

УДК 004.912

ТЕСТИРОВАНИЕ МЕТОДОВ АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА, ОСНОВАННЫХ НА СЛОВАРЯХ

Е.В. Тутубалина¹, В.В. Иванов¹, М.А. Загулова¹, Н.Р. Мингазов¹,
И.С. Алимова¹, В.А. Малых²

¹Высшая школа Информационных технологий и информационных систем
Казанского федерального университета

²Институт системного анализа РАН, г. Москва

elvtutubalina@kpfu.ru, vivanov@kpfu.ru, lolmariya@gmail.com,
icrotek547@gmail.com, alimovallseyar@gmail.com, alem.mipt@gmail.com

Аннотация

Технологии анализа тональности текста развиваются интенсивно, что обусловлено ростом объемов открытых источников, представляющих мнения пользователей интернета по различным вопросам. В статье описаны методы для анализа тональности текстов отзывов и коротких сообщений (твитов), приводятся результаты оценки их качества, которая производилась в рамках российского семинара SentiRuEval-2015.

Ключевые слова: извлечение информации, анализ тональности, классификация текстов, машинное обучение с учителем

1. ВВЕДЕНИЕ

Задача анализа мнений в документах или отзывах пользователей (opinion mining) является актуальным направлением исследований в области обработки естественного языка, активно развивающимся в настоящее время. Это связано с ростом объема открытых текстовых источников, представляющих данные о мнениях пользователей интернета по различным вопросам. Поскольку учет информации о мнениях потребителей продукта или услуги имеет ценность, как для других пользователей, так и для поставщиков и производителей данного продукта или услуги, задача автоматического извлечения такой информации (задача анализа тональности текста) является актуальной.

Различные постановки задачи анализа тональности (sentiment analysis) можно классифицировать по конечной цели, которая стоит перед системой обработки текста. Обычно задачу классификации текстов в целом по тональности отделяют от задачи определения тональных высказываний относительно аспектных терминов (далее аспектов), о которых высказывание было сделано (aspect-based sentiment analysis). Современные методы решения задач анализа тональности текста принято сравнивать друг с другом в рамках научных семинаров, таких, как SEMEVAL [1, 2], где любому участнику доступны одни и те же наборы данных для обучения методов и постановки задач. Для систем обработки русскоязычных текстов такие «соревнования» проводятся, начиная с 2011 года. Исторически методы решения задачи классификации текстов по их тональности рассматривались ранее в серии семинаров SentiRuEval (в рамках ROMIP [3]). В 2014–2015 годах перед участниками SentiRuEval ставились задачи анализа тональности отзывов относительно аспектов о ресторанах и автомобилях, а также классификации коротких сообщений (твитов) из социальной сети Twitter. Ниже приведен список задач анализа тональности и «звездочкой» отмечены те методы, решения которых представлены в данной статье. Подробное описание исходных данных и задач можно найти в работе [4]. Список задач анализа тональности отзывов, поставленных в рамках SentiRuEval-2015:

- А) Задача извлечения явных аспектов;
- В) Задача извлечения всех аспектов и тональных фактов;
- С) Задача определения тональности относительно явных аспектов (*);
- Д) Задача категоризации явных аспектов по аспектным категориям (*);
- Е) Задача определения тональности по каждой из аспектных категорий.

Кроме того, в рамках SentiRuEval решалась задача анализа тональности твитов о банках и телекоммуникационных компаниях. В следующем разделе дан краткий обзор состояния исследований в области анализа тональности. В разделах 3 и 4 приводятся описания методов и полученные результаты.

2. СОВРЕМЕННОЕ СОСТОЯНИЕ ИССЛЕДОВАНИЙ

Извлечение аспектных терминов (аспектов). Это широко известная задача в области анализа тональности, которая требует глубокого понимания каждого аспекта продукта. Существует несколько наиболее популярных методов, реша-

ющих задачу извлечения аспектов как бинарную задачу классификации [5], как задачу классификации последовательностей (sequential classification) [6–8], как задачу тематического моделирования или традиционную задачу кластеризации [9, 10]. Цель классификации – определить, являются ли термины, существительные и словосочетания искомым объектом (аспектом, относительно которого высказывается некоторое мнение). В [5] используются синтаксические шаблоны, соотнесённые с тональностью из словаря общего назначения, для определения наиболее частых словосочетаний с существительными. В [11] предлагается подход, основанный на правилах, используются деревья зависимостей для предложений. Эти подходы не совершенны и дают низкие результаты для аспектов, встречающихся в тексте редко, и для более сложных случаев извлечения аспектов. В работах 2014 года [7, 8] были предложены две модификации метода условных случайных полей (Conditional Random Fields, CRF) для классификации последовательностей в задаче извлечения аспектных терминов.

Анализ тональности аспектов. Большинство ранних подходов к классификации аспектов полагались на вручную созданные словари, содержащие списки позитивных и негативных слов. В работе [12] был предложен метод обучения без учителя, основанный на подсчёте поточечной взаимной информации между фразой и двумя словами-индикаторами для каждой фразы. Методы машинного обучения широко применялись для аспектно-ориентированной классификации [7, 13–15]. Могаддам и Эстер в [9] для извлечения аспектов и их тональности предложили модификацию вероятностных тематических методов, основанных на скрытом распределении Дирихле (Latent Dirichlet Allocation, LDA), используя синтаксические связи зависимостей между аспектами, позитивными и негативными словами в предложении. Однако тематические модели показывают худшие результаты классификации тональных высказываний относительно аспектных терминов по сравнению с классификатором на основе метода опорных векторов (Support Vector Machine, SVM) [16].

