

**МЕТОД КОНВЕРСАЦИОННОГО АНАЛИЗА ДЛЯ ОЦЕНКИ АКТИВНОСТИ И ТОНАЛЬНОСТИ СООБЩЕСТВ В СОЦИАЛЬНЫХ СЕТЯХ**© 2020 И.А. Рыцарев<sup>1,2</sup>, А.В. Куприянов<sup>1,2</sup>, В.Г. Литвинов<sup>1</sup><sup>1</sup> Самарский государственный аэрокосмический университет имени академика С.П. Королёва (национальный исследовательский университет), г. Самара, Россия<sup>2</sup> Институт систем обработки изображений – филиал ФНИЦ «Кристаллография и фотоника» РАН, г. Самара, Россия

Статья поступила в редакцию 01.12.2020

В данной работе предлагается использовать метод конверсационного анализа для анализа активности и тональности сообществ. Данные для исследования были собраны из социальных сетей. Апробированная технология анализа активности сетевых групп по предвыборной тематике была использована на практике для определения активности и тональности сообществ. Полученные результаты показывают, что в активных группах существует сильная взаимосвязь между новостью и тематическими комментариями, связанными с ней.

*Ключевые слова:* Компьютерная обработка текстов, конверсационный анализ, социальные сети.  
DOI: 10.37313/1990-5378-2020-22-6-87-91

**1. ВВЕДЕНИЕ**

Компьютерная лингвистика — направление прикладной лингвистики, ориентированное на использование компьютерных программ для моделирования функционирования языка в тех или иных условиях, а также сфера применения компьютерных моделей языка в лингвистике и близких ей дисциплинах. Дисциплина зародилась в 60-е годы XX века и прежде всего представляла собой разработку методов для общения человека с ЭВМ на естественном или ограниченно естественном языке [1].

Интонация и подтекст живого языка, его разговорный контекст, может быть косвенно измерен только при соотношении с другими словами, содержащимися в высказывании [2,3]. Слово, вырванное из высказывания, может диссонировать со смыслом, которое оно имело в контексте. Поэтому слово не может быть единицей однозначного анализа для текстов, содержащихся в социальной многоуровневой коммуникации. Как показали исследования, более емкой и более показательной единицей содержания может служить, так называемый, фиксированный семантический код. Семантический код – это единица анализа, состоящая из пятидесяти наиболее часто встречающихся слов в коммуникации виртуального сообщества по-

сле отделения местоимений, союзов, предлогов и числительных. Семантический код может интерпретировать как вектор содержания групповой коммуникации.

**2. МЕТОД КОНВЕРСАЦИОННОГО АНАЛИЗА**

Данное исследование посвящено разработке метода векторизации неструктурированных текстовых данных для анализа контекстных связей на основе конверсационного анализа.

Конверсационный анализ – изучение структур и формальных свойств языка, рассматриваемого в его социальном использовании.

Исходя из предпосылок, изложенных в статье [4], особенности конверсационного анализа можно сформулировать следующим образом. Во-первых, этот метод следует за данными, т.е. анализ базируется на эмпирии без привлечения (по возможности) заранее сформулированных гипотез. Во-вторых, мельчайшие детали текста рассматриваются как аналитический ресурс, а не как помеха, которую надо отбросить. В-третьих, важен порядок в организации деталей высказываний [5].

Обобщая определение конверсационного анализа можно отметить, что основной задачей метода является анализ контекста сообщения, а не его отдельных частей, что является отличительной чертой метода [6].

В настоящей работе целью исследования является разработка метода оценки активности социальных сообществ для анализа тональности публикуемого контента с применением методов конверсационного анализа. В качестве объекта исследования выбраны новостные сообщения и комментарии к ним в крупнейших информа-

*Рыцарев Игорь Андреевич, аспирант кафедры технической кибернетики. E-mail: rytsarev\_igoryan@mail.ru*

*Куприянов Александр Викторович, доктор технических наук, заведующий кафедрой технической кибернетики. E-mail: akupr@ssau.ru*

*Литвинов Владимир Геннадьевич, кандидат технических наук, доцент кафедры информационных систем и технологий. E-mail: litvinov.vg@ssau.ru*

ционных группах социальной сети. Источником данных для исследования была выбрана социальная сеть Вконтакте.

