Item Recommendation for Word-of-Mouth Scenario in Social E-Commerce

Chen Gao, Chao Huang, Donghan Yu, Haohao Fu, Tzh-Heng Lin, Depeng Jin, *Member, IEEE*, and Yong Li, *Senior Member, IEEE*

Abstract—Social commerce, which is different from traditional e-commerce where people purchase products via initiative searching or recommendations from the platform, transforms a social community into an inclusive place to do business by enabling people to share products with their friends. A user (*sharer*), can share a link of a product to their social-connected friends (*receiver*). Once a receiver purchases the product, the sharer can earn commission provided by the platform. To promote sales, the platform can also assist sharers by providing product candidates which are more likely to be purchased during the social sharing. We define this task of generating sharing suggestions as item recommendation for word-of-mouth scenario, and to the best of our knowledge, this is a new task that has never been explored. In this article, we propose a *TriM* (short for **Tri**ad based word-of-**M**outh recommendation) model that can capture both the sharer's influence and the receiver's interest at the same time, which are two significant factors that determine whether the receiver will buy the product or not. Furthermore, with joint learning on two parts of interaction data to address data sparsity issue, our proposed TriM-Joint further improves the recommendation performance. By conducting experiments, we show that our proposed models achieve the best results compared to state-of-the-art models with significant improvements by at least $7.4\% \sim 14.4\%$ respectively.

Index Terms—Recommender systems, word-of-mouth, social E-commerce

1 Introduction

Social commerce [1], defined as a subset of e-commerce that involves social media, is one of the most popular topics nowadays. In social commerce, users are encouraged by the platform via financial rewards to share/diffuse/broadcast various products sold in the platform to social media. This is the most significant difference compared with traditional e-commerce, such as Amazon[1] and Taobao,[2] where people purchase items through initiative finding or passive recommendation of the platform. In other words, this model creates a new channel of advertisement with the help of users' social network.

More specifically speaking, a user could be a consumer and an advertiser at the same time. That means, he/she (*sharer*) not only purchase products, but also shares items to his/her friends (*receiver*) and earns rewards once the deal closes. In such a product-sharing scenario, ternary interactions among three roles (a sharer, a receiver, and an item) are involved. Under this new business model, social commerce platforms like Beidian[3] and Yunji[4] are developing rapidly [2]. In such platform, the phenomenon of *word-of-mouth* [3], which means the passing of information from person to person, is extremely intensive. According to a Nielsen study [4], 92 percent of consumers believe suggestions from friends and family play a more important role than advertisements since people tend to trust what their friends recommend. Based on this phenomenon, word-of-mouth marketing aims to recommend products to users via those sharing activities instead of direct advertisements. Therefore, systematically analyzing and modeling user behaviors of word-of-mouth marketing has become a significant and urgent task.

In this paper, we investigate the problem of *Word-of-Mouth (WoM) Recommendation*. Given a sharer and a receiver on the social commerce platform, we are asked to provide sharing suggestions (products candidates) which would be more likely preferred by the receiver. Traditional recommendation methods, which only consider user-item binary interactions, cannot model this sharer-item-receiver ternary relation. The most related task is social recommendation [5], [6], [7], [8], [9], [10], [11]. However, these social recommendation methods solely rely on social connections between users instead of specific sharing behaviors. Therefore, they are not suitable for our problem. In fact, the task of WoM recommendation is a further study compared with all existing social recommendation researches, as it tries to understand how social influence affects user decision with finer granularity.

1. https://www.amazon.com/
2. https://www.taobao.com/

- *Chen Gao, Chao Huang, Donghan Yu, Tzh-Heng Lin, Depeng Jin, and Yong Li are with the Beijing National Research Center for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China. E-mail: {gc16, c-huang15, yudh14, linzh14}@mails.tsinghua.edu.cn, {jindp, liyong07}@tsinghua.edu.cn.*
- *Haohao Fu is with the College of Letters and Science, UC Berkeley, Berkeley, CA 94720-5800 USA. E-mail: clintfhh@berkeley.edu.*

Manuscript received 26 Nov. 2019; revised 11 June 2020; accepted 31 July 2020. Date of publication 20 Aug. 2020; date of current version 29 Apr. 2022. (Corresponding author: Yong Li.)
Recommended for acceptance by M. Wang.
Digital Object Identifier no. 10.1109/TKDE.2020.3017509

3. http://www.beidian.com/
4. http://www.yunjiweidian.com/

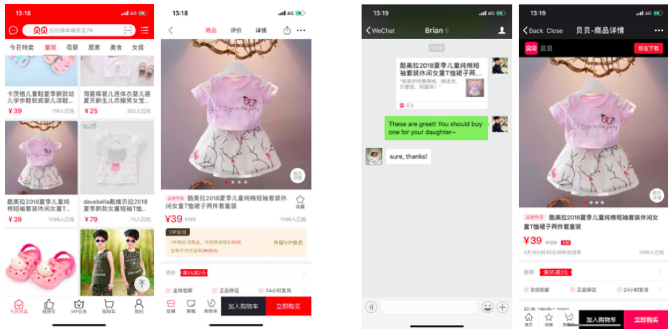
1041-4347 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

The problem of word-of-mouth recommendation is brand new but important, with the following three challenges:

- *Ternary Relation Modelling:* Different from traditional recommendation problem with only user-item binary relation, word-of-mouth recommendation is based on ternary relation composed of sharer, item and receiver. As a result, existing approaches can hardly be applied to this new research problem. How to model the ternary relation and subsequently utilize it to design a recommender system is the first challenge we need to address.
- *User Behaviors Complexity:* In this ternary relation, user behaviors are more complex and complicated. On the one hand, receivers exhibit their own interests when interacting with items. On the other hand, sharers may have their item-relevant influence on receivers, which can change the behaviors of receivers. How to take these complicated factors into consideration is also very challenging.
- *Data Sparsity Issue:* The final challenge is about data sparsity. In real-world scenarios of word-of-mouth recommendation, data of sharing behaviors is hard to obtain and usually very sparse. In addition, compared with traditional recommendation tasks, it requires more data for training due to the complexity of ternary relations. Thus, how to develop an effective model based on such sparse data is difficult and worth studying.

To address the above three challenges, we first use multi-relation learning in knowledge graph to consider the sharer, item and receiver as head node, relation and tail node, respectively. Then we propose a novel method named *TriM* (short for **Tri**ad based word-of-**M**outh recommendation) to model the ternary relation, and fuse the two factors into our model for capturing the complexity of user behaviors. To deal with the data sparsity problem, we introduce the user-item binary interaction data from traditional e-commerce platform to better capture receivers' interests, and propose a joint learning method *TriM-joint* by performing a co-optimization task. Our main contributions are summarized as follows.

- To the best of our knowledge, we are the first to study the problem of WoM recommendation from the perspective of designing machine learning algorithms. By analyzing the influence from sharers and interest of receivers in such ternary relation, we convey an in-depth study about social recommendation.
- We propose a novel model to learn the ternary relation by unifying the factors of both sharers and receivers, and utilize joint learning to help capture receivers' interests and handle the data sparsity problem.
- We carry out extensive experiments demonstrating our model performs better than state-of-the-art models by a relative improvement of $7.4\% \sim 14.4\%$. With joint learning, our model can be further improved by $15.5\% \sim 16.0\%$. What's more, a thorough investigation explaining why our proposed method outperforms the baselines is given to validate the rationality of our algorithm design further.



(a) Traditional Commerce (App) (b) Social commerce (Wechat)

Fig. 1. Online shopping scenarios through traditional e-commerce platform, and through social commerce platform.

2 Problem Formulation

Fig. 2 illustrates the problem of word-of-mouth recommendation. In the traditional commerce platform, let u denote the user, i denote the item and (u, i) denote an interaction between user u and item i, including clicking, buying or other type of interaction. In the social commerce platform, let u^s denote the sharer, u^r denote the receiver, and i denote the item. $\langle u^s, i, u^r \rangle$ denotes an event that user u^s once shared item i to her friend u^r, and u^r purchased it. The key problem here is how to predict items with a high probability of future consumption by u^r after u^s's sharing. In social commerce, user can still consume in the same way as on traditional e-commerce platform, which means users' *traditional interactions* (consuming products via initiative finding or passive recommendation of platform) are also accessible.

Input. Records of item sharing and purchase, ⟨sharer, receiver, item⟩.

