

doi: 10.17586/2226-1494-2022-22-6-1159-1165

УДК 004.9

Применение волновой модели текста к задаче сентимент-анализа

Анастасия Сергеевна Груздева^{1✉}, Родион Николаевич Юрьев²,
Игорь Александрович Бессмертный³

^{1,2,3} Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

¹ prog.anastasia@gmail.com✉, <https://orcid.org/0000-0003-4963-0823>

² rodion@juryev.ru, <https://orcid.org/0000-0003-1146-2617>

³ bessmertny@itmo.ru, <https://orcid.org/0000-0001-6711-6399>

Аннотация

Предмет исследования. Исследована волновая модель представления коротких текстов на русском языке. Модель является одной из реализаций дистрибутивной семантики. В модели учтены не только частоты встречаемости слов в тексте, но и их взаимное влияние. Итогом реализации модели служит повышение точности анализа тональности коротких текстов. **Метод.** Основу определения взаимосвязей между текстом и термином составляет расчет амплитуды вероятности близости текста к термину с использованием волновой модели. Термин, обладающий наибольшей амплитудой вероятности, считается наиболее точно соответствующим смыслу текста. Волновая модель позволяет учесть тот факт, что известные методы определяют антонимы как семантически близкие лексические единицы. **Основные результаты.** Для экспериментального исследования предложенного метода выбрано решение задачи сентимент-анализа, то есть нахождения соответствия отзывов пользователей о покупках на маркетплейсе классам «позитивный» и «негативный». В результате эксперимента получена точность оценки тональности текста до 76,4 %, что превышает точность классического подхода, а также известных методик сентимент-анализа для русского языка. Выявлено значительное влияние таких параметров модели, как выбор базисной дистрибутивно-семантической модели, выбор контрольной точки для расчета волновых чисел, учет влияния антонимов на точность классификации. Представленная модель показала высокую точность в выявлении взаимосвязей текста с не присутствующими в нем явно понятиями. **Практическая значимость.** Предложенный метод может успешно применяться как математическая основа решения задач сентимент-анализа. Полученные результаты показали потенциальные возможности использования волновой модели в областях, требующих классификации текстов по косвенным признакам, например, для определения элементов психологического портрета автора.

Ключевые слова

сентимент-анализ, классификация, обработка естественного языка, волновая модель, квантово-подобная модель

Благодарности

Работа выполнена в рамках магистерско-аспирантской НИР № 620164 «Методы искусственного интеллекта для киберфизических систем».

Ссылка для цитирования: Груздева А.С., Юрьев Р.Н., Бессмертный И.А. Применение волновой модели текста к задаче сентимент-анализа // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 6. С. 1159–1165. doi: 10.17586/2226-1494-2022-22-6-1159-1165

Application of the text wave model to the sentiment analysis problem

Anastasia S. Gruzdeva^{1✉}, Rodion N. Iurev², Igor A. Bessmertny³

^{1,2,3} ITMO University, Saint Petersburg, 197101, Russian Federation

¹ prog.anastasia@gmail.com✉, <https://orcid.org/0000-0003-4963-0823>

² rodion@juryev.ru, <https://orcid.org/0000-0003-1146-2617>

³ bessmertny@itmo.ru, <https://orcid.org/0000-0001-6711-6399>

Abstract

Authors researched the wave model of text representation which is one of the implementations of distributive semantics. This model takes into account not only the frequency of words occurrence in the text, but also their mutual location.

© Груздева А.С., Юрьев Р.Н., Бессмертный И.А., 2022

The purpose of the study: to increase the accuracy of the analysis of the tonality of short texts based on the wave model. The method of determining the relationship between the text and the term is based on the calculation of the probability amplitude of the text and term proximity using a wave model. The term with the highest probability amplitude is considered to correspond most closely to the meaning of the text. The wave model allowed taking into account the fact that well-known methods define antonyms as semantically close lexical units. For the experimental study of this technique, a solution to the problem of sentiment analysis was chosen, exactly, finding the correspondence of user reviews about the product to the classes “positive” and “negative”. As a result, the accuracy of the text tonality defining was obtained up to 76.4 %, which exceeds the accuracy of the classical approach as well as the well-known methods of sentiment analysis for the Russian language. In addition, authors detected significant influence on classification accuracy of such model parameters as the choice of a basic distributive semantic model, the choice of a control point for calculating wave numbers, taking into account the influence of antonyms. The presented model has shown high accuracy in identifying the relationships of the text with concepts that are not explicitly present in it and can be successfully used as a mathematical basis for solving problems of sentiment analysis. In addition, the results obtained indicate the potential use of the wave model in other areas that require the classification of texts by indirect signs, for example, to determine the elements of author psychological portrait.