Был разработан собственный программный комплекс на языке программирования Python, содержащий модуль авторизации, модуль сбора данных, модуль фильтрации и модуль контекстного анализа. Общая схема комплекса представлена на рисунке 1. Данный программный комплекс позволяет собирать данные и фильтровать их с целью выделения только необходимой информации [7].

При помощи разработанного программного комплекса было собрано более 11000 постов и более 190000 комментариев к ним из 15 наиболее популярных сообществ города в период 1 августа по 9 сентября 2018 года.

### 3. АНАЛИЗ АКТИВНОСТИ И ТОНАЛЬНОСТИ СООБЩЕСТВ

В рамках исследования удалось выявить наиболее активные группы, в которых происходит обсуждение нового содержания (группы, в которых участники состоят, читают и активно обсуждают публикуемые сообщения).

В таблице 1 представлены результаты мониторинга разных сетевых групп в течение августа 2019 года.

Технология анализа активности сетевых групп по предвыборной тематике была апробирована на практике в течение месяца избирательной кампании. По каждой из анализируемых групп были получены данные по ежедневной активности упоминания основных тем, связанных с предвыборной кампанией. Для оценки тональности сообщества по отно-

шению к ключевым словам были сформированы частотные словари употребления текстовых единиц. Данные словари прошли процедуру нормировки в пределах  $[0;1]$ , где 0 – слово редко употребляется, 1 – слово часто употребляется по следующей формуле:

$$y(x) = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (1)$$

где  $x_{min}$  – минимальное значение среди элементов в векторе,  $x_{max}$  – максимальное значение среди элементов в векторе. На основании подсчета частот разных частиц был определен общий эмоциональный настрой участников групп.

Были проанализированы результаты с точки зрения динамики общественного мнения и настроения. Полученные результаты показывают, что в активных группах существует сильная взаимосвязь между новостью и тематическими комментариями, связанными с ней. Также результаты позволили выявить эмпирический факт, что в результате публикации новости, где фигурировали фамилии кандидатов, резко возросло количество отрицательных частиц в тексте, что может свидетельствовать о попытках изменения общественного мнения, а также привлечения общественного внимания при помощи негативных комментарии в социальных сетях.

Для наглядности приведены графики рассчитанной тематической активности (в основных сообщениях и в комментариях) в двух самых активных группах: «Группа 8» (Рисунок 2) и «Группа 6» (Рисунок 3).

Полученные результаты наглядно показывают разницу в динамике двух групп, приведенных в качестве примера. Если в наиболее активной группе «Группа 9» активность обсуждения и комментирования предвыборных тем суще-

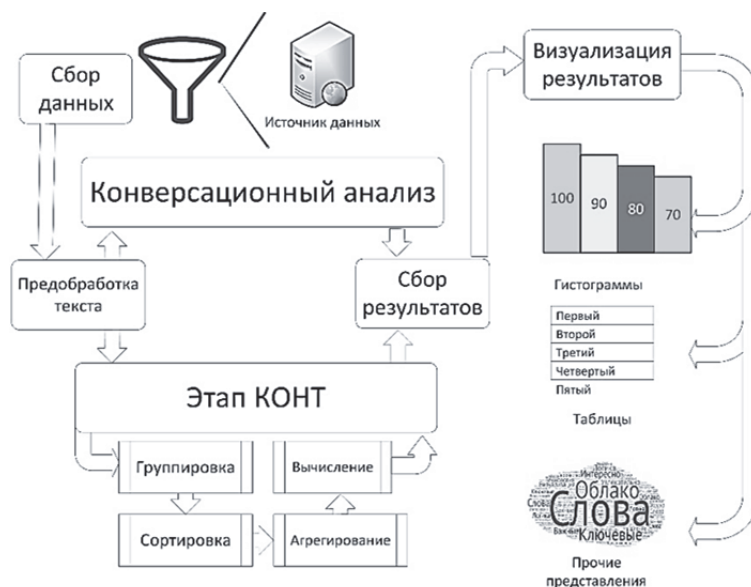
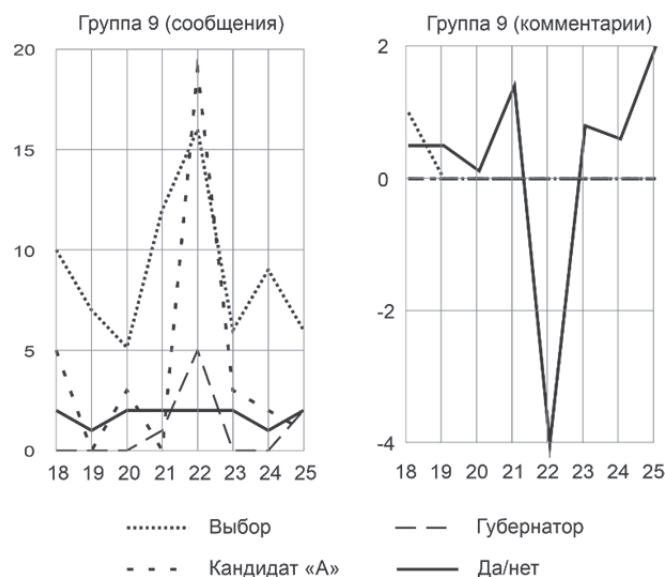


