Н.В. Лукашевич, И.И. Четверкин

Открытое тестирование систем анализа тональности на материале русского языка

Аннотация. В статье описан опыт проведения открытой оценки методов анализа русскоязычных текстов по тональности на базе семинара РОМИП в 2011-2012 годах. В рамках проведения дорожки было создано несколько обучающих коллекций, которые теперь находятся в свободном доступе. Приводится обзор текущего состояния дел в обработке оценочных текстов на русском языке, описание основных задач, характеристик коллекций, а также мер для измерения качества.

Ключевые слова: анализ отзывов, обработка текстов, классификация по тональности, РОМИП.

Введение

Автоматический анализ тональности текстов, т.е. автоматическое выявление отношения автора текста к обсуждаемым в тексте объектам и ситуациям, является одним из динамично развивающихся направлений автоматического анализа текстов на естественном языке. Большинство предлагаемых подходов к анализу тональности тестируются, прежде всего, на англоязычных текстах, также для английского языка создано множество разнообразных словарных ресурсов и инструментов. В последнее время было проведено большое количество исследований в области анализа тональности и для других языков [1, 16, 23].

В настоящее время развитие методов анализа тональности вызывает большой интерес и в России, как среди исследователей, так и со стороны компаний и государственных организаций. В 2011-2012 гг. было организовано два независимых тестирования систем анализа тональности русскоязычных текстов, которые проводились в рамках семинара по информационному поиску РОМИП (www.romip.ru) [6, 8].

Семинар по информационному поиску РОМИП был организован в 2002 г. для систем информационного поиска по примеру известной конференции по оценке методов информационного поиска TREC. В течение прошедших лет в рамках этого семинара тестировались подходы к задачам поиска документа по запросу, классификации текстов, поиска ответов на вопросы, автоматическому аннотированию, поиску изображений и др. Новые направления тестирования РОМИП, связанные с анализом тональности текстов, продолжают традиции уже прошедших международных тестирований такого рода систем [2, 17, 25, 32], но сосредотачиваются именно на анализе тональности русскоязычных текстов.

В этой статье мы представим обзор заданий, предложенных в рамках данных тестирований, опишем данные, которые предоставлялись участникам тестирования и которые теперь могут быть получены и другими заинтересованными исследователями, применяемые методы и результаты, полученные участниками. Проведенные тестирования продемонстрировали уровень результатов, достигаемых разными методами, качество их работы в конкретных условиях поставленных задач.

Данная статья имеет следующую структуру. В разделе 1 будут кратко представлены основные подходы к анализу тональности. В разделе 2 будут рассмотрены походы к анализу тональности русскоязычных текстов, не относящиеся к проведенным тестированиям РОМИП-2011 и

РОМИП-2012. В разделе 3 мы представим задания тестирований, данные, методы и результаты.

1. Методы анализа тональности

Автоматический анализ тональности осуществляется с помощью следующих основных методов [20]:

- методы машинного обучения, когда система «обучается» на коллекции размеченных текстов, т.е. текстов, которым человек явно приписл некоторую тональность. Как и во многих других задачах классификации текстов, обычно лучшие результаты показывает метод опорных векторов (SVM). Применяются также такие методы, как логистическая регрессия, наивный байесовский классификатор и др. [20];
- инженерно-лингвистические методы, которые заключаются в использовании специально создаваемых словарей оценочных слов и выражений и применении лингвистических правил, с помощью которых учитывается контекст употребления слов [22, 31].

Для английского языка было показано, что методы машинного обучения дают более высокие результаты классификации текстов по тональности при наличии достаточной обучающей коллекции. В некоторых областях такая обучающая коллекция может возникнуть естественным путем, например, когда пользователи при написании отзыва о каком-то продукте одновременно проставляют числовые баллы, обозначающие их отношение (положительное или отрицательное) к данному продукту. В остальных случаях создание обучающей коллекции требует значительных усилий по разметке тональности.

Во многих работах отмечается, что в различных предметных областях используется свой набор оценочной лексики. Так, в предметной области фильмов может встретиться такое оценочное выражение как «рекомендую посмотреть», но не будут использоваться такие явные оценочные выражения, применяемые по отношению к людям как «подлец» или «наглец».

