DOI: 10.17117/na.2016.05.03.095

<http://ucom.ru/doc/na.2016.05.03.095.pdf>

Поступила (Received): 27.05.2016

Козловская К.В.**Реализация алгоритма выявления наиболее
влиятельных пользователей в социальных сетях****Kozlovskaya K.V.****The implementation of the algorithm to identify the most
influential users of social networks**

В данной статье рассматривается проблема распространения информации в социальных сетях. Приводится пример реализации современных способов, моделей и алгоритмов для решения проблемы, а также возможности для практического применения реализованного алгоритма

Ключевые слова: социальная сеть, распространение информации, модели распространения влияния, жадный алгоритм, ВКонтакте

The article discusses the problem of spreading of information in social networks. It describes an example of the implementation of modern methods, models and algorithms for the solution of problems and opportunities for practical application of the implemented algorithm

Key words: social network, spreading of information, diffusion models, greedy algorithms, VKontakte

Козловская Ксения Васильевна

Магистрант

Национальный исследовательский Нижегородский
государственный университет им. Н.И.

Лобачевского

г. Нижний Новгород, пр. Гагарина, 23

Kozlovskaya Ksenia Vasilevna

Master

State university of Nizhni Novgorod named N.I.

Lobachevsky

Nizhny Novgorod, Gagarin ave., 23

Стремительное развитие информационных технологий и различных он-лайн сетевых структур (социальные сети, форумы, блоги) значительно отразилось на социальном и экономическом взаимодействии людей. Люди и организации общаются и поддерживают отношения друг с другом с помощью платформ, которые им предоставляют социальные сети.

Сети – это фундаментальное средство для появления и проникновения информации. Поэтому, несомненно, возникает такая проблема, как распространение влияния в социальных сетях. Влияние здесь представляется в качестве новостей, идей, информации и т.п. В социальной сети интерес представляет не только решение пользователя-агента, но и его «ценность», влияние на остальных агентов социальной сети, т.к. их действия довольно часто основываются на мнении других агентов или на решениях, принимаемых другими членами, т.е. зависят от «социального влияния». Поэтому, появляется необходимость выяв-

ления небольшого количества агентов, которым необходимо предоставить информацию с целью последующего распространения по всей сети. Это бы позволило распространять информацию крайне быстро и затронув минимум пользователей. Такие возможности привлекают внимание множества специалистов, особенно маркетологов.

К настоящему времени исследователями опубликовано довольно много работ, посвященных распространению информации, влияния и идей между членами социальных сетей. После изучения данных исследований, было обнаружено, что для решения проблемы распространения информации и максимизации влияния в сети, учёными Д. Кэмпе, Д. Кляйнберг и Тардос предложен жадный приближенный алгоритм. [4] Данный алгоритм работает на основе следующего правила: на каждом шаге он выбирает множество, которое покрывает максимальное число элементов, из еще не покрытых. Было решено реализовывать именно этот алгоритм.

Работа проводилась в социальной сети «ВКонтакте» и языком для написания алгоритма являлся язык PHP. Социальная сеть «ВКонтакте» была выбрана из-за простоты её интерфейса, API, а также из-за отсутствия ограничений на анализ информации со стены пользователей.

Для разработки был выбран язык PHP, так как он является лидером на рынке веб-приложений. Язык PHP дает широкие возможности по увеличению производительности ресурса. Более того, сам сайт «ВКонтакте» написан на данном языке, что значительно упрощает предстоящую работу.

Прежде всего, для реализации алгоритма необходимо определиться с моделью распространения влияния. Были проанализированы несколько существующих моделей [1], а именно модель независимых каскадов, линейная пороговая модель, модель просачивания и заражения, а также модель коммуникаций. Каждая модель отражает основные свойства социальных сетей и обеспечивает распространение информации по сети. Однако, чтобы понять, какую модель лучше всего использовать при решении поставленной задачи, был проведен тщательный анализ моделей по различным факторам, таким, как: учет мнения агентов сети, учет изменения мнения агента под влиянием других членов социальной сети, различная значимость мнений агентов, наличие порога чувствительности агентов и прочие факторы.