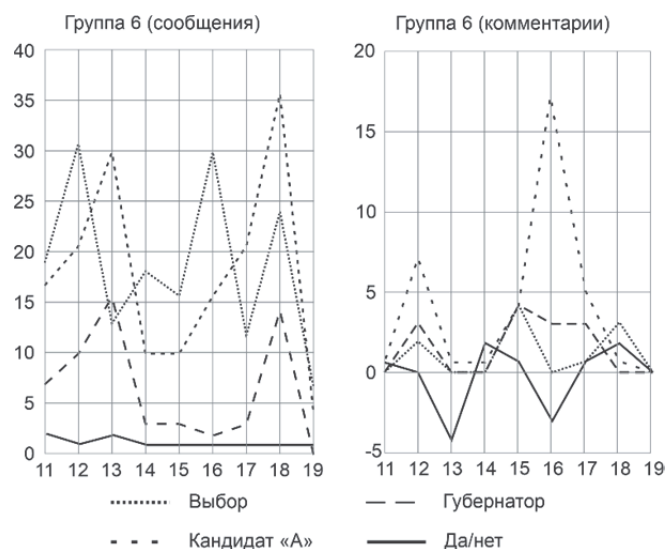
Рис. 1. Технология КОТ с модулем конверсационного анализа

Таблица 1. Активность исследуемых сообществ

Название группы и индекс активности (число)	Параметр	Значение
Группа 1	комментарии	974
1,921104536	основной текст	507
Группа 2	комментарии	3302
7,521640091	основной текст	439
Группа 3	комментарии	234
0,696428571	основной текст	336
Группа 4	комментарии	0
0	основной текст	456
Группа 5	комментарии	529
1,005703422	основной текст	526
Группа 6	комментарии	77580
23,10303752	основной текст	3358
Группа 7	комментарии	1781
5,801302932	основной текст	307
Группа 8	комментарии	881
0,943254818	основной текст	934
Группа 9	комментарии	86916
65,44879518	основной текст	1328
Группа 10	комментарии	2974
3,857328145	основной текст	771
Группа 11	комментарии	880
1,644859813	основной текст	535
Группа 12	комментарии	4489
9,142566191	основной текст	491
Группа 13	комментарии	7883
10,04203822	основной текст	785
Группа 14	комментарии	3242
7,318284424	основной текст	443
Группа 15	комментарии	329
5,771929825	основной текст	57



**Рис. 2.** Пример ежедневной динамики упоминания основных тем предвыборной кампании в группе «Группа 9» в основных текстах и комментариях к ним в период 18-25.08.18



**Рис. 3.** Пример ежедневной динамики упоминания основных тем предвыборной кампании в группе «Группа 6» в основных текстах и комментариях к ним в период 11-19.08.18

ственно сократилась, то в группе «Группа 6» этого не произошло. Результаты показывают, что снижение активности обсуждения предвыборной тематики в первой группе связано с тем, что в основных текстах группы в течение последней недели перед выборами темы, связанные с самими выборами и кандидатом «А» просто перестали публиковаться, что и снизило количество соответствующих упоминаний и, что более важно, общую негативность высказываний по поводу предстоящих выборов и участников предвыборной гонки. В то время как в сообществе «Группа 6» тематические новости по данным тематикам продолжали публиковаться, что привело даже к некоторому росту комментариев и обсуждения соответствующих тем и спровоцировало более негативные высказывания и обсуждения.