Такая ситуация приводит к тому, что для каждого типа методов необходимы дополнительные настройки на предметную область. Для методов, основанных на знаниях, необходимо пополнение словарей. Системы классификации тональности, основанные на машинном обучении, при переносе на другую предметную область также необходимо специальным образом модифицировать, иначе происходит резкое падение качества классификации [9, 19].

2. Анализ тональности русскоязычных текстов

В России публикации, посвященные анализу тональности русскоязычных текстов, до 2012 г. не многочисленны. В работе [10] описывается система анализа тональности отзывов об автомобилях на материалах блога http://avto-ru.livejournal.com/. Подход основывается на детальном описании марок автомобилей, их деталей и характеристик, а также синтаксико-семантических шаблонах оценочных высказываний. Эта статья была первой статьей по анализу тональности в России, в которой были представлены результаты тестирования предложенного подхода – точность 84%, полнота 20%.

В зарубежных исследованиях анализ тональности русскоязычных текстов производится, в основном, в многоязычном контексте.

В работе [33] анализируются два сравнимых корпуса отзывов о книгах на русском и английском языках, что позволяет авторам этой работы изучать особенности способов выражения тональности.

В работе [30] описывается способ создания общих словарей оценочной лексики для нескольких языков. Для этого были взяты два исходных словаря оценочной лексики: английский (2400 слов) и испанский (1737 слов). Оба списка были переведены Google-переводчиком на целевые языки. Только те слова, которые появились в переводах обоих списков, были взяты для дальнейшей работы. Набор целевых языков состоял из шести языков, включая русский. Полученный словарь оценочной лексики для русского языка включал 966 слов, точность (accuracy) списка была оценена как 94.9%.

В статье [7] рассматривается формирования русскоязычного словаря оценочных слов для обобщенной области продуктов и услуг, описывается модель извлечения оценочных слов для конкретной предметной области, основанная на совокупности признаков слов и их комбинировании алгоритмами машинного обучения. После обучения модели на области фильмов, полученная модель применяется еще

к нескольким предметным областям и полученные списки оценочных слов суммируются формулой специального вида, которая присваивает больший вес словам, имеющих высокий оценочный вес в нескольких предметных областях. В результате был получен список из 5000 оценочных слов ProductSentiRus, который в настоящее время находится в свободном доступе (http://www.cir.ru/SentiLexicon/ProductSentiRus.txt).

3. Задания для систем анализа тональности на семинарах РОМИП-2011, 2012

Задачи двух проведенных тестирований систем анализа тональности включали:

– задачу классификации отзывов пользователей в трех областях (фильмы, книги, цифровые камеры) по нескольким шкалам;

– задачу классификации новостных цитат, т.е. фрагментов прямой или косвенной речи, извлеченных из новостных сообщений;

– поиск постов в блогах, содержащих оценку товара или произведения, заданного в запросе.

В РОМИП-2011 приняли участие 12 групп, приславших более 200 прогонов своих систем, в РОМИП-2012 - 17 групп с более 150 прогонами, что подтвердило значительный интерес к данной задаче.

Далее задачи и полученные результаты будут рассмотрены более подробно.

3.1. Классификация отзывов по тональности

Единственным заданием тестирования РОМИП-2011 и одним из заданий РОМИП-2012 была классификация по тональности отзывов пользователей в трех областях: фильмы, книги, цифровые камеры. Приведем пример отзыва о фильме.

Неожиданная развязка и новые герои делают этот фильм непохожим на предшественника.

Обучающая коллекция для этого тестирования была основана на двух источниках. Во-первых, использовались отзывы с портала Imhonet (imhonet.ru) (фильмы – 15718 отзывов, книги – 24159 отзывов), эти отзывы были снабжены оценкой пользователей по 10-балльной шкале. Во-вторых, обучающая коллекция отзывов о цифровых камерах с оценкой пользователей по 5-балльной шкале была получена с сайта Яндекс-маркет (http://market.yandex.ru/).

