

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 622.331:628.25

DOI: 10.22213/2410-9304-2019-4-53-62

ИЗУЧЕНИЕ ДОСТУПНЫХ СИСТЕМ ДЛЯ АНАЛИЗА ЭМОЦИЙ, ИЗВЛЕЧЕННЫХ ИЗ ТЕКСТА, И ОБЕСПЕЧЕНИЯ МЕХАНИЗМА ДЛЯ УЛУЧШЕНИЯ ВЗАИМОДЕЙСТВИЯ ЧЕЛОВЕКА С МАШИНОЙ

М. М. Аббаси, аспирант, Удмуртский государственный университет, Ижевск, Россия

А. П. Бельтюков, доктор физико-математических наук, профессор, Удмуртский государственный университет, Ижевск, Россия

Развитие информационных технологий позволяет нам разрабатывать системы для анализа и обработки данных. Сегодня основным источником доступных данных является интернет. Исследователи разработали онлайн- и офлайн-системы для анализа этих данных. Этот анализ может быть использован для различных целей. Данная работа сосредоточена на изучении систем, которые используются для анализа эмоций, извлеченных из текста. В этой статье анализируются такие системы и, основываясь на авторских исследованиях, предлагаются механизмы для повышения их характеристик и улучшения области применения, а также сравнивается их производительность.

Анализ эмоций, извлеченных из текста, может быть использован для прогнозирования будущих событий, отзывов людей о продукте или услуге, идентификации группы людей по интересам и разработки машины, которая может имитировать поведение человеческих эмоций. Основной целью статьи является усовершенствование механизма человеко-машинного взаимодействия при общении с использованием текста. В статье предлагается механизм, который улучшает взаимодействие человека и машины путем определения психолингвистических характеристик текста, которые представляют поведение человека. Данный механизм будет изучать отношения между эмоциями и психолингвистическими характеристиками текста. Это облегчит процесс взаимодействия человека с машиной.

Ключевые слова: текст, эмоции, блог, общение, тема, анализ.

Введение

Системный анализ является важной темой современных исследований. Он позволяет анализировать механизмы, которые используются для обработки данных. Это помогает определить возможности для улучшения их функциональности. В этой статье рассматриваются наиболее популярные системы, которые используются для анализа эмоций онлайн и офлайн, и предложены механизмы для их совершенствования.

Эмоциональный анализ приобрел наибольшую популярность за последнее десятилетие. Основным источником данных для анализа эмоций являются веб-сайты социальных сетей, блоги, группы, используемые людьми для выражения своего мнения. Эмоции и их анализ могут быть использованы для выявления онлайн-группы людей, групп ненависти, прогнозов на будущее, отзывов клиентов и т. д. Это требует от систем быть эффективными, действенными и надежными. Со временем появляются разные программы, разрабатываемые для анализа эмоций. Система общего опроса в 1966 году была, вероятно, первой системой, созданной для анализа эмоций [1].

Для анализа эмоций в Интернете применяется огромное количество различных системы, каждая из них использует свой механизм анализа. Исследованию подверглись системы, которые обычно используются онлайн для анализа мнений и эмоций. Подробное описание таких систем представлено в последующих разделах.

Целью публикации данной работы является возможность предложить механизм, который способен анализировать эмоции, извлеченные из текста, луч-

ше, чем доступная система, и использовать уникальный подход для улучшения взаимодействия между человеком и машиной.

Связанные работы

После системы общих запросов основной вклад в дальнейшее развитие направления внесли Джейнс Вибе, Питер Терни и Василейос Хатзивассилоглу в начале 90-х годов. Джейнс Вибе в 1990 году (Wiebe, [2]) определяет термин «субъективность» для исследования поиска информации. Позже, в 1997 году (Hatzivassiloglou et. Al., [3]), была определена семантическая ориентация прилагательных в тексте. Спустя несколько лет Питер Терни (2002, [4]) предложил свой революционный подход Thumbs Up и Thumbs Down для классификации положительных и отрицательных отзывов. Панг (et. Al., 2002, [5]) предложил строить лексикон настроений вручную для домена. Denecke (2009, [6]) сообщил об интересном исследовании по нескольким доменам, чтобы продемонстрировать полезность предыдущих оценок полярности из эмоциональной базы данных под названием SentiWordNet. Предварительная обработка текста перед анализом эмоций повышает производительность и точность результатов [7].

Кластеризация рассматривалась как метод, основанный на обобщениях разбиения граф, которые не требуют заранее заданных специальных дистанционных функций и способны автоматически обнаруживать сходства или ассоциации документов [8]. Логические формулы используются для определения наиболее распространенных методов анализа эмоций, извлеченных из текста. Логические свойства эмо-

циональных модальностей, логика эмоциональных оценок и определение различных модальностей позволяют проводить аффективный анализ эмоций [9].

Исследования, посвященные анализу настроений в русском языке до 2011 года, не многочисленны. В 2009 году А. Ермаков (2009, [10]) предложил систему анализа настроений, извлекающую мнения об автомобилях из российского сообщества блогов. Шаблоны зависели от языка и предметной области, это означает, что шаблоны должны отражать лексические, синтаксические и стилистические особенности анализируемого текста. Было невозможно напрямую перевести или отобразить английскую базу шаблонов в русский шаблон [11]. Первоначально анализ эмоций, извлеченных из текста на русском языке, проявляется, главным образом, в многоязычных экспериментах. Загibalов (2010, [12]) сравнивает совокупности обзоров, относящихся к одним и тем же книгам на английском и русском языках. Steinberger (2011, [13]) описывает построение словарей общих эмоций для нескольких языков. Четвёркин и Лукачевич (2012, [14]) описали генерацию русских настроений и эмоциональный словарь для обобщенной области продуктов и услуг и т. д.

Для идентификации и анализа эмоций на русском языке большинство исследователей используют методики, основанные на правилах, что совершенно необходимо, поскольку правила являются ключевыми для разработки грамматики, которая может быть использована для разработки синтаксического и семантического анализатора [15, 16]. Было замечено, что методы машинного обучения работают лучше в категоризации и классификации текста. Использование синтаксиса для анализа эмоций текста, написанного на русском языке, в последнее время приобрело популярность [17].

