

# МЕТОДИКА ОБРАБОТКИ И ВИЗУАЛИЗАЦИИ СООБЩЕНИЙ ПОЛЬЗОВАТЕЛЕЙ В СОЦИАЛЬНЫХ СЕТЯХ

В.М. Саклаков

Научный руководитель – А.О. Савельев

Томский политехнический университет

romanov\_ky@mail.ru

## Введение

Социальные сети – достаточно крупный источник эмпирических данных, потенциально интересный многим исследователям. С одной стороны, огромное количество пользователей оставляет «цифровые следы» [1], при анализе которых формируется возможность моделирования общественных процессов и явлений разного уровня [2-3]. С другой стороны, количество данных, генерируемых пользователями в ежедневном режиме в достаточной степени затрудняет проведение их комплексного анализа.

Проблемная ситуация заключается в том, что исследователи, как правило, идут по пути анализа сообщений и иной активности отдельных пользователей и групп [4-5]. При этом не уделяется достаточного внимания систематизации описания процесса распространения данных в общественных системах, их преобразования в информацию и, в конечном счете, в действия или бездействия отдельных людей или групп людей в определенных «точках» общественных процессов.

Целями настоящей работы является (1) краткий обзор традиционного подхода к визуальной аналитике данных и (2) разработка собственной методики интеллектуальной обработки сообщений пользователей социальных сетей и визуализации полученных данных.

## Существующие методы визуализации данных

Традиционным подходом для облегчения процедуры анализа данных является их визуализация, как правило в виде графов [6-7]. Данную модель можно разделить на два уровня: (1) «объект-связь-объект» (иногда «процесс-связь-процесс») и (2) «класс-кластер-множество связей-множество других кластеров-множество других классов». Первый уровень обобщения (практический) является более низким – на нем мы можем увидеть лишь связи между отдельными объектами. При переходе на второй уровень (системный) для каждого объекта на необходимом исследователю уровне проводится классификация – выделяются свойства каждого класса и определяется принадлежность всех объектов к определенному из них. Затем, согласно заданным свойствам проводится кластеризация – упорядочивание объектов в сравнительно однородные группы.

Данный подход дает определенные результаты, однако имеет известные ограничения для применения процессного подхода к анализу общественных систем. Главным его недостатком с точки зрения классического системного анализа [8] является использование только моделей «состава» и «структуры». При этом отсутствие инструмента модели «черного ящика» не позволяет получать комплексное понимание протекающих в обществе процессов и явлений.

## Методика обработки сообщений пользователей

Предлагаемая модель, на уровне отдельных сообщений так же является двухуровневой, декомпозируется по следующему принципу: (1) «объект с вектором-связь- объект с вектором» и (2) «класс с вектором-кластер с общим вектором-множество межкластерных связей-множество других кластеров с вектором-множество других классов с вектором».

Для апробации методики из социальной сети Twitter программным способом был осуществлен сбор 360 сообщений («tweet») без учета повторных публикаций («retweet») для дальнейшего анализа. Ниже приведено одно из таких сообщений, на котором будет показана методика его обработки.

(Sun Sep 09 19:29:52 +0000 2018) «Судя по сообщениям, вот такие «вежливые люди» пытались сегодня провоцировать на «экстремизм» в Оренбурге. Рядом в темных очках оперативник ЦПЭ в штатском. Странно, что нормальные менты «в форме» не привлекли их, ведь в масках митинговать запрещено. #9сентября #Оренбург <https://t.co/SYZVIWdfnA> Likes:7

В первую очередь необходимо выделить объекты, описанные в сообщении: (1) вежливые люди, (2) «объекты, симулирующие экстремизм», (3) «Центр противодействия экстремизму» (ЦПЭ), (4) «нормальные менты».

На втором этапе установим связи между объектами. Очевидным образом группы (1, 3-4) из предыдущего абзаца связаны между собой. Группы «объектов, симулирующих экстремизм» связаны между собой через набор хештегов «#9сентября» и

«#Оренбург». Данная связь, по сути, является событийной «точкой» в одном из множества протекающих в обществе процессов.

Далее выделим инструменты воздействия: (1) симуляция экстремизма, (2) снижение уровня целостности, а значит и устойчивости целевой системы.

Представим концептуальную визуализацию полученных данных – рисунок 1.

