

АДАПТИВНАЯ ЛИЧНОСТНАЯ ДИАГНОСТИКА ПОДРОСТКОВ НА ОСНОВЕ ПСИХОЛИНГВИСТИЧЕСКОГО АНАЛИЗА

DOI: 10.25629/НС.2019.04.18

Абушкин Б.М.¹, Мещанкин А.В.¹, Машкова Л.А.²¹Московский городской педагогический университет. Россия, Москва²Московский государственный университет имени М.В. Ломоносова. Россия, Москва

Аннотация. В статье рассмотрен методический подход организации адаптивной личностной диагностики и активного диалога с подростком, направленный на осознание и развитие им своих личностных качеств, а также социальных интересов и возможностей. В основу этого подхода положена методология психолингвистического анализа письменной речи подростков для исследования особенностей их эмоционального реагирования в разных социальных ситуациях. Рассмотрен опыт зарубежных исследований, обосновывающий актуальность работы в данном направлении. Представлен программный комплекс для эмоциональной разметки прилагательных русского языка, включающий более 4000 прилагательных, который позволяет в автоматическом режиме рассчитывать эмоциональный вектор слов. Описана методика исследований, реализующая модель смысловой близости слов, при которой смысловая близость между двумя словами строится исходя из частоты появления данных слов с одинаковым лексическим окружением. Была проделана работа по обоснованию набора базовых эмоций и на их основе осуществлена разметка испытуемыми заранее выбранных прилагательных русского языка, набрана статистика по предложенным прилагательным. В статье представлена технология автоматического формирования информации об эмоциональном содержании слов на основе методов дистрибутивной семантики. Показана эффективность данного подхода по результатам зарубежных и наших исследований. Приведены результаты исследования на основании ответов 416 респондентов.

Ключевые слова: дистрибутивная семантика, рейтинг слов, базовые эмоции, разметка прилагательных русского языка, эмоциональный вектор слов.

Введение

Для современной психологии актуальны поиски новых методических подходов для системной организации личностной диагностики и активного диалога с подростком, направленный на осознание и развитие им своих личностных качеств, а также социальных интересов и возможностей [1, 2]. Опыт многих исследований показывает, высокую перспективность методических разработок подхода на основе психолингвистического подхода [5, 6, 7, 13]. В последние десятилетия, одной из основных целей психолингвистики является объяснение того, как слова, которые используют люди в своей повседневной социальной практике, отражают то, кто они такие и каковы их интересы. Многие исследования были сосредоточены на сборе рейтингов, связанных с такими свойствами слов, как частота, сложность, конкретность, образность, аффективные состояния [6, 13]. Сбор рейтинга предполагает привлечение большого количества респондентов (носителей данного языка), которым предлагалось определить эмоциональную окраску слов. Эти рейтинги в психолингвистике предназначены для количественной оценки различных свойств большого объема текста на естественном языке, например, при анализе лексического выбора между демографическими группами [5] или предпочтений в музыке [7].

Благодаря обширному распространению компьютеров и возможности общаться в социальных сетях, специалисты получили доступ к большому количеству текстов разного смыслового наполнения, что позволяет исследователям значительно расширить фронт работ по поиску взаимосвязей между языком и личностью. При проведении тестовой оценки валидности моделей широко используются эталонные суждения людей о семантической близости слов [9]. Было

доказано, что на множестве лексических единиц можно задать метрику (или полуметрику, квазиметрику), измерив для каждой пары слов «семантическое расстояние».

Краткий обзор исследований

Основными единицами в построении эмоционального рейтинга слов являются валентность (разные эмоции) и возбуждение (аффект – интенсивность эмоции). Используя данные единицы, мы имеем несколько более расширенную модель циркумплекса Рассела [12], который утверждает, что все состояния представлены в виде линейной комбинации этих двух независимых величин. Тем не менее, сбор рейтинга слов занимает очень длительное время и является дорогой по стоимости процедурой. Кроме того, хотя сбор рейтинга происходит под контролем, часто опрашиваемые люди дают абсолютно противоположные оценки, что свидетельствует о различном предыдущем личностном опыте. Независимое автоматическое построение подобных рейтингов стало единственным выходом, обеспечивающим исследователей данными с достаточной полнотой и точностью. Автоматический подход к построению эмоциональных рейтингов слов был изучен на основании интуитивного предположения, что слова с одинаковой семантической дистрибуцией будут иметь схожие значения рейтинга [10].

