

Раздел II. Моделирование процессов и систем

УДК 0004.912

DOI 10.18522/2311-3103-2022-5-106-116

Е. М. Герасименко, В.В. Стеценко

АНАЛИЗ ТОНАЛЬНОСТИ ТЕКСТОВЫХ ОТЗЫВОВ С ПРИМЕНЕНИЕМ ТОНАЛЬНЫХ СЛОВАРЕЙ И КАРДИНАЛЬНОСТИ НЕЧЕТКОГО МНОЖЕСТВА

Анализ тональности или мнений направлен на определение полярности мнения людей в отношении к какому-либо продукту, услуге, мероприятию или любому человеку. Один из самых распространенных методов, применяемый при анализе тональности текстового контента – обработка естественного языка. Тональность текста на естественном языке может быть оценена с помощью многочисленных методологий, таких как алгоритмы машинного обучения и статистические инструменты, в то время как применение нечеткой логики не является распространенным явлением. Использование нечеткой логики было выбрано по следующим причинам. Во-первых, нечеткая логика хорошо справляется с лингвистической неопределенностью. Такой способ определения проблемы приводит к уменьшению предвзятости как с положительной, так и с отрицательной стороны. Во-вторых, подходы к обучению на основе нечетких правил коренным образом отличаются от тех подходов к обучению, которые широко используются при классификации настроений, таких как метод опорных векторов, метод наивного Байеса и т. д., так как относятся к генеративному обучению, т. е. целью обучения является оценка степени принадлежности экземпляра к каждому отдельному классу. Предлагаемая модель для анализа тональности текстовых обзоров основана на использовании тональных словарей с применением нечеткой логики и состоит из четырех основных этапов. Этапы включают в себя токенизацию, формулировку модели мешка слов, формулировку нечеткой оценки тональности и присвоение полярности. В предложенной модели используется мощность нечеткого множества как мера оценки показателей полярности слов. Значения полярности слов получены путем применения двух тональных словарей: SentiWordNet и AFINN. Созданы две версии модели в зависимости от типа используемого словаря: на основе SentiWordNet и AFINN. Сравнение представленного подхода на основе нечеткой логики с другими методами на основе словарей демонстрирует превосходство разработанных моделей, основанных на применении нечеткой логики.

Анализ тональности; обработка естественного языка; нечеткая логика; оценка тональности; SentiWordNet; AFINN.

E.M. Gerasimenko, V.V. Stetsenko

SENTIMENT ANALYSIS OF TEXT REVIEWS USING TONE DICTIONARIES AND FUZZY SET CARDINALITY

Sentiment or opinion analysis aims to determine the polarity of people's opinions in relation to any product, service, event or any person. One of the most common methods used in sentiment analysis of text content is natural language processing. Sentiment analysis of natural language text can be assessed using numerous methodologies such as machine learning algorithms and statistical tools, while the application of fuzzy logic is not common. The use of fuzzy logic was chosen for the following reasons. First, fuzzy logic handles linguistic uncertainty well. This way of defining the problem leads to a reduction in bias, both positively and negatively. Secondly, learn-

ing approaches based on fuzzy rules are fundamentally different from those learning approaches that are widely used in sentiment classification, such as support vector machines, naive Bayes, etc., as they relate to generative learning, i.e. the goal of learning is to assess the degree to which an instance belongs to each individual class. The proposed model for sentiment analysis of text reviews is based on the use of tone lexicons using fuzzy logic and consists of four main stages. The steps include tokenization, word bag model formulation, sentiment fuzzy score formulation, and polarity assignment. In the proposed model, the power of the fuzzy set is used as a measure of the evaluation of the indicators of the polarity of words. Word polarity values are obtained by applying two sentiment lexicons: SentiWordNet and AFINN. Two versions of the model were created depending on the type of vocabulary used: based on SentiWordNet and AFINN. Comparison of the presented approach based on fuzzy logic with other dictionary-based methods demonstrates the superiority of the developed models based on the application of fuzzy logic.

Sentiment analysis; NLP; fuzzy logic; sentiment score; SentiWordNet; AFINN.