Kruengkrai и Jaruskululchi определили заголовок текста [18]. Их подход использует преимущества как локальных, так и глобальных свойств предложений. Они использовали группы значимых слов в каждом предложении, чтобы вычислить локальное свойство предложения и отношения всех предложений в документе, чтобы определить глобальное свойство текстового документа.

В 2012 году Maruam Kiabod et al. [19] представил технику для обобщения текста путем определения значимых слов в нем. Он использовал экстрактивный метод, чтобы выбрать подмножество предложений, которые содержат основную концепцию текста.

Использование пространства слов – это новый механизм измерения и определения силы эмоций и их частоты в тексте [20, 21].

Современные системы анализа эмоций

Рассмотрим одни из самых популярных современных систем, используемых для анализа эмоций.

1. Приложение Power BI

Система используется для анализа эмоций людей с помощью смайликов, которые обычно применяются на Facebook. Человек может добавлять темы и сообщения, которые его интересуют. Другие люди и подписчики могут использовать смайлики, чтобы показать свои эмоции по поводу этой конкретной темы. Затем

схожие смайлики объединяются в группы, и с помощью графического интерфейса отображается количество смайликов. Каждый человек может высказать свое мнение, используя сравнение, и увидеть последние результаты. Существует проблема с областью применения системы, которая очень ограничена. Использование текстовых эмоций может быть лучшим решением, поскольку можно использовать слова гораздо большие, чем смайлики для эмоций. Также можно создать способы наблюдения взаимосвязи между эмоциями и их сильными сторонами. На рис. 1 показан графический интерфейс пользователя Power BI.

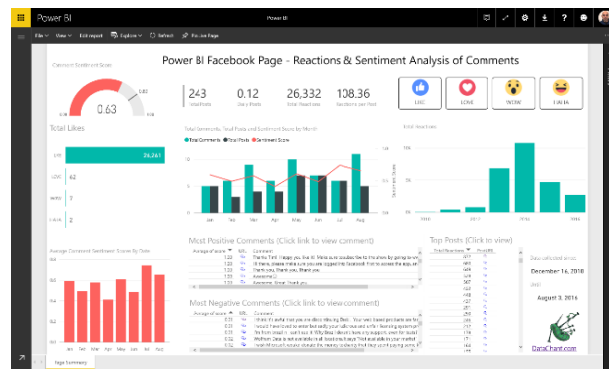


Рис. 1. Отображение графического интерфейса пользователя для Power BI [22]

2. Приложение Tableau

Это онлайн-приложение, которое можно загрузить на персональные компьютеры. Блогер может добавлять новые темы, а разные люди могут реагировать и высказывать свое мнение по этому поводу. Основными категориями для написания постов являются: блогеры, журналисты, спортивные болельщики, «политические наркоманы».

Приложение предоставляет публике основные фразы и термины, которые используются для выражения эмоций. Панель мониторинга анализа настроений выражает результаты темы несколькими различными способами. Она выражает анализ настроения, категории, степень удовлетворенности людей при высказывании их мнения по теме, оценка ODC, смайлики, текст. У каждого из них есть свои подклассы, чтобы выразить эмоции более подробно. На рис. 2 показан графический интерфейс пользователя Tableau.

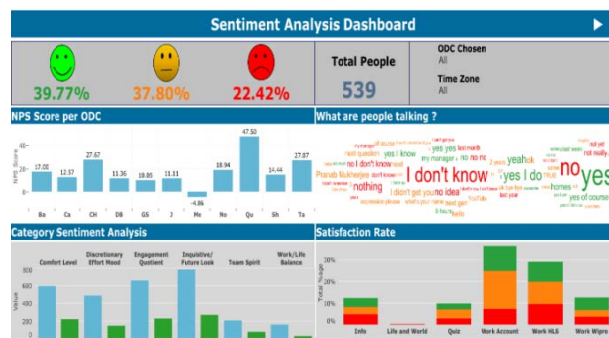


Рис. 2. Отображение графического интерфейса пользователя для Tableau [23]

3. Приложение Qemotion

Qemotion просит пользователей представить текст с использованием API, и алгоритм определяет основные эмоции и соответствующие эмоции в терминах температуры (буквально температура). Такие, как от -20 до -15 °C представляют собой отвращение, от -14 до -6 °C – злость, от -5 до 5 °C – грусть и разочарование, от 6 до 10 °C – страх, от 11 до 20 °C – спокойствие, от 21 до 30 °C – сюрприз, а от 31 до 40 °C – счастье. Диапазон температур в каждой эмоциональной категории составляет около 5, каждое число в этом диапазоне не обеспечивает степень счастья. Эмоции могут быть выражены на подуровне, поскольку доступен диапазон из 5 цифр для каждой эмоции.



Рис. 3. Отображение главной страницы приложения Qemotion [24]

4. Приложение MoodPatrol API

MoodPatrol помогает анализировать тексты и извлекать из них основные эмоции. Этот API использует сложный анализ текста, чтобы найти шаблоны эмоций в тексте, помогает обнаружить 8 основных эмоций в тексте.

Эти основные эмоции – гнев, страх, грусть, отвращение, удивление, предвкушение, доверие и радость. Кроме основных, существует большое количество дополнительных эмоций. На рис. 4 показан графический интерфейс пользователя MoodPatrol.

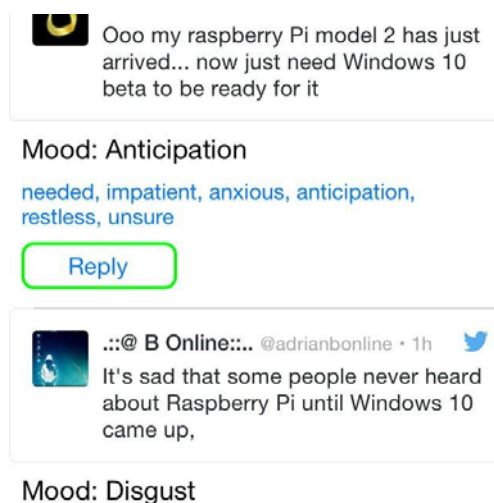


Рис. 4. Пример MoodPatrol API [25]

Для сравнения систем используется модель измерения производительности DMSM. Модель была разработана в 1992 году Delone и Mclean как модель успеха информационных систем Delone и Mclean (DMSM) [26]. Первоначально они предусмотрели шесть показателей для производительности системы, которые включают в себя качество системы, качество информации, использование, удовлетворенность пользователей, индивидуальное влияние системы на общество и влияние на организации. Позже некоторые компоненты были добавлены, а некоторые удалены другими исследователями [27–30]. В данной работе используются отдельные компоненты этих систем для анализа их производительности, как указано в табл. 1.