Часто рейтинг неизвестного слова вычисляют как среднее соответствующих величин его k -ближайших соседей из низкоразмерного семантического пространства [3]. Однако, недостатком такого подхода, является то, что антонимы также семантически похожи, что, как ожидается, уменьшает точность этих методов. Орфографическое сходство показало немного лучшие результаты [11].

Исходя из работ, посвященных выделению дескрипторов (слов, явно отражающих ту или иную эмоцию), а также последних исследований в области психолингвистики, известно, что наибольшая эмоциональная составляющая передается в письменной речи посредством прилагательных [4].

Общепринятым в лингвистике является следующий способ: используя разнообразные тексты как можно большего размера собираются сведения о совместной дистрибуции слов в одинаковом окружении, далее каждому слову присваивается вектор w , являющийся вектором частот разных слов в его контекстных окружениях заданной ширины. Частоты могут быть помноженными на веса, приписанные словам-признакам (элементам базиса) в соответствии с тем, насколько тесна их связь с w . Величина семантической близости между w и произвольным словом w' вычисляется как расстояние в какой-либо метрике между векторами, описывающими их [7].

Психолингвистические исследования показали, что данный метод показывает хороший результат при оценке синонимичных слов [14]. В работе [12] была использована модель близости слов, составленная на основании корпуса из текстов. Далее добровольцев просили определить синонимы, для некоторой подгруппы слов. После того, как синонимы были названы, результат сравнивался с синонимами, найденными моделью. Данные показали хорошую сходимость. Среди прочих таким пугём идет Л. Ли [4], в работе которого исследовано поведение целого ряда весовых функций и метрик относительно эталонов.

Основной идеей предлагаемых методов является так называемая дистрибутивная гипотеза, состоящая в том, что о грамматических и семантических свойствах языковой единицы (морфемы или лексемы) можно многое узнать из её окружения другими языковыми единицами в тексте, не привлекая никаких дополнительных сведений [7].

Научная новизна нашего подхода автоматического предсказания эмоционального вектора прилагательных заключается в использовании методов дистрибутивной семантики и анализа графов, что позволяет избежать вышеуказанных трудностей [15]. Такой подход впервые использован для русского языка. Нами разработан программный комплекс для эмоциональной разметки прилагательных русского языка, позволяющий в автоматическом режиме рассчитывать эмоциональный вектор слов.

Методика исследований

На первом этапе было выполнено сопоставление слов определенным эмоциям вместе с количественной мерой аффекта. Данные были собраны при помощи онлайн-опросника, который предлагался для прохождения респондентами. Задача заключалась в выборе для каждого слова одной из 7 эмоций (гнев, предвосхищение, отвращение, страх, радость, удивление, доверие) [8]. Словами в опросе являлись различные прилагательные русского языка. Алгоритм предлагал для разметки наиболее частотные в употреблении естественной речи прилагательные таким образом, чтобы каждое слово имело возможность получить статистически значимое значение эмоции, либо признавалось эмоционально неопределенным. В итоге, каждое слово определялось 7-ми мерным вектором эмоций.

На втором этапе использовалась модель смысловой близости слов, получаемой в ходе анализа большого количества текстов русского языка. Так же, как и смысл слова был рассчитан эмоциональный контекст слова. На основе контрольной группы заранее размеченных испытуемыми слов осуществлялось автоматическое распространение эмоционального вектора на близкие им неразмеченные слова.

Распространение эмоционального вектора между словами происходило по следующему алгоритму:

С помощью заранее обученной модели для размеченного респондентами слова находились его ближайшие смысловые соседи

Найденные слова проверялось на соответствие всех лексических правил исходному.

При соответствии лексических правил, эмоциональный вектор передавался как его произведение на скалярную величину, которой являлась мера близости слов согласно п.1.

Для слова, которому был спрогнозирован эмоциональный вектор, записывалось количество ссылающихся на него изначально размеченных слов.

Если слово, на которое ссылалось первоначально размеченное, уже имело эмоциональный вектор, то он поэлементно складывался с предыдущим.

После завершения всего цикла распространения, предсказанный вектор эмоций, поэлементно был разделен на количество ссылок из п.4.

Результаты и их обсуждение

Нами был собран эмоциональный рейтинг слов на основании ответов 416 респондентов. В таблице 1 представлены результаты работы одного респондента из общей выборки с рядом конкретных прилагательных.

