

МЕТОДИКА ВЫЯВЛЕНИЯ СЕМАНТИЧЕСКИХ ДИФФЕРЕНЦИАЛОВ ДЛЯ АВТОМАТИЗАЦИИ ОЦЕНКИ ПСИХОСЕМАНТИЧЕСКОГО ПРОФИЛЯ ПОЛЬЗОВАТЕЛЯ СОЦИАЛЬНОЙ СЕТИ

Монахов Ю. М., Семенова И. И., Медведникова М. А., Костина Н. В.

¹ФГБОУ ВПО «Владимирский государственный университет имени А. Г. и Н. Г. Столетовых», Владимир, Россия (600000, г. Владимир, ул. Горького 87, 1 корпус), e-mail: osobaii@gmail.com

Для исследования поведения пользователя в социальной сети и настройки под психосемантический профиль пользователя социально значимых сообщений была поставлена задача – создать автоматизированную систему оценки профиля. В качестве основополагающей выдвинута гипотеза о возможности определения в автоматизированном режиме без привлечения эксперта психосемантического профиля пользователя на основе анализа его сообщений и комментариев на стене в социальной сети. В данной работе представлена методика выявления семантических дифференциалов на основе работ Артемьевой, Петренко и описан эксперимент, результаты которого дали необходимые знания в виде построенной онтологии с «эмоциональными шкалами», содержащими психосемантические дифференциалы, для дальнейшей оценки профиля. На базе созданной онтологии и индекса PMI-IR создана программная система, работа с которой подтверждает выдвинутую гипотезу.

Ключевые слова: автоматизация оценки профиля узла социальной сети, онтология, пользователь социальной сети, эмоциональное шкалирование, семантический дифференциал, психосемантический профиль.

METHODOLOGY OF SEMANTIC DIFFERENTIAL IDENTIFICATION FOR ASSESSMENT AUTOMATION OF PSYCHOSEMANTIC PROFILE OF SOCIAL NETWORK USERS

Monakhov Y. M., Semenova I. I., Medvednikova M. A., Kostina N. V.

Vladimir State University named after A. and N. Stoletovs, Vladimir, Russia (600000, Russia, Vladimir, 87 Gorky street), e-mail: osobaii@gmail.com

For the study of user behavior in a social network and configure a user profile psychosemantic socially relevant messages was a task – to create an automated system for evaluating profile. Was defined hypothesis- "The possibility of determining psychosemantic user profile in an automated mode. Performs work on the basis of his analysis of posts and comments on the wall of a social network." This paper presents a method for identifying semantic differentials on the basis of the works Artemievoy, Petrenko and describes an experiment, the results of which were given the necessary knowledge in the form of ontology built for further evaluation profile. Based on established ontologies and PMI-IR index created the software system, the work which confirms the hypothesis put forward.

Keywords: automating the assessment profile social networking site, ontology, user of social network, emotional scaling, semantic differential, psychosemantic profile.

Введение

Пропаганда всегда оказывала большое влияние на общество. Одним из наиболее быстрых каналов ее распространения в настоящее время является Интернет, неотъемлемой частью которого для значительной части населения являются социальные сети.

Чем больше людей оказывается вовлечено в сетевое общение, тем больше появляется желающих воздействовать на формирование общественного мнения. Цели, методы и формы такого воздействия могут быть самыми разными – от прямого слова конкретного человека, действующего из идейных (бескорыстных) соображений и высказывающего свое мнение в специальном блоге или сообществе, до сложных комбинированных схем, разработанных для применения в ряде социальных ресурсов и предусматривающих использование множества

самых разных агентов влияния с особыми программами поведения. Важной разновидностью насильственного формирования общественного мнения и является информационная война. Таким образом, выявление механизмов ориентирования информации на потребителя становится важной задачей.

Анализ существующих направлений в выявлении механизмов ориентирования информации на пользователя сети

Авторы работы [2] осветили вопросы моделирования динамики репутации членов социальной сети и исследовали роль репутации в осуществлении информационных воздействий. Рассмотрен ряд моделей сетей, позволяющих ставить и решать задачи формирования репутации с целью последующего ее использования при осуществлении информационного управления. В социальной сети агенты часто не имеют достаточной для принятия решений информации или не могут самостоятельно обработать ее, поэтому их решения могут основываться на наблюдаемых ими решениях или представлениях других агентов (социальное влияние). Социальная сеть играет большую роль в распространении информации, идей между ее членами.