Введение. Анализ тональности – это интерпретация и изучение мнения, отношения и эмоции людей по отношению к какой-либо сущности. Сущность может быть представлена отдельными лицами, событиями или темами. Это способ анализа письменной или устной речи, чтобы определить, является ли сказанное благоприятным, неблагоприятным или нейтральным, и в какой степени. В настоящее время применение анализа тональности востребовано в таких областях как: маркетинг, электронная коммерция, фильмы [1], новости [2], политика [3–5], отели [6–8], рестораны, социальные сети [9] и т.д. Интернет стал неотъемлемой частью современной жизни. Цифровая вселенная на начало 2020 г., по оценкам, состояла из 44 зеттабайт данных. В 2019 г. Google обработал 3,7 миллионов запросов, в Facebook (организация, деятельность которой запрещена на территории РФ) был осуществлен миллион входов в систему, и YouTube зафиксировал 4,5 миллионов видео, просматриваемых каждые 60 секунд. Количество интернет-пользователей выросло с 3,7 миллиардов в 2018 году до 4,9 миллиардов в 2021 году. Объем производства данных высок, они создаются каждую минуту интернет-приложениями, такими как электронная почта, Google-приложения, WhatsApp, музыкальными приложениями и т.д., а также соцсетями. Это огромное количество данных может быть использовано в качестве входных данных для осуществления анализа тональности с целью прогнозирования, маркетинга, всевозможных исследований, интеллектуального анализа данных и многого другого. В маркетинге необходимо анализировать мышление клиентов. Что именно клиенты хотят от продукта или услуги может быть проанализировано путем проведения опроса или путем сбора отзывов о товаре или услуге [10]. Области применения анализа тональности показаны на рис. 1.



Рис. 1. Сферы применения анализа тональности

Интернет и социальные сети являются неиссякаемым источником информации, выраженной в форме отзывов (обзоров). Все эти отзывы (обзоры) могут быть использованы в качестве входных данных в процессе извлечения мнений, выходом же будет являться тональность: положительная либо отрицательная, если речь идет о бинарной классификации мнений. Общественные настроения по любому социальному вопросу также могут легко анализироваться с помощью применения методов анализа тональности. Основные этапы процесса тонального анализа включают в себя: сбор данных (ввод), подготовку текста (предварительная обработка), извлечение признаков, выбор признаков, обнаружение мнений, классификацию мнений и представление результатов. Анализ тональности можно проводить с помощью естественной обработки языка (NLP). Исследователи используют NLP для сбора информации и знаний о том, как люди понимают и используют язык, для разработки модели, инструментов и методов, применяемых при анализе тональности. Большая часть данных, используемых при анализе тональности, связана с реальными проблемами. Эти проблемы сложны, поэтому они требуют умной и интеллектуальной системы для интеграции знаний, методов и методик из разных источников. Предполагается, что данные системы будут обладать такой мощностью, что смогут имитировать процесс человеческого мышления и будут являться экспертами в определенной области; адаптироваться и учиться самосовершенствоваться в постоянно меняющейся среде. Для решения реальных проблем была создана концепция нечеткости. Нечеткое множество состоит из таких элементов, которые не имеют четких границ [11]. В данной работе нечеткая логика применяется к онлайн-обзорам, чтобы вычислить оценку настроений. Для вычисления оценки тональности слов применялись словари тональности SentiWordNet и AFINN. Данные словари представляют собой лексические ресурсы, содержащие список слов и их оценку полярности.

Обзор источников. Анализ тональности является одной из самых молодых и перспективных областей исследования. Самый простой способ определить настроение – классифицировать его как положительное либо же отрицательное. В [10] процесс анализа тональности классифицируется по пяти признакам. Прежде всего, он может быть разделен по уровням: уровень документа, уровень предложения и уровень аспекта. Второй способ – это полярность настроений. Существует несколько типов классификации: бинарный подход, многоуровневый подход и контекстный или нечеткий подход. В [12] пользователи мобильных телефонов высказывают свое мнение о мобильных телефонах на основе их функций. Есть множество функций, но только те характеристики, которые имеют решающее значение и необходимы для выполнения мобильным телефон своих функций, выбраны для определения настроений. В статье [13] обсуждались общие перспективы, связанные с анализом тональности, в основе которых лежит анализ текста и эмоций. Количество веб-сайтов и блогов с отзывами об отелях постоянно увеличивается. Данные обзоры с этих веб-сайтов и блогов помогают людям определиться с выбором при планировании своего следующего отпуска. В ряде исследований предлагались различные подходы к вычислению тональности отзывов и рейтингов [14] об отелях. В работе [15] были проанализированы реакция и отношение людей, выраженные в сети интернет во время естественного бедствия (наводнение в Керале) с использованием Наивного Байесовского классификатора. В [16–17] авторы исследовали влияние демонетизации на мировом и индийском финансовом рынках. Общественное мнение о демонетизации было собрано из твитов по всей стране.

