

Мультимодальная кластеризация социальных сетей в социальном демпфировании с использованием BIG DATA (получение знаний из данных)

И.Н. Хаймович^{1,2}, В.М. Рамзаев¹, В.Г. Чумак¹

¹Самарский университет государственного управления «Международный институт рынка», Г.С. Аксакова 21, Самара, Россия, 443030

²Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Разработанная методика обеспечивает решение двух принципиально значимых задач, раскрывая тем самым гносеологический потенциал технологии Больших данных: социальное прогнозирование в трех наиболее значимых сферах информационного общества с опорой на модель выявления условий возникновения социального резонанса; успешную реализацию процедуры социального демпфирования на основе применения соответствующих вариантов управленческого воздействия с использованием мультимодальной кластеризации социальных сетей на основе технологии Big Data. В статье предложен инструмент для повышения эффективности работы в области социального демпфирования в регионе. Предложенный способ регулирования может быть эффективен при управлении процессами социального демпфирования региона, для которых характерны многообразие форм и широкий спектр составляющих и факторов, а также свойственна динамика развития и активная трансформации жизнедеятельности. При этом использование современных программно-аппаратных средств позволяет производить оценку и визуализацию изменений фактически в режиме реального времени, что может быть полезно органам власти на местах.

1. Введение

Рассмотрим анализ гносеологического потенциала Big Data с позиций синергетики, акцентирующей свое внимание преимущественно на неравновесных, неупорядоченных системах, формирующихся как в природных, так и в социальных средах, обретающих равновесие лишь на определенные моменты, зачастую крайне быстротечные, но от этого не менее важные и требующие всестороннего осмысления.

Безусловно, выход системы, и прежде всего системы социальной, из равновесия крайне интересен с диалектических позиций как познания причин данного процесса, так и в плане изучения возможности обеспечения системе устойчивого состояния. Однако прагматическая реальность убеждает нас в том, что многочисленные обыватели, являющиеся непосредственными элементами различного рода систем, так же, как и политическая элита, стремящаяся удерживать позиции ведущей социальной силы, обладающей определенными рычагами влияния на социальную динамику, все же заинтересованы прежде всего в устойчивости системных структур. Отклонение от усредненных значений социальных

показателей с очевидностью воспринимается как девиантность. В этих условиях технические возможности Big Data позволяют рассматривать их как инструмент индикации уровня отклонения самых разнообразных социальных процессов от оптимальной для данного общества модели, способствуя выработке механизма возврата к усредненной, равновесной форме как отдельно взятого процесса, так и целой их совокупности.

По нашему мнению, для проведения научного анализа социальных процессов необходимо выявить сущность многоплановых отклонений, которые возникают в тех или иных ситуациях в социуме. Анализ, проводимый в данном направлении, показывает возможность выявления двух основных типов изменений социальных процессов. Первый тип характеризуется достаточно медленным отклонением социальных процессов от устоявшихся в обществе норм и принципов, описываемых аддитивной функцией. В данном случае социальное сглаживание, реагирование можно рассматривать как когнитивное воздействие на каждый социальный субъект в отдельности, независимо от его природы: индивидуальной или коллективной. Здесь проявляется доминирование элементаристского подхода к описанию действительности, при котором качество системы определяется качеством входящих в неё элементов, вполне доступных для всестороннего анализа при помощи Big Data.

Сущность второго типа изменений социальных процессов определяется свойственными ей взрывными изменениями и прежде всего социальными катаклизмами, выводящими социум из равновесия. Второй тип в своей сущности обнаруживает отклонения мультипликативного типа, усиливающиеся благодаря крайне низкой эффективности процессов взаимодействия социальных субъектов. Здесь, в отличие от первого типа, социальный взрыв происходит в более резкой форме, гораздо более сконцентрированной во времени и определяющейся во многом недостатками межгруппового взаимодействия, подобными социальному резонансу. Данная ситуация представляет собой пример проявления в социальной практике системного подхода к действительности, исходя из принципов которого качество, а следовательно, и устойчивость социальной системы зависит от связей входящих в её состав элементов.

Для анализа классифицируемых указанным выше образом социальных процессов, обеспечения управления ими и приведения социальных отклонений в допустимую норму авторы предлагают ввести в понятийный аппарат термин «социальное демпфирование», принципиально новое для социально-гуманитарного знания. Данное понятие призвано обозначать стремление общества к усредненности в показателях проявления своей активности, следовательно, в стабильности. Подобное стабильное существование общества целесообразно характеризовать как социальное равновесие, особое состояние социальной системы, сложившееся благодаря исторической преемственности, проявляющей себя в целой совокупности факторов экономического, социального, этнического и т.п. характера.

Отклонение от критериев социальной стабильности составляет сущность социального баланса и приводит к «вспучиванию» кривой социального равновесия, что обуславливает необходимость социального демпфирования как значимого проявления процессов управления, способного, в том числе, решать проблему когнитивного управления сознанием, отличающимся особенностями своего ментального проявления.