Output. A model that generates a list of items given a user as the sharer.

3 Data Analytics

Our dataset is collected from Beibei,[5] one of the largest e-commerce platform for maternal and infant products in China. As shown in Fig. 1, users have two main channels to buy products on Beibei, namely through traditional e-commerce app or social media. For the latter one, by clicking on a product link shared by their friends through Wechat,[6] users can also directly purchase this product on Wechat. Note that there may be a slight difference between chatting and social media clear. In our collected dataset, most users share the item through group chat or Moments,[7] while person-to-person chat takes up a small percentage. Thus, we can consider this channel as a kind of social media to some extent.

Our dataset is collected within the time period from September 20, 2017 to October 22, 2017. The statistics of it are shown in Table 1. To conclude, we collect data of *traditional interaction* from App as an aid for the WoM recommendation task on Wechat. We will release this precious dataset upon acceptance.

5. http://www.beibei.com
6. Wechat is one of the largest Social Network Service Platform in China. https://www.wechat.com
7. A function in Wechat through which users can share texts and pictures to all friends.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

TABLE 1
Statistics of the Dataset (Bridging User and Item are Defined as Overlapped User and Item Across Two Platforms)

Social commerce	Sharer#	Receiver#	Item#	Events#
Beibei-Wechat	175827	380639	75464	574273

Traditional e-commerce	User#	Item#	Events#
Beibei-App	3011253	1231307	43273133

Bridging User#	Bridging Item#
40011	46340

In this section, we analyze the necessity of considering sharers' influence and receivers' interests into word-of-mouth recommendation by conducting empirical study on the Beibei.

3.1 Sharers' Influence

In traditional recommender system, which is mainly composed of binary relation between users and items, users interact with items due to their own interest and feature of items. In social commerce, receivers' behaviors may be influenced by who the sharer is. For example, if a sharer knows well about clothes, then the receivers are very willing to purchase the clothes recommended by the sharer. We conducted an empirical study to check whether such influence exists. First, we utilize BPR [12] method to obtain the latent feature vectors of users and items from data in the app platform, a traditional system without sharing behaviors. Then we utilize the learned latent vectors to calculate the preference (dot product) of receiver u^r on two item set: $item^+_{u^r}$ and $item^*_{u^r}$, where $item^+_{u^r}$ contains the received and finally purchased items by receiver u^r, while $item^*_{u^r}$ contains items shared by friends but not interacted by u^r. Mathematically, we define

$$Value^+_{u^r} = mean_{i \in item^+_{u^r}} (l(u^r)^\top l(i)) \ ,$$
$$Value^*_{u^r} = mean_{i \in item^*_{u^r}} (l(u^r)^\top l(i)) \ ,$$

where $l(\cdot)$ means the latent feature vector, obtained by learning on traditional commerce. Then we plot the distribution of $Value^+$ and $Value^*$ in Fig. 3a, and we calculated the mean value of $Value^* = 3.924$, which is significantly higher than that of $Value^+ = 1.932$. This empirical study shows that receiver does not only follow their own interests when interact with items in social commerce, which demonstrates the influence of sharer.

3.2 Receivers' Interests

Receivers' actions will also partially follow their interests in items. For example, if a user is very interested in sports, he/she still intends to click the shared links related to sports items. To show that users' own interests are also reserved in the social platform, we conduct analysis in a similar manner. We consider two item set: $item^+_{u^r}$ and $item^-_{u^r}$, where $item^+_{u^r}$ is same with Section 3.1, while $item^-_{u^r}$ contains items that have not been interacted by receiver u^r. More specifically, we define

$$Value^-_{u^r} = mean_{i \in item^-_{u^r}} (l(u^r)^\top l(i)).$$

Then, we plot the distribution of $Value^+$ and $Value^-$ in Fig. 3b, and we calculated the mean value of $Value^+ = 1.932$, which is significantly higher than $Value^- = 0.660$.[8] This empirical study indicates the receivers' own interest also is an important factor. Note that there is a suitable manner to validate these two factors. More specifically, for validating the sharers' influence, it is better to study the case where different sharers send the same item to the same receiver. However, such case seldom occurs in our dataset. Thus, we choose an indirect manner to study these two factors.

In short, these results provide one useful conclusion: three kinds of items, share, and purchase, share but not-purchase, not-purchased, have different matching scores. The conclusion motivates our method design in considering the effect of both sharer and receiver.

8. Here there is a slight difference between two figures for $Value^+$, which is due to different figure settings.

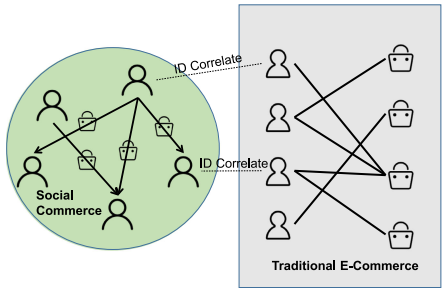


Fig. 2. Problem illustration for word-of-mouth recommendation. The left part is social commerce platform where users can share items to their friends, and the right part is traditional e-commerce platform with only user-item binary interaction. ID correlation paves the way for joint learning with two platforms.

4 Method

In this section, we propose a multi-relation learning model called TriM to capture the ternary relation of sharer, item and receiver. In order to solve the data sparsity problem on social commerce platform, we upgrade our model to a joint-learning one called TriM-joint using data from traditional commerce platform. In the following part, we denote the total number of sharers, items and receivers as l, m and n, and the set of all sharers, items and receivers as U_1, V and U_2.

In our model, we model receivers' interests and sharers' influence with the latent factor model at the same time. Then we propose to optimize our model with pairwise loss and jointly train it with traditional interaction data to alleviate the data sparsity issue. In what follows, we present our method by elaborating on the above designs. The used symbols are listed in Table 2.

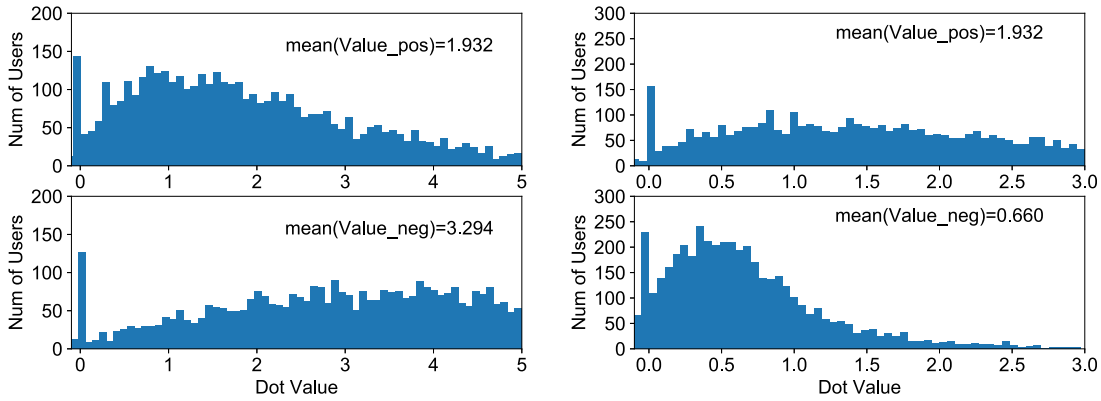
4.1 Modeling Receivers' Interests

In order to learn preferences of receivers, we employ matrix factorization (MF) to model the receiver-item interactions.

TABLE 2
Symbols Used in This Article

Symbol	Meaning
$Value^+_{u^r}$	Set of shared and purchased items
$Value^*_{u^r}$	Set of shared but not-purchased items
$Value^-_{u^r}$	Set of not purchased items
l	Latent vector
u^r, u^s, i	Receiver, sharer, item
X, Y	Observed preference
H, T, R	Model parameters

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.



(a) Analysis of sharers' influence (b) Analysis of receivers' interest

Fig. 3. Motivation analysis by conducting empirical study to show the existence of sharer's influence and receivers' interests in ternary relation.