Для тестирования систем была собрана другая коллекция отзывов, которая изначально не имела проставленных оценок пользователей. Данная коллекция состояла из отзывов пользователей в блогах и была получена посредствам исполнения запросов в поисковой машине Яндекс-блоги (http://blog.yandex.ru). Таким образом, в данной задаче мы пытались моделировать одну из существующих практических постановок задач, когда имеющиеся данные для обучения несколько отличаются от реальных данных, на которых должна работать система. Кроме того, такая постановка задачи ставила всех участников в равные условия. В очном обсуждении результатов этого задания участники подчеркивали, что условия задания были сложнее, чем обучение и тестирование на тех же данных, однако соглашались, что такая постановка задачи более реалистична.

Для каждой предметной области был вручную составлен список запросов о соответствующем товаре (произведении) и далее посредством поисковой машины были извлечены посты блогов, релевантные этим запросам. Все полученные посты были объединены в единую коллекцию для тестирования, которая и была послана участникам. Системы-участники должны были классифицировать коллекцию по тональности.

Для проверки ответов систем тестовая коллекция поступила на оценку экспертам. В их задачу входило отобрать из всех имеющихся постов такие, которые релевантны заданным предметным областям и содержат оценку упоминаемых объектов, а также классифицировать отобранные посты по трем шкалам: двухбалльной (позитивный, негативный), трехбалльной (позитивный, негативный, удовлетворительный), пятибалльной (отлично, хорошо, средне, плохо, ужасно).

В процессе проставления оценок эксперты столкнулись с ситуацией, когда коллекция содержала большое количество нерелевантных постановке задаче постов. Так, например, по запросу «Джеймс Бонд» (подразумевались отзывы о фильмах из этой серии) было получено огромное количество постов вида «вот это да … ты похож на Джеймса Бонда», что потребовало дополнительных усилий по отбору постов

для оценки. Также было выявлено, что среди постов имеется значительный перевес положительных отзывов – 85-90% по разным предметным областям.

Для того, чтобы выяснить уровень согласия между экспертами, в 2011 г. была выполнена двойная разметка постов двумя экспертами. Каппа-статистика, т.е. уровень совпадения между экспертами по сравнению со случайным совпадением, представлена в Табл. 1. Каппа вычисляется по формуле:

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}, \qquad (1)$$

где *Pr(a)* – наблюдаемое согласие между экспертами, *Pr(e)* – вероятность согласия между экспертами, если бы они проставляли свои оценки случайным образом.

Из Табл. 1 видно, что при росте числа классов классификации несогласие между экспертами резко возрастает.

Основными метриками качества в данной задаче были правильность классификации (accuracy) и F-мера в варианте макро-усреднения (Табл.2). Макро-усреднение здесь означает, что сначала точность и полнота вычисляются для каждого класса в отдельности, затем находится среднее для значения каждой метрики [6]. Макро-меры позволяют лучше оценить насколько хорошо системы различают объекты разных классов в условиях несбалансированной коллекции.

$$P = \frac{tp_x}{tp_x + fp_x} \qquad (2)$$

$$R = \frac{tp_x}{tp_x + fn_x} \qquad (3)$$

$$Macro\_P = \frac{1}{|S|} \cdot \sum_{x \in S} \frac{tp_x}{tp_x + fp_x} \qquad (4)$$

$$Macro\_R = \frac{1}{|S|} \cdot \sum_{x \in S} \frac{tp_x}{tp_x + fn_x} \qquad (5)$$

$$F - мера = \frac{2 \cdot P \cdot R}{P + R} \qquad (6)$$

$$Accuracy = \frac{tp_x + tn_x}{tp_x + tn_x + fp_x + fn_x} \qquad (7)$$

Для задачи классификации отзывов по пяти классам использовалась также метрика Евклидова расстояния, которое представляет собой среднее квадратов разностей между оценками, проставленными системой, и оценками, проставленными экспертами *p*.

$$Dist = \sqrt{\frac{\sum_{i=1}^{n} (q_i - p_i)^2}{n}} \qquad (8).$$

Лучшие результаты, полученные участниками РОМИП-2011 и РОМИП-2012, представлены в Табл. 3 и Табл. 4 соответственно.

