World Scientific Research Journal (WSRJ) ISSN: 2472-3703
www.wsr-j.org

Research on Information Propagation Model in Online Social Networks

Yong Wang[a], Meiling Cui[b, *], Zhonghuai Pan[c]

College of Computer Science and Technology, Harbin Engineering University, Harbin

[a]wangyongcs@hrbeu.edu.cn, [b]cuimeiling@hrbeu.edu.cn, [c]panzhonghuai@hrbeu.edu.cn

---

Abstract: In recent years, with the development of information technology and the popularization of the Internet, the online social networks represented by Weibo, WeChat, Twitter, Facebook and so on have developed rapidly. Information has become faster, has a wider scope of influence, and has more forms of communication, which has become an important medium for people to communicate with each other. Studying the models and laws of online social network information dissemination has a high scientific value and important practical significance for the monitoring and early warning of public opinion and the social network marketing. This article summarizes the information dissemination models proposed by relevant scholars in recent years and introduces 9 types of models from the perspectives of theory-based dissemination model and real data-based dissemination model. The article also makes a comparative analysis of the principle and key points of the model to reveal the microscopic dissemination of information in social networks. Finally, combined with the problems and challenges the current online social network information dissemination model facing, the future research direction is prospected.

Keywords: information dissemination, online social network, theory-based dissemination model, real data-based dissemination model.

---

1. INTRODUCTION

With the rapid development of Internet technology and the advent of Web 2.0 era, online social networks have become an important part of people's daily life in information exchange. They have gradually become an important medium of communication between people and are profoundly changing people's way of thinking, behavior patterns and social patterns [1-2].

Information is an important vehicle for people to communicate through online social networks, to study the process of information dissemination in online social networks, to find out the rules and characteristics of information dissemination effectively, and to use the laws and characteristics to solve the spread of bad information such as rumors and information dissemination forecast, maximize the influence of information dissemination and information traceability [3]. At the same time, the above research results can be applied to the fields of forecasting, dissemination of situational awareness, personalized information recommendation,

precise advertisement placement and the like, as well as discovering bad users and information, ensuring network and information security.

In summary, the study of online social network information dissemination is necessary, with good theoretical significance and application value. This article first introduces the concept and process of information dissemination model, and then illustrates and compares the principles and key points of 9 types of models from the perspectives of theory-based dissemination model and real data-based dissemination model respectively, revealing that information in social networks microscopic dissemination process. Finally, the full article is summarized, and the future research direction of online social network information dissemination model is prospected.

2. ONLINE SOCIAL NETWORK INFORMATION DISSEMINATION PROCESS AND CLASSIFICATION

2.1 Online Social Network Information Dissemination Process

There are some significant differences between online social networks and traditional modes of information dissemination. Information can be disseminated on a large scale through the interaction between users. As a complex network structure, graph can be used to express its structural features. The nodes in the graph represent users, and the nodes directly connected with them are called their neighbors, edges represent the interaction between users. When a node receives a message, its neighbor node decides whether to transmit the message with a certain probability. At the next moment, the neighbor who has just received the message still decides whether to broadcast the message with a certain probability. In summary, information in the graph nodes through the edges of the process of diffusion can be expressed in the online social network of information dissemination process.

2.2 Theory-based information spreading models

2.2.1 Linear Threshold Model

Linear Threshold Model (LT) was firstly proposed by Kempe to study the influence of social networks and to describe the process of information spreading [4]. The formal definition of a linear threshold model can be described as:

(1) During the process of information spreading, there are only two states of node, which are active state and inactive state.

(2)There is an activation threshold θ_v for each node v , and for all the active set of nodes A_v , the influence of each neighbor u_i in the active state on node v is $b(u_i, v)$. That is, when the combined influence of the neighbor node in the activated state of the node v exceeds the

activation threshold of v, $\sum_{u_i \in A_v} b(u_i, v) \geq \theta_v$,the non-activated state of the node v becomes the active state.

The linear threshold model mainly studies the influence of each node v on the neighbor node in the process of information transmission. It should be noted that the linear threshold model describes the "memory effect" and "cumulative impact" phenomena in information dissemination very well. That is, the influence of each neighbor nodes in the active state when the node v is in the inactive state Will be recorded and accumulated until the activation threshold of the node is exceeded and node v becomes active.

Beriberi et al. considered that the characteristics of information also affect the process of propagation in the process of information dissemination [5]. Therefore, a new linear model of topic awareness is proposed by introducing information characteristic parameters into the calculation of the influence of neighbor nodes $b(u_i, v)$ on user nodes.