Let $X_{u^r i}$ represent the preference of receiver u^r for item i, and $T \in \mathbb{R}^{m\times d}$, $R \in \mathbb{R}^{n\times d}$ denote latent receiver and item feature matrices, with column vectors T_{u^r} and R_i representing the d-dimensional latent feature vectors of receiver u^r and item i respectively. Following a common setting of MF models [12], [13], the likelihood function of latent parameters is defined as

$$p(X|T,R,\sigma_X^2) = \prod_{u^r=1}^{m}\prod_{i=1}^{n}[N(X_{u^r i}|S(T_{u^r}^\top R_i),\sigma_X^2)]^{I_{u^r i}^X}, \tag{1}$$

where $N(x|\mu,\sigma^2)$ is the probability density function of Gaussian distribution with mean μ and variance σ^2, $I_{u^r i}^X$ is the indicator function that is equal to 1 if receiver u^r interacted with item i and otherwise 0, and $S(\cdot)$ is the logistic *sigmoid* function. The zero-mean spherical Gaussian priors are also placed on latent feature vectors of receiver and item expressed as follows:

$$\begin{aligned} p(T|\sigma^2) &= \prod_{u^r=1}^{m} N(T_{u^r}|0,\sigma^2 I),\\ p(R|\sigma^2) &= \prod_{i=1}^{n} N(R_i|0,\sigma^2 I). \end{aligned} \tag{2}$$

Hence, through a Bayesian inference, we have

$$\begin{aligned} p(T,R|X,\sigma_X^2,\sigma^2) &\propto p(X|T,R,\sigma_X^2)p(T|\sigma^2)p(R|\sigma^2)\\ &= \prod_{u^r=1}^{m}\prod_{i=1}^{n}[N(X_{u^r i}|S(T_{u^r}^\top R_i),\sigma_X^2)]^{I_{u^r i}^X}\\ &\quad\times \prod_{u^r=1}^{m} N(T_{u^r}|0,\sigma^2 I)\times\prod_{i=1}^{n} N(R_i|0,\sigma^2 I), \end{aligned} \tag{3}$$

which represents the method on how to derive the receivers' latent feature space solely based on the receiver-item interaction matrix without considering sharers. In the next section, we will systematically illustrate how to model the influence of sharers.

4.2 Modeling Sharers' Influence

We also employ matrix factorization to model the sharer-item interactions from sharing data. Let $Y_{u^s i}$ represents the interactions of sharer u^s and item i. We also denote $H \in \mathbb{R}^{l\times d}$ and $R \in \mathbb{R}^{n\times d}$ as latent sharer and item feature matrices, with column vectors H_{u^s} and R_i representing the d-dimensional sharer-specific and item-specific latent feature vectors of sharer u^s and item i, respectively. Similar to previous section, we have

$$\begin{aligned} p(H,R|Y,\sigma_Y^2,\sigma^2) &\propto p(Y|H,R,\sigma_Y^2)p(H|\sigma^2)p(R|\sigma^2)\\ &= \prod_{u^s=1}^{l}\prod_{i=1}^{n}[N(X_{u^s i}|S(H_{u^s}^\top R_i),\sigma_Y^2)]^{I_{u^s i}^Y}\\ &\quad\times \prod_{u^s=1}^{l} N(H_{u^s}|0,\sigma^2 I)\times\prod_{i=1}^{n} N(R_i|0,\sigma^2 I). \end{aligned} \tag{4}$$

4.3 Our Proposed Model: TriM

Based on the above derivation, given the receiver-item interaction matrix, the observed preference $X_{u^r i}$ is interpreted by the the latent feature vectors of receiver u^r and item i, while given the interaction matrix of sharers and items, the observed interaction $Y_{u^s i}$ is realized as the influence of sharer u^s on item i. Then, given the ternary relation of sharer u^s, item i and receiver u^r, both the preference and influence should be considered. When interacting with an item shared by a friend, a user not only has her own tastes, but may also be influenced by her friend. Hence, in order to define the model more realistically, every observed interaction of sharer, item and receiver should reflect both of these two factors. Based on this, we model the conditional distribution over the observed ratings as follows:

$$\begin{aligned} &p(H,R,T|X,Y,\sigma_Z^2,\sigma^2)\\ &= \prod_{u^s=1}^{l}\prod_{i=1}^{n}\prod_{u^r=1}^{m}[N(Z_{u^s i u^r}|S(\alpha H_{u^s}^\top R_i\\ &\quad+(1-\alpha)T_{u^r}^\top R_i),\sigma_Z^2)]^{I_{u^s i u^r}^Z}\times\prod_{u^s=1}^{l} N(H_{u^s}|0,\sigma^2 I)\\ &\quad\times\prod_{i=1}^{n} N(R_i|0,\sigma^2 I)\times\prod_{u^r=1}^{m} N(T_{u^r}|0,\sigma^2 I), \end{aligned} \tag{5}$$

where parameter α balances the trade-off between the receivers interests and sharer's influence on items. It controls to what extent users rely on themselves or their trusted friends when doing online shopping. Maximizing the logarithm of the posterior distribution for the recommendations

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

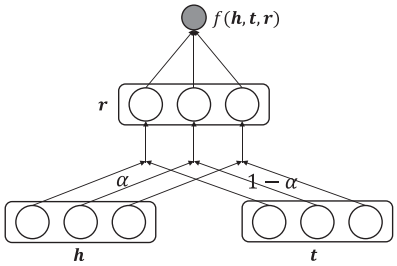


Fig. 4. Method illustration for our proposed TriM model, where **h**, **t**, and **r** are the latent vectors of sharer (head node), receiver (tail node) and item (relation), respectively. The head node, tail node and relation here refer to sharer, receiver and sharing relation, respectively. α is a tuning parameter controlling the weight of sharers' influence and receivers' interests in the ternary relation.

with fixed hyper-parameters (i.e., the observation noise variance and prior variances) is equivalent to minimizing the following objective function:

$$\begin{aligned}&\mathcal{L}(Z,H,R,T)\\&=\frac{1}{2}\sum_{u^s=1}^{l}\sum_{i=1}^{n}\sum_{u^r=1}^{m}I^Z_{u^s i u^r}[Z_{u^s i u^r}-S(\alpha H_{u^s}^\top R_i\\&\quad+(1-\alpha)T_{u^r}^\top R_i)]^2+\frac{\lambda}{2}||H||_F^2+\frac{\lambda}{2}||R||_F^2+\frac{\lambda}{2}||T||_F^2,\end{aligned}\tag{6}$$

where $\lambda=\sigma_Z^2/\sigma^2$, and $||\cdot||_F^2$ denotes the Frobenius norm. However, pairwise loss usually performs better than the loss function above in implicit recommender systems. Therefore, we introduce our final TriM model using pairwise loss in the next section.

4.4 TriM With Pairwise Loss

Borrowing the concepts from knowledge graph embedding [14], the score function of TriM is defined as below:

$$f_{TriM}(u^s,i,u^r)=R_i^\top(\alpha H_{u^s}+(1-\alpha)T_{u^r}),\tag{7}$$

where $\alpha\in[0,1]$ is a tuning parameter which controls the weight of sharer and receiver in the ternary relation. The graphical representation of TriM is illustrated in Fig. 4. Since $f_{TriM}(u^s,i,u^r)\neq f_{TriM}(u^r,i,u^s)$ when $\alpha\neq 0.5$, this score function is able to model the asymmetric relation. Then we utilize pairwise learning to model the implicit feedback in social commerce platform. We denote the set of items shared by user u^s and clicked by user u^r as $V^+_{(u^s,u^r)}$. For each item in this set, we assume the user pair (u^s,u^r) prefers it over all other non-observed items. For two items that both belong to $V^+_{(u^s,u^r)}$, we cannot infer any preference. The same is true for two items that are not in the set. To formalize this we create training data $D_s: U_1\times U_2\times V\times V$ as

$$D_s:=\{(u^s,i,j,u^r)|i\in V^+_{(u^s,u^r)}\cap j\in V\backslash V^+_{(u^s,u^r)}\}.\tag{8}$$

Then we define the preference using minus form as follows:

$$x(u^s,i,j,u^r)=f(u^s,i,u^r)-f(u^s,j,u^r).\tag{9}$$

To this end, we propose the pairwise loss function as

$$\begin{aligned}\mathcal{L}_s=&-\sum_{(u^s,i,j,u^r)\in D_s}\ln S(x(u^s,i,j,u^r))\\&+\frac{\lambda}{2}||H||_F^2+\frac{\lambda}{2}||R||_F^2+\frac{\lambda}{2}||T||_F^2.\end{aligned}\tag{10}$$

As we can see in Equation (10), the error function does not focus on predicting the right value, but on the ranking of the preference for observed and unobserved items.

