

УДК 004.89+004.514+004.52+159.942

ВЫЯВЛЕНИЕ ПСИХОЛОГИЧЕСКОГО ПОРТРЕТА НА ОСНОВЕ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ СООБЩЕНИЙ ДЛЯ АНТРОПОМОРФНОГО СОЦИАЛЬНОГО АГЕНТА

А.А. Алексеев¹, В.В. Кугуракова², Д.С. Иванов³

Казанский (Приволжский) федеральный университет

¹aleksanton5@gmail.com; ²vlada.kugurakova@gmail.com;

³denis.999.ivanov@gmail.com

Аннотация

Исследованы аспекты выявления психологического портрета респондента и генерация отношения к нему социальным агентом на основе анализа тональности диалога. Рассмотрены принципы генерации отношения социального агента к респонденту и изменение его эмоционального настроения на протяжении общения. Реализация поставленной задачи осуществлена с помощью языка программирования Python и работы с реальными данными. Проведенный анализ алгоритмов классификации, основанных на подходах машинного обучения, подтвердил практическую значимость работы.

Ключевые слова: *социальный агент, тональность, эмоциональное отношение, машинное обучение*

ВВЕДЕНИЕ

Компьютерные технологии развиваются огромными темпами, уже сейчас они окружают нас практически везде. Компьютеры могут управлять как домашним телевизором, так и запуском ракеты в космос. Такое бурное развитие информационных технологий привело к потребности в новых интерфейсах, которые должны упростить взаимодействие компьютера и человека, проникнуть в нашу жизнь и автоматизировать её ещё сильнее. Одним из решений данной проблемы можно считать антропоморфные интерфейсы (интерфейсы, по своему облику похожие на человека), позволяющие полностью или частично (в зависимости от реализации) отказаться от стандартных компьютерных интерфейсных устройств

(клавиатура, мышь, джойстик и т. д.) и за счет этого упростить взаимодействие с пользователем.

Разработка антропоморфного социального агента – это кросс-дисциплинарная, трудоемкая и ресурсозатратная задача. Вот лишь несколько составляющих человеко-машинного интерфейса: распознавание, анализ и симуляция человеческой речи; симуляция мимики, движений человека, его психоэмоциональных состояний и реакций; рассуждения на основе здравого смысла, поддержание контекста, диалоговый режим [1].

В настоящей работе исследованы такие аспекты создания антропоморфного пользовательского интерфейса, как выявление психологического портрета респондента и генерация отношения к нему социальным агентом. Эта задача далеко нетривиальна. Определить такой портрет собеседника можно разными способами: по интонации речи, мимике лица, текстовому диалогу. Ниже решается проблема определения эмоционального состояния собеседника на основе текста. Ее можно свести к задаче анализа тональности текста, учитывая особенности, присущие выбранной теме. Кроме того, рассмотрены принципы генерации отношения социального агента к респонденту и изменение его эмоционального настроения на протяжении общения.

АНАЛИЗ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ ТЕКСТА

Анализ эмоциональной окраски текста – это раздел компьютерной лингвистики, занимающийся автоматическим определением в тексте эмоционально окрашенной лексики и выявлением мнений. В семантическом анализе принято считать, что текстовую информацию можно разделить на два класса: мнения и факты [7]. Главным в нашей работе будет понятие определения мнения. В анализе тональности текста простым мнением называется кортеж из пяти элементов:

(entity, feature, sentiment value, holder, time),

где entity – объект, feature – аспект, sentiment value – эмоция, holder – автор, time – время. Далеко не всегда приходится искать все пять названных элементов сложными методами, зачастую некоторые из них бывают известны сразу. Например, при исследовании отзывов о каком-нибудь фильме могут быть изначально известны такие параметры, как объект, автор и время.

В задаче определения психологического портрета респондента главной

проблемой является правильная классификация такого элемента, как эмоция. Именно по ней будет проходить классификация текста.