Таблица 1 – Эмоциональная разметка одним респондентом прилагательных русского языка

Слово	гнев	предвосхищение	отвращение	страх	радость	удивление	доверие
модный		1					
расширенный					1		
лимфатический						1	
плачущий		1					
модный		1					
бестолковый		1					
альтернативный				1			

Как правило, ответы респондентов, отражали их личностную позицию, с учетом их предыдущего опыта. Разные респонденты по-разному размечали одни и те же предложенные слова. После этого для каждого конкретного слова, суммировалась разметка от всех респондентов. В большинстве случаев респонденты выделяли одну, реже две-три эмоции, характеризующие, по их мнению, конкретное слово.

Однозначность выбора эмоции зависит от предыдущего опыта респондентов. Тем не менее, могут встречаться слова, которые не вызывают однозначной эмоции либо имеют слишком слабый аффект, что человеку становится трудно идентифицировать однозначность. Это означает, что мы можем использовать только одну эмоцию для характеристики слова только тогда, когда с ее разметкой будет согласно абсолютное большинство респондентов, что является недостижимым на практике условием. Именно поэтому нами было введено понятие «эмоционального вектора» слова. «Эмоциональный вектор» - вектор, базисом которого являются рассматриваемые базовые эмоции, а величина равна аффекту каждой из рассматриваемых эмоций. Так, используя именно векторное представление, мы можем точнее проводить операции автоматического прогнозирования эмоций в словах и не терять важной информации о не доминантных эмоциях слова.

В таблице 2 представлен пример содержания эмоциональных векторов для отдельных прилагательных. Цифры в таблице – это коэффициенты представленности конкретной базовой эмоции в слове. Эти коэффициенты являются мерой аффекта (интенсивности) эмоции и вычисляются как отношение количества размеченных этой эмоцией слов к количеству ссылающихся на слово семантически близких слов. Этот показатель изменяется от 0 до 1. Расчетные коэффициенты менее 0,2 округлялись до 0 как не существенные для определения эмоционального вектора.

Таблица 2 – Пример автоматического прогнозирования эмоциональных векторов для отдельных прилагательных

Слово/эмоция	Гнев	Предвосхищение	Отвращение	Страх	Радость	Удивление	Доверие
британский	0	0	0	0	0,37	0,37	0
французский	0	0	0	0	0	0,79	0
кремовый	0	0,37	0	0	0,41	0	0,77
черный	0	0	0	0,76	0	0,76	0
безбрежный	0	0	0,81	0,71	0	0	0
меньший	0	0,86	0	0	0	0,86	0
небольшой	0	0	0	0	0	0,71	0,71
жизнерадостный	0	0	0	0	0,78	0	0
вероятный	0	0	0,77	0	0,77	0	0
субботный	0	0	0	0	0,76	0	0,76
квалифицированный	0	0,80	0	0	0	0	0,79

На основе автоматической разметки слов, был собран словарь прилагательных с их эмоциональными векторами. Данный словарь дает возможность анализировать эмоциональную составляющую любого содержательного текста.

Полученный словарь является инструментом, позволяющим найти отклонения в опыте испытуемого относительно усредненного опыта контрольной выборки. Например, если во время свободного ответа испытуемого на заданную тестом ситуацию, анализ его письменного ответа покажет смещение эмоционального вектора в сторону эмоции «страх», в то время как у контрольной выборки данная ситуация вызывала эмоцию «радость», то это может свидетельствовать об особенном опыте испытуемого или даже об отсутствии опыта в данной ситуации. На основании этого выдвигается предположение о его личностных характеристиках и, например, делается вывод о его отношении к представленной в тексте проблеме или ситуации.

Для проведения подобного анализа важна высокая точность определения эмоционального вектора слов. Если в случае с эмоциональной разметкой респондентов точность будет пропорциональна количеству респондентов, разметивших одно и то же слово, то в случае автоматической разметки данный вопрос может быть решен за счет дополнительных фильтров.