Определение категории аспекта. Автоматическая категоризация явных аспектов изучается в рамках задачи резюмирования аспектных терминов по тематическим категориям. Данная задача была исследована в работе [9] как часть задачи аспектного латентного анализа: выделением групп заранее известных ключевых терминов и предсказанием рейтинга каждой из групп. На SemEval-

2014 также была проделана работа по разбиению аспектных терминов из отзывов на группы для анализа тональности. Лучшие результаты по F-мере были достигнуты подходами, в которых использовались классификаторы SVM с моделью «мешок слов» и информации из коллекции неразмеченных отзывов [7, 17].

Анализ тональности коротких сообщений. Получение информации из коротких неформальных сообщений, таких, как твиты и смс, вызывает исследовательский интерес в области анализа тональности [7, 18, 19], отслеживания событий [20], извлечения проблем [21], обнаружения сарказма [22] и пр. Традиционный подход к классификации тональности основывался на присутствии негативных и позитивных слов или пиктограмм, изображающих эмоцию (далее – эмотиконов), служащих индикаторами позитивной либо негативной окраски [12, 23, 24]. В современных исследованиях применялись также смешанные подходы, где лексические ресурсы (словари тональности) комбинировали с техниками машинного обучения [7, 25, 26]. Было показано, что биграммы и униграммы модели «мешок слов» важны для машинного обучения и что признаки, характерные для твиттера (хэштеги, ретвиты, ссылки), способствуют улучшению результатов классификации [27]. В [7] описаны эксперименты, показывающие важность определения тональных слов относительно негативного контекста в предложении и необходимость использования нескольких словарей для более точной классификации.

Многие работы по анализу тональности включают в себя формирование лексических ресурсов с указанием тональности слов [28]. Словари создаются различными способами, от ручной разметки до полностью автоматизированных подходов. В [26] был вручную расширен существующий словарь тональных слов, эмотиконов и интернет-сленга. В [25] словарь сформировали автоматически из слов, встреченных по эмоциональным хэштегам, высчитав для каждого слова поточечную взаимную информацию со словами из обучающего множества. Учитывая тот факт, что данная система победила в международном семинаре SemEval, был использован аналогичный подход к генерации словарей.

Анализ тональности сообщений на русском языке изучен на данный момент слабее. В [29] описано первое открытое тестирование анализа тональности отзывов пользователей на русском языке. В нём заняли первые места методы

машинного обучения с учителем, использующие метод опорных векторов и словари, и системы, основанные на правилах, классифицирующие отзывы о фильмах, книгах, цифровых камерах. В работе [30] предлагается подход, основанный на специальных словарях и семантических фильтрах. Другие участники использовали созданные вручную словари эмотиконов для каждой из трёх предметных областей [15]. Эксперименты подтвердили, что методы машинного обучения показывают наилучшие результаты относительно традиционных методов, основанных на знаниях, для отзывов пользователей на русском языке. Было показано, что популярные методы машинного обучения не являются универсальными, поскольку каждый классификатор показал наилучшие результаты лишь в одной из предметных областей.

3. ОПИСАНИЕ МЕТОДОВ АНАЛИЗА ТОНАЛЬНОСТИ НА ОСНОВЕ СЛОВАРЕЙ

В этом разделе описан подход к двум заданиям аспектно-ориентированного анализа тональности пользовательских отзывов о ресторанах и автомобилях и к заданию классификации твитов пользователей о банках и телекоммуникационных компаниях. При помощи свободного программного обеспечения для анализа данных были применены техники машинного обучения (см. раздел 3.2), основанные на модели «мешка слов» и признаках, которые подробно описаны в разделе 3.4. Значительная часть признаков строится при помощи тональных словарей двух видов – созданных вручную и автоматически.

3.1. Предобработка текстов

Твиты пишутся неформальным языком и содержат много неформальной лексики, ошибок и специальных символов, поэтому они обрабатывались следующим образом. Все упоминания пользователей нормализуются до @username. Ссылки заменяются @link. Слова, которые используются в отрицательном контексте (например, *не понравилось*, *не круто*, *нет пререканий*), помечаются приписываемым к ним тегом neg_. Текстовые эмотиконы заменяются индикатором той эмоции, которую выражают (например, ':-)' заменится на happyEmoticon, 'o_o' – на surpriseEmoticon, ';-)' – на winkEmoticon). Морфологический анализатор MyStem приводит все слова в их начальную форму. Слова в текстах отзывов были также нормализованы при помощи MyStem.

3.2. Методы машинного обучения для анализа тональности

Задача (С) классификации заключалась в предсказании тональной полярности каждого аспекта из отзывов о продуктах (негативный, позитивный, две тональности, нейтральный). Был применён классификатор MaxEnt с параметрами по умолчанию, основанный на модели «мешка слов» и признаках, которые описаны в разделе 3.4. При заданном контексте аспектного термина для извлечения признаков создаются два типа биграмм слов: (1) контекстные биграммы из текста внутри контекстного интервала аспектного термина; (2) аспектные биграммы как комбинация аспектного термина и контекстного слова к нему из контекстного интервала. Контекстный интервал для аспектного термина w_i – последовательность $(w_{i-4}, \dots, w_{i+4})$.

Задача D заключалась в классифицировании аспекта в одну из установленных заранее категорий. Для ресторанов этими категориями были: еда, сервис, интерьер, цены, общее впечатление. Для автомобилей: маневренность, надёжность, безопасность, внешний вид, комфорт, цена, общее впечатление. Было выполнено обучение классификатора на основе машин опорных векторов (SVM) с последовательной минимальной оптимизацией (SMO). Для каждого аспектного термина был извлечен набор признаков из его контекстного интервала (2 слова до аспектного термина, 2 слова после). Были созданы словари категорий, основанные на оценке каждого слова w в обучающем множестве.