2.2.2 Independent Cascades Model

Independent Cascades Model (IC), it was first proposed by Kempe while studying the issue of maximizing influence with the linear threshold model. The formal definition of the independent cascade model can be expressed as:

(1) During the process of information dissemination, there are only two states of node, which are active state and inactive state. Same as linear threshold model.

(2) Each active node v_i, for its inactive neighbor nodes u_j, each node v_i will attempt to activate its neighbor node u_j with an independent probability $p(v_i, u_j)$. The greater the probability $p(v_i, u_j)$ of activation, the easier the node u_j changes from an inactive state to an active state.

The independent cascade model mainly studies the influence of the nodes in the activated state on the neighboring nodes during the information dissemination. Unlike the linear threshold model, a node in an independent cascade model that fails to be attempted to activate has no memory and accumulate. The next time a node v_i is active, it attempts to activate the node u_j with an independent probability $p(v_i, u_j)$.

Zhoudong Hao et al, who take into account the characteristics of the node features and information on the impact of online social network information dissemination, proposed a social network information dissemination model based on the node information and features [6]. It extracts the characteristics of information dissemination from multiple dimensions, including the characteristics of node attributes and the characteristics of information content, and models the propagation probabilities and propagation delays among nodes. A fine-grained online social network information dissemination model is proposed.

2.2.3 Epidemic Model

The Susceptible-Infective-Removal (SIR) model was first proposed by Kermack and Mc Kendrick in 1927 to study the transmission of blacks disease [7]. Due to the high similarity between the process of information dissemination and the spread of infectious diseases, more researchers use the epidemiological model to describe the process of information dissemination. The epidemiological model describing the information dissemination process is defined as follows:

(1) In the process of information dissemination, there are three kinds of nodes, that is, unaccepted information node S, propagation information node I and immune information node R. The total number of the initial three types of nodes is N. Assume that at time t, the number of unaccepted information nodes S is S (t), the number of propagation information nodes I is I (t), and the number of immune information nodes R is R (t). In the process of information dissemination, the total number N of three types of nodes should be satisfied, ie S (t) + I (t) + R (t) = N.

(2) During the information spreading, the state transitions of the three types of nodes are shown in Fig. 1 below.

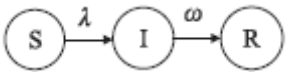

Fig. 1 SIR state transition diagram

That is, node S will become node I with a certain probability of infection, and node I will become node R with a certain probability of immunity. According to the state transition diagram of three kinds of nodes, the dynamic equation of the model can be drawn as:

$$\frac{dS(t)}{t} = -\lambda S(t)I(t) \quad , \quad \frac{dI(t)}{t} = \lambda S(t)I(t) - \omega I(t) \quad , \quad \frac{dR(t)}{t} = \omega I(t)$$

Infectious disease model mainly studies the changing rules of the number of state nodes in the process of information dissemination. And in the research process, assuming that the structure of the entire social network is uniform, the model of infectious diseases does not take into account the influence of network structure and user differences on the information dissemination process.

Tao Wang et al. believe that when users receive information, they may make comments or thumbs up, thus introducing a reviewing information mechanism. In addition, when receiving information, users may be interested in them and will conduct a collection operation [8]. Therefore, a collection mechanism is introduced Proposed IRCSS model, the state transition diagram as shown below in Fig. 2 Compared with the traditional SIR model, it simulates the most of the functions of online social network, so it is more in line with the information in the social network.

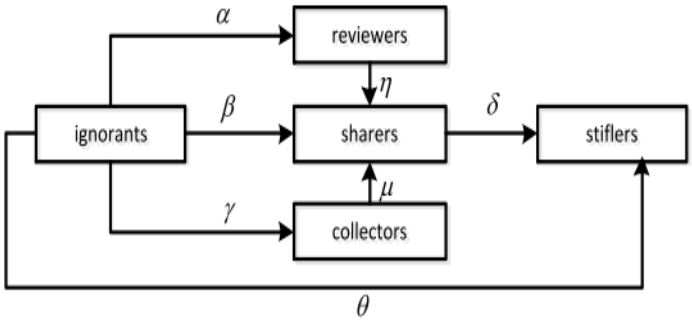

Fig. 2 IRCSS state transition diagram

2.2.4 Game Theory Model

Von Neumann in 1928 proved the basic principles of Game Theory (Game Theory), which declared the birth of game theory. The game usually takes place between two or more. Firstly, the strategy set is determined in the game theory model, and the return matrix of the game participants is determined according to the strategy set, that is, the cost and return when the participants make different choices (It's expressed by the revenue function). When the strategy can maximize the benefits of the participants, participants choose the strategy. The optimal portfolio solution for all participants is that the strategy achieves Nash equilibrium.