Участники применяли как подходы, основанные на различных методах машинного обучения, так и инженерно-лингвистические подходы. Однако подавляющее большинство лучших подходов в задачах классификации отзывов базируются на применении метода опорных векторов SVM [5, 11, 18, 24], который мо-

Табл.1. Каппа-статистика по предметным областям

Каппа	2 класса	3 класса	5 классов
Кино	0.818	0.615	0.429
Книги	0.812	0.674	0.545
Цифровые камеры	0.808	0.602	0.398

Табл.2. Обозначения для определения мер качества классификации

	Правильный класс	
Предсказанный класс	**tp_x** (true positive) Correct result	**fp_x** (false positive) Unexpected result
	fn_x (false negative) Missing result	**tn_x** (true negative) Correct absence of result

Табл. 3. Лучшие результаты участников РОМИП-2011 в задачах классификации отзывов

Предметная область	**2-класса**		**3-класса**		**5-классов**	
	F1	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**
Кино	0.786	0.881	0.592	0.754	0.286	0.602
Книги	0.747	0.938	0.577	0.771	0.291	0.622
Цифровые камеры	0.929	0.959	0.663	0.841	0.342	0.626

Табл. 4. Лучшие результаты участников РОМИП-2012 в задачах классификации отзывов

Предметная область	**2- класса**		**3- класса**		**5- классов**	
	F1	**Acc.**	**F1**	**Acc.**	**F1**	**Acc.**
Кино	0.707	0.831	0.520	0.694	0.377	0.407
Книги	0.715	0.884	0.560	0.752	0.402	0.480
Цифровые камеры	0.669	0.961	0.480	0.742	0.336	0.513

жет комбинироваться с дополнительными ресурсами вроде словарей или правил. Данный результат согласуется с результатами, полученными для задачи анализа тональности англоязычных текстов, где также было показано, что метод опорных векторов обычно порождает лучшие по качеству результаты в этой задаче.

Лучшие результаты в рамках РОМИП-2012 показал подход [5], который использовал классификаторы на основе методов опорных векторов (SVM) и максимизации энтропии (MaxEnt) наряду с набором признаков, полученных полуавтоматически с помощью словаря из работы [7]. Кроме того, авторы исследовали различные схемы присвоения весов признакам, учет доли положительных и отрицательных слов в тексте, учет знаков препинания (вопросительных и восклицательных), а также смайликов и нецензурной лексики.

Только в одном случае подход [21], который показал хорошие результаты в задаче классификации отзывов о фильмах на два класса, был основан на применении словаря оценочных слов и правил их комбинирования. В частности, учитывались: словосочетания, состоящие из нескольких оценочных слов, инверсия тональности, синтаксически связанные слова, входящие в заданные семантические списки.

Сравнивая результаты, полученные на двух тестированиях РОМИП-2011 и РОМИП-2012 можно констатировать, что они вполне согласуются с результатами, полученными для английского языка: правильность классификации (accuracy) для двухклассовой задачи - около 90%, трехклассовой - около 75%, пятиклассовой - около 50%. Однако стоит отметить, что результаты по accuracy могут быть несколько завышены из-за превалирования положительных отзывов в коллекции для тестирования.

3.2. Анализ тональности цитат

Еще одним заданием второго тестирования систем анализа тональности РОМИП-2012 была задача классификации коротких (в среднем, 1-2 предложения) фрагментов прямой или косвенной речи, извлеченных из новостных сообщений (далее цитаты). Приведем пример цитаты:

По мнению эксперта, глава белорусского государства больше всего боится, что страну все-таки лишат права провести чемпионат мира по хоккею в 2014 году.

Тематика цитат никак не ограничивалась и могла быть достаточно различной: от политики и экономики до культуры и спорта. Поэтому предполагалось, что данное задание будет достаточно сложным для подходов: подходов, основанных на знаниях, и подходов, основанных на машинном обучении.

Эксперты размечали цитаты на четыре класса: позитивные, негативные, нейтральные и смешанной тональности. После этого цитаты со смешанной тональностью были удалены из обучающего и тестового множеств. Таким образом, системы должны были классифицировать цитаты на три класса.

Похожая задача решалась в рамках семинара NTCIR-6, где одной из основных задач было извлечение предложений, содержащих мнения, из новостных сообщений на трех языках: английском, японском и китайском [29]. Такая постановка задачи похожа также на задачу классификации политических высказываний [3, 4] на позиции «за» и «против». В работе [3] авторы подчеркивают, что короткие цитаты сложны для классификации, поскольку лингвистические признаки тональности разнообразны и разрежены, и некоторые цитаты могут иметь разную полярность в разных тематиках. В нашем случае задача была даже более сложной из-за отсутствия ограничений по тематике и необходимости классификации цитат на три класса.