Задача анализа твитов состоит в определении, содержится ли в твите о телекоммуникационных/финансовых компаниях позитивная, негативная либо нейтральная тональность. Применен подход на основе машинного обучения, использующий модель «мешка слов» и вектор признаков, каждый из которых описан ниже. Были протестированы три различных алгоритма обучения (Naive Bayes, MaxEnt, SVM). Лучшие результаты показал SVM, с настройками по умолчанию библиотеки scikit-library на языке Python.

3.3. Два вида словарей тональности

Словари позитивной и негативной лексики строились двумя методами: автоматически и вручную. При создании словаря вручную были собраны отзывы

пользователей с сайта otzovik.com. Для точности в позитивный корпус вошли только тексты *Преимущества*, с оценкой 5, а в негативный корпус – только *Недостатки*, с оценкой 1 или 2. *Преимущества* и *Недостатки* – это те части отзывов, в которых пользователь пишет только то, что он оценил позитивно в продукте, либо, соответственно, только то, что не понравилось. Для каждой предметной области в каждом корпусе было выбрано определённое число глаголов, существительных, прилагательных и наречий, встречающихся максимально часто. Для предметной области телекоммуникаций и банков удалялись те слова, которые являются явно нейтральными для данной предметной области (например, *связь, услуга, платёж, скорость, сотрудник*). Корпусы были расширены формами присутствующих в них слов. В Таблице 1 собрана информация об объёме полученных словарей в сравнении с числом отзывов, для каждой предметной области.

Таблица 1. Информация о словарях, сгенерированных вручную

	Число собранных отзывов	Число позитивных слов, вошедших в словарь	Число негативных слов, вошедших в словарь
Рестораны	7526	741	362
Автомобили	4951	1576	741
Банки	3357	139	131
Телекоммуникации	1928	68	168

Для подсчёта значений словарных признаков для задач C и D использовались взвешенные оценки: каждое слово предложения взвешивается степенью своей удалённости от конкретного аспекта:

$$score(w) = \frac{sc(w)}{e^{|i-j|}}, \quad (1)$$

где i, j – позиции аспектного термина и слова w , $sc(w)$ – тональность слова w , равная 1 для позитивных слов и -1 для негативных, определяется по словарю. Словарь, генерируемый автоматически для твитов, использует обучающее множество размеченных твитов (предоставленное SentiRuEval-2015). Как и в работе

[25], подсчитывается тональный балл (*score*) для каждого слова (*w*) в этом обучающем множестве:

$$score(w) = PMI(w, pt) - PMI(w, nt), \quad (2)$$

$$PMI(w, pt) = \log_2 \frac{p(w, pt)}{p(w) * p(pt)}, \quad (3)$$

где *PMI* – поточечная взаимная информация, *pt* – число позитивных твитов, *nt* – число негативных твитов, *p(w)*, *p(pt)* и *p(w, pt)* – вероятности появления *w* в позитивном корпусе. Тональность слова максимально наглядна: слово «сомнительный» имеет значения -19.68 и 0.28, тогда как «спустя» – только -0.80 и 0.14. Поскольку в твитах встречаются слова, бесполезные с точки зрения задачи, редко встречающиеся слова игнорировались, если они встречались в обучающем множестве менее трёх раз. Автоматический словарь категорий аспектов работает по следующей формуле:

$$score(w) = PMI(w, cat) - PMI(w, oth), \quad (4)$$

$$PMI(w, cat) = \log_2 \frac{p(w, cat)}{p(w) * p(cat)}, \quad (5)$$

где *PMI* – поточечная взаимная информация, *cat* – все контексты аспекта в конкретной категории, *oth* – контексты аспекта во всех прочих категориях, *p(w)*, *p(cat)* и *p(w, cat)* – вероятности появления слова *w* в контексте аспекта конкретной категории.

3.4. Признаковое пространство

Каждый отзыв преобразуется в вектор признаков, которые зависят от аспекта и его контекста в предложении. Каждый твит – в вектор лексических, словарных и характерных для твиттера признаков. Кратко опишем признаки, использованные во всех задачах:

- **n-граммы слов:** униграммы (отдельные слова) и биграммы (словосочетания) из текста твита, если встречаются в множестве твитов/отзывов более двух раз;
 - **n-граммы символов:** $n=2, \dots, 4$ строчных букв, встречающихся более двух раз;
 - **признаки словаря, построенного вручную:** количество совпавших позитивных слов; негативных слов;
 - **признаки автоматического словаря:** количество слов с со значением *score*,
-

большим нуля; количество слов с score, меньшим нуля; максимальный тональный score; минимальный тональный score; сумма тональных score; сумма позитивных score; сумма негативных; если с тональным словом в тексте употреблено отрицание, то его тональность меняется на противоположную.

При анализе отзывов дополнительно использовались следующие признаки:

– **контекстные n-граммы:** униграммы и биграммы из контекстного интервала; извлекаются следующими несколькими комбинациями: замена аспектного термина словом *aspect*; замена тональных слов на их тэг тональности (*pos* или *neg*), на часть речи слов.

– **аспектные биграммы:** биграммы как комбинации самого аспектного термина и слова из контекстного интервала; извлекаются теми же комбинациями, что описаны выше.