Keywords

sentiment analysis, classification, natural language processing, wave model, quantum-like model

Acknowledgements

The work was carried out within the framework of the project no. 620164 (Artificial intelligence methods for cyberphysical systems).

For citation: Gruzdeva A.S., Iurev R.N., Bessmertny I.A. Application of the text wave model to the sentiment analysis problem. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 6, pp. 1159–1165 (in Russian). doi: 10.17586/2226-1494-2022-22-6-1159-1165

Введение

Проблемы современных подходов к сентимент-анализу связаны с высокой размерностью данных и неизбежно возникающей неоднозначностью в текстах. Результативность техник сентимент-анализа ограничена в настоящее время для английского языка точностью около 70–80 % в зависимости от применяемого инструментария [1]. Для русского языка точность распознавания продемонстрирована в отдельных исследованиях от 60 до 73 % [2]. В то же время востребованность сентимент-анализа в мировом и российском сегментах растет с каждым годом, что показывает статистика запросов на популярном ресурсе Google Trends¹. График роста числа запросов за последние годы показан на рисунке.

Цель работы — совершенствование методов сентимент-анализа для выявления элементов онтологий в поведении пользователей

Состояние проблемы и постановка задачи

Сентимент-анализ — одна из популярных задач компьютерной лингвистики. Первые методы сентимент-анализа базировались на размеченных словарях оценочных слов [3, 4]. Применение словарей позволяет не только упростить оценку тональности текста, но и скорректировать его при необходимости [5]. Ограничением является словарь, составление которого требует времени, а одни и те же слова в разных контекстах могут иметь разную эмоциональную окраску. Альтернатива словарному подходу — корпусная лингвистика, и, в частности, дистрибутивная гипотеза Харриса [6], согласно которой слова, часто встречающиеся в похожем контексте, имеют схожий смысл. В данном случае слова и тексты представлены векторами

в гильбертовом пространстве, а близость слов и текстов определены евклидовой мерой или косинусным расстоянием. Последние исследования в области сентимент-анализа описаны в работах [7–10], где общей проблемой является семантическая неразличимость антонимов.

Около 30 лет назад к анализу текстов стали применять квантовую теорию вероятностей [11–16]. Основное отличие квантовой теории от колмогоровской — наличие дополнительного параметра — фазы, за счет которой происходит интерференция амплитуд вероятностей. В связи с этим предложено использовать аппарат квантовой теории для классификации текстов, чтобы в отличие от известных методов, базирующихся на «мешке слов», учитывать взаимное расположение слов в тексте. В работе [17] разработана волновая модель, которая позволила учесть интерференцию при анализе текстов. Задачи сентимент-анализа часто связаны с необходимостью выявления неявных взаимосвязей и классификации текстов по косвенным признакам. В тексте могут отсутствовать явные указания на то, к какому эмоциональному полюсу — позитивному или негативному — он ближе. Тем не менее, набор используемых слов, строение предложений позволяют человеку понять общую эмоциональную тональность текста. Задача настоящей работы — проверить, может ли введение дополнительного параметра — фазы — оказать положительное влияние на точность классификации текстов по косвенным признакам.

Краткое описание волновой модели

Рассмотрим возможность применения квантово-подобной волновой модели представления текстовой информации для выявления тональности текста. Отметим основные моменты волновой модели, подробное описание которой приведено в работе [17].

В рамках волновой модели текст представлен в виде ансамбля элементарных частиц, где каждое сло-

¹ Google Trends. [Электронный ресурс]. URL: <https://trends.google.es/trends/explore?date=all&q=sentiment%20analysis> (дата обращения: 14.09.2022).

