посвящена обзору различных задач и методов анализа тональности текстов. Рассматриваются методы глубокого обучения в смежных задачах. В исследованиях задача анализа тональности может ставиться как задача классификации, в качестве *выходных данных* которой предполагается набор *классов тональности*. При этом задача извлечения оценочных отношений исследовалась недостаточно: есть только небольшое число работ для английского языка, для русского языка исследований не было. Поэтому важным является исследование методов автоматического сбора обучающей коллекции. Таким образом подтверждается актуальность разработки методов и проведение исследований для русского языка.

Вторая глава посвящена исследованию методов машинного обучения для извлечения оценочных отношений из аналитических текстов.

Формальная постановка задачи. Источником анализа является текстовая коллекция, состоящая из n документов $\{D_i\}_{i=1}^n$, где D_i – грамматически организованная последовательность слов с передачей

мнения автора по отношению ко множеству упомянутых именованных сущностей. Каждый документ представлен последовательностью символов: $D_i = \{c_1, c_2, \ldots, c_{|D_i|}\}$. Для каждого документа D_i коллекции предоставляется список упомянутых именованных сущностей $E_i = [e_1, \ldots, e_{|E_i|}]$. *Именованная сущность* (e_i) – слово или словосочетание, указывающие на объект реальности; представлена кортежем $\langle v_i, t_i \rangle$, где $v = [c_b, \ldots, c_e]$ – слово/словосочетание; t – категория: *личности* (PER), *организации* (ORG), *места* (LOC); *геополитические места* (GEO). Сущность может иметь множество вариантов наименований (например: Россия$_e$ – РФ$_e$), поэтому дополнительно вводится *список синонимов* S, состоящий из *групп* $[g_1, \ldots, g_{|S|}]$, где произвольная группа g_k представлена значениями сущностей: $\{v_1, \ldots, v_{|g_k|}\}$. Каждое из значений групп списка является уникальным в рамках всего списка S; пересечение любых двух различных групп является пустым множеством. Под *извлечением оценочных отношений* из D_i подразумевается составление списка кортежей $\mathbf{a} = \{\langle v_{1,j}, v_{2,j}, l_j \rangle\}_{j=1}^{|\mathbf{a}|}$, где $v_{1,j}$ – субъект, $v_{2,j}$ – объект, l_j – оценка, позитивная (POS) либо негативная (NEG). Например, в следующем тексте (именованные сущности подчеркнуты): *«При этом Москва$_e$ неоднократно подчеркивала, что ее активность на Балтике$_e$ является ответом именно на действия НАТО$_e$ и эскалацию враждебного подхода к России$_e$ вблизи ее восточных границ ...»* результатом является список отношений: [(НАТО, Россия, NEG), (Россия, НАТО, NEG)].

Контексты с отношениями. Основное предположение наличия оценочного отношения между парой сущностей $\langle v_1, v_2 \rangle$ – относительно короткое расстояние между этими сущностями в тексте. Под *контекстом* понимается текстовый фрагмент, включающий две и более именованных сущностей. *Контекст соответствует отношению*, когда v_1 и v_2 (их синонимы в том числе) присутствуют в контексте. Таким образом, для каждого отношения можно выделить множество соответствующих контекстов. Под *размеченным контекстом* понимается контекст с выделенной парой «субъект-объект». Размеченный контекст рассматривается как *оценочный*, если соответствующая пара присутствует в разметке коллекции. В противном случае размеченный контекст рассматривается как *нейтральный* (принадлежит дополнительному классу NEU). Таким образом, процесс извлечения оценочных отношений на уровне контекстов может быть сведен к классификационной задаче. Такой процесс подразумевает отбор таких контекстов, которые не являются нейтральными.

Классические методы. Классификация контекстов может быть выполнена классическими методами машинного обучения: KNN, SVM, Naïve Bayes, Random Forest, Gradient Boosting. Признаки, используемые для классификации разделены на группы [5], характеризующие: (1) именованные сущности; (2) контекст с отношением.

Модель извлечения оценочных отношений. Для решения задачи разработана модель, состоящая из *классификатора размеченных контекстов* с блоком агрегации результатов нескольких контекстов в единую оценку. В качестве классификаторов рассматривается и исследуется применение сверточных и рекуррентных нейронных сетей, языковых моделей семейства BERT. Общая архитектура классификатора включает: (1) кодировщик размеченного контекста, (2) классификационный слой.

Входной информацией для кодировщика на основе языковых моделей являются *размеченные контексты*. В случае нейронных сетей, размеченные контексты предварительно конвертируются в *вектора признаков* для *термов* (атомарные элементы, выделенные из размеченного контекста, следующих типов: ENTITIES (вхождения именованных сущностей), TOKENS (знаки препинания, URL-ссылки, числа), FRAMES (вхождения в сторонний лексикон), WORDS (прочие подпоследовательности контекста, разделенные пробелами)) контекста. Признаки термов контекста: (1) *вектор основного представления терма* из модели $\texttt{news}_{2015}$ (размер: 1000); (2) *вектор расстояния* [Zeng и др., *Distant supervision for relation extraction via piecewise convolutional neural networks*, 2015] – расстояние в термах от рассматриваемого терма до участников отношения $\langle v_1, v_2 \rangle$ (размер: $5 \cdot 2 = 10$); (3) *вектор частей речи* терма (размер: 5) (вычисляется для соответствующего слова токена, с помощью пакета `Yandex Mystem`; для токенов группы TOKENS вводится дополнительный тип `UNKNOWN` – значение части речи не определено). Общая длина вектора признаков (m) равна 1015.