Поскольку размеры контекстного интервала ограничены и сложно классифицировать аспект одновременно как позитивный и негативный, для этих случаев было сформулировано своё правило: для предложений, содержащих аспектный термин и союзы *а* или *но*, если классификатор размечает аспект как нейтральный, его оценка меняется на *both* (двойственная тональность). При анализе твитов дополнительно использовались следующие признаки:

– **слова прописными буквами:** число в твите слов, набранных полностью прописными буквами;

– **пунктуация:** число подряд идущих вопросительных знаков, восклицательных знаков либо их комбинаций. Учитываются последовательности из более, чем одного знака. «Почему у дебетовой карты списали деньги просто так?!?» Для подобного твита, с нейтральным выбором слов, признак очень полезен, так как указывает на негатив;

– **последний символ:** является ли последний символ в твите скобкой или восклицательным знаком. *№Сеть прыгает из Е в 3G и обратно каждые 5 минут* (“ Встречая такой твит, классификатор не знает, хорошо ли это – прыгать в 3G, но скобки помогают определить, что твит негативный;

– **эмодиконы:** выделяются четыре признака – число позитивных эмодиконов, число негативных эмодиконов, является ли последний униграмм твита позитив-

ным эмодиконом, является ли негативным;

– **признаки, специфичные для твиттера**: три бинарных значения – содержит ли твит упоминания юзера, является ли ретвитом, содержит ли твит ссылку.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Данный раздел описывает результаты, достигнутые классификаторами твитов, аспектов и категорий аспектов. Каждая из таблиц содержит также *базовые* результаты (baseline) и результаты победителей дорожки SentiRuEval-2015, определённых по макро F-мере, которая выбрана основным критерием соревнования [4]. Раздел содержит эксперименты с удалением различных признаков классификаторов с целью определения значимости каждой группы в рамках задач анализа мнений.

4.1. Результаты и эксперименты в рамках задачи анализа коротких сообщений пользователей

В Таблице 2 приведены результаты экспериментов, проведенных в рамках SentiRuEval-2015, по задаче классификации определения тональности относительно компаний в твитах о банках и телекоммуникациях. Макро F-мера считалась как среднее между F-мерой для позитивного класса и F-мерой негативного класса, нейтральный класс исключается. Предложенные методы показали 4 и 2 результаты среди 10 и 7 команд в «дорожке» по банкам и телекоммуникациям соответственно.

Таблица 2. Метрики качества в задаче классификации твитов в банковской сфере и твитов о телекоммуникационных компаниях

Участник	Телекоммуникации		Банки	
	Micro F,%	Macro F,%	Micro F,%	Macro F,%
Лучший	53.6	48.8	34.3	35.9
Предложенный подход	52.8	44.8	33.7	35.2
Офиц. baseline	33.7	18.2	23.8	12.7

Результаты экспериментов по классификации твитов о компаниях описаны

в Таблице 3.

Таблица 3. Результаты экспериментов при удалении различных признаков

Набор признаков	Телекоммуникации			Банки		
	macro P, %	macro R, %	macro F, %	macro P, %	macro R, %	macro F, %
Полный вектор	44.3	47.1	44.7	53.8	27.9	35.2
Без n-грамм слов	39.0	41.2	37.3	50.7	31.6	37.3
Без n-грамм символов	44.7	41.3	40.5	44.4	23.3	30.1
Без пунктуации	42.9	42.9	41.2	52.2	28.6	35.0
Без прописных	44.6	44.7	43.6	49.8	29.3	34.9
Без эмотиконов	41.3	45.0	40.6	48.9	27.4	33.5
Без последнего символа	45.8	37.9	39.0	50.9	27.4	34.0
Без твиттер-признаков	44.7	44.1	44.3	49.1	28.9	35.1
Без словаря, составленного вручную	37.9	50.5	43.2	51.6	27.0	34.0
Без словаря, составленного автоматически	42.7	56.9	48.8	42.6	29.2	34.3
Без обоих словарей	41.9	55.3	47.5	49.6	27.6	33.7

Самыми полезными признаками с точки зрения решения задачи оказались n-граммы символов, словари и эмотиконы. Метод показывает улучшение в 0.021% F-меры после удаления n-грамм слов в твитах о банках, и улучшение в 0.041% F-меры после удаления слов автоматического лексикона в области телекоммуникаций. Объяснением этому может служить динамичность контекста сообщений о банках, поскольку твиты обучающего множества датированы 2014 годом, а твиты для тестового множества собирались в 2013 году. Вследствие этого многие положительные и негативные слова из обучающей выборки либо получили противоположную тональность, либо не были найдены в тестовом корпусе.

После проведения подробного анализа текстов ошибочно классифицированных твитов и того, в какие классы они были ошибочно занесены, были определены следующие типы наиболее частых ошибок классификации:

- орфографические ошибки и транслитерация;
- хэштеги, склеивающие несколько слов;
- эмоциональное обсуждение нейтральных тем;
- недостаточный размер тональных словарей (не замечены ярко-окрашенные тональные слова, незнакомые классификатору).

Таблица 4. Распределение типов ошибок

Предметная область	Орфография и транслитерация	Хэштеги из нескольких слов	Эмоции на нейтральные темы	Слова не из тонального словаря
Телекоммуникации	20.4%	8%	14.9%	43%
Банки	9%	1%	11%	64%

Таблица 4 демонстрирует, что большая часть ошибок возникла по причине недостаточного покрытия эмоционально-окрашенных слов соответствующим словарём тональности. В твите *«Билайну труба короче»* слово «труба» содержит негативное значение, неизвестное классификатору. Негативные твиты, такие, как *«Самый безалаберный банк!»*, классифицированы неверно, поскольку такие негативные слова, как «безалаберный», редко употребляются в речи и ни разу не появились в обучающем множестве.