В качестве обучающей коллекции было размечено и выдано участникам 4260 цитат. Для тестирования было разослано более 120 тыс. цитат, но реальное оценивание производилось на 5500 цитатах.

В этой задаче распределение цитат по классам было значительно более сбалансированным: 41% негативных цитат, 32% позитивных и 27% нейтральных цитат. Для оценивания подходов также применялись ранее использованные метрики: макро F-мера и правильности классификации. Результаты участников приведены в Табл. 5. Результат «baseline» соответствует классификации по наиболее частотному классу. В противоположность задаче классификации отзывов все лучшие подходы были основаны на лингвистических знаниях (словарь + правила), что связано с отсутствием большой обучающей коллекции и широтой тематик цитат.

В системе, получившей лучшие результаты в этой задаче, использовался большой словарь, включающий около 15 тыс. негативных слов и

Табл.5. Лучшие результаты участников РОМИП-2012 в задаче классификации цитат

Run_ID	Macro_P	Macro_R	Macro_F1	Accuracy
xxx-4	0.626	0.616	0.621	0.616
xxx-11	0.606	0.579	0.592	0.571
xxx-15	0.563	0.560	0.562	0.582
Baseline	0.138	0.333	0.195	0.413

выражений, 7 тыс. позитивных слов и выражений, 120 так называемых операторов – слов, которые меняют тональность стоящих рядом с ними слов (например, *не, очень, снизить*), и 200 нейтральных стоп-выражений, в состав которых входят оценочные слова (например, *Фонд эффективной политики)* [12]. Также в этой работе [12] было показано, что дополнение набора правил, используемых в системе, при неизменном словаре позволяет дополнительно повысить качество работы системы. Система ATEX, получившая второй и третий результат в этом задании, имеет относительно небольшой словарь, но больший набор правил [21].

Интересным вопросом в создании систем анализа тональности, основанных на словарях и правилах, является собственно набор правил, которые считаются полезными при решении этой задаче. В работах [12, 31] указывается, что наиболее распространенными правилами являются правила, учитывающие слова-операторы: слова-отрицания (*не, нет, отсутствие)* меняют тональность следующих за ними слов; слова-усилители (*очень, чрезвычайно*..) увеличивают тональность следующих за ними слов. Воздействие слов-операторов может быть учтено домножением веса следующего оценочного слова на заданные множители. Например, в случае отрицаний на множитель -1, в случае усилений на множитель 2.

Второе правило касается агрегации оценок слов в тексте – чаще всего, такие оценки просто суммируются.

В работе [12] на данных РОМИП-2012 было показано, что важным является учет так называемого фактора нереальности, т.е. определение того, что упоминаемая в предложении ситуация, скорее всего, не осуществилась, и, значит, вклад оценочных слов, встретившихся в таких предложениях, в общую оценку текста должен быть снижен. Маркерами учета фактора нереальности для задачи анализа тональности являются следующие случаи:

– оценочные слова, встретившиеся в вопросительных предложениях, не начинающихся со слов *почему* и *зачем*,

– оценочные слова, встретившиеся во фрагменте между знаками препинания, где встречаются также слова *если*, *бы*,

– таким же образом нужно учитывать частицу *ли в* том случае, если перед этой частицей не встретились такие слова как *чуть/то/вряд/мало/едва/что*.

Кроме того, полезными являются следующие правила [12]:

– если слова-операторы образуют последовательность, то соответствующие им множители перемножаются,

– если в фразе встречается несколько оценочных слов, и среди них одно отрицательное, то оценка всей группы становится отрицательной,

– слово-оператор применяется не к отдельному следующему за ним оценочному слову, а к синтаксической группе в целом.

3.3. Поиск оценочных постов в блогах

В течение ряда лет в рамках конференции по информационному поиску TREC проводилось тестирование Blog Track, в котором были соединены задачи поиска по блогам и анализа тональности [13, 14, 26-28]. В рамках РОМИП-2012 было поставлено похожее задание. В этой задаче участники должны были найти все оценочные посты из коллекции блогов, релевантные заданному запросу, т.е. правильный пост должен отвечать двум критериям: в нем должен обсуждаться объект из запроса, и пост должен содержать мнение об этом объекте.