4.5 TriM-Joint

Data sparsity issue is one of the major issues in recommender systems. In general, introducing attributes data [15], [16] or other behavioral data [17], [18] can help improve the recommendation performance. Thus, in our problem, we can take use of the interaction data from the traditional e-commerce platform. Since there are overlapping users and items between two platforms, we can share the latent feature vectors of these users and items. More specifically, in the traditional e-commerce platform, we use matrix factorization to model the user-item action as follows:

$$f_{MF}(u,i')=R_{i'}'^\top T'_u,\tag{11}$$

where $R'_{i'}$ and T'_u are the latent vectors of item i and user u. We define set of items interacted with user u as V_u^+, then create the training data as

$$D_t:=\{(u,i',j')|i'\in V_u^+\cap j'\in V\backslash V_u^+\}.\tag{12}$$

Then we define the preference using minus form

$$y(u,i',j')=f_{MF}(u,i')-f_{MF}(u,j').\tag{13}$$

The loss in traditional e-commerce platform can be expressed by

$$\begin{aligned}\mathcal{L}_t=&-\sum_{(u,i',j')\in D_t}\ln S(y(u,i',j'))\\&+\frac{\lambda}{2}||R'||_F^2+\frac{\lambda}{2}||T'||_F^2.\end{aligned}\tag{14}$$

To transfer the knowledge of traditional e-commerce platform, we jointly learn the two losses together, by defining the optimization problem as

$$\begin{aligned}&\text{minimize }\mathcal{L}=\beta\mathcal{L}_s+(1-\beta)\mathcal{L}_t,\\&\text{subject to }R_i-R'_{i'}=0\text{ and }T_{u^r}-T'_u=0,\\&\qquad\text{for any } i\triangleq i',u^r\triangleq u',\end{aligned}\tag{15}$$

where β is the tuning parameter to control the weight of social commerce platform loss and traditional e-commerce platform loss. $i\triangleq i',u^r\triangleq u'$ means item i and receiver u^r in the social commerce platform are the same with item i' and user u in the traditional e-commerce platform. So when adopting joint learning in two platforms, we share the latent feature vectors of receivers and items with corresponding users and items in traditional e-commerce platform.

With joint-learning, not only the relation between users' preferences and influence on friends become clear, but also the data sparsity issue on social e-commerce websites can be alleviated.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1. Optimization of TriM-Joint Model

Input: $D_s, D_i, d, \alpha, \beta, \lambda, epoch, num, B_s, B_i, optimizer$
Output: H, R, T, R', T'
$H, R, T, R', T' \Leftarrow H_0, R_0, T_0, R'_0, T'_0$;
$i = 0$;
for $k = 1; k < epoch; k++$ **do**
  for $m = 1; m < num; m++$ **do**
    Sample a batch $\mathcal{B}_s^m$ from D_s of size B_s;
    Sample a batch $\mathcal{B}_i^m$ from D_i of size B_i;
    Calculate $\mathcal{L}^m$ with $\mathcal{B}_s^m$ and $\mathcal{B}_i^m$;
    Calculate $\partial\mathcal{L}^m/\partial H$, $\partial\mathcal{L}^m/\partial R$,
    $\partial\mathcal{L}^m/\partial T$, $\partial\mathcal{L}^m/\partial R'$, $\partial\mathcal{L}^m/\partial T'$ by (17);
    $H \leftarrow optimizer(H, \partial\mathcal{L}^m/\partial H)$;
    $R \leftarrow optimizer(R, \partial\mathcal{L}^m/\partial R)$;
    $T \leftarrow optimizer(T, \partial\mathcal{L}^m/\partial T)$;
    $R' \leftarrow optimizer(R', \partial\mathcal{L}^m/\partial R')$;
    $T' \leftarrow optimizer(T', \partial\mathcal{L}^m/\partial T')$;
return H, R, T, R', T'

4.6 Optimization Method of Our Proposed TriM

For our proposed TriM, the parameters can be optimized by stochastic gradient descent (SGD) with mini-batch training. The gradients for updating parameters are given as follows:

$$
\begin{aligned}
\frac{\partial \mathcal{L}_s}{\partial H_{u^s}} &= \sum_{(u^s,i,j,u^r)\in D_s} -S(-x(u^s,i,j,u^r)) \\
&\quad \cdot \alpha(R_i - R_j) + \beta\lambda H_{u^s}, \\
\frac{\partial \mathcal{L}_s}{\partial R_i} &= \sum_{(u^s,i,j,u^r)\in D_s} -S(-x(u^s,i,j,u^r)) \\
&\quad \cdot (\alpha H_{u^s} + (1-\alpha)T_{u^r}) \\
&\quad + \sum_{(u^s,j,i,u^r)\in D_s} S(-x(u^s,j,i,u^r)) \\
&\quad \cdot (\alpha H_{u^s} + (1-\alpha)T_{u^r}) + \lambda R_i, \\
\frac{\partial \mathcal{L}_s}{\partial T_{u^r}} &= \sum_{(u^s,i,j,u^r)\in D_s} -S(-x(u^s,i,j,u^r)) \\
&\quad \cdot (1-\alpha)(R_i - R_j) + \lambda T_{u^r}, \\
\frac{\partial \mathcal{L}_t}{\partial R_{i'}} &= \sum_{(u,i',j')\in D_t} -S(-y(u,i',j')) \cdot T_u \\
&\quad + \sum_{(u,j',i')\in D_t} S(-y(u,j',i')) \cdot T_u + \lambda R_{i'}, \\
\frac{\partial \mathcal{L}_t}{\partial T_u} &= \sum_{(u,i',j')\in D_t} -S(-y(u,i',j')) \\
&\quad \cdot (R_{i'} - R_{j'}) + \lambda T_u,
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial H_{u^s}} &= \beta \frac{\partial \mathcal{L}_s}{\partial H_{u^s}}, \\
\frac{\partial \mathcal{L}}{\partial R_i} &= \beta \frac{\partial \mathcal{L}_s}{\partial R_i} + I(i' \triangleq i)(1-\beta)\frac{\partial \mathcal{L}_t}{\partial R_{i'}}, \\
\frac{\partial \mathcal{L}}{\partial T_{u^r}} &= \beta \frac{\partial \mathcal{L}_s}{\partial T_{u^r}} + I(u \triangleq u^r)(1-\beta)\frac{\partial \mathcal{L}_t}{\partial T_u}, \\
\frac{\partial \mathcal{L}}{\partial R_{i'}} &= (1-\beta)\frac{\partial \mathcal{L}_t}{\partial R_{i'}} + I(i \triangleq i')\beta \frac{\partial \mathcal{L}_s}{\partial R_i}, \\
\frac{\partial \mathcal{L}}{\partial T_u} &= (1-\beta)\frac{\partial \mathcal{L}_t}{\partial T_u} + I(u^r \triangleq u)\beta \frac{\partial \mathcal{L}_s}{\partial T_{u^r}},
\end{aligned}
\tag{17}
$$

where $I(\cdot)$ stands for the indicator function and $S(\cdot)$ stands for sigmoid function. To reduce the time complexity of model training, we adopted a mini-batch gradient descent approach to learn the parameters. With sampling, the cost of the gradient update no longer grows linearly in the number of sharers, items and receivers, but in the number of entries sampled. The hyper-parameters are set by cross validation. Algorithm 1 illustrates our SGD-based joint training algorithm for learning the model.

5 Experiments

In this section, we conduct experiments on a precious real-world dataset to answer the following research questions:

- *RQ1:* Can the performance of our proposed model outperform other state-of-the-art methods in the task of topK recommendation? What's the effect of factors, i.e., sharers' influence and receivers' interests?
- *RQ2:* Can our proposed TriM capture the ternary relation better than other multi-relation learning models and why? What's the effect of parameter α?
- *RQ3:* Is joint learning helpful or not? If yes, what's the effect of joint learning parameter β, and how much improvement it can bring about under different data sparsity levels?
- *RQ4:* What's the effect of other hyper-parameters, such as dimension d and regularization factor λ?

5.1 Experimental Settings

5.1.1 Evaluation Datasets

Our dataset is collected from Beibei,[9] one of the largest e-commerce platforms for maternal and infant products in China. Users have two main channels to buy products on Beibei, namely through traditional e-commerce app or social media. For the latter one, by clicking on a product link shared by their friends through Wechat,[10] users can also directly purchase this product on Wechat. Besides collecting logs of sharing events from Wechat, we also collect interaction data of *traditional interaction* from Beibei App as an aid for the WoM recommendation task. Our data was collected from September 20, 2017, to October 22, 2017.