Однако меры для обработки орфографических ошибок не были приняты, и поэтому слова *ацтой* (отстой) и *чорд* (чёрт) не были узнаны классификатором, хотя их грамотные версии содержатся в словаре. Транслитерированный твит *«Билайн. Дисконнектинг пипл.»* содержит слова, имеющие яркую негативную окраску на английском языке, но для русскоязычного словаря сбор таких лексических единиц представляет сложную задачу. Заметим, что орфографические ошибки способствовали меньшему числу неверных классификаций, чем слова удлинённые (*«ненавижуууу»*, *«ураа»*), слова транслитерированные и те ненормативные, в которых часть букв заменялась автором твита на символ «*» или другие специальные символы.

Хэштеги *#отстойсвязь*, *#мтсумри*, *#люблюего* содержат тональную окраску, которая остаётся нераспознанной. Около 8% ошибочных классификаций

твитов о телекоммуникациях могло быть исключено путём разделения хэштегов на отдельные слова.

Четвёртый тип ошибок – твиты о событиях, нейтральных для репутации компании, которые, однако, написаны крайне эмоционально (это могут быть твиты о дресс-коде компании, флирте с работником компании, исповедь о потерянной карточке, дружеская беседа). *Похожая ситуация с твитами о празднике или мероприятии, проводимом в офисе компании. «Матч штаб-квартиры Вымпелком – Сибирь. Пока ведем!!! :)»* Во всех приведённых случаях инструкция предписывает размечать твиты как нейтральные. Это стоило классификатору 14.9% ошибок в области телекоммуникаций, 11% – в предметной области банков.

4.2. Результаты и эксперименты в рамках задачи мнений пользователей

В таблицах 5а и 5б приведены результаты экспериментов по задаче определения тональности относительно аспектов. В рамках данной задачи макро F-мера высчитывается как среднее между F-мерами позитивного класса, негативного класса и класса двойственной тональности (без нейтрального класса). В таблицах 5а и 5б классификаторы на основе обучения с учителем показывают вторые результаты по значению макро F-меры со значительными улучшениями классификации относительно baseline методов. Классификатор на основе метода максимальной энтропии показал улучшение над базовыми результатами 14.1% и 13.5% по макро F-мере в рамках задачи для ресторанов и автомобилей соответственно.

Таблица 5а. Метрики качества в задаче классификации аспектов (С), отзывы о ресторанах

Участник	Micro P,%	Micro R,%	Micro F,%	Macro P,%	Macro R,%	Macro F,%
Официальный baseline	71.04	71.04	71.04	32.09	25.06	26.71
1_1	61.94	61.94	61.94	25.17	24.54	23.79
1_2	61.94	61.94	61.94	25.17	24.54	23.79
3_1	66.96	66.96	66.96	32.23	24.30	26.96
4_1	82.49	82.49	82.49	58.72	55.69	55.45
Предложенный подход	76.71	76.71	76.71	45.82	37.29	40.81

Таблица 5б. Метрики качества в задаче классификации аспектов (С), отзывы об автомобилях

Участник	Micro P,%	Micro R,%	Micro F,%	Macro P,%	Macro R,%	Macro F,%
Офиц. baseline	61.92	61.92	61.92	29.49	26.85	26.48
1_1	64.71	64.71	64.71	33.99	31.94	32.93
1_2	65.31	65.31	65.31	35.63	32.97	34.22
3_1	55.89	55.89	55.89	30.16	26.21	27.94
4_1	74.28	74.28	74.28	57.25	56.67	56.84
1_3	62.52	62.52	62.52	35.07	32.62	33.45
Предложенный подход	71.11	71.11	71.11	44.81	37.61	40.01

В Таблице 6 приведены результаты предложенного в статье метода, результаты метода, занявшего второе место по макро F-мере, а также официальные базовые результаты в рамках задачи категоризации явных аспектов по аспектным категориям. Предложенный метод показал наилучшие результаты среди 4 систем в обеих предметных областях. Лучший подход даёт улучшения над базовыми результатами 6,57% и 8,85% по макро F-мере для ресторанов и автомобилей, соответственно.

Таблица 6. Метрики качества в задаче категоризации аспектов (D)

Участник	Рестораны			Автомобили		
	Macro P,%	Macro R,%	Macro F,%	Macro P,%	Macro R,%	Macro F,%
Предложенный подход	89.60	84.14	86.53	68.54	63.55	65.21
Второе место	86.27	79.63	81.10	71.46	57.50	60.77
Офиц. baseline	87.42	77.37	79.96	66.72	51.90	56.36

Результаты классификации тональных высказываний относительно аспектов на основе различных наборов признаков представлены в Таблице 7 по следующей схеме: каждый эксперимент по классификации проведен на всем наборе признаков, за исключением признака, указанного в строке таблицы. Самым эффективным признаком показали себя аспектные биграммы, которые являются комбинацией аспектного термина и слова из контекстного интервала.

Таблица 7. Результаты экспериментов с признаками определения тональности относительно явных аспектов (C)

Набор признаков	Рестораны			Автомобили		
	Macro P,%	Macro R,%	Macro F,%	Macro P,%	Macro R,%	Macro F,%
Все признаки	45.82	37.29	40.81	44.81	37.61	40.01
Без n-грамм символов	44.79	36.59	40.00	44.80	37.50	39.94
Без словарных униграмм	42.59	36.51	39.21	42.13	36.69	38.69
Без аспектных биграмм	42.61	33.96	37.28	43.80	37.46	39.51
Без контекстных n-грамм	43.55	35.86	39.06	43.70	37.17	39.41
Без словарных score	46.29	36.81	40.50	43.74	37.47	39.59

Эксперименты в рамках задачи категоризации явных аспектов по аспектным категориям описаны в Таблице 8. Как показывают результаты, самыми важными признаками являются признаки, основанные на поточечной взаимной информации для категорий и включающие в себя максимальную и минимальную оценки, среднее значение оценок и сумму оценок контекста для аспектного термина.