Примеры запросов включали:

– для фильмов: “Девушка с татуировкой дракона” и “Диктатор”;

– для книг: Агата Кристи "Десять негритят";

– для цифровых камер: Canon EOS 1100D Kit.

На Рис. 1 показан пример выдачи по заданной конкретной цифровой камере, документы, содержащие оценочные суждения о данной камере обведены. В задачу участников данного задания входило поставить на первые места в выдаче именно такие документы.

Поскольку оказалось, что в данном задании участвовал только один участник, то нами как организаторами был реализован простой метод, чтобы поддержать проведение этого тестирования. Реализованный метод поиска оценочных

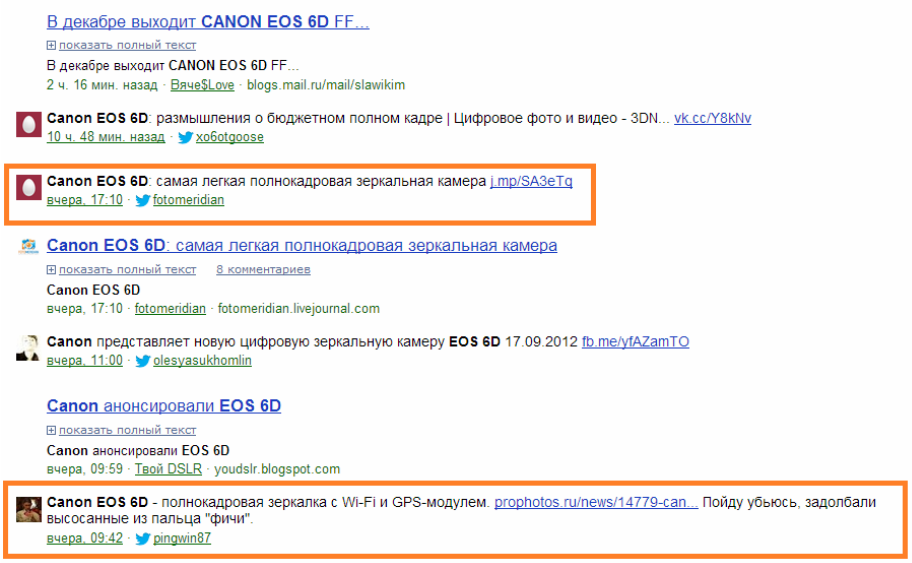

Рис. 1. Выдача оценочных постов по запросу о конкретной цифровой камере подчеркнуты документы, содержащие оценочные суждения о данной камере

постов был построен на трех компонентах: tfidf[1] сходства запроса с заголовком поста, tfidf сходства запроса с текстом поста, доля оценочных слов в посте. Для вычисления последнего компонента использовался упомянутый выше список оценочных слов ProductSentiRus. Таким образом, для каждого запроса посты были упорядочены по следующей величине веса:

$$Weight = \alpha \cdot (\sum_{w \in q} tfidf_w + \sum_{w \in q} tfidf_w^{header}) + \\ + (1-\alpha) \cdot SentiWeight \cdot \quad (9)$$

Мы экспериментировали с различными величинами α = {0.2, 0.4, 0.5, 0.6, 0.8}. Лучшие результаты были получены для значения α = 0.6 во всех предметных областях, которые в итоге и оказались лучшими в этой задаче. Чтобы избежать недооценки результатов, оценка качества была выполнена только на постах, оцененных экспертами. Для оценки качества исполнения этой задачи использовались две основные метрики: точность (precision) на уровне *n* (P@n) и метрика NDCG@n. Метрика P@n обозначает долю правильных ответов в первых

Табл. 6. Лучшие результаты в задаче поиска мнений по блогам

Run_ID	Object	P@1	P@5	P@10	NDCG@10
xxx-0	book	0.3	0.32	0.286	0.305
xxx-8	book	0.25	0.31	0.332	0.298
Yyy-9	camera	0.402	0.313	0.302	0.305
Yyy-1	camera	0.402	0.328	0.325	0.226
Zzz-3	film	0.494	0.449	0.438	0.338
Zzz-8	film	0. 494	0.448	0.444	0.332

n выданных результатах. Метрика NDCG@n измеряет полезность, информационную значимость (gain) очередного документа в списке выдачи [15]. Главными метриками в этой задаче были NDCG@10 и P@10 (Табл. 6).