Постановка задачи. В основе многих подходов к анализу тональности лежат такие классификаторы, такие как наивный Байес (Naïve Bayes), метод опорных векторов (SVM) и Максимальная энтропия. Но мало кто использует классификато-

ры на основе нечетких множеств. Важность нечеткости вступает в игру при работе с естественным языком из-за наличия двусмысленности в языке. Понятие нечетких множеств было сформулировано Лотфи Заде. Нечеткие множества могут быть применены к определению степени того насколько положительным или отрицательным является слово при помощи нечетких членств для оценки настроений. В [18] предложенная модель для определения тональности показывает, что не все положительные или отрицательные слова можно рассматривать как равные, поскольку некоторые слова имеют более положительную или отрицательную окраску по сравнению с другими словами. Концепция нечеткости помогает справляться с реальными проблемами. Анализ тональности был применен к обзорам продуктов, чтобы классифицировать их как положительные, отрицательные или нейтральные с помощью нечеткой модели. В [19] системы на основе нечетких правил при осуществлении анализа тональности показали лучшие результаты, чем обычно используемые Наивный байесовский алгоритм, деревья решений и SVM. Задача этих систем состоит в том, чтобы определить значение степени классификации полярности. Существуют и системы, основанные на нечеткой логике, для классификации настроений онлайн-обзоров с использованием систем нечеткого вывода на основе правил. Методы машинного обучения [20–21] и методы глубокого обучения наряду с нечетким подходом в текстовых обзорах показали хорошие результаты. В отдельных работах нечеткая энтропия и кластеризация k-средних были применены для выделения краткого списка важных слов из обзоров с высоким коэффициентом полярности. Нечеткие оценки этих слов, включенных в короткий список, вычислялись с использованием словаря SentiWordNet; далее полученные результаты отправлялись в нейронную сеть LSTM для классификации тональности. А в предыдущей работе авторов данной работы применялся метод обучения с учителем – SVM для набора данных, состоящего из видеобзоров, размещенных в социальных сетях. В этом методе применялось сочетание акустических и лингвистических особенностей для классификации настроений в мультимодальном анализе тональности.

В Интернете содержатся миллионы онлайн-обзоров на различные темы, события, продукты или услуги. Эти обзоры были проанализированы многочисленными исследователями с применением методов тонального анализа, но они не смогли получить достоверные результаты для неопределенных или неоднозначных данных, присутствующих в языке. Анализ тональности – это процесс определения и вычисления мнений, установок, чувств, выраженных людьми на естественном языке. Слова – это основной строительный блок языка. Каждый язык устный или письменный состоит из слов. Подход к анализу тональности на уровне слов в NLP является первым шагом к пониманию естественного языка. Компьютер должен понимать, как вещи работают в реальном мире, но эта попытка, хоть и очень прогрессивная, имеет ограничение. Существует пробел в интеллекте между человеком и машиной. Нечеткую логику можно использовать чтобы машина смогла восполнить этот пробел лучшим образом, потому что она имеет дело с неопределенностью, расплывчатостью и факторами неточности, присутствующими в языке.

Таким образом, это побудило нас включить нечеткую логику в модель анализа тональности, чтобы добиться лучших результатов при классификации. В данной статье мы применили подход на основе нечеткой логики для определения и классификации настроений в отзывах на уровне слов. Большинство существующих работ в области тонального анализа используют алгоритмы обучения с учителем. Мы применили подход без учителя, потому что в отличие от обучения с учителем, в процессе обучения без учителя нет необходимости в заранее размеченных данных, однако, длительность обучения и вычислительные сложности при таком подходе также присутствуют. В следующем разделе описывается предлагаемая методология.