«Выравнивание» количественных (цифровых) характеристик должно осуществляться в приемлемых для данного общества показателях. Использование Big Data как максимально чувствительного инструмента измерения и детализации биогеосоциальных процессов позволяет проводить необходимые итерации, которые невозможно осуществлять при помощи обычного научного инструментария, определяя «доминанту, что направляет вектор развития конкретного явления или процесса» [1] и тенденции их изменений.

Постановка проблем представляемого авторами исследования направлена также на разъяснение вопроса, связанного с необходимостью определения вектора самого алгоритма социального управления, а именно: если мы задаемся вполне очевидной целью обеспечить грамотное управление социальными системами, то возникает вопрос, относительно чего (в статусе социальной цели) нам следует осуществлять подобного рода управление. Применение технологий Big Data обеспечивает новые возможности решения подобных задач. Это, в свою очередь, означает необходимость определения факторов и критериев стабильного состояния

структур, а точнее, объектов управления, в системе социальных координат. Частота отклонения от данных показателей, а также способность их нивелирования, по мнению авторов, и будет основополагающим моментом в понимании качества социального управления, что актуализирует необходимость разработки методик оценки степени отклонения социальных процессов и качественного состояния систем, в них участвующих, от приемлемого обществом их равновесного состояния. Означенная методика должна в обязательном порядке предусматривать возможность как прямого, так и косвенного воздействия на социальные системы, что, по сути, дополнительным образом раскрывает указанное выше понятие «социальное демпфирование» как принципиально значимую теоретическую основу формирования и последующей реализации управленческих воздействий на социум, позволяющую эффективно уменьшать путем просчитанного с помощью Big Data пиковые, выходящие за рамки социальной приемлемости значения социальных показателей, игнорирование которых способно не только частично разбалансировать социальную систему, но сделать ее в полной мере социально неуправляемой и подверженной социальным взрывам. Вместе с тем, помимо социальных процессов, приводящих к отклонению среды от социальной «нормали», которое может регулироваться путем социального демпфирования, необходимо подробнее остановиться на выявлении сущности процесса социального резонанса. Социальный резонанс представляет собой такое состояние взаимодействующих факторов и критериев, характеризующих социальную среду, которое приводит к взрывному отклонению от «нормали». «Снятие» подобных социальных обострений возможно интерпретировать как процесс возвращения от социального резонанса к норме, что представляется авторам крайне важной задачей социального управления, успешное решение которой определяется необходимостью учета эффекта пассивного взаимодействия, понимаемого как своеобразная «активность, проявляющаяся в стремлении индивида к цели (часто - иллюзорной) и в способности к сверхнапряжениям и жертвенности ради достижения этой цели», при этом «жертвенность понимается... как отказ от удовлетворения ближайших потребностей, иногда жизненно важных, ради доминирующей социальной или идеальной потребности, осознаваемой как цель» с преобладанием «потребности развития» [1].

Таким образом, по мнению авторов, Big Data позволяют сформировать модель «снятия» социальных отклонений, в том числе социального резонанса, возникающего на основе сочетания целого ряда факторов, взаимодействие которых является весьма сложным для анализа и часто невидимым исследователю сквозь призму традиционных для науки инструментов познания. В свою очередь, реализация на практике подобного рода модели социального демпфирования позволяет составить комплексное понимание социальных процессов как нейробиологического проявления человеческой активности и формирует необходимые возможности для преодоления её негативных последствий.

Скрывая анализ аддитивных и мультипликативных эффектов за фасадом масштабных компьютерных вычислений, потенциал Big Data позволяет обеспечить принятие объективных решений, особенно значимых в сфере управления многофакторными социальными и природными системами [2,3], что, благодаря реализации на практике процедуры социального демпфирования, позволяет гарантировать стабильность, управляемость и предсказуемость биогеосоциальных процессов.

Проведенное нами исследование со всей очевидностью убеждает, что методология подобного рода должна предусматривать наличие в своей структуре двух смысловых блоков: собственно методологического, структурирующего алгоритм исследования социальных систем и процессов с помощью технологии Больших данных и управленческого, определяющего возможные методы управленческого воздействия на основе проведенного анализа.

Первый, методологический, блок предполагает прежде всего вычленение, с последующим анализом и оценкой, в сфере информационного пространства мультимодальных кластеров, наиболее взрывоопасных с точки зрения их потенциальной возможности дестабилизировать социальную систему, выведя ее из относительного равновесия. Здесь, в качестве глобальных кластеров, обращают на себя внимание три классических сферы общества: политическая, экономическая и социальная. Неоспоримая масштабность данных сфер позволяет вычленить в

каждой из них подкластеры с возможностью их дальнейшей детализации на группы (в том числе по интересам) и IP-адреса.

Указанную детализацию вполне возможно проводить на основании ключевых слов, которые задаются в соответствии с исследовательской задачей и используются технологией Больших данных для вычленения подкластеров и элементов их внутренней структуры. В дальнейшем, практическая реализация данного этапа предлагаемой нами методологии потребует параллельной разработки алгоритма «обучения системы», позволяющей программе самостоятельно вычленять слова, термины, понятия и т.д. для последующей кластеризации и выявления значимых связей.