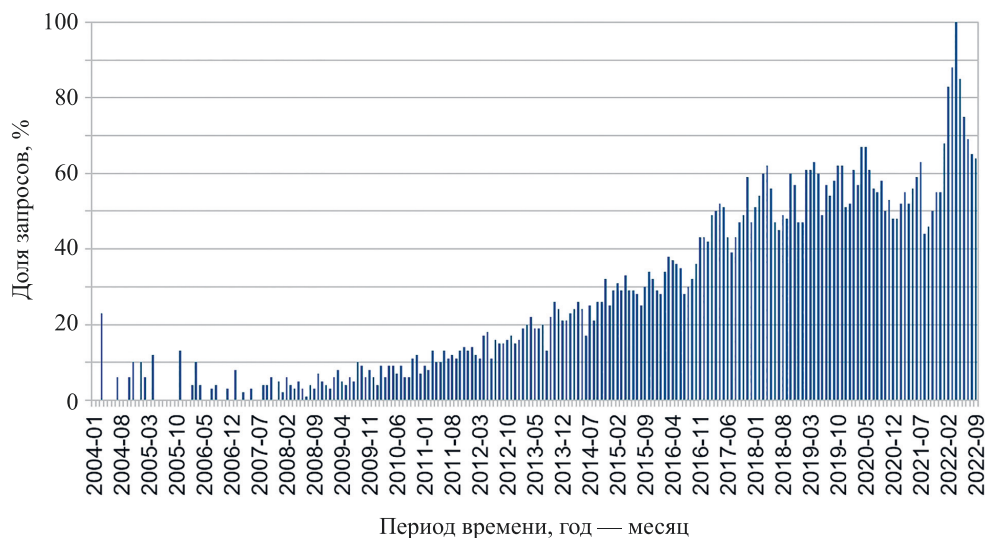


Рисунок. Диаграмма роста запросов «сентимент-анализа» с 2004 по 2022 годы (во всем мире). Чередование цветов использовано для улучшения зрительного восприятия диаграммы. Доля запросов рассчитана относительно максимального числа запросов в месяц за исследуемый период

Figure. Diagram of “sentiment analysis” requests growth from 2004 to 2022 (worldwide). The share of the requests is calculated relative to the maximum number of requests per month for the study period

во — отдельная частица. Поведение частицы описано волновой функцией, соответствующей сферической волне [18], интенсивность которой определяет амплитуду вероятности детектирования частицы в точке пространства. Поведение ансамбля элементарных частиц описано пакетом сферических волн, суммарная интенсивность которого отражает амплитуду вероятности обнаружения всего ансамбля в точках пространства. Такой принцип лежит в основе классификации текстов с использованием волновой модели. Формула расчета амплитуды вероятности принадлежности текста к некоторому классу l имеет вид [17]

$$I_l = \sum_{j=1}^M \left(\frac{A_j}{r_{lj}} \right)^2 + 2 \times \sum_{j=1}^{M-1} \sum_{n=j+1}^M \frac{A_j}{r_{lj}} \frac{A_n}{r_{ln}} \cos(k_j r_{lj} - k_n r_{ln} + \varphi_{j0} - \varphi_{n0}), \quad (1)$$

где I_l — интенсивность волнового пакета в точке пространства, соответствующей классу l ; M — число волн в пакете (число слов в тексте); A_j, A_n — число вхождений соответствующих слов в исходный текст; r_{lj} — семантическое расстояние между текстом l и словом j (величина, обратная семантической близости, рассчитанной на предобученной дистрибутивно-семантической языковой модели [19]); k_j, k_n — некоторые волновые числа.

В рамках используемых базисных моделей семантическая близость принимает значения в интервале от -1 до 1 . Семантические расстояния изменяются в диапазонах от 1 до $+\infty$ и от -1 до $-\infty$. Рассчитанная амплитуда вероятности может быть больше 1 , что не противоречит принципам квантовоподобных моделей.

Пренебрежем начальными фазами $\varphi_{j0}, \varphi_{n0}$ и для расчетов используем упрощенную модификацию уравнения (1):

$$I_l = \sum_{j=1}^M \left(\frac{A_j}{r_{lj}} \right)^2 + 2 \times \sum_{j=1}^{M-1} \sum_{n=j+1}^M \frac{A_j}{r_{lj}} \frac{A_n}{r_{ln}} \cos(k_j r_{lj} - k_n r_{ln}). \quad (2)$$

Исследуем применение нескольких моделей, предоставляемых интернет-ресурсом RusVectors¹: НКРЯ + Википедия (Национальный корпус русского языка + русскоязычная Википедия) за ноябрь 2021 года; Новостной корпус и корпус Тайга за 2019 год. Причина выбора данных моделей — наличие словарей достаточно большого объема (около 249 000 слов каждый). Кроме того, на портале RusVectors реализован программный интерфейс приложения, который упрощает обработку данных для анализа.

