

ОЦЕНКА ВЛИЯТЕЛЬНОСТИ ПОЛЬЗОВАТЕЛЕЙ СОЦИАЛЬНОЙ СЕТИ В ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ

Замятина В. С., Лунева Е. Е.
Научный руководитель: Лунева Е.Е.
Томский политехнический университет
zamiatina.v@gmail.com

Введение

В настоящее время социальные сети довольно часто используются для выражения пользователями своего личного мнения по интересующей их теме. Мнение некоторых пользователей оказывает информационное влияние на мнение других. Существуют сервисы, которые позволяют оценить популярность и влияние пользователей, однако любая информация об используемых методах вычисления является закрытой. Более того, большинство сервисов не предоставляет какие-либо данные о влиятельности пользователей в пределах какой-либо определенной предметной области [1, 2].

Цель данной работы – разработать способ, вычисления показателя влияния пользователей некоторой социальной сети в заданной предметной области.

Теоретический анализ

В результате теоретического обзора области вычисления информационного влияния пользователей друг на друга в социальных сетях было обнаружено, что подобного рода задача является задачей поиска «ключевых игроков» («keyplayers problem», KPP). В качестве математической модели, представляющей пользователей и структуры связей между ними, может быть использован социальный граф. В таком графе вершины представляют собой социальные объекты – пользовательские профили или сообщества, а ребра – связи между этими объектами [3].

Согласно С. Боргатти[4], существует два типа ключевых игроков: KPP-NEG (Negative) и KPP-POS (Positive). Для решения поставленной в данной работе задачи основной интерес представляют игроки второго типа, KPP-POS.

Существует ряд методов и подходов, позволяющий решить задачу поиска ключевых игроков [4-6]:

- Метод, основанный на вычислении показателей центральности социального графа. Недостаток такого подхода для решения представленной в данной работе задачи – нечувствительность показателей центральности, в частности, при их вычислении для избыточных графов.

- Метод, использующий комбинаторную оптимизацию и жадный алгоритм. Алгоритмы

такого метода требуют больших вычислительных затрат.

- Подход, основанный на вычислении информационной энтропии.
- Подход, основанный на вычислении коммуникационной эффективности.

Для решения поставленной задачи был выбран последний подход. Это обусловлено его относительно простой реализацией, низкими требованиями к вычислительному ресурсу, помимо этого, согласно работам, описывающим этот подход, он дает стабильные результаты. Коммуникационная эффективность социального графа G может быть вычислена следующим образом [5]:

$$E(G) = \frac{\sum_{i \neq j \in G} \varepsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} * \sum_{i \neq j \in G} \frac{1}{d_{ij}},$$

где N - количество узлов графа G , ε_{ij} - коммуникационная эффективность, обратно пропорциональная d_{ij} – кратчайшему расстоянию между узлами i и j .

При вычислении эффективности E узлы последовательно исключаются из графа. Узел, при удалении которого коммуникационная эффективность минимальна, представляет собой наиболее влиятельного пользователя. Однако при вычислении показателя эффективности требуется определить правила, согласно которым будет строиться социальный граф, а самое главное – определить вес его ребер, на основе которого вычисляются кратчайшие расстояния между узлами.

В современных социальных сетях информация распространяется через механизм подписки пользователей. Однако факт подписки пользователя еще не означает, что он действительно заинтересован в публикациях на какую-либо тему или подвержен информационному влиянию автора публикаций.

Интерес одного пользователя к мнению другого можно проследить по следующим характеристикам: количество комментариев к публикации, репостов этой публикации и упоминаний автора публикаций. Таким образом, мера заинтересованности пользователя A пользователем B может быть выражена функцией $f(x, y, z, l)$, где x - количество репостов с комментарием пользователя A , y – количество репостов, z – количество комментариев, l – количество упоминаний пользователя B пользователем A .