To generate the evaluation dataset, in the social commerce platform, we filter out sharers with too few sharing actions from the raw dataset (less than 5), and make sure every receiver has interaction logs in the traditional e-commerce platform. In the traditional e-commerce platform, we remove users and items with too few interaction logs (less than 5). Note that such manner of preprocessing data is widely used in existing researches [19], [20]. The details of the evaluation dataset are shown in Table 3. We will release this precious dataset upon acceptance.

5.1.2 Evaluation Protocols

To evaluate the performance, we adopt the leave-one-out evaluation method (to predict the rank of left item based on historical interactions) with two ranking metrics widely

9. http://www.beibei.com
10. Wechat is one of the largest social network services in China. https://www.wechat.com

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

TABLE 3
Statistics of Our Evaluation Datasets

Social Commerce	Sharer#	Receiver#	Item#	Events#
Beibei-Wechat	1864	4609	3911	11331
Traditional e-commerce	**User#**		**Item#**	**Events#**
Beibei-App	13368		13670	294597

used in literature [10], [21], *Hit Ratio* (HR) and *Normalized Discounted Cumulative Gain* (NDCG), to judge the performance of the ranking list. HR@K measures whether the test item is contained in the top-K item ranking list (1 for yes and 0 for no); NDCG@K complements HR by assigning higher scores to the hits at higher positions of the ranking list.

5.1.3 Baselines

We compare the performance of our proposed TriM and TriM-joint with twelve baseline methods, which can be divided into two groups, namely single-platform and cross-platform. Seven single-platform baseline methods, which only unitize data in social commerce platform, are introduced as follows:

- *ItemPop*. This method ranks items based on their popularity, which is defined by the number of historical interactions. This is a non-personalized recommendation method.
- *MF-logloss* [22]. This method extends MF with negative sampling to handle implicit dataset.
- *BPR* [12]. This is a widely-used method optimizing a pairwise loss when dealing with implicit feedback data.
- *SocialBPR* [6]. This method optimizes a pairwise loss with a distance-of-weighted-sum social regularization term, which is one of the most commonly used methods in social recommendation.
- *TransE* [23]. This is the most representative translational distance model for multi-relational learning.
- *DistMult* [24]. This model is one of the most commonly used semantic matching models for multi-relational learning.
- *HolE* [25]. This method extends the expressive power of DistMult with higher computing complexity for learning multi-relation.

Note that for TransE, DistMult, and HolE, we consider sharer, item, and receiver as head entity, relation, and tail entity, respectively. The details are presented in Fig. 6. For social-recommendation baselines, we do not compare with the recent advances of neural network based models [26], [27], [28] for two reasons. First, our proposed method is a general solution. The matrix factorization in TriM can be extended by complicated interaction functions such as MLP in [21] and GCN in [29]. Second, we found neural methods are easy to over-fit in our utilized dataset.

The second group contains five methods that consider cross-platform characteristics with jointly learning from user behaviors of two platforms, including *BPR-joint*, *SocialBPR-joint*, *TransE-joint*, *DistMult-joint* and *HolE-joint*. They are extended from BPR [12], SocialBPR [6], TransE [23], DistMult [24] and HolE [25], respectively. These methods and our proposed one share the latent feature vectors of receivers and items in the social commerce platform with the same users and items in the traditional e-commerce platform.[11] For all these models except ItemPop and MF-logloss, we adopt pairwise learning, a widely used technique in implicit recommender systems [9], [12].

TABLE 4
Performance Evaluation for Different Models

Model	HR@5	NDCG@5	HR@10	NDCG@10
ItemPop	0.0446	0.0207	0.1091	0.0418
MF	0.0602	0.0300	0.1059	0.0447
BPR	0.0682	0.0342	0.1102	0.0477
SocialBPR	0.0704	0.0350	0.1171	0.0501
HolE	0.0602	0.0305	0.0973	0.0424
DistMult	0.0704	0.0351	0.1128	0.0490
TransE	0.1596	0.0807	0.2176	0.0997
TriM	**0.1714**	**0.0896**	**0.2230**	**0.1062**
BPR-joint	0.1161	0.0580	0.1838	0.0801
SocialBPR-joint	0.1209	0.0612	0.1934	0.0847
HolE-joint	0.0913	0.0447	0.1429	0.0612
DistMult-joint	0.1290	0.0663	0.1875	0.0851
TransE-joint	0.1740	0.0905	0.2413	0.1126
TriM-joint	**0.1988**	**0.1035**	**0.2783**	**0.1292**

Performance is measured in terms of the ability to predict items with highest probability of future interactions given the sharer and receiver, with metrics HR and NDCG.

5.1.4 Parameter Settings

To make experiments fair and reasonable, the regularization factor λ and dimension d for all methods are searched carefully in $\{1e\text{-}1, 1e\text{-}2, 1e\text{-}3, 1e\text{-}4, 1e\text{-}5, 1e\text{-}6, 1e\text{-}7, 1e\text{-}8\}$ and $\{8, 16, 32, 64, 128, 256, 512\}$, respectively. The transfer learning weight β for all joint learning model and the weight α of sharer and receiver in ternary relation are all searched in $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. Then we adopt Adam [30], a SGD-based optimizer, and searching learning rate in {0.0001,0.001,0.01} to optimize all the methods. The number of epochs is set as 500 to make sure the convergence of all the models. In the following sections, we report the results under optimal parameter setting for each model. We will release our precious and carefully collected dataset and our implementation with optimal parameter setting after paper being accepted.

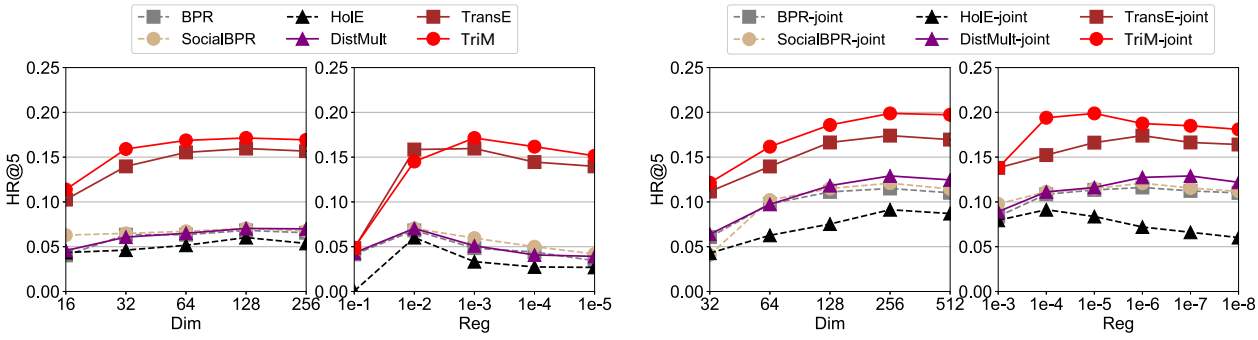
5.2 Performance Comparison (RQ1)

Table 4 presents the performance comparison among all methods. From the results, we have the following findings.

- *Modelling sharers' influence is useful*. Compared with models such as BPR that can only capture the interaction between receivers and items, models dealing with ternary relation such as DistMult, TransE and TriM can obtain much better results, with improvement of 3.2, 134.0 and 151.3 percent on HR@5 respectively. As for cross-platform models, DistMult-joint, TransE-joint and TriM-joint can improve BPR-joint with 11.1, 49.9 and 71.2 percent on HR@5

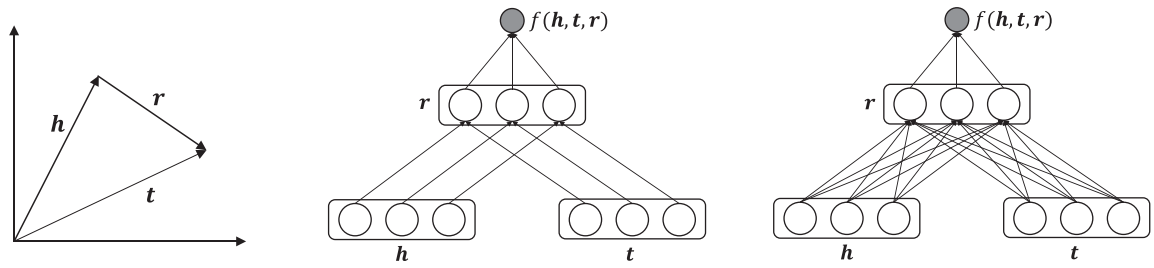
11. These users and items can be considered as *bridge* users and items across two platforms.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.