Крестинина Е. С., Чернышов Ю. Г. в работе [3] в разделе «Социальные сети и российская публичная политика» говорят о влиянии социальных сетей на политику. Сетевая структура данных систем позволяет, с одной стороны, организовать личное коммуникативное пространство, с другой – дает возможность отследить контакты человека и при наличии общих знакомых установить с ним неформальные связи. Критичность восприятия информации в таком случае будет понижена. Также речь идет о том, что «...отрицать значимость политизированных групп «в контакте» нельзя: на сегодняшний день они являются своеобразным аналогом подписного листа. Присоединяясь к подобной группе, человек декларирует свою приверженность определенной идее. Сам факт вступления говорит об актуальности поднятой в группе проблемы, даже если дальше вступления активность большинства участников не идет. Помимо этого участие в подобных группах создает видимость действия и позволяет канализировать негативную энергию» [3].

Черниковым Б. В. в [5] в 2011 г. была поднята проблема влияния социальных сетей на бизнес. Выявлена возможность формирования социальных связей и управления ими внутри рабочей группы, что необходимо для эффективного функционирования бизнеса. Рассмотрены способы управления топологией человеческих сетей, направленные на повышение их продуктивности. Определена основа потенциала производительности социальных сетей – как использование и смешивание любого полезного материала с получением в результате новых комбинаций, составленных из разных кусочков информации. В работе [10] показано, как люди реагируют на новые сведения через динамические образцы

их поведения по ведению блога. Данная сфера – легковозбудимая социальная среда, в которой могут быть прослежены волны коллективных волнений. Изучая размер событий, заметили аналогии с сейсмологией. Интенсивность первых и последующих шоков следуют за законом Омори, распределение размеров событий определяется законом Гутенберга – Рихтера. Авторы предположили, что имеются существенные корреляции между динамичными толчками.

Также в работах показано, что, зная возраст узлов в сети, можно определить восприимчивость узлов, которые будут заражены. Например, недавно прибывший узел более восприимчив к заражению, чем тот, который присутствовал в сети долгое время. Такой узел, возможно, развил необходимую неприкосновенность, чтобы противостоять инфекции. В работе [8] выдвинута гипотеза о сегрегации индивидов с экстремальными точками зрения. Результатом работы является вывод о том, что наличие таких индивидов может гарантировать множественность точек зрения только в том случае, если эти индивиды будут формировать открытые для взаимодействия структуры. В статье [7] показано, что более сильные связи менее ответственны за распространение информации. Это означает, что слабые связи узлов могут играть доминирующую роль в распространении информации «онлайн». Авторы в работе [9] предлагают метод искусственного продления продолжительности обсуждения «онлайн», который основывается на энтропии распределения количественных характеристик эмоционального содержания сообщений.

По результатам обзора публикаций можно говорить о том, что исследования механизмов распространения и влияния пропаганды в социальных сетях только начинают развиваться, особенно в России. В основном в работах приводятся различные агентные модели влияния, описывается, как связаны узлы между собой, то, что социальная среда легковозбудимая, но практически не учитываются психологические особенности узлов.

Таким образом, выработка методов анализа и противодействия пропаганде в Интернете, в частности, в социальных сетях, в настоящее время является актуальным вопросом. Для того чтобы знать, насколько узел социальной сети подвержен влиянию пропаганды, возникает необходимость оценки его психосемантического профиля, исходя из публикуемых им сообщений.

В результате была поставлена задача – создать методику автоматизированной оценки психосемантического профиля пользователя социальной сети. Конечная цель данного исследования заключается в повышении эффективности предоставления агитационных и прочих материалов, которые будут коррелировать с интересами пользователя.

Выбор базового метода

Объектом исследования являются микроблоги пользователя социальной сети. Предмет

исследования – методы автоматизированной оценки и построения психосемантического профиля пользователя социальной сети. Результаты таких исследований помогут координировать социальные работы (например, агитационные, продвигаемые бизнес-идеи и т.д.) в социальных сетях. Оценку профиля предлагается выполнять, исходя из следующих принципов: принцип текстоцентричности; принцип объективности. Первый принцип означает, что данные будут выявляться только на основе публикуемых пользователем сообщений. Вводя принцип объективности, авторы стараются минимизировать роль экспертных оценок.

Для создания методики необходимо выявить параметры психосемантического профиля узла социальной сети. Артемьева Е. Ю. в своей работе [1] отталкивается от теории образа мира и одна из первых в России начинает использовать психосемантические методы. Она описывает понятие семантического пространства, методики их получения и подтвердила экспериментально ряд гипотез о существовании субъективных семантик реальных объектов мира, о первичности регуляции восприятия семантическими характеристиками объектов и т. д. Петренко В. Ф. в [4] описывает методы экспериментальной психосемантики и излагает результаты своих исследований категориальных структур индивидуального и общественного сознания. В результате, опираясь на работы [1, 4 и др.], был разработан план эксперимента и сформирована общая методика оценки психосемантического профиля пользователя социальной сети.