Предлагаемая модель. В основе предлагаемой модели для анализа тональности текстовых обзоров лежит применение тональных словарей, но для расчета итогового результата мы применили кардинальность нечеткого множества, что, как показали результаты экспериментов, позволило повысить точность определения тональности. Модель состоит из четырех основных этапов. Этапы включают в себя токенизацию, формулировку модели мешка слов, формулировку нечеткой оценки тональности и присвоение полярности. Созданы две версии модели в зависимости от типа используемого словаря: на основе SentiWordNet и AFINN. Ниже приводится описание шагов.

1. Токенизация и лемматизация.

Отзывы содержатся в документе. Для работы с документами необходимо сначала разбить текст документа на отдельные предложения. Разделение абзацев на предложения называется токенизацией предложений. Токенизация предложения – это процесс разбиения предложения на список слов (токенов). Вывод процесса токенизации хранится в динамическом списке. Каждое слово предложения подвергается процессу лемматизации.

2. Мешок слов.

Термин «мешок слов» в анализе тональности относится к тем ключевым словам, которые важны при извлечении отзывов, мнений и т.д. Технически, это модель, которая преобразует документ в векторную (числовую) форму, где каждому слову в документе присваивается некоторая оценка в диапазоне от 0 до 1. Данный процесс можно интерпретировать как нечеткую принадлежность к нечетким множествам Pos и Neg. В предложенной модели использован также NLTK Part of Speech (POS) Tagger для извлечения слов, которые являются существительными, прилагательными, глаголами или наречиями. Нечеткое множество A можно представить в виде $A = \{(x, \mu_A(x)), x \in U\}$, где x — элемент из универсального множества и μ_A является принадлежностью элемента x . Нечеткие множества Pos и Neg представлены в виде:

$$Pos = \{(a, \mu_{Pos}(a)), a \in X_i\}. \quad (1)$$

$$Neg = \{(a, \mu_{Neg}(a)), a \in X_i\}, \quad (2)$$

где a – слово, X_i – i -й набор мешка слов. Если общее количество отзывов равно n ; тогда для каждого отзыва создается мешок слов. Переименуем функции множества μ_{Pos} и μ_{Neg} в μ_{swnPos} и μ_{swnNeg} для вычисления при использовании словаря тональности SentiWordNet и в μ_{afPos} и μ_{afNeg} при использовании словаря AFINN

$$\mu_{swnPos}(a) = \frac{\sum_{synsets} [syn.pos_score()]}{length(synsets)}. \quad (3)$$

$$\mu_{swnNeg}(a) = \frac{\sum_{synsets} [syn.neg_score()]}{length(synsets)}. \quad (4)$$

Уравнения (3) и (4) представляют собой нечеткие функции принадлежности слова для словаря SentiWordNet, где $syn.pos_score()$ и $syn.neg_score()$ – это баллы, полученные из SentiWordNet; $synsets$ – это набор синонимов каждого слова, присутствующего в SentiWordNet.

Второй вариант предполагает использование словаря AFINN. В уравнении (5) вычисляется оценка μ_{af} каждого слова с использованием AFINN и делится на пять, потому что оценка по AFINN находится в диапазоне между -5 и +5. Если полученное в результате число больше или равно нулю – это положительная оценка – μ_{afPos} , в противном случае – это отрицательная оценка – μ_{afNeg} .

$$\mu af(a) = \frac{af.score(a)}{5} \quad (5)$$

В уравнениях (6) и (7) представлены вычисления нечетких функций принадлежности слова для словаря AFINN.

$$if(\mu af(a) \geq 0) then(\mu afPos(a) = \mu af(a)). \quad (6)$$

$$if(\mu af(a) < 0) then(\mu afNeg(a) = -\mu af(a)). \quad (7)$$

3. Мощность (кардинальность) нечеткого множества.