(a) Models without joint learning (b) Models with joint learning

Fig. 5. Performance under different hyper-parameter values. (ItemPop does not have these hyper-paramters).



(a) TransE (b) DistMult (c) HolE

Fig. 6. Simple illustrations of three commonly used multi-relation representation models: TransE, DistMult, and HolE.

respectively. These results show that capturing sharers' influence is indeed helpful for the recommendation scenarios.

- *Modelling receivers' interest is also useful.* Users' behavior in the traditional e-commerce platform reflects their own interests for items. Our results show that when utilizing transferred knowledge from traditional e-commerce platform with joint learning, models such as BPR, SocialBPR, HolE, DistMult, TransE, and TriM can be improved by 70.2, 71.7, 51.7, 83.2, 9.0 and 16.0 percent on HR@5 respectively.
- *TriM and TriM-joint outperform single-platform and cross-platform baselines, respectively.* We can observe that our proposed TriM outperforms all baseline methods substantially and consistently w.r.t all HR@K and NDCG@K metrics. The relative improvement for TriM to the best baseline is 7.4 and 11.0 percent for HR@5 and NDCG@5, respectively. As for cross-platform models, the relative improvement is 14.2 and 14.4 percent. This justifies the effectiveness of our model. It is worth mentioning that ItemPop's performance is poor, and thus in further experiments we do not consider it anymore.

To conclude, capturing the sharers' influence and receivers' interest together can improve the performance of recommendation significantly, and our proposed model can achieve better results than state-of-the-art models.

5.3 Ternary Relation Learning (RQ2)

In this section, we study the multi-relation learning model in detail, and analyze the effect of parameter α in our proposed method.

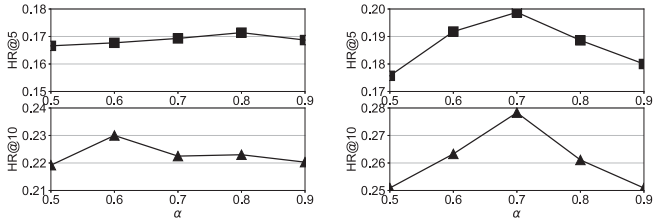
5.3.1 Comparison of Multi-Relation Learning Model

From the results listed in Table 4, we can see that for HR@5, the performance order is TriM $>$ TransE $>$ DistMult $>$ HolE. Since the sharing data is very sparse, models with complex structure such as HolE will suffer from the overfitting issue. HolE first calculates the circular correlation between sharer latent feature vector and receiver latent feature vector, and then obtains the cosine similarity between the result vector and item vector. However, from the perspective of representation learning in recommender system, each dimension of embedding represents a corresponding attribute of user or item. Such interactions across different dimensions in latent space violate the physical meaning of embedding. DistMult is simpler compared to HolE, but motivated by logical chain rule in Knowledge Graph such as $BornInCity(a, b) \cap CityInCountry(b, c) \longrightarrow Nationality(a, c)$. However, this rule may not hold true in sharing network for the task in this paper. TransE simply assumes that the latent feature vector of the item is close to the difference between sharer and receiver, which may be not accurate enough since a pair of sharer and receiver may interact with different kinds of items. What's more, in recommender system, it would be better to measure the similarity between user and item in latent space rather than distance.

5.3.2 Effect of Parameter α

Then we analyze the impact of hyper-parameter α, which controls the weight of sharer and receiver in the ternary relation. As shown in Fig. 7, for TriM, the optimal α for HR@5 and HR@10 is 0.8 and 0.6, respectively. While for the joint learning model, the optimal α is 0.7. This indicates that the effectiveness of our model changes under different

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.



(a) HR value of TriM (b) HR value of TriM-joint

Fig. 7. Performance of our proposed model under different α values (other hyper-parameters are searched to be optimal).

α. What's more, the range of HR@5 of TriM is 0.048 with α ranging from 0.5 to 0.9, while the range of HR@5 for TriM-joint is 0.231, which shows that the performance of our model is more sensitive to α when jointly learned with traditional e-commerce data in order to learn receivers' interests better.

5.4 Effect of Joint Learning (RQ3)

In this section we'll discuss the effectiveness of joint learning.

5.4.1 Weight Parameter in Joint Learning

In Fig. 8, we present the performance of under different β settings. It can be observed that with the rise of β, the performances of all models first increase and then decrease. When β is too small, the social e-commerce platform information cannot contribute much to the loss function. When β is too large, the social e-commerce platform dominates the loss function. With $\beta = 0.7$, our proposed model achieves best performance. Also we can see that under different values of β, our model always outperforms any other models.

5.4.2 Data Sparsity

We explore how these models perform when the training data is at different sparsity levels. In this work, the whole dataset is very sparse already, and it is not so suitable to split users into sparse and dense groups. Thus, we follow a more suitable manner [31] to study the data sparsity problem. Specifically, we use different ratios of training data from 20 to 100 percent to evaluate the methods. The results are shown in Fig. 8. We can observe that the performances of all these models are very sensitive to data sparsity. For

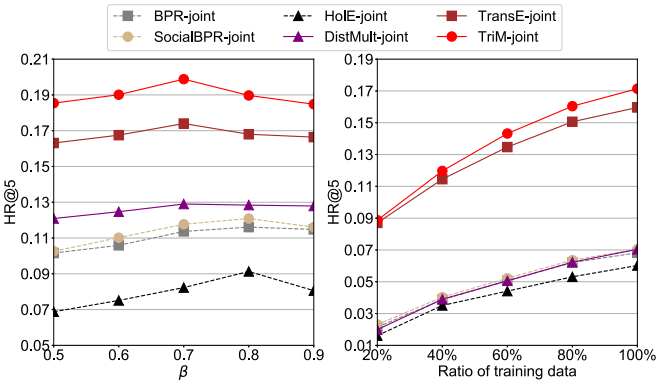


Fig. 8. Performance of models under different β and different ratios of training data (β is carefully tuned for different data ratio).

TABLE 5
Performance Improvement Rate of Different Models Under Different Ratios of Training Data

Model	20%	40%	60%	80%	100%
BPR-joint	350.7%	172.1%	117.8%	82.3%	70.2%
SocialBPR-joint	327.7%	168.5%	117.2%	84.6%	71.7%
HolE-joint	126.7%	63.2%	63.5%	55.7%	51.7%
DistMult-joint	394.0%	182.3%	134.2%	101.0%	83.2%
TransE-joint	52.5%	34.7%	24.2%	13.5%	9.0%
TriM-joint	52.6%	37.3%	26.7%	20.0%	16.0%

example, when using only 20 percent training data, BPR-joint, SocialBPR-joint, HolE-joint, DistMult-joint, TransE-joint and TriM-joint only achieve 0.0215, 0.0231, 0.0161, 0.0199, 0.0870 and 0.0887 for HR@5 respectively, decreased by 68, 67, 73, 72, 45 and 48 percent compared with HR@5 under 100 percent training data. Table 5 shows the improvement rate of HR@5 for different models under different ratios of training data by joint learning. It can be observed that the sparser the training data is, the larger improvement can be achieved via joint learning. What's exciting is that, when the ratio of training data is only 20 percent, joint learning can improve the results by 350.7, 327.7, 126.7, 394.0, 52.5 and 52.6 percent for BPR-joint, SocialBPR-joint, HolE-joint, DistMult-joint, TransE-joint, and TriM-joint respectively.

5.5 Effect of Other Hyper-Parameters (RQ4)

In this section we'll discuss how other hyper-parameters, i.e., dimension d and weight of regularization λ, influence the performance of our proposed model and baseline methods.

5.5.1 Effect of Dimension

In terms of parameter d (i.e., number of latent features), intuitively, increasing it could add more flexibility to the model. However, after reaching the peak, further increasing d degrades the performance, which may be caused by over-fitting with redundant parameters. This is consistent with existing works' finding in the impact of embedding sizes, including representation learning on graph data [32], interaction data [21], etc. The results in Fig. 5a and Fig. 5b for different values of d show that our proposed method performs the best when varying d values.