Таблица 1. Показатели эффективности различных систем, используемых для анализа эмоций

Показатель эффективности	Power BI	Tableau	Qemotions	MoodPatrol
Пользователи	200,000	86 000	10000	20000
Анализ типов эмоций	6	8	12	11
Компонент анализа эмоций (в %)	15%	10%	70%	60%
Отзывы пользователей	115	85	25	10
Среднее время отклика при загрузке страницы	3,96 sec	4,5 sec	2,5 sec	3,02 sec
Доступность	Низкая	Низкая	Высокая	Низкая
Надежность	Низкая	Высокая	Высокая	Средняя
Точность	Высокая	Средняя	Средняя	Средняя
Эффективность	Высокая	Средняя	Средняя	Средняя

Обсуждение

В наблюдаемых системах в основном используется эмоциональный анализ только как часть всех функций, которые они предоставляют. Они не доступны бесплатно в интернете, однако в некоторых случаях пробную версию можно использовать в те-

чение определенного времени. Они выявляют эмоции и классифицируют их по основным категориям. Углубленный анализ, показывающий связь между эмоциями, их силой, частотой и другими характеристиками, не изучен. Доступные системы не дают заключения, основанного на эмоциях или обобщении

текста относительно эмоций. В большинстве случаев нет предварительной обработки текста перед анализом, что может повлиять на эффективность такого анализа эмоций, если текст плохо написан и содержит ненужные символы. До представленного исследования не существовало реальных механизмов для улучшения взаимодействия человека с машиной.

Методология (наша работа)

Сегодня возникает большая заинтересованность в создании системы, которая может анализировать эмоции лучше, чем доступные системы. Приведенный выше анализ систем и их характеристик помогает определить компоненты, требующие особого внимания для повышения их производительности.

В процессе исследования было замечено, что предварительная обработка текста является важной

фазой анализа текста. Система, которая анализирует динамический текст без надлежащей предварительной обработки, может привести к сбоям в работе и повлиять на эффективность анализа, если текст содержит нежелательные символы, с ошибками, неправильными окончаниями, отрицаниями и т. д. Кроме того, предварительная обработка текста определяет его структуру, улучшает систему производительности.

Созданный алгоритм предварительной обработки был применен к блогу из интернета, и проведено наблюдение влияния на различные компоненты текста до и после его обработки. Результаты приведены в табл. 2.

Таблица 2. Влияние на различные компоненты текста до и после его обработки

Компонент текста	Перед предварительной обработкой	После предварительной обработки	Процент изменений
Количество букв в тексте	9560	11052	15,61%
Количество слов в тексте	2152	2300	6,88%
Количество предложений	530	570	7,55%
Количество эмоций	426	467	9,62%
Типы эмоций	9	10	11,11%
Количество пунктуации	55	40	27,27%
Типы пунктуации	8	6	25,00%
Количество пробелов	1121	1110	0,98%
Количество символов	40	52	30,00%
Нераспознанные символы	21	14	33,33%

После наблюдений был выявлен процент изменений в разных компонентах текста, но без предварительной обработки результаты анализа не будут точными. Чтобы проанализировать взаимосвязь между двумя группами наблюдений, к ним был применен Т-тест. Результаты приведены в табл. 3.

Таблица 3. Результаты Т-теста на две группы

Статистические показатели	Группа 1	Группа 2
Mean	1392,10	1562,00
Variance	8707721,43	11646441,11
Observations	10,00	10,00
Pearson Correlation	1,00	
Hypothesized Mean Difference	0,00	
Df	9,00	
t Stat	-1,15	
P(T<=t) one-tail	0,14	
t Critical one-tail	1,83	
P(T<=t) two-tail	0,28	
t Critical two-tail	2,26	

Т-тест используется для определения существенной разницы между двумя группами наблюдений. В рамках Т-теста наблюдается его Р-значение. Если Р-значение (P value) меньше порогового значения 0,05, это означает, что между двумя группами существует большая разница. Если Р-значение высокое, это означает, что между двумя наблюдениями нет существенной разницы. Различия между двумя группами обусловлены только случайностью, а не какой-то другой переменной. Данная предварительная об-

работка была применена, чтобы избежать случайных ошибок в расчетах. Математический расчет Т-теста выполняется по формуле, приведенной ниже:

$$t = \frac{(\sum D)/N}{\sqrt{\frac{\sum D^2 - \left(\frac{(\sum D)^2}{N}\right)}{(N-1)(N)}}}$$

Были проанализированы психолингвистические характеристики текста, чтобы создать интеллектуальный лингвистический интерфейс для взаимодействия человека с машиной. Основываясь на этих характеристиках текста и эмоций, машина понимает особенности человеческого поведения и эффективно реагирует на них.

Основная идея работы – улучшение взаимодействия человека и машины. Эта область становится реальностью в 1970-х годах. Карлайл и др. в работе по оценке влияния автоматизации делопроизводства на общение высшего руководства впервые использовали этот термин [31]. Позже, в 1980-х годах, Стюарт К. Кард, Аллен Ньюэлл и Томас П. Моран использовали термин в своей работе над моделью для определения времени работы пользователя с интерактивными системами [32]. С тех пор эта область развивалась вместе с развитием компьютеров и увеличением роли машин в жизни человека.

