Neurocomputing 293 (2018) 18–28

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Sociability-based Influence Diffusion Probability Model to evaluate influence of BBS post

Lei Li[a], Xin Lin[a], MengChu Zhou[b,c,*], LiLi Fu[d]

[a] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Department of Electrical and Computer Engineering, New Jersey Institute of Technology (NJIT), Newark, NJ 07102, USA
[c] Institute of Systems Engineering, Macau University of Science and Technology, Macau 999078, China
[d] Business College of Beijing Union University, Beijing 100025, China

ARTICLE INFO

Article history:
Received 6 November 2017
Revised 12 January 2018
Accepted 26 February 2018
Available online 9 March 2018

Communicated by Dr Xin Luo

Keywords:
Bulletin Board System (BBS)
Post influence
Influence diffusion model
Sociability
Probability model

ABSTRACT

A Bulletin Board System (BBS) yields user-generated posts, which has enjoyed fast spreading speed. Significant events are often revealed by a post. It may then spread widely, thereby producing large influence in some specific social circles and sometimes the whole society. Hence, evaluating post influence becomes important. It can help Web service providers locate quickly those influential posts, users or communities, place right advertisements, expand an event's influence, and explode a hot topic's discussions. Recently, BBS has grown to have some new features, e.g., sociability. The existing studies use an Influence Diffusion Model (IDM) and its expanded versions for the analysis of influence. However, they suffer from such drawbacks as identical treatment of every comment or reply, and complete ignorance of relationships among users, thereby leading to the inaccurate assessment of post influence. To overcome the limitations, inspired by our prior user model for user participation in virtual communities, we propose a behavioral model for user participation in a post and give a Sociability-based Influence Diffusion Probability Model (S-IDPM) by utilizing user relationship and reply-chains to measure the responses of different users and evaluate post influence. Experiments with real data collected from a popular BBS. Our results show that S-IDPM outperforms IDM and its expanded version called Influence Diffusion Probability Model (IDPM). S-IDPM can be helpful to achieve better post influence diffusion evaluation than IDM and IDPM do.

1. Introduction

After nearly twenty years of development, Bulletin Board System (BBS) has become mature. Nowadays many significant events are revealed by posts in BBS first. They are then spread widely via BBS platforms, producing huge influence in some social circles and sometimes the entire society. BBS platforms have powerful spreading energy. Most of them provide some social functions, such as *forward* (re-posting with or without comments) and *follow* (being a follower of some users), which can push information to users easily, and thus speed up the spreading of a post to a large extent. As a result, post influence assessment becomes very important and can impact many applications, such as information collection, retrieval [1] and BBS management. Due to the enormous amount and fast growth of posts, people can hardly process all of them. It is essential to identify good posts with strong influence and high quality. Assessment of post influence can also help Web service providers locate quickly those posts, users or communities with strong influence, increase an event's influence, and deepen the discussions of a hot topic.

Meanwhile, the online virtual life of users is affecting their real life more and more. E-marketing for e-business or e-commerce has emerged and posed an important question: what the influence of marketing activities in a network is. Analysing post influence helps answer it. Given a limited budget, we can choose some post that is effective to spread information or to place appealing advertisements for users, thus achieving our desired objectives. Therefore, this study is highly meaningful for the research and practice of e-marketing, e-business and information systems.

It is noted that BBS in Web 2.0 has grown with some new features similar to those of other social networks, e.g., sociability. A social network represents a set of human beings or their digital representations who are linked by their direct links and relationships extracted from the data about their activities and communications [2]. Some BBS managers with strong influence have already appeared and attracted a large number of fans or users with simi-

* Corresponding author.
E-mail addresses: leili@bupt.edu.cn (L. Li), lx1989831@qq.com (X. Lin), zhou@njit.edu (M. Zhou), lili.fu@buu.edu.cn (L. Fu).

https://doi.org/10.1016/j.neucom.2018.02.087
0925-2312/

lar interests. In other words, BBS is no longer simply a website to publish information; instead, it becomes a system of many social networks, in which various relationships among users and contents can have more important effects on post influence diffusion. Yet we cannot find any proper model to characterize BBS to our best knowledge. Hence, we intend to build up a model to describe these new features of BBS and thus perform a better influence diffusion evaluation of posts.

Currently Influence Diffusion Model (IDM) [17] and its extended versions, e.g., Influence Diffusion Probability Model (IDPM) [19], are usually used for the evaluation of information and user influence. But they suffer from some deficiencies, e.g., neglecting some important factors that can impact accurate evaluation. For instance, they do not consider the relationships among users. Thus the weights of every comment or reply are set to be the same. We intend to improve IDPM by utilizing the user network of relationships including social networks and reply-chains to measure the responses of different users and evaluate post influence more accurately. Inspired by our model for user participation in virtual communities [3], we formulate a new behavioral model for user participation in a post with standard path coefficients that can reflect different weights of different comments or replies. Our focus is to propose and verify that users have different relationships in social networks and these relationships really affect the influence diffusion of a post they participate in. Based on it, we then propose a Sociability-based IDPM (S-IDPM), which applies the user network of relationships and structure of reply-chains to compute the response weights of different users. This idea can also be applied to the evaluation of other information sources whose users have certain social and interest relationships.