Прежде всего определим критерий точности для автоматического предсказания эмоционального вектора. В первом приближении таким критерием может быть однозначность выбора эмоции (наличие доминантной эмоции). В свою очередь, однозначность выбора эмоции зависит от точности определяемого нами аффекта эмоций, а аффект эмоций будет тем точнее, чем больше размеченных респондентами слов были использовано для предсказания эмоционального вектора предсказуемого слова. Целесообразно ввести величину, характеризующую точность предсказанных эмоциональных векторов слов – плотность однозначности выбора эмоций, которая определяется как количество доминирующих эмоций в некотором количестве слов, например, в десятке. Ясно, что при полной однозначности выбора (при 100% точности алгоритма), данная величина будет равна 1, при уменьшении точности, данная величина будет расти.

На графике плотности однозначности выбора эмоций (рисунок 1), где по оси Y отложена плотность однозначности выбора, а по оси X – номер каждой десятки слов, сгруппированных по убыванию точности аффекта, показано, что чем больший номер декады (чем меньше точность аффекта), тем меньшую точность имеет алгоритм предсказания.

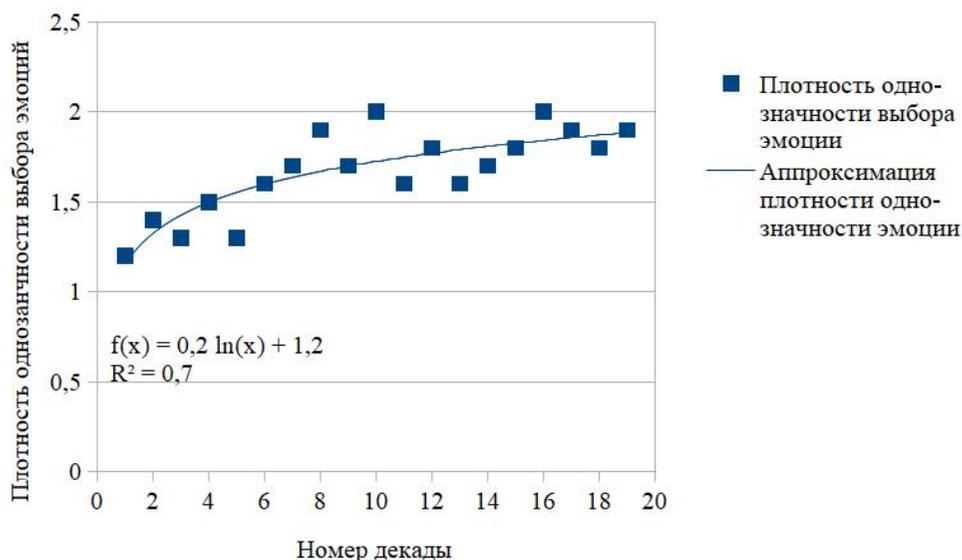


Рисунок 1 – Распределение плотности однозначности выбора эмоций

Аппроксимация исходных данных логарифмической кривой показала неплохое совпадение с распределением данных ($R^2 = 0,7$, где R – коэффициент корреляции). Полученные результаты свидетельствуют, что для увеличения точности алгоритма предсказания эмоционального вектора в словах, достаточно увеличить количество изначальной разметки слов респондентами.

Для проведения работ по определению эмоциональные вектора слов со всеми прилагательными русского языка разработан программный комплекс, включающий более 4000 прилагательных (с разной частотой употребления). Данный комплекс, размещен в сети Интернет и ведет сбор данных в режиме автономной работы на основе привлечения респондентов – пользователей Интернета

Заключение

Рассмотренные в работе психолингвистические исследования с использованием прилагательных русского языка показали, хороший результат при эмоциональной разметке прилагательных. Разработанный программный комплекс позволил в автоматическом режиме определять эмоциональные вектора слов. Полученные результаты свидетельствуют, что эмоциональные векторы слов представлены в основном 2 базовыми эмоциями. Доказано, что для увеличения точности алгоритма предсказания эмоционального вектора в словах, необходимо увеличить количество изначальной разметки слов респондентами. Показано, что полученный словарь является инструментом, позволяющим осуществлять адаптивную личностную диагностику подростков на основе выявления особенностей личного опыта испытуемого относительно усредненного опыта контрольной выборки.