Классификация текста – задача информационного поиска, которая заключается в определении, к какой категории относится текст. Она производится на основании содержания этого текста. Чтобы произвести правильную классификацию текста, необходимо: выбрать подход к классификации; определить количество классов и структуру классификации; выбрать признаки, относительно которых производится классификация.

ВЫЯВЛЕНИЕ ПСИХОЛОГИЧЕСКОГО ПОРТРЕТА РЕСПОНДЕНТА

В диалоге с респондентом антропоморфному социальному агенту особо важно почувствовать его эмоциональное состояние, чтобы правильно вести диалог: ведение диалога с собеседником на необходимом эмоциональном уровне позволит вызвать доверие к агенту [1]. В психологии существует огромное множество различных классификаций психотипов. Для определения некоторых из них применяются довольно сложные методы, которые трудно или совершенно невозможно реализовать, используя один лишь текст. Поэтому, исследуя тему выявления психотипа респондента, нам предстоит выбрать либо существующую систему классификации, либо выработать собственную, которая сможет с достаточной четкостью определять эмоциональный настрой собеседника в текстовом диалоге. Было принято решение понимать под психотипами окраску эмоций и их силу. Таким образом, предстоит выработать собственную классификацию. Для решения этой задачи выбрана классификация из пяти классов, которых должно быть достаточно для точного описания, в каком настрое находится собеседник социального агента. Решено разделять текст по следующим эмоциональным категориям: сильно положительный; средне положительный; нейтральный; средне отрицательный; сильно отрицательный.

ГЕНЕРАЦИЯ ОТНОШЕНИЯ СОЦИАЛЬНОГО АГЕНТА

Генерация отношения, как и выявление, может быть реализовано различными способами, например, можно всегда отвечать в нейтральном стиле, однако такой способ не будет антропоморфным – человекоподобный социальный агент должен реагировать на изменение эмоционального настроения его собеседника. В связи с этим необходимо разработать алгоритм, который позволит социальному

агенту изменять стиль общения в зависимости от психотипа его собеседника. Таким образом, первостепенная задача – правильно определить эмоциональную окраску сообщений собеседника.

Большинству людей приятно общаться с положительными людьми и на одном эмоциональном уровне [10]. На этом основании был разработан собственный принцип генерации отношения антропоморфного социального агента, который реализован по следующему алгоритму:

- если психотип собеседника находится в положительно-нейтральной группе (сильно положительный, средне положительный, нейтральный), то отношение к нему со стороны агента будет на том же уровне;
- если психотип собеседника находится в отрицательной группе (сильно отрицательный, средне отрицательный), то отношение к нему со стороны агента – нейтральное с попыткой выяснить причину отрицательного настроения.

ОПРЕДЕЛЕНИЕ СИСТЕМЫ КЛАССИФИКАЦИИ СООБЩЕНИЙ

Ранее мы приняли, что для определения тональности текста будем использовать пять классов: сильно положительный, средне положительный, нейтральный, средне отрицательный, сильно отрицательный. Теперь необходимо подобрать наиболее подходящую структуру, чтобы показать лучшие результаты при классификации. Для нашего случая классификацию на пять классов лучше реализовать иерархическим методом. Обучать будут несколько классификаторов для каждой группы классов, причем каждый классификатор будет производить бинарное деление. Это должно дать выигрыш в качестве по сравнению с плоской классификацией. Первый классификатор будет разделять тексты на два класса: нейтральные и эмоционально окрашенные слова. Затем другим классификатором будет производиться деление на классы: положительные и отрицательные. Последние два уточняющих классификатора будут разбивать класс положительных текстов на средне положительные и сильно положительные, а класс отрицательных – на средне отрицательные и сильно отрицательные.

Для реализации эффективного алгоритма определения психотипа респондента было принято решение использовать методы машинного обучения. Реше-

ние данной задачи состоит из следующих этапов: оценка качества работы классификатора; выбор признаков, по которым производится классификация; реализация алгоритмов классификации; выбор подходящего обучающего корпуса; тестирование эффективности.