Таблица 8. Результаты экспериментов с признаками для категоризации аспектов (D)

Набор признаков	Рестораны			Автомобили		
	P,%	R,%	F,%	P,%	R,%	F,%
n-граммы слов	76.50	71.93	73.88	65.54	60.60	62.19
n-граммы слов + единая кумулятивная оценка	81.85	77.05	79.14	68.00	62.96	64.61
n-граммы слов + все оценки	89.60	84.14	86.53	68.54	63.55	65.21

ЗАКЛЮЧЕНИЕ

В статье описаны методы анализа тональности текстов отзывов и коротких сообщений (твитов), приведены результаты оценки их качества, которая производилась в рамках российского семинара SentiRuEval-2015. Предложенные классификаторы на основе метода опорных векторов показали четвертый и второй результаты среди 10 и 7 систем в «дорожке» по задаче классификации твитов о банках и телекоммуникациях соответственно. В рамках задачи анализа мнений пользователей о ресторанах и машинах в статье описаны два метода машинного обучения: (i) метод определения тональности о конкретном объекте (аспекте), упоминание которого содержится в отзыве пользователя; (ii) метод категоризации аспектов по категориям. Классификатор на основе метода максимальной энтропии показал улучшение над базовыми результатами 14.1% и 13.5% по макро F-мере в рамках задачи определения тональности для ресторанов и автомобилей соответственно. Классификатор на основе метода опорных векторов в рамках задачи категоризации показал наилучшие результаты среди других систем, участвующих в «дорожке».

Результаты экспериментов по анализу мнений относительного значимости признаков в задаче классификации аспектов показали, что наиболее эффективными признаками являются аспектные биграммы, которые являются комбинацией аспектного термина и слова из контекстного интервала. В классификации твитов лучшие результаты показывают классификаторы, использующие следующие признаки: n-граммы символов, словари и набор символов, изображающие эмоцию.

В статье приведены анализ и классификация ошибок классификации твитов, среди которых наиболее частотными являются орфографические ошибки, транслитерация, наличие хэштегов, а также эмоциональное обсуждение нейтральных тем в твиттере.

СПИСОК ЛИТЕРАТУРЫ

1. *Nakov P., Kozareva Z., Ritter A., Rosenthal S., Stoyanov V., Wilson T.* Semeval-2013 Task 2: sentiment analysis in Twitter // Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval-2013). 2013. P. 312-320.

2. *Rosenthal S., Ritter A., Nakov P., Stoyanov V.* SemEval-2014 Task 9: sentiment analysis in Twitter // Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014. P. 73-80.

3. *Chetviorkin I.I., Braslavski P.I., Loukachevitch N.V.* Sentiment analysis track at ROMIP 2011 // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог». 2011. С. 739-746.

4. *Loukachevitch N.V., Blinov P.D., Kotelnikov E.V., Rubtsova Yu.V., Ivanov V.V., Tutubalina E.* SentiRuEval: testing object-oriented sentiment analysis systems in Russian // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». 2015. Вып. 14 (21). С. 2-13.

5. *Popescu A.M., Etzioni O.* Extracting product features and opinions from reviews // Natural language processing and text mining. ACL, 2007. P. 9-28.

6. *Jakob N., Gurevych I.* Extracting opinion targets in a single-and crossdomain setting with conditional random fields // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. ACL, 2010. P. 1035-1045.

7. *Kiritchenko S., Zhu X., Cherry C., Mohammad S.M.* NRC-Canada-2014: Detecting aspects and sentiment in customer reviews // SemEval 2014. NAACL, 2014. P. 437-442.

8. *Chernyshevich M.* IHS R&D Belarus: cross-domain Extraction of product features using conditional random fields // SemEval 2014. Dublin, 2014. P. 309-313.

9. *Moghaddam S., Ester M.* On the design of LDA models for aspect-based opinion mining // Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, 2012. P. 803-812.

10. *Zhao Y., Qin B., Liu T.* Clustering product aspects using two effective aspect relations for opinion mining // Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer International Publishing, 2014. P. 120-130.

11. *Poria S., Cambria E., Ku L. W., Gui C., Gelbukh A.* A rule-based approach to aspect extraction from product reviews // Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP). 2014. P. 28-37.

12. *Turney P.D.* Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews // Proceedings of the 40th Annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002. P. 417-424.

13. *Pang B., Lee L., Vaithyanathan S.* Thumbs up?: sentiment classification using machine learning techniques // Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. – Association for Computational Linguistics, 2002. P. 79-86.

14. *Pang B., Lee L.* Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. V. 2, No 1-2. P. 1-135.

15. *Blinov P., Klekovkina M., Kotelnikov E., Pestov O.* Research of lexical approach and learning methods for sentiment analysis // Computational Linguistics and Intellectual Technologies. 2013. No 2 (12). P. 48-58.

16. *Lu B., Ott M., Cardie C., Tsou B.K.* Multi-aspect sentiment analysis with topic models // Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference. IEEE, 2011. P. 81-88.

17. *Pontiki M., Papageorgiou H., Galanis D., Androutsopoulos I., Pavlopoulos J., Manandhar S.* Semeval-2014 task 4: aspect based sentiment analysis // Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014. P. 27-35.

18. *Go A., Bhayani R., Huang L.* Twitter sentiment classification using distant supervision // CS224N Project Report, Stanford. 2009. V. 1. P. 12.

19. *Sidorov G., Miranda-Jimenez S., Viveros-Jimenez F., Gelbukh A., Castro-Sanchez N., Velasquez F., Gordon J.* Empirical study of machine learning based approach for opinion mining in tweets // *Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2013. P. 1-14.