5.5.2 Effect of Regularization

As for parameter λ (i.e., weight of regularization), increasing it could prevent over-fitting of models. However, after reaching the peak, further increasing λ degrades the performance, which may be caused by insufficient learning from training data. The results in Figs. S5a and S5b for different values of λ show that our proposed method performs the best under varying λ values.

In summary, by conducting experiments, we show that modeling sharers' influence and receivers' interest are both helpful for recommendation performance, and our proposed model outperforms other baseline models for better capturing ternary relation in word-of-mouth recommendation. We also demonstrate the usefulness of joint learning under different sparsity levels. What's more, the effect of

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

hyper-parameters has also been analyzed to obtain a comprehensive understanding of our models.

6 Related Work

6.1 Social Recommendation

Social recommendation aims to leverage social relations to improve recommendation performance. Existing social recommendation methods can be generally classified into three groups [5]. The first group relies on regularization techniques to limit the distance of embedding vectors in latent space between a user and her friends [6], [7], [8], [9], [10], which is most prevalent in social recommendation. The second group extends CMF [33] to co-factorizing the user-item matrix and social matrix [11], [34]. The third group assumes a user's decision should be a combination of her own characteristics and her trusted friends' recommendations [35], [36], [37]. Yu *et al*. [28] proposed to utilize generative adversarial networks to identifying reliable friends, in both the observed and unobserved social networks, to enhance social recommendation with addressing data sparsity problem. Sun *et al*. [38] took the temporal factor into consideration in social recommendation and approached it with recurrent neural networks. Recently, Wu *et al*. [27] and Fan *et al*. [26] propose to utilize graph neural network for social recommendation based on social-trust assumption.

Some works utilize social network as the auxiliary information to help solve the data sparsity issue in recommender systems, including event recommendation [39], video recommendation [40], group recommendation [41], etc.

Our defined WoM recommendation task focuses on a deeper level compared with social recommendation since we model how users' consumption decision is affected by friends.

6.2 Cross-Domain Recommendation

Data sparsity issue is of great concern in recommendation. A typical solution is cross-domain recommendation, which aims to exploit data from multiple domains to help to solve the data sparsity problem in a single domain. The main category of cross-domain recommendation methods [33], [42], [43], [44] are based on collective matrix factorization (CMF), which jointly factorize behavioral matrices from multiple domains. In these methods, either user embedding vectors of item embedding vectors are shared, while the unshared ones are trained separately. Some other works [10], [45], [46], [47], [48] make use of more diverse data from other domains, including review data, user profiles, item attributes, etc. Although compelling success has been achieved by previous work, little attention has been paid to recommendation across heterogeneous domains. In our settings of joint learning, the source domain contains traditional user-item binary interactions, while the target domain is a social commerce domain with sharer-item-receiver ternary interactions. As a result, existing approaches cannot be directly applied to this new research problem.

6.3 Knowledge Graph Based Recommendation

With rapid growth in knowledge graph (KG) construction and application, a new research direction, Knowledge Graph Embedding, has been proposed and quickly gained massive attention. Such embedding techniques can be split into two groups [14]: translational distance models [23], [49], [50] and semantic matching models [24], [25], [51]. Since the user-item interactions can be very sparse, how to improve the performance of recommender systems using KGs is an important task. More recently, some work has focused on using external KGs to supplement content-based recommendation [52], [53], [54], [55]. Zhang *et al*. [54] recently proposed a hybrid recommendation framework which leverages three types of heterogeneous information in a KG, including structural knowledge, textual knowledge, and visual knowledge, to derive semantic representations for items. [55] investigates three methods for constructing KG based recommendations using a general-purpose probabilistic logic system. Recently, Wang *et al*. [56] proposes to unify the knowledge graph and interaction graph in one graph and leverage graph neural networks for learning representation. Wang *et al*. [57] propose to jointly learn representation in recommendation task and knowledge graph embedding task in a multi-task learning manner. Note that all the existing works utilize external KGs as auxiliary information for improving the performance of recommender system, while our work borrows the concept of KG Embedding to model the ternary relation between the user, item and sharer directly, which is significantly different.

7 Conclusion and Future Work

In this work, we systematically investigated the problem of word-of-mouth recommendation in social media, which, to the best of our knowledge, is a practical task but has rarely been studied previously. We proposed a novel model TriM, effectively modeling sharer-item-receiver ternary relation and captures sharers' influence and receivers' interests simultaneously. To evaluate our proposed model, we conducted extensive experiments on real-world dataset, showing that TriM significantly outperforms existing state-of-the-art methods. A limitation of this work is that we only adopt one private dataset for evaluation. As far as we know, no other publicly available data is suitable for the WoM recommendation. For future work, we will conduct online A/B test to evaluate our model further. Since WoM is an in-depth study of social recommendation with the help of this dataset, we may release our dataset upon paper accepted and benefit the community, especially researchers focusing on social recommendation.

Acknowledgments

This work was supported in part by The National Key Research and Development Program of China under grant SQ2020AAA010130, the National Nature Science Foundation of China under U1936217, 61971267, 61972223, 61941117, 61861136003, Beijing Natural Science Foundation under L182038, Beijing National Research Center for Information Science and Technology under 20031887521, and research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology.

References

[1] Z. Huang and M. Benyoucef, "From e-commerce to social commerce: A close look at design features," *Electron. Commerce Res. Appl.*, vol. 12, no. 4, pp. 246–259, 2013.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

[2] Reuters, "Beidian debuted in times square and the rise of social e-commerce in China," 2018. [Online]. Available: https://www.reuters.com/brandfeatures/venture-capital/article?id=34917

[3] J. J. Brown and P. H. Reingen, "Social ties and word-of-mouth referral behavior," *J. Consum. Res.*, vol. 14, no. 3, pp. 350–362, 1987.

[4] Nielsen, "Consumer trust in online, social and mobile advertising grows," 2012. [Online]. Available: http://www.nielsen.com/us/en/insights/news/2012/consumer-trust-in-online-social-and-mobile-advertising-grows.html

[5] J. Tang, X. Hu, and H. Liu, "Social recommendation: A review," *Soc. Netw. Anal. Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.

[6] L. Guo, J. Ma, H.-R. Jiang, Z.-M. Chen, and C.-M. Xing, "Social trust aware item recommendation for implicit feedback," *J. Comput. Sci. Technol.*, vol. 30, no. 5, pp. 1039–1053, 2015.

[7] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proc. ACM Conf. Recommender Syst.*, 2010, pp. 135–142.

[8] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. Int. Conf. Web Search Data Mining*, 2011, pp. 287–296.

[9] L. Du, X. Li, and Y.-D. Shen, "User graph regularized pairwise matrix factorization for item recommendation," in *Proc. Int. Conf. Adv. Data Mining Appl.*, 2011, pp. 372–385.

[10] X. Wang, X. He, L. Nie, and T.-S. Chua, "Item silk road: Recommending items from information domains to social users," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 185–194.

[11] H. Ma, H. Yang, M. R. Lyu, and I. King, "SoRec: Social recommendation using probabilistic matrix factorization," in *Proc. Int. Conf. Inf. Knowl. Manage.*, 2008, pp. 931–940.

[12] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme , "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. Conf. Uncertainty Artif. Intell.*, 2009, pp. 452–461.

[13] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2008, pp. 1257–1264.

[14] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.

[15] X. He and T.-S. Chua, "Neural factorization machines for sparse predictive analytics," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 355–364.

[16] S. Rendle, "Factorization machines," in *Proc. IEEE Int. Conf. Data Mining*, 2010, pp. 995–1000.

[17] H. Yin, H. Chen, X. Sun, H. Wang, Y. Wang, and Q. V. H. Nguyen, "SPTF: A scalable probabilistic tensor factorization model for semantic-aware behavior prediction," in *Proc. IEEE Int. Conf. Data Mining*, 2017, pp. 585–594.

[18] B. Loni, R. Pagano, M. Larson, and A. Hanjalic, "Bayesian personalized ranking with multi-channel user feedback," in *Proc. ACM Conf. Recommender Syst.*, 2016, pp. 361–364.

[19] R. He, W.-C. Kang, and J. McAuley, "Translation-based recommendation: A scalable method for modeling sequential behavior," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 5264–5268.