Планирование эксперимента

В качестве психосемантического параметра было предложено использовать «эмоциональное шкалирование», что позволит оценить эмоциональный окрас текстов. Определять значения данного параметра предлагается по следующему алгоритму:

1. Задается некоторое количество шкал, которые будут служить для оценки текста. Для каждой шкалы выбирается одно или несколько слов (семантических дифференциалов), наиболее полно ее описывающих и связанных с социально-демографическими признаками пользователя социальной сети. Например, для шкалы «радость» подходящим выбором выглядят слова «ура» или «наслаждение».

2. Для каждого слова текста вычисляется набор индексов PMI-IR [6], формула которых в общем виде выглядит так: $PMI(\text{слово}, \text{шкала}) = \log_2 \left(\frac{f_1(\text{слово}, \text{шкала})}{f_2(\text{слово}) \cdot f_3(\text{шкала})} \right)$, где f_1, f_2, f_3 – количество результатов в поисковой системе по заданному запросу «слово и шкала», «слово», «шкала» соответственно; «шкала» – поисковый запрос по выбранной нами шкале (все слова внутри самого запроса, входящие в шкалу, соединяются логическим оператором «или»); «слово» – поисковый запрос по выбранному нами слову. Для увеличения

точности работы из запросов также убираются все результаты с отрицанием слов запроса. Полученный индекс отражает то, насколько часто люди употребляют искомое слово и любое из слов шкалы вместе, относительно того, насколько часто они употребляют их по отдельности.

3. Индекс для текста представляет собой усредненное значение индексов слов этого текста (с учетом частоты их употребления в нем).

Для реализации алгоритма необходимо предварительно получить множества семантических дифференциалов, характеризующих наиболее распространенные эмоции, отражаемые в текстах пользователей социальной сети. Была поставлена задача – провести эксперимент для выявления ассоциативных рядов (с последующим определением семантических дифференциалов) ответной реакции пользователей социальной сети в зависимости от эмоций, которые он испытывает. Проводился он следующим образом: испытуемым были предложены слова-стимулы, которые, по мнению авторов, наиболее распространены и отражают полный спектр для формирования «эмоциональных шкал»: вдохновение, радость, равнодушие, страх, недовольство, злость, неясность, расстройство. Предлагалось в течение двух минут указать слова (свободные ассоциации), которыми они в социальной сети выражают чувства, указанные в шкалах. Опрашиваемые имели право в качестве ассоциативного ряда вписывать слова, словосочетания и символы.

Средствами, предоставляемыми сайтом [google.ru](https://sites.google.com/site/nasanauka/home), была разработана форма для опроса, расположенная по адресу: <https://sites.google.com/site/nasanauka/home>. Анкета также содержала информацию, отражающую социально-демографические признаки: возраст, образование, пол, семейное положение, – и вопросы, направленные на выявление ассоциативных рядов. Включение социально-демографических признаков основано на гипотезе, что ассоциативные ряды будут различны для групп, отличающихся по полу, по образованию и т. п. А также для реализации проверки гипотезы о том, что можно выявить семантические дифференциалы для каждой социально-демографической группы, и в последующем решать обратную задачу: определение на основе анализа текстов пользователя его принадлежности к определенной социально-демографической группе, даже если признаки указаны заведомо ложно. Стоит также отметить, что использовать нецензурную лексику во время эксперимента не запрещалось.

Обработка данных, полученных в ходе эксперимента

Информация о проведении эксперимента и предложении поучаствовать в опросе была распространена среди пользователей социальных сетей «В контакте», «Одноклассники», «Facebook», а также среди студентов. Анализ активности показал, что женщины охотнее откликаются на предложение участия в опросе. Соотношение женщин и мужчин,

проходящих опрос, соответственно 68 % и 32 %.

Для обработки результатов опроса данные были распределены по группам, состав которых формировался на усмотрение авторов в соответствии с возрастом, полом и количеством респондентов: 1) Мужчины 16–18 лет со средним образованием (СО); 2) Мужчины 19–21 лет с незаконченным высшим образованием (НВО); 3) Мужчины 22–25 лет с высшим образованием (ВО); 4) Женщины 16–18 лет со СО; 5) Женщины 19–25 лет с НВО; 6) Женщины 22–25 лет с ВО; 7) Женщины 19–25 лет со средним специальным образованием; 8) Мужчины и женщины в возрасте от 25 лет с любым образованием.

Следующим этапом был перенос результатов в базу данных по каждой шкале, которая представляет собой совокупность связанных таблиц, сформированных по видам групп с набором ответов респондентов, относительной и абсолютной частотами встречаемости слов. Например, в группе № 8 в шкале «страх» самое часто употребляемое слово – «ужас», относительная частота которого равна 0,45.