При взаимодействии человека с машиной возникает особенный интерес к некоторым областям, которые требуют доработок с обеих сторон. Со стороны машины актуальны методы в компьютерной графике, операционных системах, языках программирования и средах разработки. С человеческой стороны актуальны теория коммуникации, дисциплины графического и промышленного дизайна, лингвистика, общественные науки, когнитивная психология. Усилия направлены на создание подходящего интерфейса для этого взаимодействия. Существует несколько проектов в процессе взаимодействия человека с машиной, таких как Lumiere Project, который является адаптивной системой для рассуждения об изменяющихся во времени целях пользователей компьютеров из их наблюдаемых действий и запросов [33]. Аналогичным образом интерфейс AVANTI взаимодействует с незрячими людьми или людьми с серьезными нарушениями двигательной активности [34]. Еще одним примером таких проектов является интерактивное рабочее пространство, которое представляет собой экспериментальную исследовательскую лабораторию iRoom, расположенную в здании Gates Information Sciences в Стэнфорде. Она предоставляет виртуальную среду для стола в конференц-зале [35]. Исследователями не уделялось особого внимания в области создания интерфейса для взаимодействия человека с машиной для анализа эмоций из текста на основе изучения человеческого поведения и личности.

Сегодняшняя цель – сделать машины технически способными изучать поведение людей и их эмоции,

выделенные из текста, а также разумно реагировать на задаваемые вопросы, отвечая всем потребностям. Для этого машины снабжаются средствами, которые помогают им классифицировать поведение человека на основе его эмоций. Исследование эмоционального общения проводится для того, чтобы исследовать природу людей. Данные, полученные при этом, можно использовать для классификации индивидуальной человеческой личности, чтобы улучшить реакцию машины на нее.

При взаимодействии «человек – машина» с учетом потребностей человека и его интеллектуальных и личностно-чувствительных характеристик можно предсказать наиболее вероятный выбор и решение, к которому склоняется человек, после получения знания о его типе характера, определяемого через эмоции. Эти данные можно использовать для адаптации способа дальнейшего диалога и других видов взаимодействия так, чтобы максимизировать релевантность ответов человека.

Выявление человеческих потребностей также необходимо для развития интеллектуальных систем, которые ориентированы на формирование знаний и повышение интеллекта.

Согласно исследованиям, одни люди открыты для выражения своих эмоций во время общения, тогда как другие более закрыты. В связи с этим людей можно разделить на два типа: экстраверты и интроверты. В приведенной ниже табл. 4 показаны различные характеристики людей экстравертного типа, а также слова, используемые ими в тексте, указывающие на эти характеристики.

Таблица 4. Характеристики и слово, фазы, используемые экстравертом

Экстравертный тип	
Характеристики	Пример использования слов и фраз в общении
Доминирующий тип	Я, мой, хочу, нужно, решить, выбрать, управлять, вести, рекомендовать, мощный, выбор, большой, огромный, мой собственный, мне нравится, мое мнение, все, каждый, красивый, замечательный, исключительный, революционный, идея, сюрприз, неожиданный, вдохновляющий, экстра-, супер-, мега-, гипер-, ультра-, создать, мечтать, странный, объяснить, проверить, сравнить, понять, технология, тайна, отношение, исследовать, я хотел бы знать, неправильно, искренне, не согласен, ошибка, падение, разрыв, замечание, уведомление, наблюдение, нечестность, жесткий, острый, критический, соперник, конкурентный, обострить, сожалеть, помочь, жалеть кого-то, вызов, трудный, смелый, мужество, риск, опасность, герой, защищать, дать, передать, подарок, помощь, благотворительность пожертвовать, уникальный, любимый, особенный, выборочный, выбранный, награда, продвижение по службе, разнообразить, акцент, другой
Максималистский тип	
Вдохновляющий тип	
Обнаружения тип	
Проверки тип	
Зачочки тип	
Эмпатический тип	
Смелый тип	
Щедрый тип	
Предпочтительный тип	

Обобщенные эмоции предыдущих этапов используются для машинного обучения. Наряду с эмоциональными характеристиками текста, машина запрограммирована на психолингвистические характеристики поведения человека. Человек при общении использует определенные слова и фразы, основанные на его поведенческих характеристиках. Люди с разным психологическим типом используют в своем общении с окружающими определенные поведенческие шаблоны. Философы и психологи перечислили слова, используемые различными группами людей во время общения. Машина изучает эти слова и после

выявления, классификации и обобщения эмоций, выделенных из текста, сравнивает их, чтобы понять поведение пользователя системы на основании его характеристик поведения. В приведенной табл. 4 показаны различные характеристики людей экстравертного типа, а также слова, используемые в тексте для отображения этих характеристик.

В табл. 5 показаны различные характеристики людей интровертного типа, а также слова, которые используются в тексте для отображения этих характеристик.

Таблица 5. Характеристики и слова, фазы интроверта

Интровертный тип	
Характеристики	Пример использования слов и фраз в общении
Подчиненный тип	Он, она, ваши, их, эксперты, мнения, статистика, другие, что бы вы мне посоветовали, предложение, просто, любое, минимум, неважно, меньше, средний, нормальный, стандартный, общий, типичный, предыдущий, следующий, последний, один за другим, по порядку, заметки, разделить, разложить, по времени, дате, крайнему сроку, тем же, доказанным, традицией, постоянная, терпимость, стабильная, конвенция, доверие, приказы, апелляция, все, присоединиться, объединить, классифицировать, мало, почти, легко, круг, примирить, уменьшить, соответствовать, я думаю об этом, вы не возражаете, воздержитесь, закончить, быстро, кратко, спешить, конкретно, точно, производительность, эффективно, сомневаться, осторожно, опасаться, предупредить, ограничить, угроза, безопасная охрана, слишком много, процент, ставка, специальные предложение, вес, уровень, оценка, сортировка, судья, закон, правило, погашение, последовательно, экономия, банк, комиссия, экономия топлива, энергосбережение, минимизация, выравнивание, оценка, справедливость, равномерно, правильно, дешево, продвижение по службе, переработка, исключение, рентабельно, энергоэффективно.
Минималистский тип	
Систематический тип	
Консервативный тип	
Не замечающий тип	
Гармоничный тип	
Фактический тип	
Защитный тип	
Экономный тип	
Тип балансировки	

Характеристики человеческого поведения вместе со словами и выражениями, которые приведены в таблицах, рассматриваются как психолингвистический словарь для данного исследования. Процесс начинается с определения психолингвистических характеристик текста. Программа для психолингвистического словаря сравнивает и идентифицирует компоненты текста. Затем фразы извлекаются из текста, используя N-граммы.