Рис. 1. Схема классификации сообщений

Для оценки качества классификации существуют два подхода. Первый – это сравнение классификаторов между собой, второй подход – абсолютная оценка качества. Считается, что точно оценить качество классификации очень сложно: порой даже опытные эксперты не могут сойтись во мнении, в какую категорию отнести определенный документ. На практике чаще всего используют метрику оценки качества классификации, которая состоит из таких двух характеристик классификатора, как точность и полнота.

Точность системы в рамках класса – это та часть документов, которые верно принадлежат данному классу относительно всех документов, которые система отнесла к этому классу.

Полнота системы – это та часть найденных классификатором документов,

которые принадлежат классу относительно всех документов этого класса в тестовой выборке.

В табл. 1 представлена информация о том, какое количество раз классификатор принял верное и какое количество раз – неверное решение. Здесь TP – истинно-положительное решение; TN – истинно-отрицательное решение; FP – ложноположительное решение; FN – ложноотрицательное решение.

Таблица 1. Бинарное разбиение текстов на классы по оценке классификатора

	Классифицированы как +	Классифицированы как -
Класс +	TP	FP
Класс -	FN	TN

На основе описанных данных происходит подсчет метрик, которые позволяют не только оценить качество работы классификатора, но и произвести сравнение классификаторов между собой. Для оценки качества были использованы следующие метрики:

- полная точность или аккуратность (accuracy)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

– доля корректно классифицированных документов в общем числе документов, подаваемых на вход классификатора;

- точность (Precision); для положительных примеров

$$Precision = \frac{TP}{TP + FP}$$

– это отношение количества документов, являющихся позитивными, к общему количеству документов, классифицированных как позитивные, или, иначе, доля положительных документов среди всех документов, которые классификатор отнес к положительным; для негативных документов

$$Precision = \frac{TN}{TN + FN}$$

– это отношение количества документов, являющихся негативными, к общему количеству документов, классифицированных как негативные;

- полнота (Recall); для положительных примеров

$$Recall = \frac{TP}{TP + FN}$$

– это отношение количества правильно классифицированных позитивных примеров к общему количеству позитивных примеров или, иначе, доля действительно положительных документов из всех документов, распознанных как положительные; для отрицательных примеров

$$Recall = \frac{TN}{TN + FP}$$

– это отношение количества правильно классифицированных негативных примеров к общему количеству негативных примеров или, иначе, доля действительно негативных документов из всех документов, распознанных как негативные;

- F-мера (F-measure)

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

– это мера, комбинирующая точность и полноту.

Применим метод классификации, который не использует априорные предположения о том, какие символы и слова содержат сообщения, принадлежащие к какому-либо классу. Таким образом, все признаки априорно равнозначны. При решении задачи определения эмоциональной окраски текстов частым решением являются N-граммы слов. Ниже будут использованы униграммы, биграммы и их комбинация. Будем считать словом любую последовательность букв русского алфавита. N-грамм порядка n – это разделенная пробелами последовательность из n слов. Например, предложение «Светит яркое солнце» состоит из следующих N-грамм первого порядка: «Светит», «яркое», «солнце»; второго порядка: «Светит яркое» и «яркое солнце»; комбинация униграмм и биграмм: «Светит», «яркое», «солнце», «Светит яркое», «яркое солнце».

РЕАЛИЗАЦИЯ АЛГОРИТМА КЛАССИФИКАЦИИ

Для подготовки текста к процедуре классификации необходимо произвести несколько действий по его обработке, таких, как токенизация, стемминг, лемматизация и удаление стоп-слов:

- токенизация – это процесс выделения слов из предложений;
 - стемминг – процесс нахождения основы слова для заданного исходного слова или, иначе говоря, удаление окончаний;
-

- лемматизация – процесс приведения слова к его начальной форме; например, для существительных – именительный падеж, единственное число, для глаголов – инфинитив.

Цель этих действий – уменьшение размерности задачи, другими словами, если в текстах присутствуют одинаковые слова, имеющие разные окончания, то в результате стемминга и лемматизации они сведутся к одному виду. Стоп-слова (также называемые шумовыми) – это слова, знаки, символы, которые не несут никакой информационной нагрузки.