Game theory is mainly the analysis and prediction of node behavior in the process of information dissemination. Game theory in the dissemination of information usually occurs in both the information dissemination and the recipient, both of which choose whether or not to disseminate information based on whether the proceeds are received or not [9]. The game in information dissemination can also appear in the disseminators of information, whether the communicators conspire as a strategy set. Games are generally divided into pure strategy game and mixed strategy game. Pure strategy game, assuming that participants are rational, participants will only choose a strategy; mixed strategy game, assuming that participants are irrational, the user will have different probabilities for strategy selection.

KP Krishna Kumar et al. proposed a game theory model based on the theory of evolution based on the influence of user strategy selection on user interaction [10]. The users are divided into three types, that is, users who are always forwarded, users who do not forward, and users who decide whether to forward or not according to the revenue. The policy of the user is {forward, not forward}. In this paper, they establish a game between three kinds of users and simulate the forwarding of information between users through experiments.

2.3 Real Data-based Information Spreading Models

2.3.1 Kernel Function Model

Kernel Function (KF), dating back as early as 1909 was proposed. It is mainly used in the field of machine learning and can solve high-dimensional computing problems. The information

dissemination model established by using kernel function maps the process of information dissemination on social networks into energy diffusion process on the feature space. The distance between points in the feature space is used to represent the order of information dissemination on social networks.

The basic idea of the kernel function model is to map the observed information propagation process into the feature space. It can effectively solve the problem of analyzing and forecasting the information dissemination process with incompletely known network structure.

Weihang Shi et al. considered the process of information transmission in social network which is not completely known in structure and proposed a model of information transmission based on kernel function [11]. In this paper, the nodes in the network are mapped into the continuous feature space, and the node's propagation order is reflected by the distance between the nodes. The propagation of information in the network is described as the process of energy diffusion in the feature space. At the same time Shi Wei Hang et al. Added the content features of the information to the target feature space, using gradient descent method to optimize the solution.

2.3.2 Markov Model

The Markov Model, named after the Russian mathematician Markov, is a statistical model. In characterizing the process of information dissemination, the Markov model steps are:

(1) Determining the state space. In the process of information dissemination, the division of state space can be the trend of information dissemination, it can also be the distribution of information dissemination of various types of personnel.

(2) Determining the state transition probability matrix. Based on the historical information, the transition matrices of various states are calculated, and the probability matrices of all kinds of state transitions are further obtained.

(3) Forecasting the next moment of information dissemination. That is based on the current information dissemination and the state transition probability matrix to predict the next moment of information dissemination.

The main idea of the Markov model research is to predict and estimate the future (next moment) information dissemination based on the current information dissemination, and the future information dissemination has nothing to do with the past information dissemination.

Zhang et al. consider the impact on the network topology information dissemination, based on Markov theory, they proposed a new online social network information dissemination model [12]. In this paper, users are divided into four states: unacceptable state, propagation state, indifference state and immune state. Four dynamic equations of mutual transformation of states are established, and the depth and breadth of information dissemination are predicted.

2.3.3 Logistic Regression Model

The Logistic Regression Model was first developed by Luce in 1959 on the basis of IIA characteristics. It is a classical classification algorithm in the field of machine learning. The use of logistic regression established information dissemination model, including the following steps:

(1) According to the history of information in the process of information and data, determine the impact of information dissemination of the factors.

(2) The historical information data, feature extraction and model training.

(3) Predicting the future information dissemination according to the trained model.

The logistic regression model is modeled based on the real data in the process of information dissemination, and mainly studies the influence of different factors on the process of information dissemination.

From the perspective of users and network structure, Guille et al. considered the influence of individual social factors, message semantic factors and time factors on information dissemination [13]. In this paper, we use the method of naive Bayesian logistic regression to model the thirteen factors that affect the information dissemination and predict the probability of the information dissemination.

2.3.4 Neural Network Model

The Neural Networks Model (NNM) was first proposed by the psychologist W.S. Mc Colloch in 1943. It mainly consists of three components, input layer, hidden layer and output layer. It is a multi-input and single-output model. In characterizing the process of information dissemination, the main process of neural network model establishment includes:

(1) Determining the input of neural network model. That is to determine the factors that affect the information dissemination, as a neural network model input.