Мы использовали мощность нечеткого множества для оценки тональности каждого отзыва. Нечеткие множества Pos и Neg содержат положительные и отрицательные оценки слова в отзыве. Положительная мощность множества Pos и отрицательная мощность множества Neg вычисляется путем суммирования всех элементов в соответствующих наборах. Эта мера показывает мощность нечеткого множества. Ниже приведены формулы для расчета положительной и отрицательной кардинальности:

$$Pos_cardinality = \sum_{j=1}^l \mu Pos(a), a \in X_i. \quad (8)$$

$$Neg_cardinality = \sum_{j=1}^l \mu Neg(a), a \in X_i. \quad (9)$$

где l – длина отзыва, a – слово, X – множество мешка слов, принадлежащее i – мешку слов.

4. Присвоение полярности.

Бинарная классификация полярности отзывов предполагает разделение их на «положительные» и «отрицательные». Функции принадлежности μPos и μNeg являются мощностями множества (уравнения (8) и (9)) и вычисляются для определения общей полярности каждого текстового обзора. Если количество положительных отзывов больше или равно количеству отрицательных, это означает, что отзыв положительный, в противном случае, это отрицательный отзыв.

$$Polarity(i) = \begin{cases} P, & Pos_Cardinality \geq Neg_Cardinality \\ N, & otherwise \end{cases} \quad (10)$$

где, $Polarity(i)$ – это полярность i -го обзора, а P и N – метки для положительной и отрицательной полярности.

Таким образом, с помощью приведенного выше сравнения все обзоры делятся на два класса: Положительные (P) и Отрицательные (N). Технологический процесс предлагаемого нечеткого подхода изображен на рис. 2.



Рис. 2. Технологический процесс предлагаемого подхода

Экспериментальные данные. Предлагаемая модель для определения тональности текстовых отзывов, основанная на тональных словарях с применением нечеткой логики была реализована на языке Python. Эксперименты проводились на трех наборах данных, которые содержат онлайн-обзоры пользователей. Два набора данных включают в себя отзывы на фильмы: набор данных v2.0 от Pang-Lee [22] и IMDB. Третий набор данных содержит обзоры отеля. Эти отзывы были предоставлены путешественниками, которые останавливались в этом отеле. Все наборы данных находятся в свободном доступе в сети Интернет. Первый набор данных содержит 2000 отзывов, набор данных с IMDB содержит 50 000 отзывов, а набор данных отзывов об отеле – 38 932. Распределение всех наборов данных в соответствии с каждым классом тональности представлен в табл. 1.

Таблица 1

Распределение отзывов по классам тональности

Набор данных	Положительные	Отрицательные	Всего
Pang-Lee Movie	1000	1000	2000
IMDB Movie	25000	25000	50000
Hotel Reviews	26521	12411	38932

В данной работе применены два тональных словаря: SentiWordNet и AFINN для вычисления полярности слов. В табл. 2 показаны положительные и отрицательные оценки некоторых слов, с применением словаря SentiWordNet. Положительные слова имеют положительную оценку выше, чем отрицательную оценку, и наоборот для отрицательных слов. Есть слова, которые имеют нулевую оценку – это нейтральные слова, не содержащие эмоций, например, «Hotel» и «Staff». «Nice», «Helpful», «Clean» и «Beautiful» – положительные слова, в то время как «Filthy», «Difficult», «Disgusting» и «Accident» - отрицательные слова.

Таблица 2

Оценки слов тонального словаря SentiWordNet

Слово	Pos оценка	Neg оценка
Hotel	0.0	0.0
Staff	0.0	0.0
Nice	0.15	0.0
Helpful	0.25	0.0
Clean	0.0278	0.0
Beautiful	0.3125	0.0
Filthy	0.0417	0.25
Difficult	0.0	0.3125
Disgusting	0.0625	0.3125
Accident	0.0	0.125

Таблица 3

Оценки слов тонального словаря AFINN

Слово	Оценка	Оценка/5	Pos оценка	Neg оценка
Hotel	0.0	0.0	0.0	-
Staff	0.0	0.0	0.0	-
Nice	3.0	0.6	0.6	-