Для вычисления волновых чисел k_j, k_n используем следующие соображения. Найдем контрольную точку пространства, в которой суммарная интенсивность волнового пакета будет предположительно максимальной. Для достижения максимума интенсивности необходимо, чтобы фазы всех волн в данной точке были равны — волновые числа должны быть обратно пропорциональны расстояниям до контрольной точки. В работе [17] в качестве контрольной точки был использован центроид ансамбля частиц (текста). Здесь также применим расчет волновых чисел относительно центроида текста. Дополнительно, в качестве контрольной точки, используем заранее определенный контекст повествования, что позволит повысить точность классификации на 3,7 %.

В исходный алгоритм внесем изменение для коррекции проблемы высокой семантической близости антонимов в базисных дистрибутивно-семантических

¹ RusVectors [Электронный ресурс]. URL: <https://rusvectors.org/> (дата обращения: 05.05.2022).

моделях. Заметим, что часто пары противоположных по значению слов показывают высокие положительные величины семантической близости вместо ожидаемых значений, близких к -1 . Так, например, в модели НКРЯ+Википедия слова «белый» и «черный» обладают семантической близостью 0,732, а «плохой» и «хороший» — 0,770, т. е. они интерпретируются как синонимы, что противоречит человеческой логике. Очевидно, что высокая смысловая близость терминов, близких к понятиям «плохой» и «хороший», может особенно негативно сказаться на результатах анализа текста с целью определения отношения пользователя к качеству товара. Эта проблема отмечена во всех примененных в настоящей работе базисных дистрибутивно-семантических моделях. Предварительный анализ моделей одного из лучших проектов Russian Distributional Thesaurus [20] показал, что они также не лишены указанного недостатка. Для частичной компенсации проблемы синонимичности антонимов использован простой прием инверсии. Если пара слов обладает положительной семантической близостью и в то же время распознается как пара антонимов, знак семантической близости меняется на противоположный. Такой подход позволил увеличить точность классификации текстов.

Классификация отзывов пользователей

Для экспериментального исследования возможностей применения волновой модели в решении задач оценки общественного мнения отобраны комментарии пользователей, посвященные различным моделям смартфонов в интернет-магазине «Wildberries»¹. Комментарии распределены по двум классам: негативный и позитивный. Изначально распределение выполнено в соответствии с количеством «звезд», выставленных пользователями. Четыре и пять «звезд» расценивались как позитивный комментарий, одна и две — как негативный. Отзывы, отмеченные тремя «звездами», были отнесены к условному классу «нейтральный», не участвовавшему в классификации. В процессе классификации к «нейтральному» классу были присоединены комментарии, которые не удавалось отнести к позитивному или негативному классам в соответствии с заданными критериями. Изначальное распределение по классам пришлось пересмотреть, так как «звезды» выставлялись пользователями не столько за сам продукт, сколько за работу сервиса в целом, поэтому в целом позитивный комментарий могла сопровождать одна «звезда» из-за проблем, например, с доставкой. Окончательное распределение по классам было скорректировано в соответствии со смыслом отзывов. Всего было отобрано 100 позитивных отзывов (четыре и пять «звезд») и 100 негативных (одна и две «звезды»). Встретившиеся при отборе 16 отзывов, отмеченные тремя «звездами» были отнесены к «нейтральному» классу. В целом нейтральные отзывы наблюдались редко. Человек скорее всего не будет писать отзыв, если ему нечего сказать, а товар не произвел

ни положительного, ни отрицательного впечатления. После уточнения объективных классов в соответствии со смыслом текстов получено следующее распределение: 102 позитивных отзыва, 106 — негативных и 8 — нейтральных. Длина отзывов варьировалась от двух до 124 слов.

В процессе классификации выполнен расчет амплитуды вероятности принадлежности каждого отзыва к негативному и позитивному классам с использованием волновой модели. На основании полученных амплитуд вероятности для каждого класса, осуществлен расчет вероятности принадлежности текста:

$$P_l = \frac{I_l}{\sum_{i=1}^C I_i}, \quad (3)$$

где P_l — вероятность принадлежности текста к классу l ; I_l — амплитуда вероятности принадлежности текста к классу; C — количество классов. При предварительных расчетах обнаружено, что модель стабильно относит к позитивному классу почти в два раза больше отзывов, чем к негативному, при том, что объективно каждому из классов соответствовало примерно равное число комментариев. В связи с этим при использовании волновой модели классам присвоен понижающий коэффициент и учтен в расчете вероятностей. При этом уравнение (3) приняло вид

$$P_l = \frac{(1 - D_l) \times I_l}{\sum_{i=1}^C (1 - D_i) \times I_i},$$

где D_l — понижающий коэффициент класса l . Расчеты показали, что оптимальные значения понижающего коэффициента для позитивного класса располагаются в интервале от 0,4 до 0,6, для негативного — 0 при использовании волновой модели. Для классической модели (евклидова мера семантической близости (ЕМ)) применение понижающих коэффициентов не привело к повышению точности, поэтому для нее расчет вероятности выполнен по формуле (3). По результатам расчетов исследованный текст отнесен к классу l , при условии, что вероятность принадлежности к нему превысила пороговую величину 50,5 %. Если такой класс не обнаружен, то текст считается неклассифицированным и относится к «нейтральному» классу.