20. *Sakaki T., Okazaki M., Matsuo Y.* Earthquake shakes Twitter users: real-time event detection by social sensors // *Proceedings of the 19th International Conference on World Wide Web*. ACM, 2010. P. 851-860.

21. *Gupta N.K.* Extracting phrases describing problems with products and services from twitter messages // *Computación y Sistemas*. 2013. V. 17, No 2. P. 197-206.

22. *Davidov D., Tsur O., Rappoport A.* Semi-supervised recognition of sarcastic sentences in twitter and amazon // *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2010. P. 107-116.

23. *Taboada M., Brooke J., Tofiloski M., Voll K., Stede M.* Lexicon-based methods for sentiment analysis // *Computational Linguistics*. 2011. V. 37, No 2. P. 267-307.

24. *O'Connor B., Balasubramanyan R., Routledge B.R., Smith N.A.* From tweets to polls: linking text sentiment to public opinion time series // *ICWSM*. 2010. V. 11. P. 122-129.

25. *Mohammad S.M., Kiritchenko S., Zhu X.* NRC-Canada: building the state-of-the-art in sentiment analysis of tweets // *Second Joint Conference on Lexical and Computational Semantics (* SEM)*. 2013. V. 2. P. 321-327.

26. *Evert S., Proisl T., Greiner P., Kabashi B.* SentiKLUE: updating a polarity classifier in 48 hours // *SemEval 2014*. 2014. P. 551.

27. *Barbosa L., Feng J.* Robust sentiment detection on twitter from biased and noisy data // *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010. P. 36-44.

28. *Martinez-Camara E., Martin-Valdivia M.T., Urena-Lopez L.A., Montejo-Raez A.R.* Sentiment analysis in twitter // *Natural Language Engineering*. 2014. V. 20, No 01. P. 1-28.

29. *Chetviorkin I., Loukachevitch N.* Evaluating sentiment analysis systems in Russian // *ACL 2013*. 2013. P. 12-17.

30. Frolov A.V., Polyakov P.Yu., Pleshko V.V. Using semantic filters in application to book reviews sentiment analysis // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог». 2013. No 12 (19).

SENTIMENT CLASSIFICATION OF REVIEWS AND TWITTER POSTS BASED ON DICTIONARIES

E.V. Tutubalina¹, V.V. Ivanov¹, M.A. Zagulova¹, N.R. Mingazov¹,
I.S. Alimova¹, V.A. Malykh²

elvtutubalina@kpfu.ru, vivanov@kpfu.ru, lolmariya@gmail.com,
nicrotek547@gmail.com, alimovallseyar@gmail.com, alem.mipt@gmail.com

¹ High School of Information Technology and Information Systems
of Kazan Federal University

² Institute for Systems Analysis of Russian Academy of Sciences

Abstract

Sentiment analysis and opinion mining technologies are growing fast. This is mostly due to a rapid grow of the data sources consisting a vast amount of user opinions and reviews on a wide set of topics. In this paper we describe methods for sentiment analysis of reviews and short messages (tweets), as well as evaluation of results obtained during SentiRuEval-2015.

Keywords: *information extraction, sentiment analysis, text classification, supervised learning*

REFERENCES

1. Nakov P., Kozareva Z., Ritter A., Rosenthal S., Stoyanov V., Wilson T. Semeval-2013 Task 2: sentiment analysis in Twitter // Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval-2013). 2013. P. 312-320.

2. Rosenthal S., Ritter A., Nakov P., Stoyanov V. SemEval-2014 Task 9: sentiment analysis in Twitter // Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014. P. 73-80.

3. *Chetviorkin I.I., Braslavski P.I., Loukachevitch N.V.* Sentiment analysis track at ROMIP 2011 // *Kompyuternaya lingvistika i intellektualnyie tehnologii: po materialam ezhegodnoy Mezhdunarodnoy konferentsii «Dialog»*. 2011. S. 739-746.

4. *Loukachevitch N.V., Blinov P.D., Kotelnikov E.V., Rubtsova Yu.V., Ivanov V.V., Tutubalina E.* SentiRuEval: testing object-oriented sentiment analysis systems in russian // *Kompyuternaya lingvistika i intellektualnyie tehnologii: Po materialam ezhegodnoy Mezhdunarodnoy konferentsii «Dialog»*. 2015. Vyip. 14 (21). S. 2-13.

5. *Popescu A.M., Etzioni O.* Extracting product features and opinions from reviews // *Natural language processing and text mining. ACL, 2007*. P. 9-28.

6. *Jakob N., Gurevych I.* Extracting opinion targets in a single-and crossdomain setting with conditional random fields // *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. ACL, 2010*. P. 1035-1045.

7. *Kiritchenko S., Zhu X., Cherry C., Mohammad S.M.* NRC-Canada-2014: Detecting aspects and sentiment in customer reviews // *SemEval 2014. NAACL, 2014*. P. 437-442.

8. *Chernyshevich M.* IHS R&D Belarus: cross-domain Extraction of product features using conditional random fields // *SemEval 2014. Dublin, 2014*. P. 309-313.

9. *Moghaddam S., Ester M.* On the design of LDA models for aspect-based opinion mining // *Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, 2012*. P. 803-812.

10. *Zhao Y., Qin B., Liu T.* Clustering product aspects using two effective aspect relations for opinion mining // *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Springer International Publishing, 2014*. P. 120-130.

11. *Poria S., Cambria E., Ku L. W., Gui C., Gelbukh A.* A rule-based approach to aspect extraction from product reviews // *Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP). 2014*. P. 28-37.

12. *Turney P.D.* Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews // *Proceedings of the 40th Annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002*. P. 417-424.