Кодировщик размеченного контекста выполняет преобразование входной последовательности (размеченного контекста) в *векторное представление* $s \in \mathbb{R}^h$ длины h. В случае нейронных сетей исследуются кодировщики на основе следующих архитектур:

- Свёрточных типов [4] (CNN, PCNN [3]), в которых вектор s (далее s_{cnn}) представляет собой максимальную субдискретизацию (от англ. *maxpooling*) сверточных преобразований термов входной последовательности;
- Рекуррентных типов [7] (LSTM, BiLSTM), где s – последний элемент выходной последовательности LSTM, и конкатенация пары последних векторов двух последовательностей в случае BiLSTM.

Также для таких архитектур исследуется внедрение *механизма внимания*. Механизм внимания представляет собой отдельную нейронную сеть, перед которой стоит задача определения значимости каждого элемента входной последовательности относительно каких-либо других элементов. Под *взвешиванием* элемента последовательности понимается составление его количественной оценки $u_i \in \mathbb{R}$. Преобразование количественной оценки u_i в вероятностную ($\alpha_i \in [0, 1]$) осуществляется посредством операции *softmax*. Для входной последовательности векторов термов $X = \{x_1, \ldots, x_n\}$ (n – размер контекста), сочетание механизма

внимания ($\mathbf{a} \in \mathbb{R}^n$) с промежуточным результатом кодировщика (входные представления термов (CNN, PCNN), скрытые состояния термов (LSTM, BiLSTM)) контекста $s' \in \mathbb{R}^{n \times h}$ можно представить в виде: $s = \mathbf{a} \cdot s'$, где $\mathbf{a} = f_{\theta'}(X)$. Рассмотрим кодировщики контекстов с механизмами внимания относительно: (1) *аспектов* F, где $F \subset X$, и (2) всего X.

Механизм внимания на основе многослойного перцептрона [Huang и др., *Attention-based convolutional neural network for semantic relation extraction*, 2016]. Выберем произвольный аспект $f \in F$. Рассмотрим векторное представление i-го терма (h_i) как конкатенацию $x_i \in \mathbb{R}^m$ с $f \in \mathbb{R}^m$, т.е. $h_i = [x_i, f]$. Количественная оценка релевантности (u_i) для h_i вычисляется по формуле: $u_i = W_a \left[\tanh(W_{we} \cdot h_i + b_{we})\right] + b_a$, где $W_{we} \in \mathbb{R}^{\mathbf{h} \times 2 \cdot m}$ и $W_a \in \mathbb{R}^{1 \times \mathbf{h}}$ – матрицы весов и внимания модели соответственно; $b_{we} \in \mathbb{R}^{\mathbf{h}}, b_a \in \mathbb{R}$ – векторы смещения. Параметры $\langle W_{we}, W_a, b_{we}, b_a \rangle$ являются скрытыми состояниями механизма внимания $\mathbf{a}$ и изменяются в процессе обучения модели. Далее, оценка u_i преобразуется в вероятностную α_i с помощью операции *softmax*. Векторное представление контекста $\hat{s}$ к некоторому аспекту f вычисляется по формуле: $\hat{s} = \sum_{i=1}^{n} x_i \cdot \alpha_i$. Таким образом, для каждого аспекта $f_j \in F$ ($j \in \overline{1..k}$) составим множество векторов $\{\hat{s}_j\}_{j=1}^{k}$. Результирующее векторное представление контекста s есть конкатенация s_{cnn} и $s_{asp} = \sum_{j=1}^{k} \hat{s}_j / k$.

Механизм внимания модели IAN [Ma Dehong и др., *Interactive attention networks for aspect-level sentiment classification*, 2017]. Вход кодировщика представляет собой две раздельные последовательности X и F. Результатом применения модели LSTM к таким последовательностям являются $H_c = [h_1^c, \ldots, h_n^c]$, $H_f = [h_1^f, \ldots, h_k^f]$, где $h_i^c, h_i^f \in \mathbb{R}^{\mathbf{h}}$. Далее, для H_c и H_f вычисляются средние значения $p_c = \sum_{i=1}^{n} h_i^c / n$ и $p_f = \sum_{i=1}^{k} h_i^f / k$. Количественная оценка последовательностей выполняется в направлениях: (1) аспектов по отношению к контексту, и (2) контекста по отношению к аспектам. Вычисление весов производится по соответствующим формулам (1) $u_i^c = \tanh(h_i^f \cdot W_c \cdot p_c + b_c)$, и (2) $u_i^f = \tanh(h_i^c \cdot W_f \cdot p_f + b_f)$ где $W_c, W_f \in \mathbb{R}^{\mathbf{h} \times \mathbf{h}}$ и $b_c, b_f \in \mathbb{R}$ – скрытые состояния модели IAN. Далее, оценки u^f и u^c преобразуются в вероятностные α^f и α^c с помощью операции $softmax$. Результирующий вектор контекста s есть конкатенация векторов s_f и s_c, где $s_c = \sum_{i=1}^{n} \alpha_i^c \cdot h^c$, $s_f = \sum_{i=1}^{k} \alpha_i^f \cdot h^f$.