Также были построены запросы для выявления общих понятий у разных групп. Для групп № 3 и № 8 результат в шкале «злость» (представлен фрагмент): $Zl_{\text{общ38}} = \{n_1, n_2, n_3, n_4, n_5, n_6, n_7\}$, где n_1 – «:(»; n_2 – «(» ; n_3 – «не нравится»; n_4 – «%(»; n_5 – «неприкольно»; n_6 – «недоволен»; n_7 – «бред».

Формализация полученных результатов была выполнена путем построения онтологии, в основе которой для установления правил лежит математический аппарат дескрипционной логики. Онтология становится популярной в настоящее время за счет того, что она позволяет значительно облегчить работу с данными. Благодаря такому средству разработчик может в любой момент обратиться к базе знаний для извлечения знаний с применением поисковых запросов, а также для дальнейшего ее усовершенствования.

Отношения понятий в ассоциативных рядах можно описать с помощью дескрипционной логики в виде системы аксиом вложенности и эквивалентности (ниже представлен фрагмент описания): $Thing \subseteq Ned \cup Vd \cup Rad \cup Ney \cup Zl \cup Nen \cup St \cup Rav$,

$Ned \subseteq \text{возмущение} \cup \text{ворчание} \cup \text{грустно} \cup \dots \cup \text{гадость}, \dots$

$Vd \subseteq \text{впечатляет} \cup \text{супер} \cup \text{накрыло} \cup \dots \cup \text{идея}, \dots$

$\text{впечатляет} \equiv \text{супер} \equiv \text{понеслось}, \dots, o \equiv \text{ух ты}, \dots, \dots$

где *Thing* – это корень иерархии в онтологии (корень таксономии), которая содержит 8 классов: *Ned* – недовольство; *Vd* – вдохновение; *Rad* – радость; *Ney* – неясность; *Zl* – злость; *Nen* – ненависть; *St* – страх; *Rav* – равнодушие.

Каждый класс имеет подклассы (понятия из полученных ассоциативных рядов). Стоит отметить, что для того, чтобы идентифицировать однокоренные слова и синонимы, также используется эквивалентная связь. На базе этой логики была построена онтология в системе

настрой» текстов сообщений пользователя социальной сети, зависит от его принадлежности к социально-демографической группе. Содержание разработанной онтологии может уточняться по мере проведения экспериментов и использоваться как ключ в оценке психосемантического профиля. Открытым остается вопрос по формированию базы правил для автоматизированной оценки без привлечения эксперта в условиях присутствия в одном тексте ключевых слов из различных «эмоциональных шкал», поэтому продолжение работы видится в обозначенном направлении.

Список литературы

1. Артёмьева Е. Ю. Основы психологии субъективной семантики. – М.: Наука, 1999. – 350 с.
2. Губанов Д. А., Новиков Д. А. Модели репутации и информационного управления в социальных сетях // Управление большими системами. – 2009. – № 26-1. – С. 209-234.
3. Крестинина Е. С., Чернышов Ю. Г. Использование интернет-блогов и социальных сетей в российской публичной политике // Известия АлГУ 2008. – № 4-3. – С. 288-296.
4. Петренко В. Ф. Основы психосемантики. – СПб.: Питер, 2005. – 480 с.
5. Черников Б. В. Социальные сети и их влияние на организацию бизнеса // Вестник Томского государственного университета. Экономика. – 2011. – № 1. – С. 106-109.
6. Четвериков А. А. Дифференцированное измерение эмоциональности текстов с помощью алгоритма PMI-IR // Материалы XXXVIII Междунар. филол. конф. – Ч. 1. – СПб.: Факультет филологии и искусств СПбГУ, 2009. – 53 с.
7. Eytan Bakshy, Itamar Rosenn, Cameron Marlow, Lada Adamic The role of social networks in information diffusion [Электронный ресурс]. – Режим доступа: <http://arxiv.org/abs/1201.4145> (дата обращения: 01.02.2013).
8. Floriana Gargiulo, Alberto Mazzoni Can extremism guarantee pluralism? [Электронный ресурс]. – Режим доступа: <http://arxiv.org/abs/0803.3879> (дата обращения: 01.02.2013).
9. Julian Sienkiewicz, Marcin Skowron, Georgios Paltoglou и Janusz A. Entropy-growth-based model of emotionally charged online dialogues [Электронный ресурс]. – Режим доступа: <http://arxiv.org/abs/1201.5477> (дата обращения: 01.02.2013).
10. Peter Klimeka, Werner Bayera и Stefan Thurner The blogosphere as an excitable social medium: Richter's and Omori's Law in media coverage [Электронный ресурс]. – Режим доступа: <http://arxiv.org/abs/1102.2091> (дата обращения: 01.02.2013).

Рецензенты:

Макарова Е.А., д.т.н., доцент, профессор кафедры ТК УГАТУ, г. Уфа.

Задорожный В.Н., д.т.н., доцент, профессор кафедры АСОИУ ОмГТУ, г. Омск.