Для решения поставленной задачи были выбраны алгоритмы, показавшие себя наиболее эффективными при определении полярности коротких текстовых сообщений [11, 12]: наивный байесовский классификатор; метод опорных векторов; метод опорных векторов + delta TF-IDF.

Наивный байесовский классификатор – это вероятностный классификатор, основанный на теореме Байеса и (наивном) предположении о статистической независимости случайных величин:

$$P(C/F_1, \dots, F_N) = \frac{P(C)P(F_1, \dots, F_N/C)}{P(F_1, \dots, F_N)}.$$

Метод наивной байесовской классификации строится на двух предположениях: об условной независимости термов; о независимости позиций термов. Иными словами, во внимание не берется тот факт, что появление нескольких разных слов в тексте на естественном языке часто бывает взаимосвязанным. Например, вероятнее, что слово «мяч» встретится в одном тексте со словом «футбол», чем со словом «пальто». Также не учитывается, что вероятности того, что одно и то же слово встретится на разных позициях в тексте, различны. Именно по этим причинам данный алгоритм называется наивным. Несмотря на это, рассматриваемая модель дает весьма точные результаты. Главными достоинствами данного классификатора являются низкая вычислительная сложность, а также оптимальность, при условии действительной независимости признаков.

Метод опорных векторов является линейным классификатором. Его цель – поиск гиперплоскости в пространстве признаков, которая разделит объекты на два класса. Главный принцип метода опорных векторов – это поиск такой разделяющей гиперплоскости, что ближайшие к ней точки в пространстве признаков

находятся на максимальном расстоянии. Если выборка линейно разделима, то поиск гиперплоскости можно привести к задаче оптимизации

$$\frac{1}{2} \|\omega\|^2 \rightarrow \min_{\omega, b} y_i (\omega^T x_i + b) \geq 1, \quad i = 1, \dots, m,$$

где $1/\omega$ – величина зазора между ближайшими к гиперплоскости точками двух классов; $y_i(\omega^T x_i + b)$ – произведение положения гиперплоскости и значения класса точки.

В общем случае при выборке, являющейся линейно неразделимой, алгоритм может допустить ошибки на обучающих объектах. Новая задача оптимизации учитывает минимизацию ошибки:

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l e_i \rightarrow \min_{\omega, b} y_i (\omega^T x_i + b) \geq 1 - e_i, \\ e_i \geq 0, \quad i = 1, \dots, k.$$

Переменные e_i выражают величины ошибки на примере выборки из k элементов. С помощью константы C можно найти компромисс между минимизацией суммарной ошибки на обучающей выборке и максимальной величиной зазора. Отличие данного алгоритма от предыдущего состоит в том, что в этом методе оценка веса признаков рассчитывается другим способом.

Принцип метода *дельта TF-IDF* [13] заключается в том, чтобы присвоить больший вес тем словам, которые обладают не нейтральной тональностью, поскольку именно эти слова определяют тональность всего текста. Рассчитывается вес слова V по следующей формуле:

$$V_{t,d} = C_{t,d} \log \left(\frac{|N|P_t}{|P|N_t} \right),$$

где $V_{t,d}$ – вес слова t в документе d ; $C_{t,d}$ – количество раз, которое слово t встречается в документе d ; $|P|$ – количество документов с положительной тональностью; $|N|$ – количество документов с отрицательной тональностью; P_t – количество положительных документов, где встречается слово; N_t – количество отрицательных документов, где встречается слово t . Из приведенной формулы следует, что вес слов, тональность которых положительна, будет большим положительным числом, слова с отрицательной тональностью будут иметь отрицательный вес, а нейтральные слова будут иметь вес, близкий к нулю. Данное взвешивание вектора признаков во многих случаях повышает точность классификации тональности.