Самовнимание (Self-Attention). Используется кодировщик на основе двунаправленной LSTM. Результирующее контекстное представление такого кодировщика ($H = [h_1, \ldots, h_n]$) есть поэлементная конкатенация двух последовательностей, в которой каждый i-й элемент ($i \in \overline{1..n}$) представлен как $h_i = \overrightarrow{h_i} +\!\!+ \overleftarrow{h_i}$, где $\overrightarrow{h_i}$ и $\overleftarrow{h_i}$ элементы прямой и обратной LSTM последовательностей соответственно. Количественная оценка $h_i \in \mathbb{R}^{\mathbf{h}}$ вычисляется по формуле [Peng Zhou и др., *Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification*, 2016]: $u_i = m_i^T \cdot w$,

где $m_i = \tanh(h_i)$, $w \in \mathbb{R}^{\mathbf{h}}$ – вектор скрытых состояний механизма внимания $\mathbf{a}$, изменяемый в процессе обучения модели. Результирующий вектор представления контекста вычисляется по формуле: $s = tanh(H \cdot \alpha)$.

Классификационный слой используется для преобразования $s \in \mathbb{R}^h$ в вектор классов $o \in \mathbb{R}^{\mathbf{c}}$, где $\mathbf{c}$ – число классов. Для выполнения такого преобразования используется *полносвязный слой* с параметрами $\langle W, b\rangle$ и функцией активации f_a: $o = f_a(W \cdot s + b)$, где $W \in \mathbb{R}^{\mathbf{c}\times h}$, $b \in \mathbb{R}^{\mathbf{c}}$. В случае языковых моделей BERT, s – усредненное значение выходных векторных представлений каждого токена.

Преобразование отношений на уровень документа. Для некоторой пары $\langle v_1, v_2\rangle$ и соответствующего списка размеченных контекстов, результирующая оценка определяется как среднее значение среди меток контекста (усреднение методом голосования) [4].

Корпуса. Оценка моделей проводилась на русскоязычном корпусе аналитических текстов RuSentRel-1.0 [9]. Корпус предоставляет 73 больших аналитических текстов, размеченных с выделением порядка 2000 отношений среди них. Именованные сущности автоматически размечены методом CRF по классам [`PER`, `ORG`, `LOC`, `GEO`]. Из корпуса по каждой паре сущностей были составлены множества контекстов. В 47-48% случаев отношения представлены одним контекстом.

Модели. Рассмотрено качество работы *классических методов* классификации: KNN, Naïve Bayes, Linear SVM, Random Forest и Gradient Boosting при использовании параметров по умолчанию и с перебором таких параметров по предзаданной сетке [9]. Среди *нейросетевых моделей* исследовался следующий набор различных кодировщиков: CNN, PCNN; АттCNN$_e$, АттPCNN$_e$ (модели с кодировщиками с механизмом внимания «многослойный перцептрон»); LSTM, BiLSTM; IAN; Атт-BLSTM (модель с механизмом внимания типа «Self-Attention»). Список используемых *языковых моделей* включает вариации предобученных состояний модели BERT: mBERT [Devlin и др., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2019] (мультиязыковая модель), RuBERT [Kuratov и др., *Adaptation of deep bidirectional multilingual transformers for russian language*, 2019] (дообученная версия mBERT на русскоязычных текстах энциклопедии «Википедия»), SentRuBERT (дообученная модель RuBERT посредством: (1) текстов корпуса SNLI [Bowman и др., *A large annotated corpus for learning natural language inference*, 2015] переведенных на русский язык, (2) русскоязычных текстов коллекции XNLI [Conneau и др., *XNLI: Evaluating Cross-lingual Sentence Representations*, 2018]). Модель BERT предполагает в качестве входной информации последовательность, опционально разделенную специальным символом [SEP] на две части: TextA и TextB. Варианты представления входной информации: (1) «без TextB»– последовательность без разделения, (2) TextB$_{\text{QA}}$ – дополнение TextA вопросом в

TextB, (3) $\text{TextB}_{\text{NLI}}$ – дополнение TextA выводом отношения по контексту в TextB.

Оценка в рамках корпуса. Оценка качества моделей производилась на основе метрик: P (точность), R (полнота), и F_1^{PN}-мера. Для оценки моделей принимается показатель макро-усреднений над документами по оценочным классам: $F^{PN}_{1-mean} = \frac{1}{2} \cdot (F^{P}_{1-macro} + F^{N}_{1-macro}) \cdot 100$. Оценка F^{PN}_{1-mean} производится в рамках фиксированного тестового множества (далее test) коллекции RuSentRel – $F1_t$ [9].

Среди классических методов, SVM и Naïve Bayes достигли 0.16 по F-мере; наилучший результат получен с помощью классификатора Random Forest ($F1_t = 27.0$) и Gradient Boosting ($F1_t = 28.0$). Предлагаемый подход на основе нейронной сети PCNN значительно превосходит подходы с ручными признаками [9]; наилучший результат при $F1_t = 32.6$. Наилучший результат среди моделей с механизмами внимания показывают АттPCNN_e ($F1_t = 32.6$), IAN ($F1_t = 32.2$), Атт-BLSTM ($F1_t = 32.3$) Для языковых моделей, наилучший показатель демонстрирует модель RuBERT («без TextB»), для которой $F1_t = 36.8$ [1]. Результат согласий в разметке экспертами также достаточно низкий $F1_t = 55.0$, но в то же время значительно превышает качество разметки автоматическими методами. Следует отметить, что в смежной задаче [Eunsol Choi и др., *Document-level sentiment inference with social, faction, and discourse context*, 2016] авторы работали с документами на английском языке гораздо меньшего объема и сообщают о F-мере 0.36.