Следующим шагом первого методологического блока является построение алгебраической решетки количества связей мультимодальной кластеризации, представляющей собою своеобразную систему координат конкретного проводимого анализа. Она позволит осуществить моделирование системы связей внутри кластера. Увеличение количества связей внутри кластера (либо его подсистемы) является основным показателем последующего выхода системы из равновесия – социального резонанса.

Наибольший интерес для последующего управленческого воздействия должны представлять связи, которые демонстрируют устойчивый рост. При их выявлении следует:

- конкретизировать темы, обеспечивающие актуальность существования данных связей;
- провести анализ частотной динамики;
- выявить зоны опасности возникновения социального резонанса.

Завершением первого методологического блока является дифференцирование скорости резонанса. Это предусматривает выявление направлений социальных, информационных взаимодействий по которым наметился устойчивый рост; последующее нахождение устойчивых связей для выявления темы взаимодействия. При этом, резкое увеличение количества связей в единицу времени на порядок должно интерпретироваться, в рамках данной методологии, как основное условие скорого возникновения социального резонанса.

Второй методологический блок определяет возможные варианты практического перехода от гносеологических рассуждений к непосредственному управленческому воздействию на социальную ситуацию. Точка принятия управленческого решения как проявление социального демпфирования может быть представлена на графике в месте перехода от плавного возрастания кривой, обозначающей увеличение количества социальных связей (внутри выявленного кластера, подкластера и т.п. структур) к вертикальной прямой, отображающей порядок увеличение социальных связей.

В данном случае, второй методологический блок допускает возможность двух вариантов управленческого воздействия:

- управление количеством социально-значимых связей, что предусматривает прежде всего их минимизацию для предотвращения или нивелирования социального резонанса;
- управление диссипативностью социальной среды, предусматривающее инструментальное, информационное воздействие (в том числе посредством информационных вбросов или тематической переадресации связей) на выявленный кластер, воздействие, не допускающее формирование социальных связей, способных в перспективе привести к возникновению социального резонанса.

В противном случае, совпадение «частот резонансов» в процессе межсубъектного взаимодействия, способно вызвать глобальный, неконтролируемый социальный резонанс.

2. Применение метода анализа формальных понятий для социальных сетей

Методология социального демпфирования состоит из определения социального резонанса в группах социальных сетей и поиска управляющих воздействий для снижения напряженности. Методология социального демпфирования основана на мультимодальной кластеризации социальных сетей. Кластеризация основана на методе анализа формальных понятий (Formal Concept Analysis, FCA) [4-9]. Большое количество структурированных и неструктурированных данных генерирует тривиальные данные.

По методу формальных понятий вводим следующие определения:

G - множество объектов, M – множество признаков, отношение $I \subseteq G \times M$ такое, что $(g, m) \in I$ тогда и только тогда, когда объект g обладает признаком m ; $K := (G, M, I)$ называется формальным контекстом.

Определяем операторы Галуа следующим образом: для $A \subseteq G, B \subseteq M$

$$A' \stackrel{def}{=} \{m \in M \mid g / m \forall g \in A\}, B' \stackrel{def}{=} \{g \in G \mid g / m \forall m \in B\}.$$

Формальное понятие есть пара $(A, B) : A \subseteq G, B \subseteq M, A' = B$ и $B' = A$, где A – формальный объем, B – формальное содержание.

Понятия, упорядоченные отношением $(A_1, B_1) \geq (A_2, B_2) \Leftrightarrow A_1 \supseteq A_2 (B_2 \supseteq B_1)$, формируют полную решетку, называемую контекстной решеткой $\beta(G, M, I)$.

Пример контекста социальных сетей и их контекстная решетка показан на рисунке 1.