Окончание табл. 3

Слово	Оценка	Оценка/5	Pos оценка	Neg оценка
Helpful	2.0	0.4	0.4	-
Clean	2.0	0.4	0.4	-
Beautiful	3.0	0.6	0.6	-
Filthy	-2.0	-0.4	-	0.4
Difficult	-1.0	-0.2	-	0.2
Disgusting	-3.0	-0.6	-	0.6
Accident	-2.0	-0.4	-	0.4

В табл. 3 представлены результаты вычислений для этих же слов с применением словаря AFINN. Слова «Hotel» и «Staff» нейтральные в обоих словарях. Поскольку оценка «Nice», «Helpful», «Clean» и «Beautiful» больше нуля, полярность положительная. В то время как слова «Filthy», «Difficult», «Disgusting» и «Accident» – отрицательные слова, потому что их показатели меньше нуля. Положительные и отрицательные слова имеют разные оценки в обоих словарях, например, слово «Nice» имеет 0.15 положительного балла и 0 отрицательных баллов в SentiWordNet, в то время как в AFINN оно имеет 3.0 балла. Следовательно, оба словаря можно интерпретировать по-разному для вычисления полярности и оценки слов.

Мы сравнили наши две версии нечеткого подхода с другими двумя подходами на основе словарей. Первое сравнение с методом, разработанным Кавальканти и др. [23], где оценка тональности каждого отзыва рассчитывалась путем суммирования оценок тональности каждого слова и деления итоговой суммы на длину каждого обзора l как показано в уравнении (11).

$$score(Xi) = \frac{\sum_{a \in Xi} [\mu Pos(a) - \mu Neg(a)]}{l}, a \in Xi. \quad (11)$$

Здесь оценка каждого слова – это разница между положительной и отрицательной оценкой каждого слова, рассчитанная с использованием словаря тональности SentiWordNet, как указано в уравнениях (3) и (4).

В подходе Кавальканти и др., баллы больше или равные нулю обозначают положительное оценку (P) и отрицательные баллы обозначают негативное оценку (N).

Второе сравнение с методом Гилберта и др. [24], где применялся словарь VADER. Этот метод вычисляет различные оценки для данного входного текстового предложения. Диапазон составного балла проверяется на разные классы полярности: положительный и отрицательный.

Интерпретация результатов. В данной статье был представлен подход к анализу тональности текстовых обзоров на основе тональных словарей с применением нечеткой логики. Реализовано две версии модели с нечеткой кардинальностью: на основе оценок, полученных с помощью тональных словарей SentiWordNet и AFINN. Эти версии сравниваются с двумя подходами, описанными в статьях Кавальканти и др. и Гилберт и др. Все эксперименты проводились на трех наборах данных, состоящих из онлайн-обзоров. Сравнение наших версии версии нечеткого множества с другими методами с точки зрения точности представлены в табл. 4.

Таблица 4

Сравнение двух версий представленного подхода с другими методами

Набор данных	Accuracy			
	Метод Кавальканти и др. на основе словаря SentiWordNet	Метод Гилберта и др. на основе словаря VADER	Нечеткий метод на основе словаря SentiWordNet	Нечеткий метод на основе словаря AFINN
Pang-Lee	54,8%	63%	63,5%	65,45%
IMDB	52,87%	69,43%	64,13%	70,06%
Hotel Reviews	64,54%	76,1%	72,74%	76,2%

Эксперименты с наборами данных фильмов Pang-Lee показывают, что наша версия на основе словаря AFINN и нечеткой кардинальности достигла наивысшей точности в 65,45%, за ней следует версия на основе SentiWordNet 63,5%; практически такую же точность в 63% показал подход Гилберта и др., а подход Кавальканти и др. имеет наименьшую точность – 54,8%. Для набора данных фильмов IMDB наша версия модели нечеткой кардинальности на основе AFINN достигла наивысшей точности в 70,06%, а подход Гилберта и др. достиг второй по величине точности в 69,43%. Версия на основе AFINN продемонстрировала также максимальную точность 76,2% в наборе данных отзывов об отелях, что сопоставимо с подходом Гилберта и др. Наша версия SentiWordNet с нечеткой кардинальностью показала более высокую точность по сравнению с подходом Кавальканти и др. во всех наборах данных. Из результатов в табл. 4 мы можем сделать вывод, что наш нечеткий подход, основанный на словаре AFINN, показал самую высокую точность во всех наборах данных, подход Кавальканти и др. показал самую низкую точность во всех наборах данных. Результаты подхода Гилберта и др. сопоставимы с нашими версиями нечеткой кардинальности на основе AFINN.