Третья глава посвящена исследованию опосредованного обучения моделей в задаче извлечения оценочных отношений. Приводится структура фреймовой базы знаний RuSentiFrames-2.0 [10]. Разработан метод автоматического извлечения оценочных отношений из новостей с использованием коллекции базы знаний в анализе новостных заголовков.

Формальная постановка задачи. Пусть имеется коллекция размеченных аналитических статей C. Под *размеченной* аналитической статьей понимается $\langle D_i, E_i, \mathbf{a}_i \rangle$, где D_i – текст аналитической статьи, E_i – список упомянутых именованных сущностей, $\mathbf{a}_i$ – список оценочных отношений. Пусть имеется коллекция новостных документов $N = \{D'_1, \ldots, D'_{|N|}\}$, где $|N| \gg |C|$. Под *обучением модели машинного обучения* понимается итеративный процесс оптимизации параметров относительно эталонной разметки $\{\mathbf{a}_i\}_{i=1}^{|C|}$ с целью минимизации непрерывной функции ошибки. Под *опосредованным обучением модели* машинного обучения понимается итеративный процесс оптимизации параметров на основе *объединения* либо *предобучения* с применением **алгоритма-посредника** $\mathcal{A}$ – функции, которая для произвольного документа D' возвращает кортеж из разметки именованных сущностей и оценочных отношений: $\mathcal{A}(D'_i) = \langle E'_i, \mathbf{a}_i \rangle$; значение $\mathbf{a}_i$ в процессе опосредованного обучения рассматривается как множество *эталонной разметки* оценочных отношений документа D'_i.

Таблица 1.

Пример описания фрейма «одобрить» в базе знаний RuSentiFrames-2.0

Слоты фрейма «одобрить»	Описание
ROLES	A0: тот, кто одобряет
	A1: то, что одобряется
POLARITY	A0→A1, POS
	A1→A0, POS
EFFECT	A1, POS
STATE	A0, POS
	A1, POS

Таблица 2.

Количественная характеристика вхождений в RuSentiFrames-2.0

Тип лексической единицы	Количество
Вхождений фреймов	311
Глаголы	3 239
Существительные	986
Фразы	2 551
Другие	12
Уникальных вхождений	6 788
Всего вхождений	7 034

Таблица 3.

Число вхождений отношений базы знаний RuSentiFrames-2.0 по классам

Значения слота «POLARITY»	POS	NEG
A0→A1	2 558	3 289
author→A0	170	1 578
author→A1	92	249

Необходимо разработать подход автоматической разметки оценочных отношений в новостных документах ($\mathcal{A}$). Применение такого подхода в процессе обучения, ввиду значительного превосходства числа документов новостной коллекции над объемом коллекции C, позволит существенно увеличить объем данных для моделей машинного обучения.

Фреймовая база знаний. RuSentiFrames описывает оценки и коннотации, передаваемые предикатом в устной или номинальной форме. Структура фреймов включает в себя набор специфичных для предикатов ролей и набор характеристик для описания фреймов. Для обозначения ролей семантические аргументы отдельных глаголов нумеруются, начиная с нуля (подход PropBank). Для конкретного глагола Arg0 - это, как правило, аргумент, демонстрирующий свойства прототипического агента (Agent) [Dowty и др., *Thematic proto-roles and argument selection*, 1991], в то время как Arg1 это тема (Theme). В основной части коллекции представлены следующие *слоты*:

- ROLES – отношение автора текста к указанным участникам;
- POLARITY – положительная/отрицательная оценка между участниками отношений;
- EFFECT – положительный/отрицательный эффект для участников;

– STATE – положительное/отрицательное психическое состояние участников, связанных с описанной ситуацией.

Пример формата описания фрейма «одобрить» приведен в таблице 1. Фреймы связаны также с *семейством слов и выражений* (лексических единиц), которые имеют одинаковые отношения. Лексические единицы могут быть связаны с оценочным фреймом следующими способами: отдельные слова, идиомы, конструкции глаголов и другие выражения, состоящие из нескольких слов. Для проведения разметки в методе опосредованного обучения моделей используется только измерение отношения к теме (`Theme`), переданное прототипным агентом (`Agent`). В ресурсе RuSentiFrames-2.0 описано 311 фреймов, связанных с 7034 лексическими единицами, среди которых 6788 уникальных, из которых 48% – глаголы, 14% – существительные и оставшиеся 38% – словосочетания (см. таблицу 2). Число вхождений отношений по слоту «POLARITY» приведено в таблице 3.

Используемые ресурсы. Корпус $\text{NEWS}_{\text{Large}}$ (8,8 млн. текстов) – состоит из русскоязычных статей и новостей крупных новостных источников, специализированных политических сайтов и российских сайтов информационных агентств.