[20] B. Chang, Y. Park, D. Park, S. Kim, and J. Kang, "Content-aware hierarchical point-of-interest embedding model for successive POI recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 3301–3307.

[21] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.

[22] C. C. Johnson, "Logistic matrix factorization for implicit feedback data," in *Workshop Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 1–9.

[23] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2013, pp. 2787–2795.

[24] B. Yang, W.-T. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *Proc. Int. Conf. Learn. Representations*, 2014, pp. 1–12.

[25] M. Nickel *et al.*, "Holographic embeddings of knowledge graphs," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 1955–1961.

[26] W. Fan *et al.*, "Graph neural networks for social recommendation," in *Proc. World Wide Web Conf.*, 2019, pp. 417–426.

[27] L. Wu, P. Sun, Y. Fu, R. Hong, X. Wang, and M. Wang, "A neural influence diffusion model for social recommendation," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 235–244.

[28] J. Yu, M. Gao, H. Yin, J. Li, C. Gao, and Q. Wang, "Generating reliable friends via adversarial training to improve social recommendation," in *Proc. IEEE Int. Conf. Data Mining*, 2019, pp. 768–777.

[29] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for web-scale recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 974–983.

[30] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[31] J. Bobadilla and F. Serradilla, "The effect of sparsity on collaborative filtering metrics," in *Proc. 20th Australas. Conf. Australas. Database*, 2009, pp. 9–18.

[32] M. Wang, W. Fu, S. Hao, H. Liu, and X. Wu, "Learning on big graph: Label inference and regularization with anchor hierarchy," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 5, pp. 1101–1114, May 2017.

[33] A. P. Singh and G. J. Gordon, "Relational learning via collective matrix factorization," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2008, pp. 650–658.

[34] J. Tang, X. Hu, H. Gao, and H. Liu, "Exploiting local and global social context for recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2013, pp. 2712–2718.

[35] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with social trust ensemble," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2009, pp. 203–210.

[36] J. Tang, H. Gao, and H. Liu, "mTrust: Discerning multi-faceted trust in a connected world," in *Proc. Int. Conf. Web Search Data Mining*, 2012, pp. 93–102.

[37] A. Davoudi, "Effects of user interactions on online social recommender systems," in *Proc. IEEE 33rd Int. Conf. Data Eng.*, 2017, pp. 1444–1448.

[38] P. Sun, L. Wu, and M. Wang, "Attentive recurrent social recommendation," in *Proc. 41st Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2018, pp. 185–194.

[39] H. Yin, L. Zou, Q. V. H. Nguyen, Z. Huang, and X. Zhou, "Joint event-partner recommendation in event-based social networks," in *Proc. IEEE 34th Int. Conf. Data Eng.*, 2018, pp. 929–940.

[40] L. Sun, X. Wang, Z. Wang, H. Zhao, and W. Zhu, "Social-aware video recommendation for online social groups," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 609–618, Mar. 2017.

[41] H. Yin, Q. Wang, K. Zheng, Z. Li, J. Yang, and X. Zhou, "Social influence-based group representation learning for group recommendation," in *Proc. IEEE 35th Int. Conf. Data Eng.*, 2019, pp. 566–577.

[42] C. Park, D. Kim, J. Oh, and H. Yu, "Do also-viewed products help user rating prediction?" in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 1113–1122.

[43] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, and L. Schmidt-Thieme, "Multi-relational matrix factorization using Bayesian personalized ranking for social network data," in *Proc. Int. Conf. Web Search Data Mining*, 2012, pp. 173–182.

[44] Z. Zhao, Z. Cheng, L. Hong, and E. H. Chi, "Improving user topic interest profiles by behavior factorization," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 1406–1416.

[45] A. Farseev, I. Samborskii, A. Filchenkov, and T.-S. Chua, "Cross-domain recommendation via clustering on multi-layer graphs," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2017, pp. 195–204.

[46] S. Kang, J. Hwang, D. Lee, and H. Yu, "Semi-supervised learning for cross-domain recommendation to cold-start users," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 1563–1572.

[47] Q. Do, W. Liu, J. Fan, and D. Tao, "Unveiling hidden implicit similarities for cross-domain recommendation," *IEEE Trans. Knowl. Data Eng.*, to be published.

[48] P. Sun, L. Wu, K. Zhang, Y. Fu, R. Hong, and M. Wang, "Dual learning for explainable recommendation: Towards unifying user preference prediction and review generation," in *Proc. Web Conf.*, 2020, pp. 837–847.

[49] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proc. AAAI Conf. Artif. Intell.*, 2014, pp. 1112–1119.

[50] Y. Lin, Z. Liu, X. Zhu, X. Zhu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. AAAI Conf. Artif. Intell.*, 2015, pp. 2181–2187.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.

[51] M. Nickel, V. Tresp, and H.-P. Kriegel, "A three-way model for collective learning on multi-relational data," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 809–816.
[52] X. Yu *et al.*, "Recommendation in heterogeneous information networks with implicit user feedback," in *Proc. ACM Conf. Recommender Syst.*, 2013, pp. 347–350.
[53] X. Yu *et al.*, "Personalized entity recommendation: A heterogeneous information network approach," in *Proc. 7th ACM Int. Conf. Web Search Data Mining*, 2014, pp. 283–292.
[54] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W. Y. Ma, "Collaborative knowledge base embedding for recommender systems," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 353–362.
[55] R. Catherine and W. Cohen, "Personalized recommendations using knowledge graphs: A probabilistic logic programming approach," in *Proc. ACM Conf. Recommender Syst.*, 2016, pp. 325–332.
[56] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 950–958.
[57] H. Wang, F. Zhang, M. Zhao, W. Li, X. Xie, and M. Guo, "Multi-task feature learning for knowledge graph enhanced recommendation," in *Proc. World Wide Web Conf.*, 2019, pp. 2000–2010.

Chen Gao received the BS degree in electronic engineering from Tsinghua University, Beijing, China, in 2016. He is currently working toward the PhD degree with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include recommender systems, data mining, and data privacy. His publications have appeared in several top-tier journals and conferences such as the *IEEE Transactions on Knowledge and Data Engineering*, ICDE, WWW, KDD, SIGIR, and IJCAI.

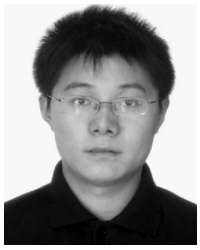
Chao Huang received the BS degree from the Electronic Engineering Department, Tsinghua University, Beijing, China, in 2019. He is currently working toward the MS degree with the Department of Computer Science, Columbia University, New York. His researches mainly focus on recommender systems and data mining.

Donghan Yu received the MS degree in electrical engineering from Tsinghua University, Beijing, China, in 2018. He is currently working toward the PhD degree with the Department of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania. His researches focus on data mining and machine learning.

Haohao Fu is currently working toward the undergraduate degree with the College of Letters and Science, UC Berkeley, Berkeley, California.

Tzh-Heng Lin received the MS degree in electrical engineering from Tsinghua University, Beijing, China, in 2018. He is currently working toward the MS degree with the Department of Computer Science, Peking University, Beijing, China. His researches focus on recommender systems and data mining.

Depeng Jin (Member, IEEE) received the BS and PhD degrees from Tsinghua University, Beijing, China, in 1995 and 1999, respectively, both in electronics engineering. Currently, he is an associate professor with Tsinghua University and vice chair of the Department of Electronic Engineering. He was awarded National Scientific and Technological Innovation Prize (Second Class), in 2002. His research fields include telecommunications, high-speed networks, ASIC design, and future internet architecture.



Yong Li (Senior Member, IEEE) received the BS degree in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2007, and the PhD degree in electronic engineering from Tsinghua University, Beijing, China, in 2012. He is currently a faculty member of the Department of Electronic Engineering, Tsinghua University. He has served as general chair, TPC chair, TPC member for several international workshops and conferences, and he is on the editorial board of two IEEE journals. His papers have total citations of more than 6,600. Among them, ten are ESI highly cited papers in computer science, and received four conference best paper (run-up) awards. He has received IEEE 2016 ComSoc Asia-Pacific Outstanding Young Researchers and Young Talent Program of China Association for Science and Technology.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.

Authorized licensed use limited to: Tsinghua University. Downloaded on July 31,2023 at 12:50:48 UTC from IEEE Xplore. Restrictions apply.