N-граммы – это алгоритм прогнозирования слов, использующий вероятностные методы для прогнозирования следующего слова после наблюдения N – 1 слов. Поэтому вычисление вероятности следующего слова тесно связано с вычислением вероятности последовательности слов. Самый простой способ оценить вероятности – это использовать оценку максимального правдоподобия (MLE), основанный на взятии отсчетов из корпуса текста и нормализации их, что лежат в интервале [0, 1]. Например, чтобы вычислить вероятность биграммы слова y , следующего за x , нужно посчитать биграммы $c(xy)$ из корпуса и нормализовать их по числу биграмм, которое начинается с x .

$$P(W_n | W_{n-1}) = \frac{C(W_{n-1}W_n)}{C(W_{n-1})}$$

В знаменателе $C(W_{n-1})$ представляет количество биграмм, начинающихся с W_{n-1} , т. к. биграммы, начинающиеся с W_{n-1} равны числу, которое происходит в W_{n-1} . Общее уравнение для оценки вероятности для MLE N-грамма имеет вид:

$$P(W_n | W_{n-1}) = \frac{C(W_{n-1}W_n)}{C(W_{n-1})}$$

После определения n -грамм фраз и слов, выделенных из текста, каждому компоненту присваивается вес на основе частоты его появления в тексте. После суммирования веса всех компонентов поведение пользователя может быть идентифицировано на основе психолингвистических характеристик текста.

Классификатор метода опорных векторов (SVM) используется для разделения экстравертных и интровертных характеристик текста. Данная модель анализирует текст разных людей, взаимодействующих с машиной. Машина визуализирует текст, как на картинке ниже (рис. 5).

Можно заметить, что рисунок и график с правой стороны представляют текст интроверта, тогда как рисунок и график с левой стороны представляют текст экстраверта. Из графика можно вычислить, что экстравертный человек выражает эмоции с различной интенсивностью. Он более выразителен и общается открыто. Тогда как частота эмоций, используемых интровертом, почти одинакова. Это указывает на его личностные характеристики. Он более осторожен в выражении своих эмоций. Машина должна перенимать психолингвистические характеристики текста и общаться соответствующим образом. В случае интроверта машина должна быть более выразительной во время взаимодействия и отвечать более кратко. Необходимо также принимать во внимание, что в случае экстраверта машина может быть сфокусирована на теме и отвечать требуемым образом.

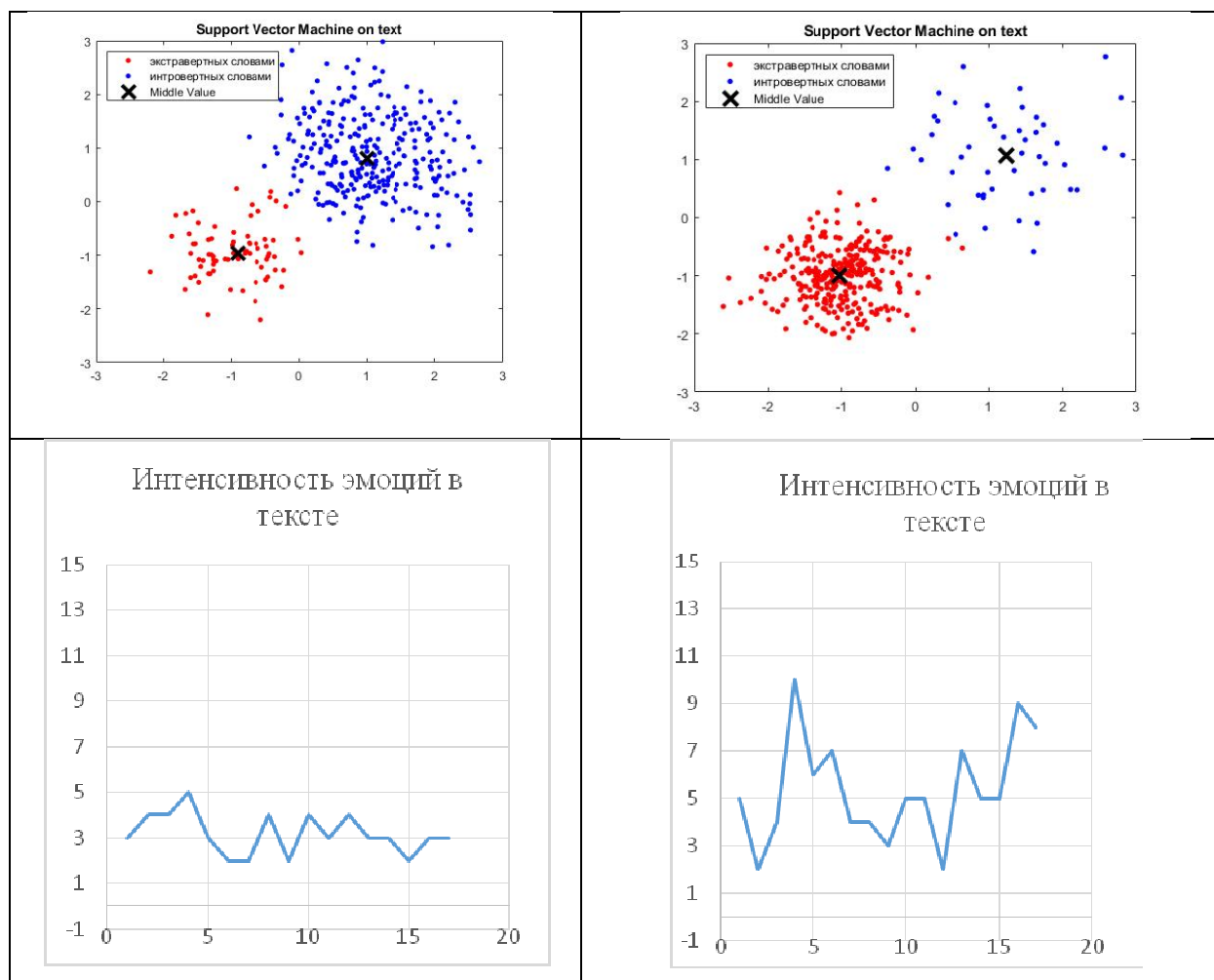


Рис. 5. Результаты применения SVM к тексту для выявления его психолингвистических характеристик

Машину можно тренировать или «кормить» словами и фразами, представляющими определенное эмоциональное состояние человека. Машина, получая запрос от клиента (человека), корректирует реакцию на основе эмоциональных характеристик слов, выраженных в запросе. Таким образом, общение будет более эффективным, дружелюбным, и человек может потребовать ответа и внимания.