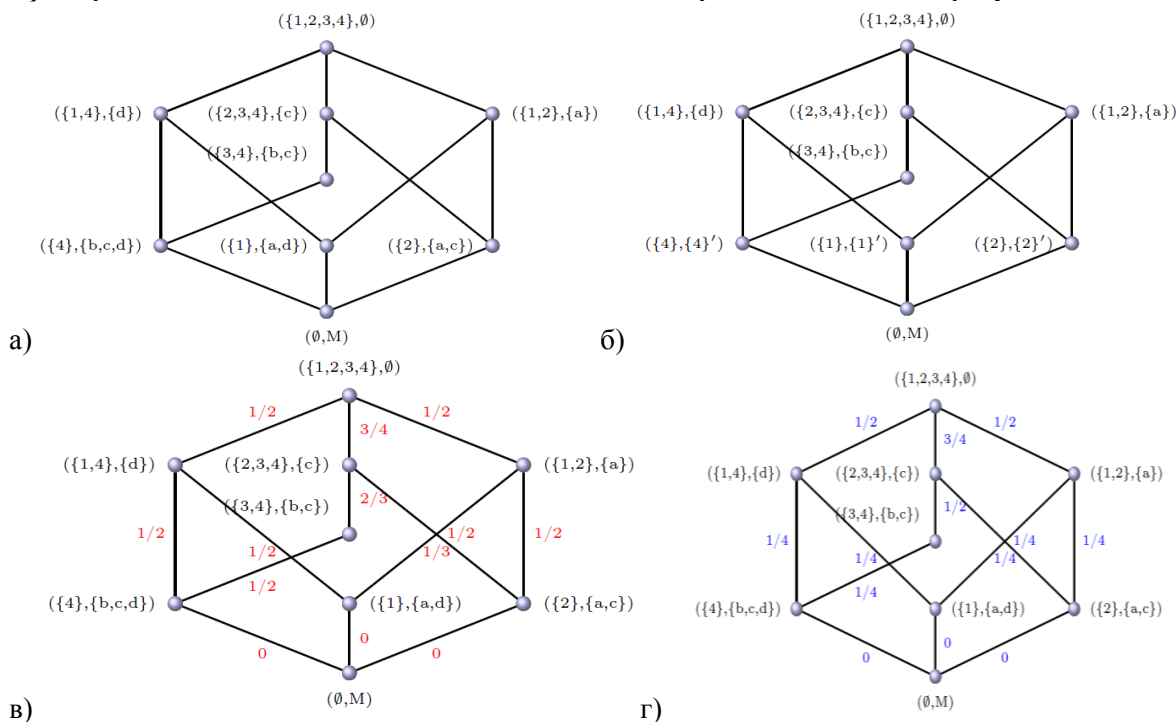


Рисунок 1. Социальные сети с контекстными решетками: а) контекстная решетка для социальных сетей; б) контекстная решетка для социальных сетей с импликацией; в) контекстная решетка с достоверностью ассоциативных связей; г) пример решетки с поддержкой ассоциативных правил.

Рассмотрим импликацию в решетках. Определим операцию «импликация» следующим образом: импликация $A \rightarrow B$, где $A, B \subseteq M$, существует, если $A' \subseteq B'$, т.е. каждый объект, который обладает всеми признаками из множества A , также обладает всеми признаками из B . Импликации соответствуют правилам Армстронга:

Рефлексивности $(\frac{A \rightarrow B}{A \cup C \rightarrow B})$; пополнения $(\frac{A \rightarrow B}{A \cup C \rightarrow B})$; псевдотранзитивности $(\frac{A \rightarrow B, D \cup B \rightarrow C}{A \rightarrow C})$.

Среди импликаций существует базис Дюкена – Гига, т.е. минимальное множество импликаций, из которых с помощью правил Армстронга выводимы все импликации. Ищется базис через методы машинного обучения: «пообъектный» алгоритм построения базиса импликаций или

через итеративную процедуру обучения. В итоге имеем контекстную решетку с импликацией (рис.2б).

Импликации: $abc \rightarrow d, b \rightarrow c, cd \rightarrow b$ позволяют переопределить вершины контекстной решетки.

Рассмотрим частичные импликации или ассоциативные правила.

Определение 1. $A \rightarrow B$ - частичная импликация (ассоциативное правило) контекста (G, M, I) ,

если $A, B \in M$; обладает поддержкой $\sup p(A \rightarrow B) = \frac{|(A \cup B)'|}{|G|}$; обладает достоверностью

$$\text{conf}(A \rightarrow B) = \frac{|(A \cup B)'|}{|A'|}.$$

Рассмотрим алгоритм нахождения ассоциативных правил:

- 1) найти все частые (с поддержкой не ниже заданной) множества признаков;
- 2) достаточно найти все частые замкнутые множества признаков содержания контекста.

Пример решетки с достоверностью ассоциативных правил (рис. 1в). Пример решетки с поддержкой ассоциативных правил (рис. 1г).

Хорошие правила с $\sup p \geq 1/2$ и $\min \text{conf} \geq 3/4$. Определены по следующему алгоритму:

1. $0 \rightarrow c, \sup p(0 \rightarrow c) = \text{conf}(0 \rightarrow c) = 3/4$.
2. $c \rightarrow b, \sup p(c \rightarrow b) = 1/2, \text{conf}(c \rightarrow b) = 2/3$.

Использование данного метода кластеризации позволит определить группы по интересам, с увеличением связей в которых потребуется принимать управленческие решения. В информационном плане необходимо будет объединять сообщения из групп с повышенным количеством связей со спокойными «стабильными» группами, в которых количество связей не подвергается резкому изменению.

3. Методы использования интеллектуального анализа данных в методологии социального демфирования

Для создания портрета и информационной модели резонансных групп необходимо использовать технологии BIG DATA [10, 11]. Метод использования интеллектуального анализа данных состоит в следующем:

1. формирование набора больших данных в hadoop из twitter по фильтру «Самарская область», выявляющему количество обращений;
2. разделение сформированного набора по различным фильтрам, связанным с базовыми факторами резонансных отклонений;
3. проведение мониторинга потокового анализа неструктурированной информации по фильтрам;
4. принятие оперативных мероприятий в случаях устойчивых «всплесков» по количеству обращений;
5. разработка программы на языке Scala для работы с фильтрацией в области Больших Данных;
6. отладка и тестирование программы с набором практических данных;
7. анализ результатов вычисления.