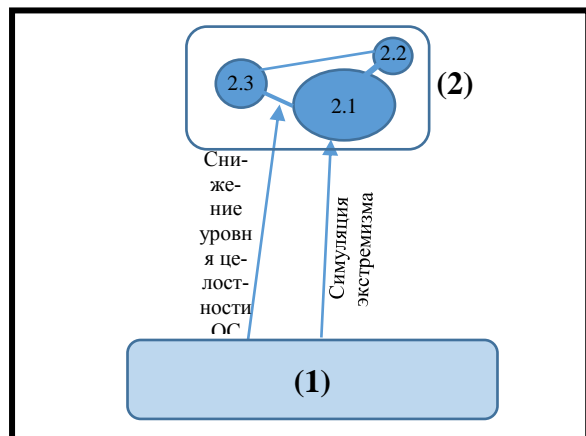


Рис. 1. Концептуальная визуализация обработанных данных

Легенда: (1) *Общественная система (ОС), «симуляторы экстремизма», (2) ОС «Силовые структуры»: 2.1 «Вежливые люди», 2.2 «Оперативник ЦПЭ», 2.3 «Менты «в форме»*

На представленном рисунке видно – множество объектов общественной системы, «симуляторы экстремизма» ведет деструктивное воздействие на другую – «Силовые структуры» с помощью перечисленных выше инструментов. Для более детального описания взаимодействия конкретных общественных систем необходим анализ существенно большего объема данных.

### Заключение

Разработанная методика обработки сообщений пользователей в социальных сетях является универсальной и подходит для выявления и прогнозирования трех видов информационного воздействия – позитивного, нейтрального и деструктивного. Разработка методики опиралась на инструментарий институциональной теории [9], процессного подхода, многомерного моделирования, а также концепции целеустремленных систем [10].

Описание процесса автоматизированного сбора данных, их обработки и анализа безусловно необходимо для проверки работоспособности гипотез, выдвигаемых в настоящей работе. В рамках продолжения работ по данной тематике на следующем этапе планируется осуществлять автоматизацию

моделирование цепочки генерации, диффузии воздействующего контента и реакции на него социальных систем. Это поможет выявить объекты и субъекты воздействия и, в конечном счете, один из значимых общественных процессов.

### Список использованных источников

1. Лушин Е. А. О термине «электронно-цифровые следы» // *Расследование преступлений: проблемы и пути их решения* №4, 2017. с. 161-163.

2. D. Xie, J. Xuy, T.-C. Lu. Automated classification of extremist twitter accounts using content-based and network-based features. *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016* 2016, vol. 7840895, Pages 2545-2549

3. Володенков С. В. Total Data как феномен формирования политической постреальности // *Вестник Омского университета. серия: исторические науки* №3 (15), 2017. с. 409-415

4. Азаров А.А., Бродовская Е.В., Дмитриева О.В., Домбровская А.Ю., Фильченков А.А. Стратегии формирования установок протестного поведения в сети интернет: опыт применения киберметрического анализа (на примере Евромайдана, ноябрь 2013 г.) // *Мониторинг общественного мнения: экономические и социальные перемены* №2 (120), 2014. с. 63-78

5. Социальные сети и деструктивный контент / Остапенко А. Г., Паринов А. В., Калашников А. О. и др.; Под ред. Чл.-корр. РАН Д. А. Новикова. – М.: Горячая линия – Телеком, 2018. – 276 с.: ил. – (Серия «Теория сетевых войн»; Вып. 3)

6. Hammer B., Hasenfuss A. Topographic mapping of large dissimilarity data sets // *Neural Computation*. Vol. 22, issue 9, 2010, p. 2229-2284. DOI: 10.1162/NECO\_a\_00012

7. J. Bollen, H. V. Sompel, A. Hagberg, L. Bettencourt, R. Chute, M. A. Rodriguez, L. Balakireva. Clickstream data yields high-resolution maps of science // *PLoS ONE* Vol. 4, Issue 3, 2009, Article number e4803. DOI: 10.1371/journal.pone.0004803

8. Тарасенко Ф. П. Прикладной системный анализ: учебное пособие / Ф. П. Тарасенко. – Москва: КРОНУС, 2017. – 220 с.

9. Норт Дуглас. Институты, институциональные изменения и функционирование экономики \ Пер. с англ. А.Н. Нестеренко; предисл. и науч. ред. Б.З. Мильнера. - М.: Фонд экономической книги «Начала», 1997. - 180 с. - (Современная институционально-эволюционная теория).

10. Р. Акофф О целеустремленных системах / Р. Акофф, Ф. Эмери – М.: Книга по Требованию, 2012. – 270 с.