Research on Opinion Leader Mining based on Influence Analysis

Chun Wang [a], Yicheng Fang [b]

School of Computer and Software Engineering, Xihua University, Chengdu 610039, China;

[a]chunxzx@qq.com，[b]fycwd@qq.com

Abstract

With the rapid development of the Internet, social network has gradually become one of the main network service platforms in people's lives. People with strong influence play a very important role in the process of information diffusion. At the same time, they have direct or indirect influence on a large number of others and be called opinion leaders. Especially, in the event of production, fermentation, communication and other aspects they are also special, which have important application value in the network marketing, public opinion guidance and other aspects. In this paper, we will study the opinion leader discovery algorithms in depth. Owning to the research background is based on the measure of social network influence. Therefore, this paper firstly summarizes the long-term measurement methods of influence, and then all the kind of models in opinion leaders will be discussed, and finally puts forward the existing shortcomings at present in the field of opinion leader mining.

Keywords

Opinion Leader, Social Network, Network Influence.

1. Introduction

In the 1840s, American scholar Lazarsfeld firstly proposed the concept opinion leader. Initially, the relevant scholars in the sociology explore the influence of the the inequality between individuals and found the differences of individual influence. In 1995, Elihu Katz [1] proposed a two-level communication theory by analyzing the intention of the voters in the US presidential election, and found that a small number of influential individuals affected the majority of the population. In 1962, Everett Rogers [2] described the definition of the influential individuals through the four typical characteristics:

(1) They can easily share their views with others;

(2) Their views often on behalf of public opinions;

(3) The point of views are novel;

(4) They own labels like Opinion Leaders, Innovators, Hubs or Mavens;

In 2006, Noah Friedkin [3] defined social influence which revealed people's roles would be diversity through social network links. With the coming of Web 2.0, researchers have the opportunity to study the complex relationship between users through large-scale networks and mass information. Online social networks provide a wealth of data for the study to verify or expand a number of theoretical models better, and a lot of social networking and related research and application were concerned.

2. Influence Analysis in Social Networks

Opinion leader discovery is based on the influence measure methods. The main task of the influence measure in the social network is to analyze and forecast the size and evolution of the social influence of the user, and to provide the theoretical basis for the research and application based on social influence. There are three types of commonly used influence measurement methods, they are based on the network topology, user behavior and the measurement of interactive information.

2.1 User Behavior

Based on the user behavior methods, the main consideration is the behavior of the users in the social network. If a user's behavior raises a lot of other users' attention or similar behavior, he can be own certain influence. For example, in Twitter network, many behaviors like forward, reply, reading, comment can be served as influential factors. If a user's many actions arouse other peoples' attention or follow, then the user should have a greater influence.

2.2 Interactive Information

The algorithms based on user interaction information are concerned with the information published by the user and its diffusion. Compared with the real social network, a significant difference is that the user in the social network mainly interacts through the flow of the network information. The information published by a user carries the record of the majority information of the user's activities, so many researchers think that the user's influence through interaction information can reflect the user's social influence. In such methods, the dissemination of information content and time is the main factors to considerate. The user's interactive information can further reflect the influence generation and the details of evolution. Therefore, in the analysis of social influence, it needs to choose the appropriate means of measurement according to the actual situation. Romero [4] analyzed the characteristics of the popular tags on Twitter in terms of propagation scope and time. Their results were consistent with those analyzed by Kempe [5], which indirectly demonstrated the importance of the time of information in the measurement of social influence.

2.3 Network Topology Analysis

2.3.1 Centrality Analysis

Centrality Degree refers to the number of nodes directly connected to the node, which can be used to analyze the direct influence of the node. The in-degree and out-degree of the nodes in social network can measure the users influential related indicators, such as the number of friends, forward, followers and so on. Closeness refers to the sum of the shortest path between the node in the social network and the other nodes. Betweeness is used to measure the influence of nodes on information dissemination, the ability of nodes to be on the shortest path of other nodes, and the size of the "mediator" role that individuals play in the network.

2.3.2 Random Walk Characteristics

In the field of influential indicators based on random walk characteristic, the common influential methods are Eigenvector center, Katz Center, PageRank and so on. Eigenvector Centrality calculates the weights of nodes based on the centrality of the nodes, the linearity of the weight of the other nodes that the current node can reach as the center the central degree of the node. Katz Center uses the walking path between the two nodes to calculate the influence between them. The farther away from the target user v on the walking path, the smaller the contribution of the Katz center to node v by the action of penalty factor α. The PageRank is an indicator of the nodes' ranking in a directed graph. If the user's influence propagation is seen as a random walk, the result can also be used to show the size of the user's influence.

3. Analysis of Opinion Leader Mining Algorithms

An important application of social influence analysis is the opinion leader mining in social network, because of the significant social influence is an important feature of opinion leaders. Currently, opinion leader mining on the social network mainly based on following three aspects.

3.1 User Attributes

Some studies construct a multidimensional model that identifies opinion leaders in micro-blog based on the user attributes which include the number of users followers, followees, posts and fans. Through the Markov network to the user's intrinsic attributes, Zhang [6] considered social properties and content attributes and analyze their relevance to mining opinion leaders.