Для получения данных используется социальная сеть «twitter», так как это «открытый» продукт, его применение не требует дополнительных инвестиций, а 50% пользователей Интернет имеют профили в данной программе. Twitter является второй по популярности сетью среди пользователей во всем мире, уступая лишь Facebook. Однако в отличие от Facebook, который не предоставляет открытый доступ к своим данным, Twitter такой доступ предоставляет, отсутствуют ограничения на доступ к наборам данных сервера. Пользователи данной социальной сети обмениваются в основном текстовой информацией, что является несомненным плюсом при обработке. Twitter не является предметной сетью и наиболее широко отражает общественное

мнение по многим интересующим вопросам, поэтому для формирования групп для анализа социального резонанса в регионе обработка данных из этой социальной сети была оптимальной.

Для работы с BIG DATA в социальных сетях используют методы сбора, обработки и анализа данных. Сбор данных осуществляется в режиме реального времени, в пределах определенной геолокации, либо в пределах всей сети, по определенным шаблонам. Информация, представляющая интерес для анализа в рассматриваемой области, это: локация, дата и время, контент, «автор» контента (пользователь), связи между пользователями. Сбор данных в социальных сетях можно осуществлять с помощью следующих инструментов: Apache Hadoop, BigInsights (IBM), Cloudera, Hortonworks, Storm. Для выполнения исследования в области социального демфирования был выбран Hortonworks. Для работы применялся Twitter Application (apps.twitter.com), в котором определялись и уточнялись ключевые параметры: API key, API secret, Access token, Access token secret.

Для сбора данных с использованием Hortonworks, Twitter App использовался конфигурационный файл сервиса flume в виртуальной машине Hortonworks Sandbox. После установки виртуальной машины Hortonworks_Sandbox версии 2.3 и настройки сервиса flume система готова к загрузке данных из twitter. Для просмотра и загрузки скаченных файлов переходим в папку HDFS, где осуществляем обработку данных. Вид файловой структуры HDFS в виртуальной машине Hortonworks при решении задачи в сфере социального демфирования показан на рисунке 2.

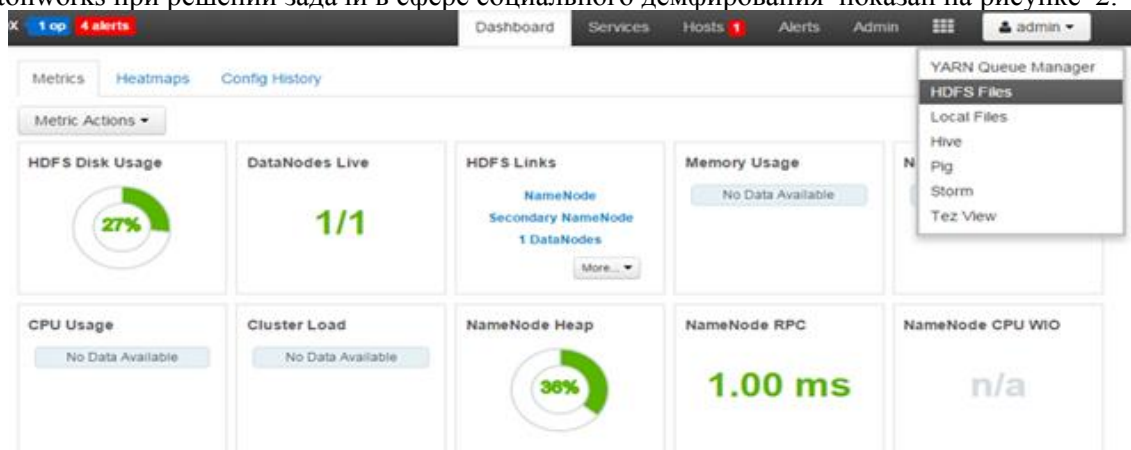


Рисунок 2. Визуализация HDFS в Hortonworks при загрузке файлов для решения задачи социального демфирования.

Собранные данные необходимо структурировать (т.е. обработать) в соответствии с парадигмой MapReduce. MapReduce — это фреймворк для выполнения распределенных задач с использованием большого количества компьютеров, образующих кластер.

Использование MapReduce позволило структурировать поток данных из социальных сетей по критериям: шрифты, размер текста, цвет, ссылка на профиль пользователя, локация, время и прочее.