ТЕСТИРОВАНИЕ

При использовании метода обучения с учителем необходима обучающая выборка. Обычно обучающее множество, на котором обучается классификатор, составляется из примеров из той области, в которой будет применяться этот классификатор. Для обучающей и тестовой выборок был составлен корпус коротких текстов, для которых была определена тональность. Набор представляет собой 10000 текстов. Часть этих данных извлечена из размеченного корпуса, который предоставлен авторами [16] в свободном доступе. Другая часть получена в результате небольшой работы с этими же текстами. Вручную было отобрано по 250 сильно положительных и сильно отрицательных текстов.

Для сравнения результатов работы алгоритмов классификации использован метод кросс-валидации. Процедура кросс-валидации происходит по следующему принципу:

- фиксируется множество разбиений обучающего множества на тренировочное и тестовое подмножества;
- каждое разбиение сначала проходит обучение на тренировочном множестве, а потом тестирование на контрольном;
- результатом кросс-валидации алгоритма является среднее значение проведенных результатов тестирования на контрольном множестве.

Будем использовать разбиение на множества, осуществленное случайным образом. Считаем, что попадание каждого предложения в одно из двух множеств равновероятно.

Таблица 2. Результаты классификации

Алгоритм	2 класса		3 класса		5 классов	
	F.	Acc.	F.	Acc.	F.	Acc.
NB	0.68	0.81	0.54	0.71	0.22	0.55
SVM	0.78	0.86	0.59	0.75	0.31	0.61
SVM+delta	0.8	0.89	0.69	0.78	0.34	0.63

Таблица 3. Результаты классификации. В качестве признаков выбраны N-граммы слов. Расчет для 5 классов

Алгоритм	униграммы		биграммы		комбинация	
	F.	Асс.	F.	Асс.	F.	Асс.
NB	0.21	0.55	0.22	0.53	0.23	0.57
SVM	0.31	0.61	0.29	0.59	0.33	0.63
SVM+delta	0.32	0.62	0.31	0.61	0.36	0.65

Как видно из таблиц 2 и 3, добавление новых классов значительно снижает качество классификации.