Алгоритм. Подразумевает применение двух подходов обработки новостных текстов: *на основе базы знаний* RuSentiFrames, *списка пар* (статистики) на основе новостных заголовков. Диаграмма рабочего процесса представлена на Рис. 1. В блоке *«разбор текста»* выполняется разметка термов групп WORDS и TOKENS в заголовке и предложениях новости. Блок «*разметки фреймов*» выполняет поиск вхождений фреймов базы знаний RuSentiFrames. Блок «*разметки сущностей*» использует предобученную модель $\text{BERT}_{\text{Mult-OntoNotes}}$ (`http://docs.deeppavlov.ai/en/0.11.0/features/models/ner.html`) для выделения сущностей типов множества O_{valid} ={`GPE`, `ORG`, `PER`}, где `GPE` = `GEO`. Модуль *«группировки сущности»* решает задачу кореференций именованных сущностей посредством использования: (1) лемматизатора `Yandex Mystem`, (2) русскоязычного ресурса RuWordNet [Loukachevitch и др., *Comparing Two Thesaurus Representations for Russian*, 2018] для поиска синонимов существительных. Результатом являются: *«коллекция синонимов»* и *«коллекция размеченных текстов»*.

Модуль *«извлечения отношений»* из коллекции размеченных текстов выполняет анализ всех новостных заголовков для составления *«списка пар»* (Рис. 1). Отношение a, извлеченное из новостного заголовка, имеет формат $a = \langle k, g_i, g_j, \mathbf{l} \rangle$, где k – индекс новости, g_i и g_j – индексы синонимичных групп начала и конца пары; $\mathbf{l}$ – назначаемый класс тональности. Отношение a передается в список пар, если:

1. Участники и все именованные сущности между ними принадлежат O_{valid}; e_i и e_j не являются синонимами; e_i упомянута раньше e_j;

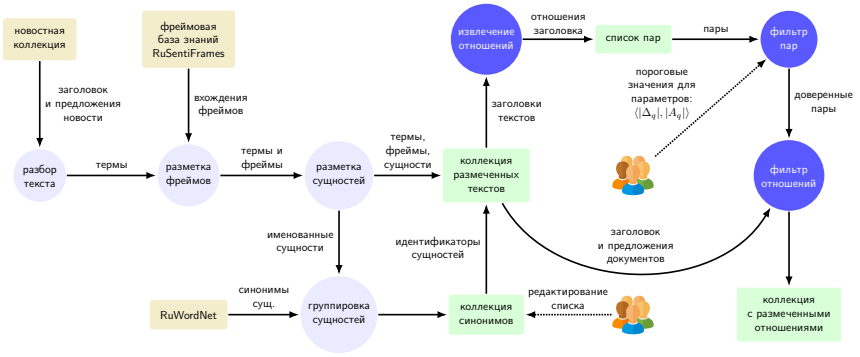

Рис. 1.
Диаграмма рабочего процесса извлечения оценочных отношений из новостных текстов; *прямоугольники* – предзаданные (желтые) и порождаемые (зеленые) источники информации; *кружки* – обработчики данных; пунктирные стрелки – интерфейс пользователя

2. Для всех фреймов, входящих между e_i и e_j, определена полярность $l_{\texttt{A0}\rightarrow\texttt{A1}} \in \{\text{POS}, \text{NEG}\}$; полярность фрейма инвертируется, если перед ним присутствует частица «не» (наиболее частый случай)
3. Отсутствуют предлоги «в» и «на» перед e_i и e_j (такие предлоги в большинстве случаев выражают отношение нахождения где-то; такие отношения не являются оценочными).

Класс тональности $\mathbf{l} \in \{\text{POS}, \text{NEG}\}$ пары a назначается следующим образом: POS (все фреймы внутри пары имеют POS оценку для A0→A1); NEG (иначе). [Kuznetsova и др., *Testing rules for a sentiment analysis system*, 2013]. Таким образом, для некоторой пары $q = \langle g_i, g_j \rangle$, имеем набор связанных отношений $A_q = \{a_1, a_2, \dots a_{|A_q|}\}$. Условная вероятность принадлежности q к классу c вычисляется по формуле: $p(q|c) = |\{\langle d, g_i, g_j, \mathbf{l}\rangle \,|\mathbf{l} = c\}|/|A_q|$. Имея список пар с условной вероятностью оценок отношения и соответствующей частоты по каждой паре, в модуле *«фильтр пар»* (Рис. 1) выбираются *доверенные пары* (q на основе предзаданных нижних пороговых значений для: (1) абсолютной разницы $|\Delta_q| = |p(q|\text{POS}) - p(q|\text{NEG})|$; (2) общего числа пар $|A_q|$. Оценка доверенной пары q определяется знаком Δ_q, т.е. POS ($\Delta_q > 0$), либо NEG ($\Delta_q < 0$). Каждая пара (q) в множестве доверенных пар A' представляется в формате триплета: $\langle g_i, g_j, \Delta_q \rangle$. Параметры $|A|$ и $|\Delta_q|$, а также возможность редактирования коллекции синонимов являются частью человеко-машинного интерфейса (пунктирные стрелки Рис. 1). Человек-оператор также может вручную отобрать доверенные пары, и таким образом управлять процессом автоматической разметки коллекции с этапа *«извлечения отношений»*.

Модуль *«фильтр отношений»* (Рис. 1) завершает процесс обработки, выполняя отбор оценочных отношений среди множества достоверных

пар. Пара $\langle e_i, e_j \rangle$ считается оценочным отношением, если соответствующая пара индексов синонимичных групп $\langle g_i, g_j \rangle$ содержится в множестве доверенных пар A' и оценка $\langle e_i, e_j \rangle$ совпадает с оценочной ориентацией доверенной пары. Дополнительно производится фильтрация предложений, которые содержат хотя бы одну оценочную пару сущностей из заголовка.

