электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель Общероссийская общественная организация "Академия инженерных наук им. А.М. Прохорова"
ISSN 2307-0609

09, сентябрь 2017

УДК 303.712

Модель оценки логической связанности потоков информации в социальных сетях

***Ефремов Е.А.**, студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Защита информации»

***Ковалевский А.Е.**, студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Защита информации»

***Скворцова М.А.**, ассистент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Компьютерные системы и сети»
magavrilova@bmstu.ru

Поведение человека в социальных сетях является отображением его реальной жизни, следовательно, в его поведении в конкретный момент времени должна быть связанность с предыдущими событиями, которые были спроецированы в социальную сеть из реальной жизни.

С давних времен люди для сокрытия информации от лишних глаз пользовались шифрованием своих сообщений. Несмотря на шифрование при передаче данных в канале связи, информация хранится в открытом виде на серверах социальной сети[1][2]. Для сокрытия каких-либо фактов, команд и т.п. в террористическом сообществе используется сленг и лексическая подмена слов. Задача поиска пользователей социальной сети, употребляющих элементы подобного сленга, достаточно тривиальна и проста в реализации. Понимая это, террористическое сообщество использует лексические подмены слов, которые и становятся сигналами к определенным действиям. Например, фраза "Сегодня в 12 встречаемся на Пушкинской и готовим праздничный пирог" на первый взгляд выглядит безобидно, однако если проследить цепочку действий пользователя, то может оказаться, что эта фраза мало того, что не имеет никакого отношения к конкретному человеку, так и никак не связана с его предшествующими действиями[3][4].

Логическая связанность - череда событий, которые произошли в конкретный момент времени и имеют некоторую логическую последовательность. Построим упрощенную модель пользователя социальной сети. Пользователь генерирует случайные события:

Событие А - пользователь пишет личное сообщение кому-либо.

Событие А1 - сообщение, которое пишет пользователь, логически связанно с одним из предыдущих.

Событие Б - пользователь пишет сообщение всем.

Событие В - пользователь просматривает контент.

Событие Г - пользователь не просматривает контент продолжительное время.

Событие Д - пользователь проявляет симпатию к конкретному элементу контента.

Событие Е - пользователь проявляет антипатию к конкретному элементу контента.

Событие Ж - пользователь пишет комментарий к конкретному элементу контента.

Рассмотрим возможные варианты наступления события X, которое заключается в том, что пользователь сообщил всем о своем намерении приготовить праздничный пирог. Введем дополнительное событие З - пользователь совершил некоторое событие вне социальной сети, о котором решил сообщить после.

Найдем вероятность события X:

```
P(X) = P(З)*P(Б) + P(-З)P(Г)*P(A)*P(A1)*...*P(A1)*P(Б)+ ... + P(-З)*P(В)*P(A)*P(A1)*...*P(A1)*P(Б)
```

Так как событие X является комбинацией случайных независимых событий, то наступление какого-либо события в социальной сети можно связать с одной из комбинаций случайных независимых событий.

Пусть мы имеем i различных вариантов наступления События X (Обозначим каждый вариант как X_i), тогда $P(X) \sim P(X_i)$.

Из этого следует, что наступление некоторого события с большой долей вероятности связано с предшествующими событиями. Будем считать P(X) коэффициентом логической связанности. Возьмем выборку из n пользователей социальной сети и для каждого из них построим распределение логической связанности наступления события X. Зададим допустимую дисперсию распределений и затем на основе полученных данных построим аналогичное распределение, которое будет соответствовать усредненной логической связанности наступления события X.

Важно отметить, что выборка пользователей которая будет браться, должна учитывать отношение того или иного пользователя к «тренду», а именно: составить

распределение того, что сейчас "нормально", т.е. в тренде. Это получается путем обработки максимально возможного количества контента (музыки, картинок, постов). Далее необходимо все это классифицировать, и ~80% от наиболее встречающегося - это тренд, это нормально. Т.е. если примерно половина людей ходит зимой по улице шортах и футболках - считаем, что это нормально, это тренд. Далее понадобится потоковый анализ и классификация новых постов (а еще лучше и всего нового контента), дабы вычислять дисперсию (отклонение) нового контента от этого самого распределения, т.е. от нормальности. Плюсы данной модели в том, что она постоянно отслеживает "тренд", нормальность и то, что степень дисперсии можно задать вручную в зависимости от потребностей.