Заключение

В данном исследовании были проанализированы доступные системы для демонстрации эмоций, извлеченных из текста, и проведено сравнение между ними, предложен механизм анализа таких эмоций и рассмотрены получившиеся результаты в блоге текста. Так, например, индивидуально характеристики или механизмы каждой существующей системы подлежат наблюдению, и обобщенные результаты из табл. 1 используются для создания системы с большим количеством характеристик и лучшей производительностью.

После изучения результатов было выявлено, что данная система обеспечивает лучший и более подробный анализ эмоций, извлеченных из текста. А также замечено, что предварительная обработка облегчает процесс анализа эмоций из текста. Наблюдается увеличение количества эмоций, выявленных

после предварительной обработки текста, примерно на 10 %. В то время как в других системах, изученных ранее, оно либо отсутствует, либо существует слабо.

Рассматриваемая методология различает личность пользователя, классифицирует ее и определяет механизм для улучшения взаимодействия между человеком и машиной. Это обеспечивает основу для развития умной, самообщающейся машины с человеком.

При сравнении с другими системами, доступными для анализа эмоций, заметно, что существует нехватка характеристик обучения человеческому поведению. Так, возможностей для таких же требований, как к человеку, для машинного общения пока недостаточно. При использовании созданной системы наблюдается 55 % увеличения взаимодействия «человек – машина».

В целом можно сказать, что система, созданная в ходе этого исследования, поддерживает характеристики, которые отсутствуют в других системах. Использование механизма взаимодействия человека с машиной дает системам преимущество над другими существующими системами.

В будущем интерес к такому анализу различных типов текста на основе темы и цели не потеряет сво-

ей актуальности, а следовательно, будет необходимость в создании новых математических моделей для повышения производительности и объема такой системы.

Библиографические ссылки

1. Stone P.J., Dunphy D.C., Smith M.S. (1966). *The General Inquirer: A Computer Approach to Content Analysis* / MIT Press - Cambridge, 519 p.
2. Wiebe, Janyce M. (1990). Identifying Subjectivity characters in Narrative // Proc. 13th International Conference on Computational Linguistics. Helsinki, pp. 401-408.
3. Vasileios H., Kathleen R. M. (1997). Predicting the Semantic Orientation of Adjectives // Proc. 8th Conference on European chapter of the Association for Computational Linguistics. Spain, pp 174-181.
4. Peter D. T. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews // Proc. of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, pp. 417-424.
5. Pang B., Lee L. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques // Proc. Conference on Empirical Methods in Natural Language Processing. Philadelphia, pp. 79-86.
6. Denecke K. (2009). Are SentiWordNet scores suited for multi-domain sentiment classification? // Proc. 4th International Conference on Digital Information Management. USA, pp. 33-38.
7. Abbasi M.M., Beltiukov A.P. (2018). Механизм предварительной обработки текста перед анализом настроений // Proc. 6th Всероссийская конференция, Информационные технологии интеллектуальной поддержки принятия решений. Уфа ; Ставрополь, Россия.
8. Boley D., Gini M., Gross R. (1999) Partition based clustering for web document categorization // Elsevier Journal for Decision Support Systems , 27 (3), 329-341.
9. Beltiukov A.P., Abbasi M.M. (2019). Logical analysis of emotions in text from natural language // Vestnik Udmurtskogo Universiteta Matematika Mekhanika Komp'yuternye Nauki, 29 (1), 106-116.
10. Ermakov A. (2009). Knowledge extraction from text and its processing: Current state and prospects // Proc. of the Computational Linguistics and Intellectual Technologies. 2009, pp. 50-55.
11. Pivovarova L., Yangarber R. (2013). Adapting the PULS event extraction framework to analyze Russian text // Proc. of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing. Sofia, Bulgaria, 8-9, pp. 100-109.
12. Zagibalov., Taras., Belyatskaya et al. (2010). Comparable English-Russian Book Review Corpora for Sentiment Analysis. Russia // Proc. of the 1st Workshop on Computational Approaches to Subjective and Sentiment Analysis, Lisbon, Portugal pp.67-72.
13. Steinberger J., Lenkova P., Kadjajov M. (2011). Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora // Proc. of Recent Advances in Natural Language Processing. Bulgaria, pp. 770-775.
14. Chetviorkin I., Braslavskiy P., Loukachevich N. (2012). Sentiment Analysis Track at ROMIP 2011 // Computational Linguistics and Intellectual Technologies. Proc. of the International Conference (Dialog). Bekasovo, pp. 1-14.
15. Sokolova M., Bobicev V. (2009). Classification of emotion words in Russian and Romanian languages // International Conference RANLP. Borovets, Bulgaria, pp. 416-420.
16. Abbasi M. M., Beltiukov A.P. (2017). Analysis of sentiment and emotion from text written in Russian language // Proc. of the 5th All Russian Conference on Information technology for intelligent decision making support (ITIDS). Ufa, Russian Federation, vol. 1, issue 1, page 42-47.
17. Abbasi M.M., Beltiukov A.P. (2019). Analysis of emotions from text in Russian Language using syntactic methods // Proc. of the 7th International Science Conference. Information Technology and Systems. Khanty-Mansiysk , Russia, pp. 137-142.
18. Jaruskululchi C., Kruengkrai. (2003). Generic text summarization using local and global properties of sentences // IEEE/WIC International conference on web intelligence. pp.13-16.
19. Kiabod M., Naderi M., Sharafi S.M. (2012). A novel method of significant words identification in text summarization // Journal of Emerging Technologies in Web Intelligence, 4 (3).
20. Abbasi M.M., Beltiukov A.P. (2018). Analyzing emotions from text corpus using WordSpace // Proc. of the 20th international workshop on computer science and information technologies CSIT. Bulgaria, Varna, vol. 3, issue 4, pp. 161-164.
21. Abbasi M.M., Beltiukov A.P. (2019). Summarizing Emotions from Text Using Plutchik's Wheel of Emotions // Advances in Intelligent systems research, In Proceedings of the 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS). Atlantis press, vol. 166, pp. 291-294.
22. Patel N. How to get actionable data from google analytics in 10 minutes // [Электронный ресурс] URL: <https://neilpatel.com/blog/how-to-get-actionable-data-from-google-analytics-in-10-minutes/> (дата обращения: 20.04.2018).
23. Gil Raviv. Sentiment Analysis in Power BI – Part 2 // [Электронный ресурс] URL: <https://datachant.com/2016/08/09/sentiment-analysis-power-bi-part-2/> (дата обращения: 10.02.2019).
24. Tableau Desktop: Start your free 14-Day trial // [Электронный ресурс] URL: <https://www.tableau.com/products/desktop/download#system-requirements> (дата обращения: 05.04.2019).
25. Q emotionapp login // [Электронный ресурс] URL: <https://twitter.com/qemotionapp> (дата обращения: 11.04.2019).
26. MoodPatrol login // [Электронный ресурс] URL: <https://twitter.com/hashtag/MoodPatrol?src=hash> (дата обращения: 19.04.2019).
27. DeLone W.H., McLean E.R. (1992). Information systems success: The quest for the dependent variable // Information Systems Research, 3(1), 60-95.
28. Rai A., Lang S. S., Welker R B. (2002). Assessing the Validity of IS Success Models : An Empirical Test and Theoretical Analysis // Information Systems Research, 13(1), 50-69.
29. Bryd T.R., Thrasher E.H., Lang T., Davidson N.W. (2006). A process-oriented perspective of IS success: Examining the impact of IS on operational cost. Omega, 34 (1), 448-460.
30. Wu J H., Wang Y M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model // Information & Management, 43(1),728-739.
31. Lai J.Y., Yang C.C., Tang W.S. (2008). Exploring the Effects of Dependability on Enterprise Applications Success in e-Business // SIGMIS-CPR'06. Claremont, California, USA. pp. 244-52. DOI:10.1145/1125170.1125229.
32. Carlisle M., James H. (June 1976). Evaluating the impact of office automation on top management