Выполним расчеты с использованием базисной модели НКРЯ + Википедия. В расчетах применим алгоритмы: расчет евклидовой меры семантической близости (ЕМ); волновая модель, в которой в качестве контрольной точки для расчета волновых чисел использован центр масс текста (ВМ ЦМТ); волновая модель, в которой волновые числа рассчитаны относительно контекста «смартфон» (ВМ К); волновая модель, в которой волновые числа рассчитаны относительно контекста «смартфон» с учетом инверсии антонимов (ВМ КИ). Полученная точность классификации приведена в табл. 1.

Наилучшую точность классификации показала волновая модель ВМ КИ, с использованием которой выполнен серию расчетов исследования влияния базисной

¹ Маркетплейс «Wildberries» [Электронный ресурс]. URL: <https://www.wildberries.ru/> (дата обращения: 22.05.2022).

Таблица 1. Точность классификации отзывов пользователей на базисной модели НКРЯ + Википедия, %
 Table 1. Precision of user reviews classification based on the distributive-semantic model NRC + Wikipedia, %

Класс	Алгоритмы			
	ЕМ	ВМ ЦМТ $D_n = 0,6$	ВМ К $D_n = 0,6$	ВМ КИ $D_n = 0,6$
Позитивный	78,4	77,5	68,6	70,6
Негативный	47,1	68,9	78,3	79,2
Нейтральный	25,0	12,5	0	0
Общий итог	61,1	70,8	70,8	72,2

Примечание. D_n — выбранные значения понижающего коэффициента позитивного класса

Таблица 2. Точность классификации для волновой модели с расчетом волновых чисел относительно контекста, %
 Table 2. Classification precision for wave model with wave number calculation based on context, %

Класс	Базисные модели			
	НКРЯ + Википедия $D_n = 0,6$	Новостной корпус $D_n = 0,45$	Корпус Тайга $D_n = 0,45$	НКРЯ + Википедия и Тайга $D_n = 0,45$
Позитивный	70,6	53,9	80,4	80,4
Негативный	79,2	87,8	68,9	78,3
Нейтральный	0	0	0	0
Общий итог	72,2	68,5	71,8	76,4

Примечание. D_n — выбранные значения понижающего коэффициента позитивного класса

дистрибутивно-семантической модели. В расчетах применим модели интернет-ресурсов: НКРЯ + Википедия, Новостной корпус, корпус Тайга и комбинацию моделей НКРЯ + Википедия и Тайга. В комбинации моделей для расчетов использовано среднее арифметическое семантических близостей. Точность классификации приведена в табл. 2.

В табл. 1 и 2 приведены наилучшие значения точностей, полученные для каждого алгоритма и модели. В рамках выбранных параметров волновой модели ВМ КИ не был верно распознан ни один из отзывов, относящихся к «нейтральному» классу. В данном исследовании не предпринимались никакие шаги для повышения точности определения данного класса, так как он одновременно и малочисленный, и малозначимый для целей sentiment-анализа. Тем не менее, вопрос об отсутствии верно распознанных данных на этом классе требует дополнительного осмысления и проверки, возможно, это дополнительно указывает на интерференционную природу самого исследуемого объекта.

Выводы

В результате работы волновая модель не только показала работоспособность, но и продемонстрировала превышение точности классификации по сравнению с традиционными моделями. Как показано в табл. 2, по отдельным типам классификации точность достигает 87,8 %, при этом в общем итоге средняя точность достигла 76,4 %, что на 3,4 % превышает аналогичные показатели традиционных моделей для русскоязычных текстов. Данные результаты получены благодаря простым математическим вычислениям, не требующим

повышенных вычислительных мощностей от пользователя, что является дополнительным преимуществом при внедрении новой системы на практике и позволяет осуществить распараллеливание процесса анализа. Алгоритмы, разработанные на базе волновой модели, могут стать хорошим дополнением к спектру существующих методик sentiment-анализа [7–10, 21–23].