Для определения портрета респондента нужны данные следующих типов: размещение, текст, язык и время. Для того, чтобы извлечь лишь эту информацию можно использовать технологию MapReduce, встроенную в инструменте Hortonworks Sandbox. Для обработки данных используем СУБД Hive в среде Hadoop, позволяющую осуществлять операции над данными и их анализ путем SQL – подобных запросов. Для этого создаем файл обработки и создания нужных таблиц hivedll.sql. Вид файла показан ниже:

```
//описание идентификаторов таблиц из twitter
CREATE EXTERNAL TABLE tweets_raw (
  id BIGINT,
  created_at STRING,
  source STRING,
  favorited BOOLEAN,
  retweet_count INT,
  retweeted_status STRUCT<
```

```

text:STRING,
usr:STRUCT<screen_name:STRING,name:STRING>>,
entities STRUCT<
urls:ARRAY<STRUCT<expanded_url:STRING>>,
user_mentions:ARRAY<STRUCT<screen_name:STRING,name:STRING>>,
hashtags:ARRAY<STRUCT<text:STRING>>>,
text STRING,
usr STRUCT< screen_name:STRING, name:STRING, friends_count:INT, followers_count:INT,
statuses_count:INT,
verified:BOOLEAN, utc_offset:STRING, -- was INT but nulls are strings time_zone:STRING>,
in_reply_to_screen_name STRING,
yearint,
monthint,
dayint,
hourint
)
CREATE EXTERNAL TABLE time_zone_map (
time_zone string,
country string,
notes string
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'
STORED AS TEXTFILE
LOCATION '/user/data/time_zone_map';
...
create table tweets_sentiment stored as orc as select
id,
case
when sum( polarity ) > 0 then 'positive'
when sum( polarity ) < 0 then 'negative'
else 'neutral' end as sentiment
from l3 group by id;
-- put everything back together and re-number sentiment
CREATE TABLE tweetsbi
STORED AS ORC
AS
SELECT
t.*,
cases.sentiment
when 'positive' then 2
when 'neutral' then 1
when 'negative' then 0
end as sentiment
FROM tweets_clean t LEFT OUTER JOIN tweets_sentiment s on t.id = s.id.

```

Запускаем данный файл командой: Hive_f hiveddl.sql. Структурированные данные будут размещены в таблице 1.

Таблица 1. Вид заголовков для анализа структурированных данных в задачах социального демфирования.

A	B	C	D	E	F
Data/Time	Time/Zona	language	Text	location	Sentiments

Для анализа данных используются следующие показатели. Общее количество твиттов (Kol_i) для каждого места размещения (R) определяется:

$$Kol_R = \sum_{i=1}^N k_i, k_i \in R,$$

где k_i - каждый следующий твитт из обрабатываемого потока.

Частота употребления уникального слова $ch(m)$ определяется из общего множества L текстовых данных:

$$ch(m) = \sum_{i=1}^N m_i, m_i \in L.$$

Отношение каждого твитта $otn(m, rez)$ может быть определен из тезауруса tez , в котором прописано отношение к данному слову:

$$otn(m, rez) = \begin{cases} 0, m - \text{негативное_значение} \\ 1, m - \text{нейтральное_значение} \\ 2, m - \text{положительное_значение}. \end{cases}$$

Для дальнейшей работы был составлен словарь, состоящий из фильтров предметной области, чтобы в дальнейшем определить количество твиттов по размещению $ch(m)$ и количество твиттов по размещению с учетом отношения $otn(m, rez)$. Определяем тезаурус с учетом фильтров по базовым факторам: зарплата, безработица, политические явления, жилищно-коммунальное хозяйство (ЖКХ). В итоге получаем 4 базовых фактора резонансных отклонений.

По фактору «зарплата» P_1 получаем количество твиттов в общем множестве текстовых данных L :

$$Kol_{otn P_1} = \frac{\sum_{i=1}^N S_i (S_i \in P_1)}{L} = 10\%.$$

По фактору «безработица» P_2 получаем количество твиттов в общем множестве текстовых данных L :

$$Kol_{otn P_2} = \frac{\sum_{i=1}^N S_i (S_i \in P_2)}{L} = 10\%.$$

По фактору «политические явления» P_3 получаем количество твиттов в общем множестве текстовых данных L :

$$Kol_{otn P_3} = \frac{\sum_{i=1}^N S_i (S_i \in P_3)}{L} = 10\%.$$

По фактору «жилищно-коммунальное хозяйство» P_4 получаем количество твиттов в общем множестве текстовых данных L :

$$Kol_{otn P_4} = \frac{\sum_{i=1}^N S_i (S_i \in P_4)}{L} = 10\%.$$

4. Результаты и обсуждение

В итоге можно сделать вывод о том, какие факторы отклонений, связаны с Самарской областью. По рисунку 3 видно, что основным фактором недовольства является сфера ЖКХ.

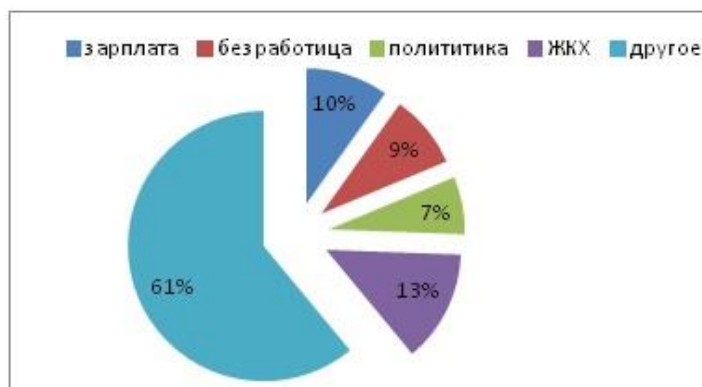


Рисунок 3. Распределение факторов социального демфирования в Самарской области.