Разметка нейтральных отношений (опционально) в заголовке и предложениях новости. Пара $\langle e_1, e_2 \rangle$ считается нейтральной, если:

- сущность e_1 упомянута раньше e_2 и имеет тип из множества O_{valid};
- сущность e_2 имеет тип `LOC` и не находится в списке стран/столиц;
- участники e_1 и e_2 не принадлежат одной синонимичной группе, а также отношения $\langle e_1, e_2 \rangle$ и $\langle e_2, e_1 \rangle$ не являются оценочными.

Результатом выполнения модуля фильтрации отношений является «*коллекция размеченных отношений*» (Рис. 1). Коллекция RuAttitudes-2.0 версии «2017-Large» [1] – результат применения такого рабочего процесса (версии 2.0) к корпусу $\text{NEWS}_{\text{Large}}$ с дополнительной разметкой нейтральных отношений.

Корпуса. Оценка производится при обучении на корпусах RuSentRel и RuAttitudes-2.0 с размеченными нейтральными отношениями. Набор используемых корпусов зависит от режима обучения модели.

Формат проведения опосредованного обучения. Опосредованное обучение выполняется в следующих форматах: (1) предварительное обучение с последующим дообучением (сверточные и рекуррентные нейронные сети, BERT), и (2) совместное обучение (сверточные и рекуррентные нейронные сети).

Оценка и анализ результатов. Оценка моделей проводится в рамках корпуса RuSentRel для экспериментов: (1) TWO-SCALE – необходимо определить оценки заведомо известных пар (классы: POS, NEG); (2) THREE-SCALE – необходимо извлечь оценочные отношения из документа (классы: POS, NEG, NEU).

Результат по метрике F^{PN}_{1-mean} фиксируется в следующих форматах: (1) $F1^a_{cv}$ – усредненный показатель F^{PN}_{1-mean} в рамках 3-кратной кросс-валидационной проверки (Разбиения проведены с точки зрения сохранения одинакового числа предложений в каждом из них.); (2) $F1_t$ – показатель F_{1-mean} на TEST множестве. Прирост качества при использовании опосредованного обучения – отношение результатов, полученных при опосредованном обучении ($F1_b$) к результатам моделей, для которых применялось обучение с учителем ($F1_a$) вычисляется по формуле: $\Delta(F1) = (F1_b/F1_a - 1) \cdot 100$. В таблице 4 приводятся результаты по моделям: (1) сверточных и рекуррентных нейронных сетей, (2) языковых моделей BERT; также приводится статистика прироста качества ($\Delta(F1)$) при использовании опосредованного обучения.

Сверточные и рекуррентные нейронные сети. В зависимости от форматов применения коллекции RuAttitudes в обучении, при дообучении

Таблица 4.

Результаты применения опосредованного обучения для: (1) моделей с кодировщиками на основе сверточных и рекуррентных нейронных сетей, моделей с механизмом внимания (объединенное обучение) (2) моделей BERT («P» – формат TextB предобучения) с дообучением; наилучший результат по каждой модели выделен жирным шрифтом

	Тип коллекции	two-scale		three-scale	
Модель	RuAttitudes-2.0	$F1^a_{cv}$	$F1_t$	$F1^a_{cv}$	$F1_t$
CNN	2017-Large	**70.0**	**74.3**	**32.8**	**39.6**
CNN	—	63.6	65.9	28.7	31.4
PCNN	2017-Large	**69.5**	**70.5**	**31.6**	**39.7**
PCNN	—	64.4	63.3	29.6	32.5
LSTM	2017-Large	**68.0**	**75.4**	**31.6**	**39.5**
LSTM	—	61.9	65.3	27.9	31.6
BiLSTM	2017-Large	**71.2**	68.4	**32.0**	**38.8**
BiLSTM	—	62.3	**71.2**	28.6	32.4
AttCNN$_e$	2017-Large	66.8	72.7	30.9	39.9
AttCNN$_e$	—	65.0	66.2	27.6	29.7
AttPCNN$_e$	2017-Large	**70.2**	67.8	**32.2**	39.9
AttPCNN$_e$	—	64.3	63.3	29.9	32.6
IAN$_e$	2017-Large	69.1	**72.6**	30.7	36.7
IAN$_e$	—	60.8	63.5	30.8	32.2
Att-BLSTM	2017-Large	66.2	71.2	31.0	37.3
Att-BLSTM	—	65.7	68.2	27.5	32.3
Среднее-$\Delta(F1)$ (прирост)	2017-Large	8.7%	8.9%	10.6%	23.4%
Среднее	—	63.5	65.9	28.8	31.8
mBERT (NLI$_P$ + «без TextB»)	2017-Large	68.9	67.7	**30.5**	31.1
mBERT («без TextB»)	—	67.0	68.9	26.9	30.0
mBERT (NLI$_P$ + TextB$_{\text{QA}}$)	2017-Large	69.6	65.2	**30.1**	**35.5**
mBERT (TextB$_{\text{QA}}$)	—	66.5	65.4	28.6	33.8
mBERT (NLI$_P$ + TextB$_{\text{NLI}}$)	2017-Large	69.4	68.2	**33.6**	36.0
mBERT (TextB$_{\text{NLI}}$)	—	67.8	58.4	29.2	37.0
RuBERT (NLI$_P$ + «без TextB»)	2017-Large	**70.0**	69.8	35.6	35.4
RuBERT («без TextB»)	—	67.8	66.2	**36.8**	37.6
RuBERT (NLI$_P$ + TextB$_{\text{QA}}$)	2017-Large	**69.6**	68.2	34.8	37.0
RuBERT (TextB$_{\text{QA}}$)	—	69.5	66.2	32.0	35.3
RuBERT (NLI$_P$ + TextB$_{\text{NLI}}$)	2017-Large	**71.0**	68.6	**36.8**	39.9
RuBERT (TextB$_{\text{NLI}}$)	—	68.9	66.4	29.4	39.6
SentRuBERT (NLI$_P$ + «без TextB»)	2017-Large	70.0	69.8	37.9	**39.8**
SentRuBERT («без TextB»)	—	69.3	65.5	34.0	35.2
SentRuBERT (NLI$_P$ + TextB$_{\text{QA}}$)	2017-Large	69.6	64.2	**38.4**	**41.9**
SentRuBERT (TextB$_{\text{QA}}$)	—	70.2	67.1	34.3	38.9
SentRuBERT (NLI$_P$ + TextB$_{\text{NLI}}$)	2017-Large	70.2	67.7	**39.0**	38.0
SentRuBERT (TextB$_{\text{NLI}}$)	—	69.8	67.6	33.4	32.7
Среднее-$\Delta(F1)$ (прирост)	2017-Large	1.8%	3.7%	13.5%	10%
Среднее	—	68.5	65.7	31.6	35.6