13. *Pang B., Lee L., Vaithyanathan S.* Thumbs up?: sentiment classification using machine learning techniques // *Proceedings of the ACL-02 Conference on Empiri-*

cal Methods in Natural Language Processing-Volume 10. – Association for Computational Linguistics, 2002. P. 79-86.

14. *Pang B., Lee L.* Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. V. 2, No 1-2. P. 1-135.

15. *Blinov P., Klekovkina M., Kotelnikov E., Pestov O.* Research of lexical approach and learning methods for sentiment analysis // Computational Linguistics and Intellectual Technologies. 2013. No 2 (12). P. 48-58.

16. *Lu B., Ott M., Cardie C., Tsou B.K.* Multi-aspect sentiment analysis with topic models // Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference. IEEE, 2011. P. 81-88.

17. *Pontiki M., Papageorgiou H., Galanis D., Androutsopoulos I., Pavlopoulos J., Manandhar S.* Semeval-2014 task 4: aspect based sentiment analysis // Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014. P. 27-35.

18. *Go A., Bhayani R., Huang L.* Twitter sentiment classification using distant supervision // CS224N Project Report, Stanford. 2009. V. 1. P. 12.

19. *Sidorov G., Miranda-Jimenez S., Viveros-Jimenez F., Gelbukh A., Castro-Sanchez N., Velasquez F., Gordon J.* Empirical study of machine learning based approach for opinion mining in tweets // Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2013. P. 1-14.

20. *Sakaki T., Okazaki M., Matsuo Y.* Earthquake shakes Twitter users: real-time event detection by social sensors // Proceedings of the 19th International Conference on World Wide Web. ACM, 2010. P. 851-860.

21. *Gupta N.K.* Extracting phrases describing problems with products and services from twitter messages // Computación y Sistemas. 2013. V. 17, No 2. P. 197-206.

22. *Davidov D., Tsur O., Rappoport A.* Semi-supervised recognition of sarcastic sentences in twitter and amazon // Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, 2010. P. 107-116.

23. *Taboada M., Brooke J., Tofiloski M., Voll K., Stede M.* Lexicon-based methods for sentiment analysis // Computational Linguistics. 2011. V. 37, No 2. P. 267-307.

24. *O'Connor B., Balasubramanyan R., Routledge B.R., Smith N.A.* From tweets to polls: linking text sentiment to public opinion time series // ICWSM. 2010. V. 11. P. 122-129.

25. *Mohammad S.M., Kiritchenko S., Zhu X.* NRC-Canada: building the state-of-the-art in sentiment analysis of tweets // Second Joint Conference on Lexical and Computational Semantics (* SEM). 2013. V. 2. P. 321-327.

26. *Evert S., Proisl T., Greiner P., Kabashi B.* SentiKLUE: updating a polarity classifier in 48 hours // SemEval 2014. 2014. P. 551.

27. *Barbosa L., Feng J.* Robust sentiment detection on twitter from biased and noisy data // Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010. P. 36-44.

28. *Martinez-Camara E., Martin-Valdivia M.T., Urena-Lopez L.A., Montejo-Raez A.R.* Sentiment analysis in twitter // Natural Language Engineering. 2014. V. 20, No 01. P. 1-28.

29. *Chetviorkin I., Loukachevitch N.* Evaluating sentiment analysis systems in Russian // ACL 2013. 2013. P. 12-17.

30. *Frolov A.V., Polyakov P.Yu., Pleshko V.V.* Using semantic filters in application to book reviews sentiment analysis // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog". 2013. No 12 (19).

СВЕДЕНИЯ ОБ АВТОРАХ



ТУТУБАЛИНА Елена Викторовна – аспирант Высшей школы информационных технологий и информационных систем Казанского федерального университета.

Elena Viktorovna TUTUBALINA received Diploma in applied mathematics and informatics from Kazan Federal University (2012). She is a graduate student at the High School of Information Technology and Information Systems of Kazan Federal University. Current scientific interests: natural language processing, opinion mining, topic modeling.

email: EIVTutubalina@kpfu.ru



ИВАНОВ Владимир Владимирович – старший преподаватель кафедры интеллектуальных технологий поиска Высшей школы информационных технологий и информационных систем Казанского федерального университета.

Vladimir Vladimirovich IVANOV is a head of Big Data and Textual Analysis Lab at the High School of Information Technology and Information Systems of Kazan Federal University. Current scientific interests: natural language processing, information extraction.

email: vivanov@kpfu.ru



ЗАГУЛОВА Мария – студент Высшей школы информационных технологий и информационных систем Казанского федерального университета.

Maria ZAGULOVA is a student at the High School of Information Technology and Information Systems of Kazan Federal University. Current scientific interests: sentiment analysis.

email: lolmariya@gmail.com



МИНГАЗОВ Никита – лаборант-исследователь НИЛ «Большие данные и анализ текста» Высшей школы информационных технологий и информационных систем Казанского федерального университета.

Nikita MINGAZOV is an assistant engineer at the Big Data and Textual Analysis Lab of High School of Information Technology and Information Systems of Kazan Federal University. Current scientific interests: information retrieval.

email: nicrotek547@gmail.com



АЛИМОВА Ильсеяр – аспирант Высшей школы информационных технологий и информационных систем Казанского федерального университета.

Ilseyar Alimova received Diploma in applied mathematics and informatics from Kazan Federal University (2014). She is a graduate student at the High School of Information Technology and Information Systems of Kazan Federal University. Current scientific interests: time series.

email: alimovallseyar@gmail.com

МАЛЫХ Валентин – аспирант Института системного анализа РАН.

Valentin MALYKH is a student at the Institute for Systems Analysis of Russian Academy of Sciences. Current scientific interests: natural language processing.

email: alem.mipt@gmail.com

Материал поступил в редакцию 15 июля 2015 года