Благодаря технологии BIG DATA можно хранить и обновлять данные в файловой системе «hadoop» по фильтру «Самарская область» (filter1= {Самарская область}). Затем необходимо данную область отфильтровать по базовым факторам социального демфирования, установив, например, следующие фильтры: Filter2 (зарплата) = {деньги, рубл*, долл*, валют*, криптовал*}; Filter3 (безработица)= {поиск работы, инж*, рабоч*, строит*}; Filter4 (политические явления)= {выборы, депутат*, пенс*, администр*}; Filter5 (ЖКХ) = {вывоз мусора, труб*, вод*, газ*}. Набор дескрипторов, по которым будет осуществляться фильтрация Интернет-дискурса определяется лексическими репрезентантами понятия, сформировавшегося в картине мира среднестатистического русскоязычного потребителя.

Для принятия решений в области социального резонанса в регионе была проведена мультимодальная кластеризация социальных сетей. Большое количество структурированных и неструктурированных данных социальных сайтов в рассматриваемой области можно представить в виде следующей тройки (пользователь, группа, интерес) (рисунок 4).

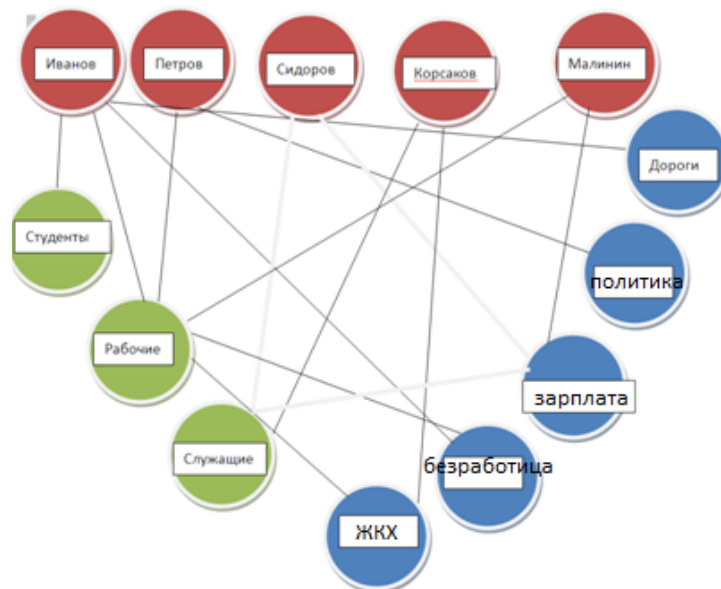


Рисунок 4. Данные для анализа социального резонанса из социальной сети «twitter» как граф.

По методу формальных понятий формируем полную решетку, называемую контекстной решеткой $\beta(G, M, I)$. Пример контекста социальных сетей в области социального демфирования и их контекстная решетка показаны в таблице 2 и на рисунке 5.

Использование данного метода кластеризации позволит определить группы по интересам, с увеличением связей в которых потребуется принимать управленческие решения. Но данный инструмент имеет ограничения по использованию. Пользователи, которые работают с социальной

сеть «Twitter», находятся в группе «студенты» и частично в группах «служащие», «рабочие» и лишь незначительно затрагивают часть группы «пенсионеры», поэтому для полноты принятия управленческих решений необходимо добавлять новые группы.

Таблица 2. Пример контекста данных для анализа социального резонанса из социальной сети (а - атрибуты по фильтру «зарплата», б - атрибуты по фильтру «безработица», в - атрибуты по фильтру «политические явления», г - атрибуты по фильтру «ЖКХ»).

	G/M	a	b	c	d
1	пенсионеры	x			x
2	служащие	x		x	
3	рабочие		x	x	
4	студенты		x	x	x

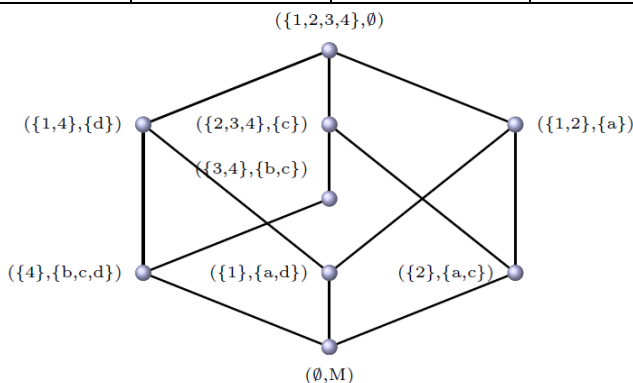


Рисунок 5. Контекстная решетка для социальной сети.

Можно получить графики зависимости количества обращений пользователей по фильтрам от времени сбора данных (рисунок 6). Время сбора данных из Интернета в технологии BIG DATA неограниченно.



Рисунок 6. График зависимости количества обращений пользователей по фильтру от времени сбора данных.