Литература

1. Рыжов Б.Н. Системная структура личности // Системная психология и социология, 2017. № 3 (23). с. 5-11.
2. Рыжов Б.Н. Психологический возраст цивилизации // Системная психология и социология, 2015. № 2 (14). с. 60-70.
3. Bestgen Y., Vincze N. Checking and bootstrapping lexical norms by means of word similarity indexes // Behavior Research Methods. 2012, No. 3, Vol.44, p 998-1006.
4. Lee L. Similarity-based approaches to natural language processing. // Ph.D. thesis, Harvard University, 1997.
5. Liu L., Preotiu-Pietro D., Samani Z., Moghaddam E., Ungar L. Analyzing Personality through Social Media Profile Picture Choice // Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016), p. 211-220.
6. Marc B., Amy B., Victor K. Concreteness ratings for 40 thousand generally known English word lemmas // Behavior Research Methods. 2014, No. 3, Vol.46, p. 904-911.
7. Maulidyani A., Manurung R. Automatic Identification of Age-Appropriate Ratings of Song Lyrics // 2015. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, p. 583-587.
8. Ortony A., Turner T. What's Basic About Basic Emotions? // Psychological Review, 1990. V. 97, No.3, P. 315-331.
9. Pelevina M., Arefyev N., Biemann C., Panchenko A. Making Sense of Word Embeddings // In Proceedings of the 1st Workshop on Representation Learning for NLP co-located with the ACL conference. Berlin, Germany. Association for Computational Linguistics (2016). P. 174-183.
10. Pennington J., Socher R., Manning C. Glove: Global vectors for word representation // Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). – 2014. P. 1532-1543.

11. Recchia G., Louwse. Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance // *The Quarterly Journal of Experimental Psychology*. 2015. №8, Vol. 68, p 1584-1598.
12. Russell J. A. A circumplex model of affect. // *Journal of Personality and Social Psychology*. 1980. № 6. Vol 39, p 1161-1178.
13. Victor K., Hans S., Marc B. Age-of-acquisition ratings for 30,000 English words // *Behavior Research Methods*. 2012, №4, Vol.44, p 978-990.
14. Ovesdotter C., Roth D., Sproar R. Emotions from text: machine learning for text-based emotion prediction // *HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, p 579-586.
15. Yadollahi A., Gholipour A., Zaiane O. Current State of Text Sentiment Analysis from Opinion to Emotion Mining // *ACM Computing Surveys (CSUR)*. 2017, №25, Vol.50.

Абушкин Борис Михайлович. SPIN: 7649-5398. E-mail: abushkinbm@mgpu.ru.

Мещанкин Андрей Вячеславович. E-mail: meschankin.andrey@physics.msu.ru

Машкова Лариса Александровна. E-mail: la.mashkova@gmail.com

Дата поступления 13.03.2019

Дата принятия к публикации 10.04.2019

**ADAPTIVE PERSONAL DIAGNOSTICS OF TEENAGERS ON THE BASIS
OF PSYCHOLOGICAL ANALYSIS**

DOI: 10.25629/HC.2019.04.18

Abushkin B.M.¹, Meshchankin A.V.¹, Mashkova L.A.²

¹Moscow City Pedagogical University. Russia, Moscow

²Moscow State University Lomonosov. Russia, Moscow

Abstract. The article describes a methodical approach to the organization of adaptive personality diagnosis and active dialogue with a teenager, aimed at the awareness and development of their personal qualities, as well as social interests and opportunities. The basis of this approach is the methodology of psycholinguistic analysis of the written speech of adolescents to study the characteristics of their emotional response in different social situations. The experience of foreign studies, justifying the relevance of work in this direction. A software package for the emotional marking of adjectives of the Russian language, including more than 4000 adjectives, which allows you to automatically calculate the emotional vector of words, is presented. A research methodology is described that implements a model of semantic proximity of words, in which semantic proximity between two words is based on the frequency of occurrence of these words with the same lexical environment. The work was done to substantiate a set of basic emotions, and on their basis, the subjects made markup of adjectives of the Russian language selected in advance, collected statistics on the proposed adjectives. The article presents the technology of automatic generation of information about the emotional content of words based on the methods of distributive semantics. Shows the effectiveness of this approach according to the results of foreign and our research. The results of the study based on the answers of 416 respondents.

Keywords: distributive semantics, word ranking, basic emotions, marking of adjectives of the Russian language, emotional vector of words.

Abushkin Boris Mikhailovich. SPIN: 7649-5398. E-mail: abushkinbm@mgpu.ru.

Meshankin Andrey Vyacheslavovich. E-mail: meschankin.andrey@physics.msu.ru

Mashkova Larisa Aleksandrovna. E-mail: la.mashkova@gmail.com

Date of receipt 13.03.2019

Date of acceptance 10.04.2019