По итогам анализа была выбрана модель линейных порогов, так как она наиболее полно отражает все признаки, присущие агентам социальной сети. В данной модели агенты (узлы) имеют два состояния – активное (принимает изменения и инновации) и неактивное. Модель работает следующим образом:

1) Выбор начального набора активных узлов A_0 . Предположим, что $\sum_{i \neq j} w_{i,j} \leq 1$, где $w_{i,j}$ – влияние узла i на узел j

2) В начале, каждый узел j случайно выбирает равномерно случайным образом порог $\varphi_j \in [0; 1]$ – он независим для каждого узла, отражает вероятность узлов принимать инновации, и представляет собой взвешенную долю соседей j , которые должны стать активными для того, чтобы j стал активным.

3) На шаге k , узел j становится активным, если он был неактивным на шаге $k-1$ и выполняется условие $\sum_{i \in A_{k-1}} w_{i,j} \geq \varphi_j$ (общий вес активных соседей к неактивному узлу j не меньше чем φ_j).

Итак, первым шагом в реализации были разработка и построение сеть взаимного влияния пользователей друг на друга. Так как объективно оценить влияние невозможно, сеть влияния строилась по следующему принципу:

- Выбор любого пользователя социальной сети и подсети его друзей;
- Получение «лайков» у последних 20 записей со стены каждого пользователя;
- Анализ «лайков» – кто из друзей кому их ставил и сколько. Важно отметить, что учитывались «лайки» только от пользователей внутри сети;
- Поиск максимального числа «лайков» в сети пользователей для задания эталона;

Распределение влияния каждого пользователя на другого по интервалу $[0,1]$, основываясь на эталоне и далее по убыванию.

В результате проведенных действий было известно влияние всех агентов выбранной подсети друг на друга. Так как была выбрана линейная пороговая модель, необходимо присвоить каждому пользователю сети порог, т.е. показатель, отражающий вероятность им принимать инновации. Данный порог выбирался случайно из интервала $[0,1]$.

После того, как была построена сеть влияния и каждому пользователю присвоены пороги, можно было переходить к непосредственной реализации самого алгоритма. Жадный алгоритм принимает локально оптимальные решения на каждом шаге работы, считая, что конечное решение при этом является оптимальным. В данном случае, алгоритм на каждом шаге должен выбирать пользователя, который мог бы сделать активными максимальное количество пользователей. Заражение происходит, только если влияние активного пользователя больше чем порог неактивного. Таким образом, после завершения работы алгоритма были получены 3 группы людей:

- те, на кого уже повлияли другие пользователи сети;
- те, кто повлиял на первую группу
- те, на кого еще нужно повлиять, т.е. пользователи, которым необходимо сообщить информацию напрямую, так как остальные пользователи в сети не имеют на них влияния.

Для анализа результатов работы алгоритма был проведен эксперимент на нескольких пользователях. Например, при запуске алгоритма на человеке с 2577 друзьями, результаты эксперимента показали, что чтобы заразить 2577 человек, достаточно оказать влияние на 1725 человек, из которых 1219 окажут влияние на оставшееся количество пользователей (851 человек).

В заключение, хотелось бы отметить, что область применения реализованного алгоритма для распространения информации в социальных сетях широка и не ограничена, начиная с исследований в различных областях и заканчивая разнообразным коммерческим использованием. Поэтому, можно сделать вывод о том, что данная работа оказалась актуальной, впоследствии может быть использована для решения различных маркетинговых задач.

Список используемых источников:

1. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Модели информационного влияния и информационного управления в социальных сетях // Проблемы управления. 2009.
2. Domingos P., Richardson M. Mining the network value of customers. P. 57–66. 2001.
3. Fan L., Wu W., Lu Z., W., Du D-Z. Influence Diffusion, Community Detection and Link Prediction in Social Network Analysis.
4. Kempe D., Kleinberg J. M., Tardos: Influential nodes in a diffusion model for social networks. P. 1127-1138, 2005.

© 2016, Козловская К.В.

Реализация алгоритма выявления наиболее
влиятельных пользователей в социальных сетях

© 2016, Kozlovskaya K.V.

The implementation of the algorithm to identify the
most influential users of social networks