Полученные экспериментальные данные позволяют сделать вывод о положительном влиянии учета интерференции на классификацию текстов по косвенным признакам. Как видно из уравнения (2) волновая модель отличается от евклидовой меры семантической близости только наличием члена, отвечающего за интерференцию, поэтому именно учет интерференции позволяет достичь заметного увеличения точности классификации.

Высокая точность, достигнутая в классификации по неявным признакам, позволяет допустить последующее расширение применимости волновой модели до любых других объектов исследования (например, в литературе представлены предложения по анализу музыкальных данных [24]).

Заключение

Безусловно, для полномасштабного применения волновой модели остаются ограничения, в первую очередь это касается правильного подбора вычислительного базиса, что во много зависит от искусства аналитика, работающего с материалом. Тем не менее, достижение вычислительной точности, продемонстрированной в работе, позволяет сделать еще один шаг в направлении решения задачи классификации текстов.

Литература

References

- Garg N., Sharma K. Text pre-processing of multilingual for sentiment analysis based on social network data // *International Journal of Electrical and Computer Engineering*. 2022. V. 12. N 1. P. 776–784. <https://doi.org/10.11591/ijece.v12i1.pp776-784>
- Корней А.О., Крючкова Е.Н. Проблемы эффективности сентимент-анализа русскоязычных текстов в социальных сетях // *Высокопроизводительные вычислительные системы и технологии*. 2018. Т. 2. № 2. С. 87–92.
- Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке // *Тезисы докладов Международной конференции по компьютерной лингвистике и интеллектуальным технологиям «Диалог 2011»*. М.: РГТУ, 2011. С. 510–522.
- Клековкина М.В., Котельников Е.В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики // *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: материалы XIV Всероссийской научной конференции (RCDL-2012)*. 2012. С. 118–123.
- Бессмертный И.А., Джалишвили З.О., Максимов В.В., Маркин Д.А. Лингвооценочное управление текстом // *Тезисы докладов X Международной конференции «Применение новых технологий в образовании»*. Троицк: Фонд новых технологий в образовании «Байтик», 1999.
- Harris Z. *Mathematical Structures of Language*. Wiley, 1968. 230 p.
- Богданов А.Л., Дуля И.С. Сентимент-анализ коротких русскоязычных текстов в социальных медиа // *Вестник Томского государственного университета. Экономика*. 2019. № 47. С. 220–241. <https://doi.org/10.17223/19988648/47/17>
- Рябыкин А.С., Сухов Е.А. Нейросетевые методы в задаче сентимент-анализа // *DSPA*. 2022. Т. 2. № 2. С. 41–57.
- Воробьев А.А., Рыбак А.М., Середкин Р.А., Андросов А.Ю., Соловьев Б.И. Методика сбора и обработки социологической информации из сети Интернет // *Известия Тульского государственного университета. Технические науки*. 2022. № 2. С. 208–213. <https://doi.org/10.24412/2071-6168-2022-2-208-214>
- Фомина И.К., Татаурова А.С. Повышение эффективности модели классификации определения тональности текста // *Актуальные проблемы экономики и управления*. 2022. № 1. С. 55–58.
- Zucco G., Azzopardi L., van Rijsbergen K. The quantum probability ranking principle for information retrieval // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2009. V. 5766. P. 232–240. https://doi.org/10.1007/978-3-642-04417-5_21
- Melucci M. *Introduction to Information Retrieval and Quantum Mechanics*. Springer-Verlag Berlin Heidelberg, 2015. 247 p. <https://doi.org/10.1007/978-3-662-48313-8>
- Хренников А.Ю. Введение в квантовую теорию информации. М.: Физматлит, 2008. 283 с.
- Blacoe W., Kashefi E., Lapata M. A quantum-theoretic approach to distributional semantics // *Proc. of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*. 2013. P. 847–857.
- Jaiswal A.K., Holdack G., Frommholz I., Liu H. Quantum-like Generalization of Complex Word Embedding: a lightweight approach for textual classification // *CEUR Workshop Proceedings*. 2018. V. 2191. P. 159–168.
- Surov I.A., Semenenko E., Platonov A.V., Bessmertny I.A., Galofaro F., Toffano Z., Khrennikov A.Yu., Alodjants A.P. Quantum semantics of text perception // *Scientific Reports*. 2021. V. 11. N 1. P. 4193. <https://doi.org/10.1038/s41598-021-83490-9>
- Груздева А.С., Бессмертный И.А. Классификация коротких текстов с использованием волновой модели // *Научно-технический вестник информационных технологий, механики и оптики*. 2022. Т. 22. № 2. С. 287–293. <https://doi.org/10.17586/2226-1494-2022-22-2-287-293>
- Френкель Я.И. Волновая механика. Ч. 1. Элементарная теория. Квантовая физика. М., 2019. 386 с.