Заключение. В данной исследовательской работе представлена модель анализа тональности, основанная на нечеткой логике и тональных словарях, примененная к онлайн-обзорам для вычисления их тональности. Для вычисления оценки тональности слов использовались два тональных словаря – SentiWordNet и AFINN. Основные моменты:

- 1) предложен подход к определению тональности текстовых обзоров без учителя на основе тональных словарей с применением нечеткой логики,
- 2) предлагаемая модель использует нечеткую кардинальность в качестве меры для оценки показателей полярности слов,
- 3) модель имеет две версии: на основе словаря SentiWordNet и AFINN,
- 4) сравнение предложенной модели с применением нечеткой кардинальности по сравнению с обычными современными методами.

Предложенная в данной статье модель анализа тональности на основе тональных словарей с применением мощности нечеткого множества превосходит стандартные методы на основе словарей. Наш подход рассчитывает мощность средних положительных и отрицательных оценок каждого слова в каждом обзоре в виде нечетких оценок. Таким образом, кардинальность нечетких множеств дает лучшие результаты, чем простые средние баллы. Применение нечеткой логики с НЛП позволяет нам получить результаты, которые соответствуют человеческой

интерпретации для анализа настроений. Наш подход применим к любому текстовому набору данных, однако, ограничение применения предложенной модели заключается в том, что количество слов зависит от словаря и некоторые слова, которых нет в словаре, не смогут быть обработаны. В данной работе мы применили анализ тональности с использованием нечеткой логики на уровне слов, можно усовершенствовать модель, применив анализ тональности на уровне предложений.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Bhoir P., Kolte S. Sentiment analysis of movie reviews using lexicon approach, *IEEE International Conference on Computational Intelligence and Computing Research (ICICR)*, 2015, pp. 1-6.
2. Li X., Xie H., Chen L., Wang J., and Deng X. News impact on stock price return via sentiment analysis, *Knowledge-Base Systems*, 2014, Vol. 69, pp. 14- 23.
3. Bose R., Dey R.K., Roy S., Sarddar D. Analyzing Political Sentiment Using Twitter Data, *Information and Communication Technology for Intelligent Systems*. Springer, Singapore, 2019, pp. 427-436.
4. Dundar B., Ozdemir S., Akay D. Opinion mining and fuzzy quantification in hotel reviews, *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, 2016, pp. 1-4.
5. Park E., Kang J., Choi D., Han J. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews, *Current Issues in Tourism*, 2020, Vol. 23, No. 5, pp. 605-611.
6. Mostafa L. Machine Learning-Based Sentiment Analysis for Analyzing the Travelers Reviews on Egyptian Hotels, *Joint European-US Workshop on Applications of Invariance in Computer Vision*, Springer, Cham, 2020, pp. 405-413.
7. Vashishtha S., Susan S. Fuzzy logic based dynamic plotting of mood swings from tweets, *International Conference on Innovations in Bio-Inspired Computing and Applications*. Springer, Cham, 2018, pp. 129-139.
8. Shivaprasad T.K., Shetty J. Sentiment analysis of product reviews: a review, *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017, pp. 298-301.
9. Liu Y., Bi J.W., Fan Z. P. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, *Information Fusion*, 2017, Vol. 36, pp. 149-161.
10. Indhuja K., Reghu RPC. Fuzzy logic based sentiment analysis of product review documents, *2014 First International Conference on Computational Systems and Communications (ICCSC)*, 2014, pp. 18-22.
11. Zadeh L.A. Calculus of Fuzzy Restrictions. Fuzzy Sets and their Applications to cognitive and decision processes, 1975, pp. 1-40.
12. Singh W. Sentiment analysis of online mobile reviews, *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017, pp. 20-25.
13. Yadav P., Pandya D. SentiReview: Sentiment analysis based on text and emoticons, *International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2017, pp. 467-472.
14. Songpan W. The analysis and prediction of customer review rating using opinion mining, *15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2017, pp. 71-77.
15. Dudani A., Srividya V., Sneha B., Tripathy B. K. Sentiment Analysis on Kerala Floods, *International Conference on Innovative Computing and Communications*, Springer, Singapore, 2020, pp. 107-124.
16. Singh P., Sawhney R.S., Kahlon K.S. Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, 2018, Vol. 4, No. 3, pp. 124-129.
17. Kumar A., Singh J.P. Demonetization in India: Good or Bad in Context of Social Media. – 2019.
18. Jusoh S., Alfaware H. M. Applying fuzzy sets for opinion mining, *International Conference on Computer Applications Technology (ICCAT)*, 2013, pp. 1-5.