3.2 Information Interaction Analysis

Information interaction analysis mainly by analysing the information sent by the user's influence and its transmission characteristics, it can reflect the influence of the user and achieve opinion leader mining. A representative study of such methods is the influence of the diffusion model that proposed by a Japanese researcher[7]. The model considers the transfer of relevant text content in the process of information diffusion and evaluate the influence of information, and the user's influence can be calculated by the sum of published information. Because the algorithm mainly considers the influence of the information published by the user only in the process of communication. However, the user's influence are reflected not only in the interaction between users but also in the user's own attributes, such as the number of fans and certified users, so the method also has shortage.

3.3 Network Structure Analysis

Compared with the above two aspects, scholars put forward the idea of using the network structure in opinion leader mining. Firstly, the network structure is created by the relationship between the users, and then they measure the importance of the users by network structure analysis. According to the study of this field, such research can be divided into two directions:

The first method analysis the center of network on the basis of complex network theory. Some researchers use the theory of complex networks to analyze the network center, such as degree of in-degree, out-degree, distance center, network density, cohesion subgroup, small world network. The second algorithm named PageRank is an indicator of node ranking in a directed graph. If the process of users' influence propagation is seen as a random walk, the indicator can also be used to measure the influence of users. In the early work of the study, many models of information dissemination and influence diffusion were given empirical values between connection of users to simplify the model and calculation.

3.4 Influence Analysis Combine With Topic Information

The three aspects of methods above are three complementary aspects in opinion leader mining. In the social network research, some scholars combine text information with the methods above. Many models used network topology analysis or their extended algorithms to analyze the importance of the users, such as Katsimpras [8] proposed Topic-Specific Random Walk depend on topic similarity that extracted by LDA model, the LeaderRank algorithm considered the emotional polarity of the forum [9] The influence of the users are different in different topics. On the basis of network structure and user interests similarity. Weng [10] studied the individual's influence in different topics of Twitter data set, the TwitterRank algorithm considered the user's attention relationship and topic similarity, the following contents can show algorithmic thinking of it.

Based on the idea of PageRank, the transfer probability from the user s_i to the user s_j is as formular (1). Where P_t is the transfer matrix of topic t. $|T_j|$ is the number of microblog posts of user s_j, $\sum_{a:s_i \, followss_a} |T_a|$ is the sum of microblog posts of neighbors of user s_i, $sim_t(i,j) = 1 - |DT_{it}' - DT_{jt}'|$ is the similarity of probability distribution of topics.

$$P_t(i,j) = \frac{|T_j|}{\sum\limits_{a:s_i \, followss_a} |T_a|} \times sim_t(i,j) \tag{1}$$

$$D_{JS}(i,j) = \frac{1}{2}(D_{KL}(D_i' \| M) + D_{KL}(D_j' \| M)) \tag{2}$$

$$M = \frac{1}{2}(D_i' + D_j') \tag{3}$$

Where $D_{JS}(i,j)$ is the Jensen-Shannon divergence between the two probability distributions. M is the average of the two probability distributions, $D_{KL}(P \| Q)$ is the Kullback-Leibler distance between Q and P, which can measure differences in the case of two probability distributions, but does not

satisfy the symmetry, so it would be better to use the Jensen-Shannon distance formula as formular (2).

4. Conclusion

After a comprehensive analysis of the influence measure methods, this paper gives its theoretical application in opinion leader mining, and analyzes the insufficiency of the algorithm in all aspects. At present, although the opinion leaders have made great achievements in the theory and application, but there are some issues to be in-depth study. Although many studies have objectively described the influence of opinion leaders, but cannot accurately define the concept, it is necessary to study the the evaluation indicator of social influence, then provide guidance for the further new models.

References

[1] Katz E, Lazarsfeld P F, Roper E: Personal influence: the part played by people in the flow of masscommunications[J]. American Journal of Sociology, 1956, 21(Volume 62, Number 1): 1583

[2] Everett Rogers: Diffusion of Innovations, Simon an Schuster, 1962

[3] Noah Friedkin: A structural theory of social influence, Cambridge University Press, 2006

[4] Romero D M, Meeder B, Kleinberg J: Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter[C]// International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April. DBLP, 2011: 695-704.

[5] Kempe D, Kleinberg J, Éva Tardos: Influential Nodes in a Diffusion Model for Social Networks[C]// Automata, Languages and Programming, International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005, Proceedings. DBLP, 2005: 1127-1138.

[6] ZHANG W, LI X, HE H, et al: Identifying Network Public Opinion Leaders Based on MarkovLogic Networks[J]. The Scientific World Journal, 2014: 1–8.

[7] MATSUMURA N, OHSAWA Y, ISHIZUKA M: Mining and Characterizing Opinion Leadersfrom Threaded Online Discussions[C]//Proceedings of the 6th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies. Crema, Italy: 2002:1267–1270.

[8] Katsimpras G, Vogiatzis D, Paliouras G. Determining Influential Users with Supervised Random Walks[J]. 2015:787-792.

[9] Yu X, Wei X, Lin X: Algorithms of BBS Opinion Leader Mining Based on Sentiment Analysis[M]// Web Information Systems and Mining. Springer Berlin Heidelberg, 2010: 360-369.

[10] Weng Jianshu, et al. TwitterRank: finding topic-sensitive influential twitterers[J]. Wsdm (2010): 261-270.