- Kutuzov A., Kuzmenko E. Webvectors: A toolkit for building web interfaces for vector semantic models // *Communications in Computer and Information Science*. 2017. V. 661. P. 155–161. https://doi.org/10.1007/978-3-319-52920-2_15
- Panchenko A., Ustalov D., Arefyev N., Paperno D., Konstantinova N., Loukachevitch N., Biemann C. Human and machine judgements about russian semantic relatedness // *Communications in Computer and Information Science*. 2016. V. 661. P. 221–235. https://doi.org/10.1007/978-3-319-52920-2_21
- Pang B., Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proc. of the 42nd*
- Garg N., Sharma K. Text pre-processing of multilingual for sentiment analysis based on social network data. *International Journal of Electrical and Computer Engineering*, 2022, vol. 12, no. 1, pp. 776–784. <https://doi.org/10.11591/ijece.v12i1.pp776-784>
- Korney A.O., Kryuchkova E.N. Performance aspects of the russian texts sentiment analysis in social networks. *High-performance computing systems and technologies*, 2018, vol. 2, no. 2, pp. 87–92. (in Russian)
- Pazelskaia A.G., Solovov A.N. Method for determining emotions in Russian texts. *Proc. of the International conference on computational linguistics and intellectual technologies “Dialogue 2011”*. Moscow, RSUH, 2011, pp. 510–522. (in Russian)
- Klekovkina M., Kotelnikov E. The automatic sentiment text classification method based on emotional vocabulary. *Digital Libraries: Advanced Methods and Technologies. Proceedings of the RCDL 2012*, 2012, pp. 118–123. (in Russian)
- Bessmertnyi I.A., Dzhaliashvili Z.O., Maksimov V.V., Markin D.A. Linguistic evaluation of the text. *Abstracts of the X International Conference “Application of New Technologies in Education”*, Troitsk, BYTIC, 1999. (in Russian)
- Harris Z. *Mathematical Structures of Language*. Wiley, 1968, 230 p.
- Bogdanov A.L., Dulya I.S. Sentiment analysis of short Russian texts in social media. *Tomsk State University Journal of Economics*, 2019, no. 47, pp. 220–241. (in Russian). <https://doi.org/10.17223/19988648/47/17>
- Riabykin A.S., Sukhov E.A. Neural network methods in the Sentiment Analysis problem. *DSPA*, 2022, vol. 2, no. 2, pp. 41–57. (in Russian)
- Vorobyov A.A., Rybak A.M., Seredkin R.A., Androsov A.Yu., Solovoyov B.I. Methods of collecting and processing sociological information from the Internet. *Izvestiya Tula State University. Technical Sciences*, 2022, no. 2, pp. 208–213. (in Russian). <https://doi.org/10.24412/2071-6168-2022-2-208-214>
- Fomina I.K., Tataurova A.S. Increasing the efficiency of the classification model for determining the text key. *Aktual'nye problemy ekonomiki i upravleniya*, 2022, no. 1, pp. 55–58. (in Russian)
- Zucco G., Azzopardi L., van Rijsbergen K. The quantum probability ranking principle for information retrieval. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5766, pp. 232–240. https://doi.org/10.1007/978-3-642-04417-5_21
- Melucci M. *Introduction to Information Retrieval and Quantum Mechanics*. Springer-Verlag Berlin Heidelberg, 2015, 247 p. <https://doi.org/10.1007/978-3-662-48313-8>
- Khrennikov A. *Introduction to Quantum Information Theory*. Moscow, Fizmatlit Publ., 2008, 283 p. (in Russian)
- Blacoe W., Kashefi E., Lapata M. A quantum-theoretic approach to distributional semantics. *Proc. of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, 2013, pp. 847–857.
- Jaiswal A.K., Holdack G., Frommholz I., Liu H. Quantum-like Generalization of Complex Word Embedding: a lightweight approach for textual classification. *CEUR Workshop Proceedings*, 2018, vol. 2191, pp. 159–168.
- Surov I.A., Semenenko E., Platonov A.V., Bessmertny I.A., Galofaro F., Toffano Z., Khrennikov A.Yu., Alodjants A.P. *Quantum semantics of text perception*. *Scientific Reports*, 2021, vol. 11, no. 1, pp. 4193. <https://doi.org/10.1038/s41598-021-83490-9>
- Gruzdeva A.S., Bessmertny I.A. Classification of short texts using a wave model. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 2, pp. 287–293. (in Russian). <https://doi.org/10.17586/2226-1494-2022-22-2-287-293>
- Frenkel I.A. *Wave Mechanics. Part I. Elementary Theory*. *Quantum physics*. Moscow, 2019, 386 p. (in Russian)
- Kutuzov A., Kuzmenko E. Webvectors: A toolkit for building web interfaces for vector semantic models. *Communications in Computer and Information Science*, 2017, vol. 661, pp. 155–161. https://doi.org/10.1007/978-3-319-52920-2_15
- Panchenko A., Ustalov D., Arefyev N., Paperno D., Konstantinova N., Loukachevitch N., Biemann C. Human and machine judgements about russian semantic relatedness. *Communications in Computer and Information Science*, 2016, vol. 661, pp. 221–235. https://doi.org/10.1007/978-3-319-52920-2_21
- Pang B., Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proc. of the 42nd*