Заключение

В данной статье мы представили обзор методов автоматического анализа тональности текстов на русском языке. Данный обзор базируется на проведенных в 2011и 2012 годах тестированиях систем анализа тональности. В качестве заданий данного тестирования участникам были предложены задачи поиска и классификации отзывов пользователей из блогов, а также задача классификации мнений, извлеченных из новостных сообщений. В результате проведенных тестирований были продемонстрированы преимущества и недостатки различ-

[1] tfidf – мера оценки значимости слова в тексте, применяемая в векторной модели информационного поиска, увеличивается с ростом частоты употребления слова в тексте (tf) и уменьшается с ростом частоты употребления слова в документах коллекции (подокументная частота – df) [15].

ных подходов к автоматическому анализу тональности, а также достигаемые в настоящее время характеристики качества автоматических систем. Высокий интерес к открытым тестированиям систем в области анализа тональности подтверждает актуальность решаемых задач и их востребованность в системах обработки информации. Вектор развития при решении задачи анализа тональности лежит в более детальном анализе текстов об объектах и их атрибутах, учете структуры связного текста, а также построении систем, которые будут устойчивы при переносе на различные предметные области.

Литература

1. Abdul-Mageed M., Diab M., Korayem M. Subjectivity and Sentiment Analysis of Modern Standard Arabic. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, 2011. pp. 587-591.
2. Amigo E., Corujo A., Gonzalo J., Meij E., Rijke Md. Overview of replab 2012: Evaluating online reputation management systems. In Proceedings of the CLEF 2012 Labs and Workshop Notebook Papers. 2012. pp. 1–24.
3. Awadallah R., Ramanath M., Weikum G. PolariCQ: Polarity Classification of Political Quotations In Proceedings of CIKM-2012, 2012. pp. 1945-1949.
4. Balasubramanyan R., Cohen W., Pierce D., Redlawsk D. Modeling polarizing topics: When do different political communities respond differently to the same news? Proceedings of ICWSM. 2012.
5. Blinov P., Klekovkina M., Kotelnikov E, Pestov O. Research of lexical approach and machine learning methods for sentiment analysis. In Proceedings of Dialog, volume 2, 2013. pp. 51-61.
6. Chetviorkin I., Braslavskiy P., Loukachevich N. Sentiment Analysis Track at ROMIP 2011. In Proceedings of International Conference Dialog-2012, volume 2, 2012. pp. 1-14.
7. Chetviorkin I., Loukachevitch N. Extraction of Russian Sentiment Lexicon for Product Meta-Domain In Proceedings of COLING 2012, 2012. pp. 593-610.
8. Chetvirokin I., Loukachevitch N. Sentiment Analysis Track at ROMIP 2012. In Proceedings of International Conference Dialog-2013, volume 2, 2013. pp. 40-50.
9. Choi Y., Cardie C. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2009. pp. 590-598.
10. Ермаков А.Е. Извлечение знаний из текста и их обработка: состояние и перспективы.// Информационные технологии, № 7, 2009. С. 50-55.
11. Котельников Е.В., Клековкина, М.В. Автоматический анализ тональности текстов на основе методов машинного обучения. Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции Диалог. Т. 2, 2012. С. 27-36.
12. Kuznetsova E.S., Loukachevitch N.V., Chetviorkin I.I. Testing rules for sentiment analysis system. Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog-2013, vol. 2, 2013. pp. 71-80.
13. Macdonald C., Ounis I., Soboroff I. Overview of the TREC 2007 blog track. In Proceedings of TREC-2007. Gaithersburg, USA, 2008.
14. Macdonald C., Ounis I., Soboroff I. Overview of the TREC 2009 blog track. In Proceedings of TREC-2009. Gaithersburg, USA, 2010.
15. Manning C. D., Raghavan P., Schütze H. Introduction to information retrieval. – Cambridge: Cambridge University Press. 2008.
16. Mihalcea R., Banea C., Wiebe J. Learning multilingual subjective language via cross-lingual projections. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 2007. pp. 976-983.
17. Morante R., Blanco E. *SEM 2012 shared task: Resolving the scope and focus of negation. In Proceedings of the First Joint Conference on Lexical and Computational Semantics, Montreal,. 2012. pp. 265-274
18. Pak A., Paroubek P. Language independent approach to sentiment analysis (LIMSI Participation in ROMIP'11) Computational Linguistics and Intellectual Technologies. Proc. of International Conference Dialog-2012, vol. 2, 2012. pp. 37-50.
19. Pan S. J., Ni X., Sun J-T, Yang Q., Chen Z. Cross-Domain Sentiment Classification via Spectral Feature Alignment. In Proceedings of the World Wide Web Conference, 2010. pp. 751-760.
20. Pang B., Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval. Now Publishers, 2008.
21. Паничева П. Система сентиментного анализа ATEX, основанная на правилах, при обработке текстов различных тематик. Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции Диалог. Т. 2, 2013. С.101-113.
22. Пазельская А. Г., Соловьев А. Н. Метод определения эмоций в текстах на русском языке. Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции Диалог, 2011. С. 510-522.
23. Perez-Rosas V., Banea C., Mihalcea R. Learning Sentiment Lexicons in Spanish. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). 2012.
24. Поляков, П.Ю., Калинина, М.В., Плешко, В.В. Исследование применимости методов тематической классификации в задаче классификации отзывов о книгах Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной международной конференции Диалог Т. 2, 2012. С. 51-59.
25. Pestian J., Matykiewicz.P., Linn-Gust M. Sentiment analysis of suicide notes: A shared task. Biomedical Informatics Insights. 2012;5 (Suppl. 1), 2012. pp. 3-16.