В итоге получаем динамическое изменение информации в режиме реального времени из системы Интернет, что позволяет с минимальными инвестициями проводить мониторинг потокового анализа неструктурированной информации (технология In-Memory Data Processing and Stream) по фильтрам. Для реализации данного метода была написана программа на языке Scala:

```

val file = spark.textFile("hdfs://... ")
val errors=file.filter(line=>line.contains("Самарская область"))
//count all the data
errors.count()
//count data mentioning Filter
errors.filter(line=>line.contains("валют* ")).count()
    
```

```
//Fetch the filter as an array of string  
errors.filter(line=>line.contains("туб*")).collect()
```

После работы программы получаем динамическое изменение параметров в среде BIG DATA, которые позволяют определять зоны социального резонанса в регионе с учетом неструктурированной информации. В случае выявления на графиках устойчивых «всплесков» данных по количеству обращений в соответствии с формами резонанса должно осуществляться управленческое регулирование по данному виду деятельности в регионе.

Таким образом, предложен инструмент для повышения эффективности работы в области социального демпфирования в регионе. Это является важнейшей задачей в современных экономических условиях, в основе решения которой лежит возможность принятия оптимальных управленческих решений. Предложенный способ регулирования может быть эффективен при управлении процессами социального демпфирования региона, для которых характерны многообразие форм и широкий спектр составляющих и факторов, а также свойственна динамика развития и активная трансформации жизнедеятельности.

При этом использование современных программно-аппаратных средств позволяет производить оценку и визуализацию изменений фактически в режиме реального времени, что может быть полезно органам власти на местах. Важно отметить, что социальное демпфирование не является жестким ограничителем социальных действий «по амплитуде». Оно лишь смягчает социальные действия, позволяя им проявляться в других сферах, не «ломает» социальную систему, а позволяет ей трансформироваться, создавая видимость органичного соответствия требованиям субъектов социального взаимодействия.

5. Литература

- [1] Bodrov, A.A. Philosophic aspects of developing new knowledge under data intellectual analysis (BIG DATA) / A.A. Bodrov, V.M. Ramzaev // CEUR Workshop Proceedings. – 2015. – Vol. 1490. – P. 338-345.
- [2] Khaimovich, I.N. Use of big data technology in public and municipal management / I.N. Khaimovich, V.M. Ramzaev, V.G. Chumak // CEUR Workshop Proceedings. – 2016. – Vol. 1638. – P. 864-872.
- [3] Khaimovich, I.N. Challenges of data access in economic research based on Big Data technology / I.N. Khaimovich, V.M. Ramzaev, V.G. Chumak // CEUR Workshop Proceedings. – 2015. – Vol. 1490. – P. 327-337.
- [4] Wille, R. Restructuring lattice theory: An approach based on hierarchies of concepts // Ordered Sets. NATO Advanced Study Institutes Series (Series C – Mathematical and Physical Sciences). – Springer, Dordrecht, 1982. – Vol. 83. – 470 p.
- [5] Ganter, B. Formale Begriffsanalyse / B. Ganter, R. Wille. – Springer, Heidelberg, 1996. – 540 p.
- [6] Ganter, B. Formale Concept Analysis / B. Ganter, R. Wille. – Springer, 1999. – 269 p.
- [7] Davey, B. Introduction to Order and Lattices / B. Davey, H. Priestly. – Cambridge University Press, Cambridge, 1990. – 460 p.
- [8] Denecke, K. Galois Connections and Applications / K. Denecke, M. Erne, S.L. Wismath. – Springer Science and Business Media, 2004. – Vol. 565. – 498 p.
- [9] Bonacich, P. Power and Centrality: A Family of Measures // American Journal of Sociology. – 2007. – Vol. 92(5). – P. 1170-1182.
- [10] Chumak, P.V. Models for forecasting the competitive growth of enterprises due to energy modernization / P.V. Chumak, V.M. Ramzaev, I.N. Khaimovich // Studies on Russian Economic Development. – 2015. – Vol. 26(1). – P. 49-54.
- [11] Khaimovich, A.I. Development of the requirements template for the information support system in the context of developing new materials involving Big Data / F.V. Grechnikov, A.I. Khaimovich // CEUR Workshop Proceedings. – 2015. – Vol. 1490. – P. 364-375.

Multimodel Clustering of Social Networks in Social Dampening Applying BIG DATA (acquiring knowledge from data)

I. Khaimovich^{1,2}, V. Ramzaev¹, V. Chumak¹

¹Samara University of Public Administration "International Market Institute", G.S. Aksakova 21, Samara, Russia, 443030

²Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The developed technique makes it possible to ensure the solution of two fundamentally significant tasks, thereby revealing the gnoseological potential of the Big Data technology: firstly, to carry out social forecasting in the three most significant areas of the information society based on a model for identifying the conditions for the social resonance uprise; secondly, to ensure the successful implementation of the social dampening procedure through applying the appropriate management options using multimodal clustering of social networks based on Big Data technology. The article proposes a tool to improve the efficiency of work in the field of social damping in the region. The proposed method of regulation can be effective in managing the processes of social damping of a region, which are characterized by a variety of forms and a wide range of components and factors, as well as an inherent dynamics of development and active transformation of life activity. At the same time, the use of modern software and hardware allows the assessment and visualization of changes in fact in real time, which can be useful for local authorities.