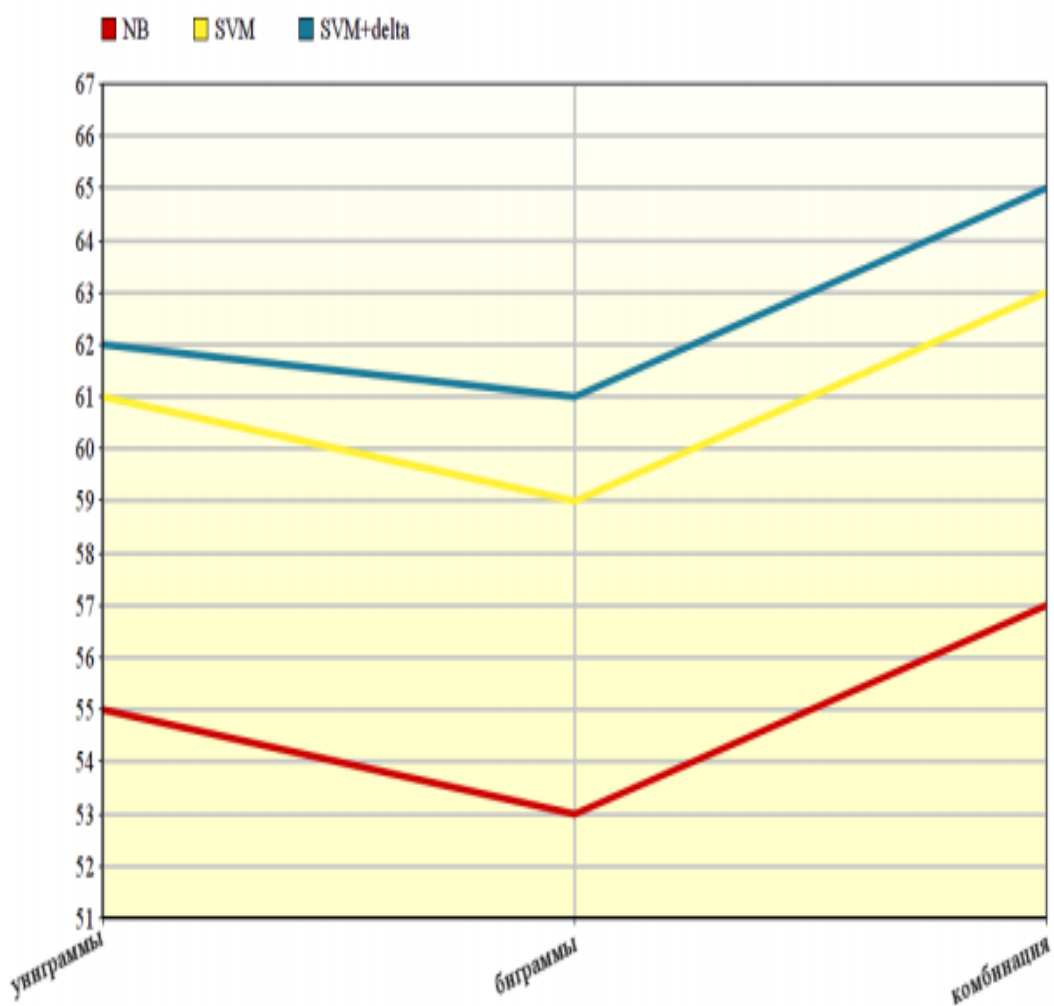


Рис. 2. График результатов классификации

Результаты показали, что лучшие показатели (рис. 2) для данной коллекции получены при использовании метода опорных векторов с функцией взвешивания дельта TF-IDF(SVM+delta). Если же использовать обычную бинарную функцию, то SVM-классификатор показывает лучшие результаты, чем NB. Комбинация униграммов и биграммов дает лучшие результаты во всех тестах.

ЗАКЛЮЧЕНИЕ

Для решения поставленной задачи и реализации системы был выбран язык программирования Python. На рис. 3 показано, в каком направлении и между какими компонентами передаются данные. С решением задач, связанных с предобработкой текста на языке Python, хорошо справляется библиотека для работ с естественными языками Natural Language Toolkit (NLTK) [17]. Эта библиотека сопровождается обширной документацией и является свободным программным обеспечением. NLTK-модуль производит разбиение входящего текста на N-граммы. На следующем этапе формируется вектор признаков из текста. В результате иерархической классификации определяется принадлежность текста к одному из рассматриваемых классов.

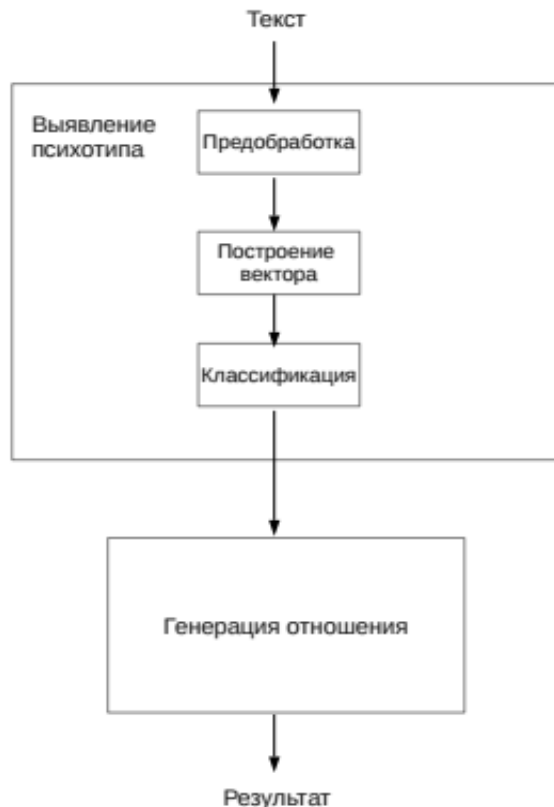


Рис. 3. Общая схема работы системы

Аппарат определения психотипа респондента и отношения его к агенту является важной составляющей внутренней логики взаимодействия машины и компьютера в рамках антропоморфных интерфейсов. Поскольку основная задача таких интерфейсов – добиться максимальной схожести своего поведения с таковой у человека, невозможно недооценить роль, которую играют в них корректно имплементированные эмоциональные реакции.