- and Information Science. 2016. V. 661. P. 221–235. https://doi.org/10.1007/978-3-319-52920-2_21
21. Pang B., Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts // Proc. of the 42nd Annual Meeting Association for Computational Linguistics (ACL). 2004. P. 271–278. <https://doi.org/10.3115/1218955.1218990>
 22. Меньшиков И.Л. Анализ тональности текста на русском языке при помощи графовых моделей // Доклады всероссийской научной конференции АИСТ'2013 «Анализ Изображений, Сетей и Текстов». Екатеринбург, 2013. С. 151–155.
 23. González F.A., Caicedo J.C. Quantum latent semantic analysis // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2011. V. 6931. P. 52–63. https://doi.org/10.1007/978-3-642-23318-0_7
 24. Dalla Chiara M.L., Giuntini R., Luciani A.R., Negri E. A quantum-like semantic analysis of ambiguity in music // Soft Computing. 2017. V. 21. N 6. P. 1473–1481. <https://doi.org/10.1007/s00500-015-1895-y>
 - Annual Meeting Association for Computational Linguistics (ACL), 2004, pp. 271–278. <https://doi.org/10.3115/1218955.1218990>
 22. Menshikov I. Sentiment analysis of the text in Russian using graph models. Reports of the All-Russian Scientific Conference “Analysis of Images, Networks and Texts, AINT (AIST)”, 2013, pp. 151–155. (in Russian)
 23. González F.A., Caicedo J.C. Quantum latent semantic analysis. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2011, vol. 6931, pp. 52–63. https://doi.org/10.1007/978-3-642-23318-0_7
 24. Dalla Chiara M.L., Giuntini R., Luciani A.R., Negri E. A quantum-like semantic analysis of ambiguity in music. Soft Computing, 2017, vol. 21, no. 6, pp. 1473–1481. <https://doi.org/10.1007/s00500-015-1895-y>

Авторы

Груздева Анастасия Сергеевна — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57674037100](https://orcid.org/0000-0003-4963-0823), <https://orcid.org/0000-0003-4963-0823>, prog.anastasia@gmail.com

Юрьев Родион Николаевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 57485730300](https://orcid.org/0000-0003-1146-2617), <https://orcid.org/0000-0003-1146-2617>, rodion@juryev.ru

Бессмертный Игорь Александрович — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [sc 36661767800](https://orcid.org/0000-0001-6711-6399), <https://orcid.org/0000-0001-6711-6399>, bessmertny@itmo.ru

Authors

Anastasia S. Gruzdeva — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57674037100](https://orcid.org/0000-0003-4963-0823), <https://orcid.org/0000-0003-4963-0823>, prog.anastasia@gmail.com

Rodion N. Iurev — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 57485730300](https://orcid.org/0000-0003-1146-2617), <https://orcid.org/0000-0003-1146-2617>, rodion@juryev.ru

Igor A. Bessmertny — D. Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, [sc 36661767800](https://orcid.org/0000-0001-6711-6399), <https://orcid.org/0000-0001-6711-6399>, bessmertny@itmo.ru

Статья поступила в редакцию 27.05.2022

Одобрена после рецензирования 27.09.2022

Принята к печати 12.11.2022

Received 27.05.2022

Approved after reviewing 27.09.2022

Accepted 12.11.2022



Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»