(2) Determining the hidden layer excitation function of neural network, and use the historical information data to learn the parameters.

(3) According to the current impact of information dissemination factors, determine the probability of information dissemination.

It’s different from the traditional models of information dissemination, the neural network model can predict the probability of information dissemination by using the factors that affect the propagation of information without providing detailed mathematical modeling only by providing the historical information data, and can well fit the non- Linear math problems.

Based on the factors affecting the information dissemination, He Yanxiang et al. improved the demographic model to make it suitable for the trend prediction of the microblog event propagation, and then mapped the improved demographic model on the neural network and used the genetic algorithm to optimize the event Trend forecast [14]. This method can effectively predict the long-term brewing event propagation in the network and is more suitable for small sample prediction.

2.3.5 Conditional Random Field Model

The Conditional Random Fields Model (GRF Model) is the first one proposed by Lafferty et al in 2001 that combines the maximum entropy model with the Markovian discriminates model. The earliest mainly used in POS tagging and other fields. Now its application in the study of information dissemination modeling, including the following steps:

(1) Defining a given information propagation modeling problem as a posteriori forwarding probability function.

(2) Determining the decisive factors affecting the dissemination of information.

(3) For various factors, define their corresponding potential functions.

(4) Using the historical data of information dissemination, we train the potential function weights.

The conditional random field model can solve the information propagation modeling problem of large scale networks and complex structure networks well.

Considering the influence of information content, network relationship and time delay on information dissemination, Peng et al. proposed a method for predicting forwarding behavior based on conditional random fields [15]. In this paper, the user's forwarding behavior is transformed into the maximum posterior probability problem, and the conditional random field is used to predict the forwarding behavior. The experimental results show that the proposed forward behavior based on conditional random field improves the effectiveness of the information dissemination prediction.

3. INFORMATION DISSEMINATION MODEL EVALUATION AND ANALYSIS

According to the difference of the basis of online social network information spreading model, the models described in this paper can be divided into theory-based spreading models and real data-based spreading models.

The theory-based models of information dissemination essentially draw on the knowledge of other fields to model the process of information spreading in online social networks. The above mainly introduces five types of theory-based information dissemination models. Table 1 compares the theoretical basis, characteristics and key points of these types of models.

The linear threshold model and the independent cascade model are often used to solve the problem of maximizing the influence in the online social network and predict the propagation probability of the nodes. Epidemic models describe the changes in the number of nodes in the information dissemination process. The game theory model considers the influence of income on the spread of information in the process of dissemination.

The model of information spreading based on real data uses the past or current status of information dissemination in online social networks to train the parameters in the model, which can predict the spread of the social network in the future or the next moment. In this paper, we give a total of 5 information dissemination models based on real data. Table 2 shows the comparison between several models.

Table 1. Comparison of theory-based spreading models

Name	Theoretical Basis	Features	Key Points	Deficiencies
Linear Threshold Model	Maximize Influence	Having a memory effect& affecting the cumulative	Influence of neighbors and activation threshold	Lacking he consideration of information features
Independent Cascade Model	Maximize Influence	The neighbor nodes are activated with independent probability each time	Probability of activation of neighbor nodes	Lacking he consideration of information features
Epidemic Model	Infectious Disease	Regardless of the differences of network structure	Establishment of dynamic equation	Lacking the consideration of the difference of network structure
Game Theory Model	Game Theory	Including the both sides of the game	The establishment of revenue matrix	Lacking the consideration of users

Table 2. Comparison of spreading models based on real data

Name	Data Set	Features	Key Points
Kernel function model	Observed information dissemination process	Aiming at non-fully known network structure	The process of information dissemination is mapped to the process of energy diffusion
Markov model	Historical statistics of various types of mutual transformation	Having Markov property	Establishing the state transition matrix
Logistic regression model	Historical information data	Consider the impact of information characteristics on the information dissemination	Extraction and Training of Features
Neural Network Model	Historical information data	No detailed mathematical modeling is required	Influencing Factors and Excitation Function of Information Propagation
Conditional field model	History of information dissemination	Solving the problem of modeling information propagation in large-scale networks and complex structured networks	Training of the weight of potential function

The kernel function model predicts which new propagation nodes in the network are to be predicted at the next moment. The Markov model predicts the transformation of the node state at the next moment in the network. Logistic regression model, neural network model and

conditional field model are used to train the corresponding parameters according to the corresponding machine learning methods, and then the probability of nodes transmitting information is predicted.