СПИСОК ЛИТЕРАТУРЫ

1. Кугуракова В.В., Таланов М.О., Манахов Н.Р., Иванов Д.С. Антропоморфный социальный агент с симуляцией эмоций и его реализация // Russian Digital Libraries Journal. 2015. Т. 18, № 5. С. 254–268.
2. Aaronson S. My conversation with "Eugene Goostman", the Chatbot that's all over the news for allegedly passing the turing test". Shtetl-Optimized // The Blog of Scott Aaronson. Archived from the original on 2014-08-07. Retrieved 2014-09-12.
3. Oppy G., Dowe D. The turing test // Stanford Encyclopedia of Philosophy, 2011.
4. Wan V., Anderson R., Blokland A., Braunschweiler N., Chen L., Kolluru B., Latorre J., Maia R., Stenger B., Yanagisawa K., Stylianou Y., Akamine M., Gales M.J.F., Cipolla R. Photo-realistic expressive text to talking head synthesis// Source of the Document Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2013. P. 2667–2669.
5. 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. P. 587–591.
6. 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.
7. Pang B., Lee L. Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. V. 2, No 1–2. P. 1–135.
8. Четверкин И.И., Лукашевич Н.В. Тестирование систем анализа тональности на семинаре РОМИП-2012. Т. 2: Доклады специальных секций РОМИП. М.: Изд-во РГГУ, 2013.
9. Liu Bing. Web Data Mining: Exploring Hyperlinks, Contents and Usage

Data. Berlin: Springer, 2011.

10. Ильин Е.П. И46 Психология общения и межличностных отношений. СПб.: Питер, 2013.

11. *Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, Ping Li.* User-level sentiment analysis incorporating social networks// Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2011.

12. *Mike Thelwall, Kevan Buckley, Georgios Paltoglou.* Sentiment in Twitter events // Journal of the American Society for Information Science and Technology Archive. 2011. V. 62.

13. *Justin Martineau, Tim Finin.* Delta TFIDF: An Improved Feature Space for Sentiment Analysis. University of Maryland, Baltimore County 1000 Hilltop Circle, Baltimore, 2013.

14. Исследование Хабрахабр. URL: <https://habrahabr.ru/post/149605/>.

15. Тональный словарь. URL: <http://linis-crowd.org/>.

16. Русский корпус для анализа тональности текстов. URL: <http://study.mokoron.com/>.

17. Open-Source библиотека NLTK. URL: <http://www.nltk.org>.

EVALUATION OF INTERLOCUTORS EMOTIONAL STATE THROUGH SENTIMENT ANALYSIS OF MESSAGES FOR THE ANTHROPOMORPHIC SOCIAL AGENT

A. Alexeev¹, V. Kugurakova², D. Ivanov³

Kazan (Volga Region) Federal University

¹aleksanton5@gmail.com; ²vlada.kugurakova@gmail.com;

³denis.999.ivanov@gmail.com

Abstract

In this paper, we study the process of evaluating the psychological portrait of a human for the purposes of enhancing the human-machine communication in the frame of the anthropomorphic social agent. We understand “psychological portrait” as the current emotional state of the interlocutor, which is evident by their way of talking and

their choice of words. By correctly identifying human's emotional state, the machine can issue an appropriate response. Our study further describes how the emotional state can be examined and how a proper response should be formed.