19. Alharbi J.R., Alhalabi W.S. Hybrid Approach for Sentiment Analysis of Twitter Posts Using a Dictionary-based Approach and Fuzzy Logic Methods: Study Case on Cloud Service Providers, *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2020, Vol. 16, No. 1, pp. 116-145.
20. Phan H.T., Nguyen N.T., Cuong T.V., Hwang D. A Method for Detecting and Analyzing the Sentiment of Tweets Containing Fuzzy Sentiment Phrases, *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 2019, pp. 1-6.
21. Phan H.T., Tran V.C., Nguyen N.T., Hwang D. Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model, *IEEE Access*, 2020, Vol. 8, pp. 14630-14641.
22. Pang B., Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, 2004, pp. 271.
23. Cavalcanti D. C., Ricardo B.C .Prudêncio, Shreyasee S. Pradhan, Jatin Y. Shah, Ricardo S. Pietrobon. Good to be bad? Distinguishing between positive and negative citations in scientific impact, *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, 2011, pp. 156-162.
24. Hutto C. J., Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text, *Eighth international AAAI conference on weblogs and social media*. – 2014.

Статью рекомендовала к опубликованию д.т.н., профессор Л.С. Лисицына.

Герасименко Евгения Михайловна – Южный федеральный университет; e-mail: egerasimenko@sfedu.ru; г. Таганрог, Россия; тел.: 88634371651; кафедра систем автоматизированного проектирования; к.т.н.; доцент.

Стеценко Валентина Витальевна – e-mail: vstecenko@sfedu.ru; кафедра систем автоматизированного проектирования; аспирант.

Gerasimenko Evgeniya Michailovna – Southern Federal University; e-mail: egerasimenko@sfedu.ru; Taganrog, Russia; phone +78634371651; the department of computer aided design; cand. of eng.sc.; associate professor.

Stetsenko Valentina Vitalievna – e-mail: vstecenko@sfedu.ru; the department of computer aided design, postgraduate.

УДК 519.712.2

DOI 10.18522/2311-3103-2022-5-116-128

А.Э. Саак, Л.А. Гладков, Н.В. Гладкова

ИНТЕГРИРОВАННАЯ МОДЕЛЬ РЕШЕНИЯ ЗАДАЧИ ДИСПЕТЧЕРИЗАЦИИ ЗАЯВОК*

Рассматривается задача диспетчеризации. Рассмотрена парадигма организации распределенных вычислений на основе Grid-компьютинга. Приведена классификация систем диспетчеризации задач. Описаны различные подходы к решению задачи диспетчеризации. Приведена модель задачи обслуживания заявок на основе принципов теории систем массового обслуживания. Сформулирована постановка задачи на основе Grid-диспетчеризации. Предложено понятие ресурсного прямоугольника. Определена среда диспетчеризации ресурсных прямоугольников. Предложена модель позволяющая формализовать заявку пользователя на обслуживание понятием ресурсного (незвклидова) прямоугольника. Вместо принципа оптимизации на основе машинного поиска лучшего распределения массива ресурсных прямоугольников, предложен эвристический принцип, что позволило снизить объем необходимых вычислений. Предложенный эвристический алгоритм диспетчеризации позволяет учитывать свойства массива и выполнять оценку качества решений. Построены модели среды спроса в фор-

* Исследование выполнено при поддержке РФФИ в рамках научного проекта № 20-01-00148.