26. Ounis I., de Rijke M., Macdonald C., Mishne G., Soboroff I. Overview of TREC-2006 Blog track. In Proceedings of TREC-2006, Gaithersburg, USA, 2007.
27. Ounis I., Macdonald C., Soboroff I. Overview of the TREC 2008 blog track. In Proceedings of TREC-2008. Gaithersburg, USA, 2009.
28. Ounis I., Macdonald C., Soboroff I. Overview of the TREC 2010 blog track. In Proceedings of TREC-2010. Gaithersburg, USA, 2011.
29. Seki Y., Evans D., Ku L., Chen H., Kando N., Lin C. Overview of opinion analysis pilot task at NTCIR-6. In Proceedings of NTCIR-6 Workshop Meeting. 2007. pp. 265-278.
30. Steinberger J., Lenkova P., Ebrahim M., Ehrmann M., Hurriyetogly A., Kabadjov M., Steinberger R., Tanev H., Zavarella V. and Vazquez S. Creating Sentiment Dictionaries via Triangulation. In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT, 2011. pp. 28-36.
31. Taboada M., Brooke J., Tofiloski M., Voll K., Stede M. Lexicon-based methods for Sentiment Analysis. Computational linguistics, 37(2), 2011. pp. 267-307.
32. Wu Y., Jin P. Semeval-2010 task 18: Disambiguating sentiment ambiguous adjectives. In Proceedings of the 5th International Workshop on Semantic Evaluation. 2010. pp. 81-85.
33. Zagibalov T., Belyatskaya K., Carroll J. Comparable English-Russian Book Review Corpora for Sentiment Analysis. In Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis WASSA, 2010. pp. 67-72.

Лукашевич Наталья Валентиновна. Ведущий научный сотрудник НИВЦ МГУ им. М.В. Ломоносова. Окончила МГУ имени М.В. Ломоносова в 1986 году. Кандидат физико-математических наук. Автор 140 печатных работ и монографий. Область научных интересов: искусственный интеллект, компьютерная лингвистика, интеллектуальный анализ данных, извлечение информации. E-mail: louk_nat@mail.ru.

Четвёркин Илья Игоревич. Аспирант факультета вычислительной математики и кибернетики МГУ им. М.В. Ломоносова. Окончил МГУ им. М.В. Ломоносова в 2010 году. Автор 15 печатных работ. Область научных интересов: искусственный интеллект, компьютерная лингвистика, интеллектуальный анализ данных, извлечение информации.
E-mail: ilia2010@yandex.ru.