То есть, говоря про взятие выборки из n пользователей, надо иметь в виду взятие пользователей наиболее близких к генеральной совокупности. Для наглядности приведём ошибочное взятие выборки (рис. 1):

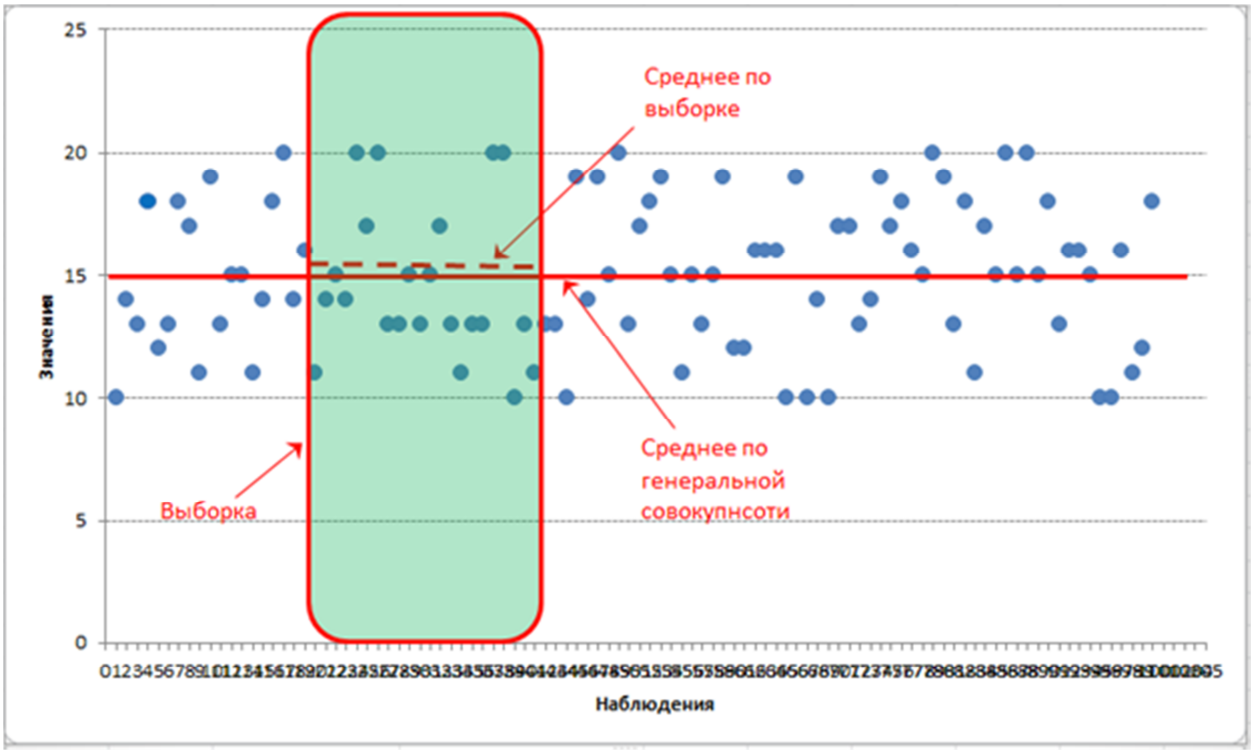

Рис. 1. Неправильная выборка

И правильной выборке (коричневым цветом на рис. 2):