Keywords: *human-machine communication, anthropomorphic social agent, emotional state, psychological portrait*

REFERENCES

1. *Kugurakova V., Talanov M., Manakhov N., Ivanov D.* Anthropomorphic social agent with simulated emotions and its implementation // Russian Digital Libraries Journal. 2015. V. 18, No 5. P. 254–268.
2. *Aaronson S.* My conversation with "Eugene Goostman", the Chatbot that's all over the news for allegedly passing the turing test". Shtetl-Optimized // The Blog of Scott Aaronson. Archived from the original on 2014-08-07. Retrieved 2014-09-12.
3. *Oppy G., Dowe D.* The turing test. Stanford Encyclopedia of Philosophy, 2011.
4. *Wan V., Anderson R., Blokland A., Braunschweiler N., Chen L., Kolluru B., Latorre J., Maia R., Stenger B., Yanagisawa K., Stylianou Y., Akamine M., Gales M.J.F., Cipolla R.* Photo-realistic expressive text to talking head synthesis// Source of the Document Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2013. P. 2667–2669.
5. *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. P. 587–591.
6. *Perez-Rosas V., Banea C., Mihailescu R.* Learning sentiment lexicons in Spanish // In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), 2012.
7. *Pang B., Lee L.* Opinion mining and sentiment analysis // Foundations and Trends in Information Retrieval. 2008. V. 2, No 1–2. P. 1–135.
8. *Chetvertkin I., Lukashevich N.* Testing of sentiment analysis systems at ROMIP-2012 seminar // T. 2: ROMIP special report. M.: RSUH press, 2013.
9. *Liu Bing.* Web Data Mining: Exploring Hyperlinks, Contents and Usage Data. Berlin: Springer, 2011.
10. *Ilyin E.* I46 Psychology of communications and interpersonal relationships.

Saint-Peterburg, 2013.

11. *Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, Ping Li.* User-level sentiment analysis incorporating social networks// Proceedings of the Sixteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD). 2011.

12. *Mike Thelwall, Kevan Buckley, Georgios Paltoglou.* Sentiment in Twitter Events // Journal of the American Society for Information Science and Technology Archive. 2011. V. 62.

13. *Justin Martineau, Tim Finin.* Delta TFIDF: an Improved Feature Space for Sentiment Analysis University of Maryland, Baltimore County 1000 Hilltop Circle, Baltimore, 2013.

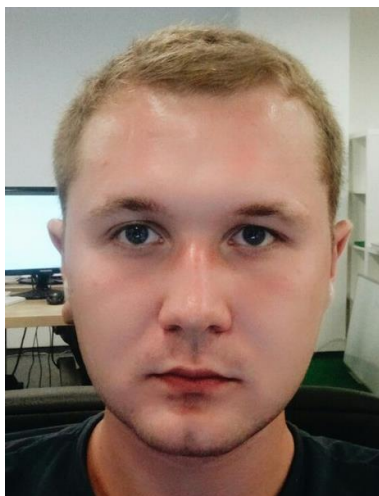
14. Habrahabr study. URL: <https://habrahabr.ru/post/149605/>.

15. Sentiment text dictionary. URL: <http://linis-crowd.org/>.

16. Russian branch for sentiment analysis. URL: <http://study.mokoron.com/>.

17. Open-Source NLTK library. URL: <http://www.nltk.org>,

СВЕДЕНИЯ ОБ АВТОРАХ



АЛЕКСЕЕВ Антон Анатольевич – выпускник 2016 года Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета.

Anton Anatolievich ALEXEEV, bachelor of Kazan Federal University.

email: aleksanton5@gmail.com



КУГУРАКОВА Влада Владимировна – старший преподаватель Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета, руководитель Научно-исследовательской лаборатории «Виртуальные и симуляционные технологии в биомедицине».

Vlada Vladimirovna KUGURAKOVA, senior Lecturer of Higher School of Information Technology and Information Systems of Kazan Federal University, Head of Laboratory “Virtual and simulational technologies in biomedicine”.

email: vlada.kugurakova@gmail.com



ИВАНОВ Денис Сергеевич – лаборант-исследователь Научно-исследовательской лаборатории «Виртуальные и симуляционные технологии в биомедицине» Высшей школы информационных технологий и информационных систем Казанского (Приволжского) федерального университета.

Denis Sergeevich IVANOV, researcher of Laboratory “Virtual and simulational technologies in biomedicine” of Kazan Federal University.

email: denis.999.ivanov@gmail.com

Материал поступил в редакцию 15 мая 2016 года