моделей прирост качества варьируется в диапазоне 3-4% и 1.5-3% для two-scale и three-scale экспериментов соответственно. При совместном обучении, показатели прироста увеличиваются в 2 раза (two-scale, прирост 7%) и в 3 и более раза (three-scale, прирост 11% при кросс-валидационном разбиении и 23% при фиксированном). *Языковые модели.* Влияние опосредованного обучения оказывает прирост в 2-5% в (two-scale), 9-13% (three-scale). Преимущество в использовании русскоязычно-ориентированных моделей перед mBERT наблюдается в three-scale эксперименте. Так, при использовании RuBERT прирост качества составил от 10% по

$F1^a_{cv}$ и 6-9% по метрике $F1_t$. Применение SentRuBERT улучшает показатели RuBERT на 6% в опосредованном обучении. SentRuBERT по качеству разметки приближается к нейронным сетям при объединенном формате обучения, задавая при этом более высокие результаты в рамках кросс-валидационного тестирования (35.6-39.0), что говорит о более стабильном результате нежели при использовании сверточных и рекуррентных нейронных сетей.

Анализ влияния опосредованного обучения. Источником контекстов являются данные множества TEST (контексты, наибольшее расстояние в термах между участниками которых не превышает значения 10 [7; 8]). Для каждого контекста анализ проводится по *вхождениям фреймов* (FRAMES), *вхождениям оценочного лексикона* RuSentiLex (SENTIMENT).

Для моделей сверточных и рекуррентных нейронных сетей. Анализируется разница между оценочными и нейтральными отношениями на основе плотности распределения веса группы g (w_g) для: *нейтральных* (ρ_{N}, отмеченные NEU) и *оценочных* (ρ_{S}, отмеченные POS либо NEG) контекстов. Под *весом* w_g группы термов g входного контекста понимается сумма весов термов контекста, принадлежащих g. Для определения разницы между ρ_{S} и ρ_{N} применяется статистика на основе теста Колмогорова-Смирнова. Наибольшую разницу по всем группам демонстрирует модель Att-BLSTM [8].

Для языковых моделей в анализе участвуют: mBERT, SentRuBERT, и SentRuBERT-NLI$_P$ (дообученная версия SentRuBERT с применением опосредованного обучения на основе коллекции RuAttitudes-2.0). Среди входных контекстов множества TEST рассматриваются только такие, которые были извлечены дообученной моделью SentRuBERT-NLI$_P$. В работе [1] приводится оценка усредненных значений весов внимания по входным контекстам для токенов (п. 3,4 – только контексты, содержащие термы соответствующих групп): (1) класса [CLS], границ последовательностей [SEP], (2) участников отношений ($\underline{\text{E}}_{subj}/\underline{\text{E}}_{obj}$), (3) групп FRAMES и внимание к ним от прочих токенов контекста; (4) группы SENTIMENT и внимание к ним от прочих токенов контекста. По результатам следует отметить высокие показатели внимания к токену класса [CLS] на ранних слоях до 31% (mBERT) и выше в случае остальных моделей. При переходе от mBERT к SentRuBERT и SentRuBERT-NLI$_P$ наблюдается увеличение внимания к [SEP] до 29% (слои с 6 по 10), $\underline{\text{E}}_{subj}/\underline{\text{E}}_{obj}$ до 12%. Внимание по отношению к термам групп FRAMES и SENTIMENT остальных токенов, в среднем составляет 4-5% и увеличивается на слоях 11-12 до 7-10%; последний показатель вдвое превышает аналогичные показатели модели mBERT.

В **четвертой главе** приводится описание архитектуры программного комплекса AREkit для извлечения оценочных отношений из документов, а также временная оценка обучения моделей, созданных на основе такого комплекса.