communication // Proc. of the June 1st conference on National computer conference and exposition on – AFIPS, 611–616. https://en.wikipedia.org/wiki/Digital_object_identifier. DOI: 10.1145/1499799.1499885.

33. Card., Stuart K., Thomas P., Moran, Allen N. (July 1980). The keystroke-level model for user performance time with interactive systems // Communications of the ACM, 23 (7), 396–410. https://en.wikipedia.org/wiki/Digital_object_identifier. DOI: 10.1145/358886.358895.

34. Horvitz E., Breese., Jack., Heckerman., et al. (July 1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users // Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, Madison, WI, San Francisco, 256-65.

35. Stephanidis C. et al. (1998). Adaptable and adaptive user interfaces for disabled users in the AVANTI project // Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 1430(1), 153-166. URL: <https://doi.org/10.1007/BFb0056962>. Johanson, B., Fox A. (2002). The Event Heap: A Coordination Infrastructure for Interactive Workspaces // Proc. of the 4th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA).

References

1. Stone P.J., Dunphy D.C., Smith M.S. (1966). The General Inquirer: A Computer Approach to Content Analysis / MIT Press - Cambridge, 519 p.

2. Wiebe, Janyce M. (1990). Identifying Subjectivity characters in Narrative. In Proc. 13th International Conference on Computational Linguistics. Helsinki, pp. 401-408.

3. Vasileios H., Kathleen R. M. (1997). Predicting the Semantic Orientation of Adjectives. In Proc. 8th Conference on European chapter of the Association for Computational Linguistics. Spain, pp. 174-181.

4. Peter D. T. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proc. of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, pp. 417-424.

5. Pang B., Lee L. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proc. Conference on Empirical Methods in Natural Language Processing. Philadelphia, pp. 79-86.

6. Denecke K. (2009). Are SentiWordNet scores suited for multi-domain sentiment classification? In Proc. 4th International Conference on Digital Information Management. USA, pp. 33-38.

7. Abbasi M.M., Beltiukov A.P. (2018). *Mekhanizm predvaritel'noi obrabotki teksta pered analizom nastroenii* [Text preprocessing mechanism before mood analysis]. Proc. 6th Всероссийская конференция, Информационные технологии интеллектуальной поддержки принятия решений. Уфа-Ставрополь, Россия.

8. Boley D., Gini M., Gross R. (1999) Partition based clustering for web document categorization // Elsevier Journal for Decision Support Systems, 27 (3), 329-341.

9. Beltiukov A.P., Abbasi M.M. (2019). Logical analysis of emotions in text from natural language. In Vestnik Udmurtskogo Universiteta Matematika Mekhanika Komp'yuternye Nauki, 29 (1), 106-116.

10. Ermakov A. (2009). Knowledge extraction from text and its processing: Current state and prospects. In Proc. of the Computational Linguistics and Intellectual Technologies. 2009, pp. 50-55.

11. Pivovarova L., Yangarber R. (2013). Adapting the PULS event extraction framework to analyze Russian text. In Proc. of the 4th Biennial International Workshop on Balto-

Slavic Natural Language Processing. Sofia, Bulgaria, 8-9, pp. 100–109.

12. Zagibalov., Taras., Belyatskaya et al. (2010). Comparable English-Russian Book Review Corpora for Sentiment Analysis. Russia. Proc. of the 1st Workshop on Computational Approaches to Subjective and Sentiment Analysis, Lisbon, Portugal pp.67-72.

13. Steinberger J., Lenkova P., Kabadjov M. (2011). Multi-lingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora. Proc. of Recent Advances in Natural Language Processing. Bulgaria, pp. 770-775.