The rest of this paper is organized as follows. Section 2 gives an overview of the related research. Section 3 analyses the deficiencies of IDPM. Section 4 presents the behavioral model for user participation in a post and S-IDPM. Section 5 reports the experiments and results of S-IDPM, IDM and IDPM on a real BBS dataset. Finally, Section 6 concludes this paper and indicates future research issues.

2. Related works

This section introduces the related literature work. First, we discuss the studies about social network content features and factors for information dissemination and influence diffusion. Second, we summarize the methods for information influence evaluation. Finally, we discuss user behavioral models related to sociability.

There are many studies focusing on various features and factors for information dissemination and influence diffusion in social networks. Karsai et al. [4] study the evolution of information dissemination in small world networks and conclude that there are two main factors slowing down the spreading of information, i.e., network topology and personal burstiness. Steeg et al. [5] analyse the dissemination of information in Diggs and find that a highly clustered architecture in Diggs network limits the scope of information dissemination. Rodrigues et al. [6] study different factors affecting information dissemination, including the shape of a dissemination tree, publishers and subscribers. They conclude that the dissemination tree passed from mouth to mouth has the widest breadth. Wang et al. [7] have found that the social and organizational contexts can greatly impact whom and how fast people forward information. They give a simple stochastic branching model for an information spreading process at the macroscopic level. Chhabra et al. [8] study the propagation form of network phishing in Twitter and reveal that a staggering 89% of references from Twitter (users) are inorganic accounts which are sparsely connected amongst themselves, but have a large number of followers and followees. They also observe that most of the phishing tweets spread by extensive use of attractive words and multiple hashtags. Liu et al. [9] propose a generative graphical model that utilizes heterogeneous link information and textual content associated with each node in the network to mine topic-level direct influence. Dou et al. [10] analyse social networks based on their structural features and confirm some properties of social networks. They study the evolution of a community structure based on an event framework and reveal that the community merge depends largely on the clustering coefficient of the graph composed of nodes that are directly connected among communities. Zhang et al. [11] study the measurement and analysis methods for information propagation in such online social networks as microblog. They find out that the Sina microblog possesses a strong “celebrity effect”, and users’ frequent posting fails to arouse much attention. All these studies have shown that influence diffusion and information dissemination are closely related to various factors, e.g., personal, social, structural and organizational contexts in social networks and words or contents. This result motivates us to consider them together in order to evaluate post influence accurately.

Currently, influence in social networks is mainly assessed with link analysis methods and IDM based methods. The former is originated from the PageRank algorithm [12]. There have been various studies based on it. Kandiah et al. [13] propose a PageRank model of opinion formation in social networks, which can be used to compute the influence of opinions. Agryzkov et al. [14] propose an algorithm for ranking the nodes of an urban network based on the concept of a PageRank vector. Liu et al. [15] analyse the information of user browsing behavior and construct a reliable Web graph according to PageRank. Jiang et al. [16] propose an indexing network model that organizes information in webpages at three levels: words, webpages, and categories to facilitate exploratory information retrieval and personalized search. Although these methods are good at utilizing various network properties, they are usually weak in integrating the factors of contents that have impact on influence diffusion in different ways from network properties.

IDM was proposed by Matsumura et al. [17] for text-based communication focusing on content factors in 2002. It uses reply-chains to compute the influence diffusion capability of an information source based on the similarity of texts according to the feature of term frequency. The influence diffusion capability of a text is the sum of the influence of every reply to this text. IDM has built up an important basis for later studies on information influence. Shi et al. [18] propose an improved IDM algorithm. They delete the denominator in the original IDM formula to avoid its induced errors in the evaluation result. Fan et al. [19] put forward an IDPM algorithm after they find two problems in the IDM algorithm. One is the influence diffusion break caused by a broken reply-chain or indirect content diffusion. Another is the illusive influence diffusion caused by flooding posts. To solve them, they define the diffusion probability influence of a single term in the same interest space, and introduce the concept of valid terms. Compared with PageRank methods, these IDM methods pay more attention to content factors and reply-chains. However, they fail to integrate other network properties. There are some other related studies like latent network structures [28], hidden diffusion networks [29] and Web-service selection [32,33], which treat sociability as an important factor for analysis. Motivated by these prior studies, this work intends to solve the post evaluation problem by improving IDM-based methods with personal and social context information.

We can obtain knowledge about personal and social contexts in social networks from user behavioral models. According to [20], the driving factors to the participating behavior of members and their active contributions to an online community are their gains from the course of participation and their motivation to contribute. Natalia and Jose [21] think that contexts and norms in Online Social Networks are important for information sharing and disseminations. They present an information model for Implicit