Предлагается определить функцию $f(x, y, z, l)$ согласно методу анализа иерархий, Т. Саати [7] посредством попарного сравнения значимости введенных характеристик x, y, z, l . По мнению авторов, а также опираясь на существующие в литературе представления, характеристики можно проранжировать следующим образом: x – очень высокая степень заинтересованности, y – высокая степень заинтересованности, z – средняя степень заинтересованности, l – некоторая степень заинтересованности. В результате чего, функция заинтересованности пользователя A пользователем B имеет вид:

$$f(x, y, z, l) = 0,68x + 0,12y + 0,1z + 0,09l. (1)$$

Таким образом, вес направленного ребра от узла A к узлу B представляет собой значение $1/f(x, y, z, l)$. В результате, предлагаемый способ вычисления влияния пользователей некоторой социальной сети состоит в вычислении коммуникационной эффективности взвешенного социального графа, ребра которого определяются функцией $f(x, y, z, l)$ в соответствии с (1).

Экспериментальный анализ

Для выполнения экспериментального анализа был реализован компонент в среде разработки Microsoft – Microsoft Visual Studio 2013, с использованием языка программирования C# и шаблона проектирования MVC (model-view-controller). Данные для анализа были загружены из социальной сети Twitter. Архитектура приложения представлена на рисунке 1.



Рис. 1 Диаграмма компонентов приложения

Сценарий работы приложения выглядит следующим образом. Пользователь вводит ключевое слово, по которому будет происходить поиск исходных данных. Программное приложение обращается к социальной сети Twitter и выгружает по заданной предметной области данные, на основе которых выстраивает социальный граф. По построенному социальному графу происходит анализ показателей «авторитетности» пользователей, результат которого предоставляется пользователю, а также передается компонентом на хранение в базе данных.

Экспериментальный анализ был произведен в две фазы: на подготовленных заранее данных, а также на реальных данных, загруженных из социальной сети Twitter. В случае с тестовыми наборами данных, где наиболее влиятельные пользователи были определены заранее, предложенный метод правильно определил влиятельных пользователей в 84% случаев. При

работе с реальными графами, тестирование было проведено на графах, содержащих от 6 до 500 узлов. Результаты эксперимента подтвердили работоспособность метода вычисления показателя влияния пользователя. Однако при обработке достаточно больших объемов данных программный компонент показывает низкую производительность. Такой недостаток может быть устранен путем модификации алгоритма вычисления кратчайших расстояний между узлами и алгоритма сортировки данных. Помимо этого, дальнейшее направление развития компонента – возможность использовать набор ключевых слов, что позволит более корректно выбирать набор анализируемых данных.

Заключение

В данной работе был представлен способ оценки влияния пользователя социальной сети, а также реализующий его компонент. Преимуществом разработанного компонента является возможность поиска влиятельных пользователей для заданной предметной области. Тестирование разработанного компонента как на подготовленных заранее данных, так и на реальных данных, загруженных из социальной сети Twitter, показало возможность применения предложенного способа и работоспособность компонента. Эксперименты на заранее подготовленных данных показали, что метод позволяет корректно определять влиятельных пользователей в 84% случаев. Дальнейшим направлением работы будет являться модификация используемых алгоритмов для повышения производительности компонента, повышение точности предложенного способа, а также предоставление возможности использования набора ключевых слов для более корректного выбора данных для анализа.

Список использованных источников

1. Del Campo-Avila J., Moreno-Vergara N., Trella-López M. Bridging the Gap Between the Least and the Most Influential Twitter Users // *Procedia Computer Science* - 2013. С 437-444
2. Cossu J-V, Dugue N, Labatut V Detecting Real-World Influence Through Twitter // *2nd European Network Intelligence Conf (ENIC)* - 2015. С 89-30
3. Luneva, E.E., Banokin, P.I., Yefremov, A.A., Tiropanis, T. Method of evaluation of social network user sentiments based on fuzzy logic // *Key Engineering Materials* – 2016. – Т. 685. – С. 847-851
4. S. P Borgatti. Identifying sets of key players in a social network // *Social Networks*. – 2006. С124-134.
5. Ortiz-Arroyo D. Discovering Sets of Key Players in Social Networks // *Computational Social Networks Analysis*. – 2010. С 32-33.
6. Stephen P Borgatti. On the Robustness of Centrality Measures Under Conditions of Imperfect Data. – 2006. С 22-34.

7. Саати, Томас Л. Принятие решений; Метод анализа иерархий : пер. с англ. //Т. Л. Саати. — Москва: Радио и связь, 1993. — 314 с.