14. Chetviorkin I., Braslavskiy P., Loukachevich N. (2012). Sentiment Analysis Track at ROMIP 2011. In Computational Linguistics and Intellectual Technologies. Proc. of the International Conference (Dialog). Bekasovo, pp. 1-14.

15. Sokolova M., Bobicev V. (2009). Classification of emotion words in Russian and Romanian languages // International Conference RANLP. Borovets, Bulgaria, pp. 416–420.

16. Abbasi M. M., Beltiukov A.P. (2017). Analysis of sentiment and emotion from text written in Russian language. In Proc. of the 5th All Russian Conference on Information technology for intelligent decision making support (ITIDS). Ufa, Russian Federation, vol. 1, issue 1, page 42-47.

17. Abbasi M.M., Beltiukov A.P. (2019). Analysis of emotions from text in Russian Language using syntactic methods. Proc. of the 7th International Science Conference. Information Technology and Systems. Khanty-Mansiysk, Russia, pp.137-142.

18. Jaruskululchi C., Kruengkrai. (2003). Generic text summarization using local and global properties of sentences. In IEEE/WIC International conference on web intelligence. pp.13-16.

19. Kiabod M., Naderi M., Sharafi S.M. (2012). A novel method of significant words identification in text summarization. In Journal of Emerging Technologies in Web Intelligence, 4 (3).

20. Abbasi M.M., Beltiukov A.P. (2018). Analyzing emotions from text corpus using WordSpace. Proc. of the 20th international workshop on computer science and information technologies CSIT. Bulgaria, Varna, vol. 3, issue 4, pp. 161-164.

21. Abbasi M.M., Beltiukov A.P. (2019). Summarizing Emotions from Text Using Plutchik's Wheel of Emotions. In Advances in Intelligent systems research, In Proceedings of the 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS). Atlantis press, vol. 166, pp. 291-294.

22. Patel N. How to get actionable data from google analytics in 10 minutes. Available at: <https://neilpatel.com/blog/how-to-get-actionable-data-from-google-analytics-in-10-minutes/> (accessed 20.04.2018).

23. Gil Raviv. Sentiment Analysis in Power BI – Part 2. Available at: <https://datachant.com/2016/08/09/sentiment-analysis-power-bi-part-2/> (accessed 10.02.2019).

24. Tableau Desktop: Start your free 14-Day trial. Available at: <https://www.tableau.com/products/desktop/download#system-requirements> (accessed 05.04.2019).

25. Q emotionapp login. Available at: <https://twitter.com/qemotionapp> (accessed 11.04.2019).

26. MoodPatrol login Available at: <https://twitter.com/hashtag/MoodPatrol?src=hash> (accessed 19.04.2019).

27. DeLone W.H., McLean E.R. (1992). Information systems success: The quest for the dependent variable // Information Systems Research, 3(1), 60-95.

28. Rai A., Lang S. S., Welker R B. (2002). Assessing the Validity of IS Success Models : An Empirical Test and

Theoretical Analysis. In *Information Systems Research*, 13(1), 50-69.

29. Bryd T.R., Thrasher E.H., Lang T., Davidson N.W. (2006). A process-oriented perspective of IS success: Examining the impact of IS on operational cost. *Omega*, 34 (1), 448-460.

30. Wu J H., Wang Y M. (2006). Measuring KMS success: A respecification of the DeLone and McLean's model. In *Information & Management*, 43(1), 728-739.

31. Lai J.Y., Yang C.C., Tang W.S. (2008). Exploring the Effects of Dependability on Enterprise Applications Success in e-Business. In *SIGMIS-CPR'06*. Claremont, California, USA. pp. 244-52. DOI:10.1145/1125170.1125229.

32. Carlisle M., James H. (June 1976). Evaluating the impact of office automation on top management communication // Proc. of the June 1st conference on National computer conference and exposition on – AFIPS, 611-616. Available at: https://en.wikipedia.org/wiki/Digital_object_identifier. DOI: 10.1145/1499799.1499885.

33. Card, Stuart K., Thomas P., Moran, Allen N. (July 1980). The keystroke-level model for user performance time with interactive systems // *Communications of the ACM*, 23 (7), 396–410. Available at: https://en.wikipedia.org/wiki/Digital_object_identifier. DOI: 10.1145/358886.358895.

34. Horvitz E., Breese., Jack., Heckerman., et al. (July 1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. Proc. of the 14th Conference on Uncertainty in Artificial Intelligence, Madison, WI, San Francisco, 256-65.

35. Stephanidis C. et al. (1998). Adaptable and adaptive user interfaces for disabled users in the AVANTI project // *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 1430(1), 153-166. Available at: <https://doi.org/10.1007/BFb0056962>. Johanson, B., Fox A. (2002). The Event Heap: A Coordination Infrastructure for Interactive Workspaces. Proc. of the 4th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA).

Studying the Systems Available to Analyze Emotions from Text and Provide Mechanism for Improving Man Machine Interaction

M. M. Abbasi, Post-graduate, Udmurt State University, Izhevsk, Russia

A. P. Beltiukov, DSc (Physics and Mathematics), Professor, Udmurt State University, Izhevsk, Russia

The development of information technology enables us to develop systems for analysis and data processing. Today, the main source of available data is the Internet. Researchers have developed online and offline systems to analyze this data. Data analysis can be used for various purposes. Our work focuses on the study of systems that are used to analyze emotions from the text. In this paper, we analyze these systems and, based on our research, propose mechanisms to increase their characteristics and improve the scope, as well as compare their performance.

An analysis of emotions extracted from the text can be used to predict future events, people's reviews of a product or service, identify a group of people by interests and develop a machine that can mimic the behavior of human emotions. Our basic goal is to improve the mechanism of man machine interaction while communication using text. We propose a mechanism that improves the interaction of man and machine by determining the psycholinguistic characteristics of the text that represent human behavior. This mechanism will study the relationship between emotions and the psycholinguistic characteristics of a text. It will facilitate the process of human-machine interaction.

Keywords: text, emotions, blogs, communication, topic, analysis.

Получено: 13.11.19