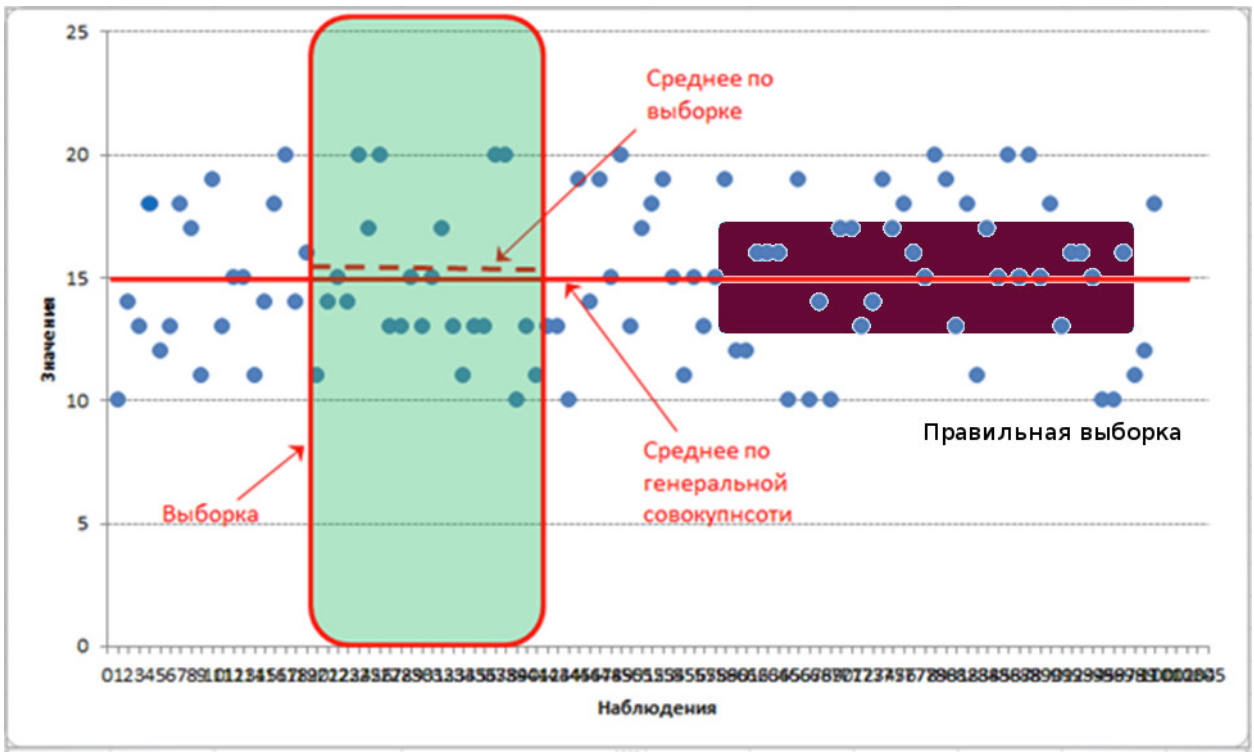

Рис. 2. Правильная выборка

Зададим допустимую дисперсию логической связанности наступления события X. (Для отсеивания крайних точек, таких как на рис. 3):

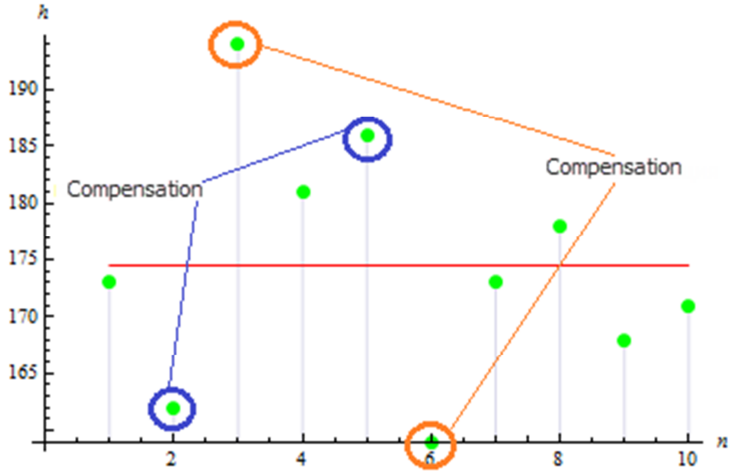

Рис. 3. Точки наиболее удаленные от медианы

То есть нас будут интересовать наиболее удаленные точки от средней по генеральной совокупности. Идея заключается в выявлении наиболее несвязанных событий

Таким образом, получим инструмент, способный анализировать логическую связанность интересующих нас событий. Наступление события, логическая связанность которого не попадает в доверительный интервал, может служить сигналом к ручной обработке данных.

Список литературы

[1]. Ressler S. Social network analysis as an approach to combat terrorism: Past, present, and future research // Homeland Security Affairs. 2006. Т. 2. №. 2.

[2]. Svenson P., Svensson P., Tullberg H. Social network analysis and information fusion for anti-terrorism // Proceedings of the conference on civil and military readiness. 2006. Т. 2.

[3]. Чураков А. Н. Анализ социальных сетей // Социологические исследования. 2001. № 1. С. 109-121.

[4]. Езепов Д. Выборочная несмещенная дисперсия. Режим доступа: https://statanaliz.info/metody/opisanie-dannyx/15-vyborochnaya-dispersiya (дата обращения 23.12. 2017)