RTPMF: Leveraging User and Message Embeddings for Retweeting Behavior Prediction

Jiguang Liang, Bo Jiang, Rongchao Yin, Chonghua Wang, JianLong Tan,* Shuo Bai

National Engineering Laboratory for Information Security Technologies,
Institute of Information Engineering, University of Chinese Academy of Sciences
{liangjiguang, jiangbo, yinrongchao, wangchonghua, tanjianlong, baishuo}@iie.ac.cn

Abstract

Understanding retweeting mechanism and predicting retweeting behavior is an important and valuable task in user behavior analysis. In this paper, aiming at providing a general method for improving retweeting behavior prediction performance, we propose a probabilistic matrix factorization model (RTPMF) incorporating user social network information and message semantic relationship. The contributions of this paper are three-fold: (1) We convert predicting user retweeting behavior problem to solve a probabilistic matrix factorization problem; (2) Following the intuition that user social network relationship will affect the retweeting behavior, we extensively study how to model social information to improve the prediction performance; and (3) We also incorporate message semantic embedding to constrain the objective function by making a full use of additional the messages' content-based and structure-based features. The empirical results and analysis demonstrate that our method significantly outperform the state-of-the-art approaches.

Keywords: retweeting behavior prediction, probabilistic matrix factorization, social, embedding

1 Introduction

Social media platform with its unique information propagation ability in people's daily life plays an increasingly important role. Every day large amounts of information are generated and spread by retweeting mechanism which is an important social function. Retweeting makes messages easy to reshare from user to user in a viral manner. Thus, exploring on user retweeting behavior not only can make us better understand information diffusion in social networks, but also help get an idea of true intentions from the perspective of human behaviors. Therefore, understanding retweeting mechanism and predicting retweeting behavior is an important and valuable task in user behavior analysis.

The task of predicting user retweeting behavior has been studied extensively over the past decade [9, 15, 12]. A common weaknesses of these methods is that they only consider the property of user or message for retweeting prediction. In fact, in addition to the retweeting behavior data, some additional sources of information can also be contained in social networks. Figure 1 illustrates the social

---

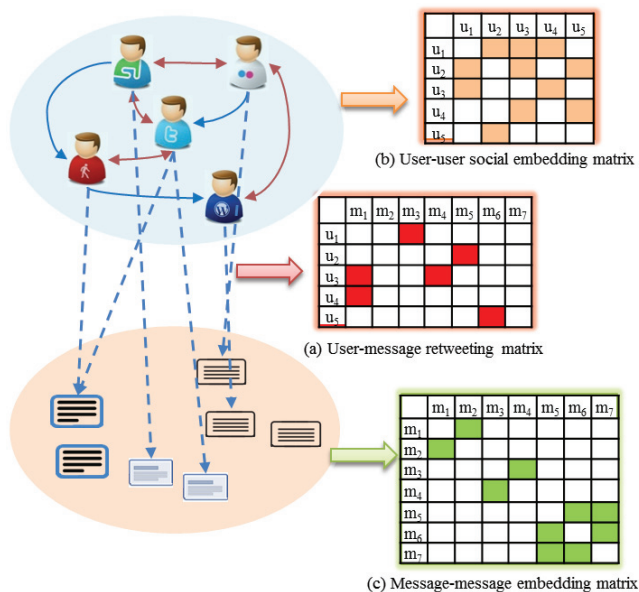*Corresponding author: Jianlong Tan, email: tanjianlong@iie.ac.cn

Figure 1: Social contextual information of user retweeting behavior.

contextual information concerning retweeting behavior. From Figure 1, we can see that except user-message retweeting matrix, there are user-user relationship matrix and message-message relationship matrix. We hold the intuition that both user social relationship and message semantic relationship can be employed to enhance retweeting prediction. The intuition behind is that given a message users first read the message and then decide whether retweet it or not. In other words, users only retweet those messages that they have read and are interested in the content of the messages. Hence, being read is the necessary prerequisite for retweeting. Typically, in Twitter user u_A is a follower of user u_B. If u_B is followed by many people. We can believe that u_B may be an authority or an icon. Then u_A is more likely to read messages issued by u_B. Moreover, if u_A follows many people, then u_A maybe only select some of messages posted by his followee to read. As a result, the probability that u_A reads and retweets u_B' message reduces. In this way, the social influence is different from traditional friendship or follwoee-follower relationship. In addition, if a message m has been read by u_A, u_A will not retweets it if u_A is not interested in the content of m. Therefore, both of user and message aspects are important for user retweeting behavior prediction. However, most of the existing methods simply ignore such side information, or intrinsically, are not capable of exploiting it.

To address this problem, we introduce a probabilistic matrix factorization model, which integrates the user-message retweeting data, user social relationship and message semantic relationship into a unified framework. More specifically, we first formulate the retweeting behavior prediction problem as a probabilistic matrix factorization problem to solve. Secondly, we incorporate user social relationship and message semantic relationship into the objective function by designing user social embedding and semantic embedding constraint regularization terms on the latent user and message feature space, respectively. Finally, we conduct extensive experiments to validate the effectiveness of our model compared with the state-of-the-art approaches. Experimental results clearly demonstrate the better effectiveness of our model for retweeting behavior prediction.

The rest of the paper is organized as follows: At first, we introduce our proposed model, and derive its inference and solution algorithms. Subsequently, we experimentally evaluate our method using