4. CONCLUSION

This paper first introduces the research background of online social network in detail, the process of information dissemination in social networks, and then summarizes the research on information dissemination model by related scholars in recent years. According to the difference of modeling basis, dissemination models are divided into theory-based information dissemination model and real data-based dissemination model. Among them, the theory-based model of information dissemination mainly draws on the knowledge of other fields to model the process of information dissemination in online social networks. The models are divided into linear threshold model, independent cascade model, epidemic model and game theory model. The real data-based propagation model mainly uses the past or current online social network information propagation conditions to train the parameters in the model, and then can predict the future or the next moment of online social network propagation. These models are divided into kernel function model, Markov model, logistic regression model, neural network model and conditional field model. Finally, analyze the principle, characteristics and key points of the above model and summarize the full artical.

Although there are many achievements in the research of online social network communication model, some problems still need to be solved urgently. For example, the dynamic evolution characteristic is the basic characteristic of online social network, how to adapt to the changes of the online social network every moment, add the dynamic characteristics of the online social network to the information dissemination model. Most of the information dissemination is within the community structure, the propagation path is relatively short, while the current more information dissemination model is defined in the entire network structure, ignoring the impact of community structure on the spread of information. The current research on the maximization of influence with strong practical significance in the field of information dissemination mainly focuses on the linear threshold. In the independent cascade model. However, in the other models, there is little research and lack of maneuverability. Therefore, it is of great research value to optimize the related model and solve the problem effectively.

ACKNOWLEDGEMENTS

The research work was supported by The Youth Foundation of Heilongjiang Province of China under Grant No. QC2016083, The Fundamental Research Funds for the Central Universities under Grant No. HEUCF170604, and Innovative Talents Research Special Funds of Harbin Science and Technology Bureau under Grant No. 2016RQQXJ128.

REFERENCES

[1] Minjung Sung, Jangsun Hwang. Who drives a crisis? The diffusion of an issue through social networks [J]. Computers in Human Behavior,2014,36 (7), p246-257.

[2] Tian Yan, Zhang Xingang. Review of Topology and Information Transmission Mechanism of Online Social Networks [J]. Journal of Changsha University,2016,30 (02),p73-75.

[3] MYERS S A and LESKOVEC J. The bursty dynamics of the twitter information network [C]. Proceedings of the 23rdInternational Conference on World Wide Web, Seoul, 2014, p913-924.

[4] Granoveeer M. Threshold models of collective behavior [J]. American Journal of Sociology, 1987,83 (6), p1420-1443.

[5] Barbieri N, Bonchi F, Manco G. Topic-aware social influence propagation models [J]. Knowledge and Information Systems,2013,37 (3), p555-584.

[6] [6] Zhou Donghao, Han Wensheng, Wang Yongjun. Communication Network Information Propagation Model Based on Nodes and Information Features [J]. Journal of Computer Research and Development, 2015,52 (01),p156-166.

[7] Kermack W O, Mckendrick A G. A contribution to the mathematical theory of epidemics [J]. Proceedings of the Royal Society of London. Series A,1927,115 (772), p700-721.

[8] Tao Wang, Juanjuan He, Xiaoxia Wang. An information spreading model based on online social networks [J]. Physica A: Statistical Mechanics and its Applications,2017.

[9] Xiao Renbin, Zhang Yaofeng. Evolutionary Game Analysis of Network Group Event Information Transmission [J]. Complex Systems and Complexity Science, 2012,9 (1), p1-7.

[10] Krishna Kumar K P, Geethakumari G. Information diffusion model for spread of misinformation in online social networks[C]. Proceeding of 2013 International Conference on Advances in Computing Communications and Informatics (ICACCI), Mysore, India,2013, p1172-1177.

[11] Shi Weihang, Lin Nan. Kernel-based information propagation model in social networks [J]. Application Research of Computers, 2016,33 (09), p2735-2737.

[12] Peng HK, Zhu J, Piao DZ, et al. Retweet modeling using conditional random fields [C]. IEEE 11th International Conference on Data Mining Workshops, 2011, p336-343.

[13] Guille A, Hacid H. A predictive model for the temporal dynamics of information diffusion in online social networks [C]. Proceedings of the 21st International Conference Companion on World Wide Web, 2012, p1145-1152.

[14] He Yanxiang, Liu Jianbo, Liu Nan, Peng Min. Anticipating Trend of Weibo Topic Based on Improved Population Model [J]. Journal of Communications, 2015,36 (04), p5-12.

[15] Zhang S, Xu K, Chen X, et al. Dynamics of Information Spreading in Online Social Networks [J]. Computer Science, 2014, ar Xiv1404.5562.