Разработанный программный комплекс AREkit (https://github.com/nicolay-r/AREkit/tree/0.20.5-rc) предоставляет возможности: (1) подготовки и обработки текстовой информации (выделение контекстов с парами упомянутых именованных сущностей) (2) проведения экспериментов с моделями машинного обучения в задаче извлечения оценочных отношений между парой упомянутых в тексте сущностей. Программный комплекс состоит из: (1) набора инструментов для работы с отношениями документа (ядро), (2) модуля запуска экспериментов на основе моделей, (3) модуля реализации моделей на основе библиотеки `Tensorflow`.

Оценка производительности проводилась на сервере с двумя процессорами Intel Xeon CPU E5-2670 v2 2.50Ггц, 80 Гб ОЗУ (DDR-3), двумя видеоускорителями Nvidia GeForce GTX 1080 Ti. Обучение моделей выполнялось под ОС Ubuntu 18.0.4 в контейнерах Docker версии 19.03.5. *Среди нейронных сетей* наилучшие показатели скорости обучения демонстрируют модели с кодировщиком на основе сверточных нейронных сетей. Добавление механизма внимания на основе перцептрона (АттCNN$_e$, АттPCNN$_e$) увеличивает время обучения примерно в 10 раз относительно PCNN. Время обучения рекуррентных нейронных сетей увеличивается в 2 раза (LSTM), и в 3-4 раза (BiLSTM) при сравнении с PCNN. Добавление механизма самовнимания (Атт-BLSTM) практически не сказалось на времени обучения модели BiLSTM, что обусловлено вычислительно более простой архитектурой механизма внимания среди прочих [1]. *Для языковых моделей* во всех форматах обучения на адаптацию русскоязычных моделей требуется меньше эпох при одинаковых настройках обучения. Так, замена mBERT на RuBERT или SentRuBERT сокращает время обучения в 3.5 раза [1].

В **заключении** приведены основные результаты работы.

Публикации автора по теме диссертации

В изданиях из списка ВАК РФ

1. Русначенко, Н. Л. Применение языковых моделей в задаче извлечения оценочных отношений // Труды Института системного программирования РАН, 33 (3). 2021. С. 199—222. (1,5 п.л.)
2. Русначенко, Н. Л., Лукашевич, Н. В. Методы интеграции лексиконов в машинное обучение для систем анализа тональности // Искусственный интеллект и принятие решений. 2017. С. 78—89. (0,75 п.л./0,5 п.л.)

В изданиях, входящих в международную базу цитирования Scopus

3. Rusnachenko, N., Loukachevitch, N. Extracting Sentiment Attitudes from Analytical Texts via Piecewise Convolutional Neural Network // In Proceedings of CEUR Workshop, DAMDID-2018 Conference. 2018. С. 186—192. (0,44 п.л./0,25 п.л.)

4. Rusnachenko, N., Loukachevitch, N. Neural Network Approach for Extracting Aggregated Opinions from Analytical Articles // International Conference on Data Analytics and Management in Data Intensive Domains. Springer. 2019. С. 167—179. (0,81 п.л./0,47 п.л.)
5. Rusnachenko, N., Loukachevitch, N. Sentiment Attitudes and Their Extraction from Analytical Texts // International Conference on Text, Speech, and Dialogue. Springer. 2018. С. 41—49. (0,56 п.л./0,22 п.л.)
6. Rusnachenko, N., Loukachevitch, N., Tutubalina, E. Distant Supervision for Sentiment Attitude Extraction // Proceedings of Recent Advances in Natural Language Processing Conference. 2019. С. 1022—1030. (0,56 п.л./0,31 п.л.)
7. Rusnachenko, N., Loukachevitch, N. Studying Attention Models in Sentiment Attitude Extraction Task // Proceedings of the 25th International Conference on Natural Language and Information Systems. 2020. С. 157—169. (0,81 п.л./0,53 п.л.)
8. Rusnachenko, N., Loukachevitch, N. Attention-Based Neural Networks for Sentiment Attitude Extraction using Distant Supervision // The 10th International Conference on Web Intelligence, Mining and Semantics, Biarritz, France. 2020. С. 159—168. (0,63 п.л./0,5 п.л.)
9. Loukachevitch, N., Rusnachenko, N. Extracting sentiment attitudes from analytical texts // Proceedings of International Conference on Computational Linguistics and Intellectual Technologies. 2018. С. 459—468. (0,63 п.л./0,25 п.л.)
10. Loukachevitch, N., Rusnachenko, N. Sentiment Frames for Attitude Extraction in Russian // Proceedings of International Conference on Computational Linguistics and Intellectual Technologies. 2020. С. 541—552. (0,75 п.л./0,28 п.л.)

Зарегистрированные патенты

11. Заявка 2021663268 Рос. федерация. Программный комплекс для извлечения оценочных отношений между именованными сущностями из коллекций новостных документов [Текст] / Н. Л. Русначенко (Российская Федерация) ; Ф. С. по Интеллектуальной Собственности. № 2021663268 ; заявл. 13.08.2021 ; Бюл. № 7 (I ч.) (Рос. Федерация). 63 с. : ил.

В сборниках трудов конференций

12. Rusnachenko, N., Loukachevitch, N. Using convolutional neural networks for sentiment attitude extraction from analytical texts // EPiC Series in Language and Linguistics. Т. 4. EasyChair, 2019. С. 1—10. (0,63 п.л./0,28 п.л.)

Русначенко Николай Леонидович

Модели, методы и программные средства извлечения оценочных отношений на основе фреймовой базы знаний

Автореф. дис. на соискание ученой степени канд. тех. наук

Подписано в печать ___.___._____. Заказ № _____
Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.
Типография _____