## 5. ЗАКЛЮЧЕНИЕ

В рамках исследования были выявлены наиболее активные сообщества социальной сети. Был реализован алгоритм векторизации на основе конвексационного анализа. На основании подсчета частот разных частиц в активных группах был определен общий эмоциональный настрой участников сообщества. Результаты позволили выявить эмпирический факт, что в результате публикации новости, где фигурировали определенные ключевые слова (фамилии кандидатов), резко возросло количество отрицательных частиц в тексте, что позволяет сделать вывод о попытке намеренного изменения отношения к кандидату при помощи целенаправленной публикации негативных отзывов.

Данная работа подтверждает применимость методов конверсационного анализа для обработки данных социальных сетей. Ведутся разработки модификаций предложенного метода с применением матриц порядка, а также частотных словарей для повышения достоверности определения тональных характеристик сообществ.

## 6. БЛАГОДАРНОСТИ

Результаты исследования были получены в рамках выполнения государственного задания Минобрнауки России (Проект № 0777-2020-0017), при частичной финансовой поддержке РФФИ в рамках научных проектов № 19-29-01135, № 19-31-90160.

## СПИСОК ЛИТЕРАТУРЫ

1. Антонова А. И. Актуальные прикладные задачи компьютерной лингвистики XXI века // Культурные инициативы. 2019. С. 38-39.
2. Rosenthal S., Farra N., Nakov P. SemEval-2017 task 4: Sentiment analysis in Twitter // Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). – 2017. – С. 502-518.
3. Rubtsova Y. Reducing the Deterioration of Sentiment Analysis Results Due to the Time Impact // Information. – 2018. – Т. 9. – №. 8. – С. 184.
4. Рыцарев И.А., Кири Д.В., Куприянов А.В. Кластеризация медиаконтента из социальных сетей с использованием технологии bigdata // Компьютерная оптика. 2018. Т. 42. № 5. С. 921-927.
5. Рыцарев, И.А. Анализ текстовых данных с применением конверсационного анализа /И.А.Рыцарев // Информационные технологии и нанотехнологии (ИТНТ-2020): сборник трудов VI Международной конференции и молодёжной школы. 2020. С. 60-63.
6. Исупова О.Г. Конверсационный анализ: представление метода // Социология: методология, методы, математическое моделирование (4М). 2002. №. 15. С. 33-52.
7. Rytsarev, I.A. Text data mining using conversation analysis // CEUR Workshop Proceedings. 2020. Iss. 2667. P. 159-161.

## APPLYING THE CONVERSATION ANALYSIS METHOD TO ASSESS COMMUNITY ACTIVITY AND SENTIMENT

© 2020 I.A.Rytsarev<sup>1,2</sup>, A.V. Kupriyanov<sup>1,2</sup>, V. G. Litvinov<sup>1</sup>

<sup>1</sup> Samara National Research University, Samara, Russia

<sup>2</sup> Image Processing Systems Institute of RAS –  
Branch of the FSRC “Crystallography and Photonics” RAS, Samara, Russia

In this paper, it is proposed to use the method of conversion analysis to analyze the activity and sentiment of communities. The research data was collected from social media. The proven technology for analyzing the activity of network groups on electoral topics was used in practice. The results show that in active groups, there is a strong relationship between news and thematic comments associated with it.

*Keywords:* text processing, conversation analysis, social networks

DOI: 10.37313/1990-5378-2020-22-6-87-91

---

Igor Rytsarev, Postgraduate Student of the Department of Technical Cybernetics. E-mail: rytsarev\_igoryan@mail.ru  
Alexander Kupriyanov, Doctor of Technics, Head of the Department of Technical Cybernetics. E-mail: akupr@ssau.ru  
Vladimir Litvinov, Candidate of Technics, Associate Professor of the Information Systems and Technologies Department. E-mail: litvinov.vg@ssau.ru