RESEARCH ARTICLE

Recommendation system in social networks with topical attention and probabilistic matrix factorization

Weiwei Zhang, Fangai Liu^{*}, Daomeng Xu, Lu Jiang

School of Information Science and Engineering, Shandong Normal University, Jinan, China

^{*} lfa@sdnu.edu.cn

OPEN ACCESS

Citation: Zhang W, Liu F, Xu D, Jiang L (2019) Recommendation system in social networks with topical attention and probabilistic matrix factorization. PLoS ONE 14(10): e0223967. <https://doi.org/10.1371/journal.pone.0223967>

Editor: Hua Wang, Victoria University, AUSTRALIA

Received: April 26, 2019

Accepted: October 2, 2019

Published: October 31, 2019

Copyright: © 2019 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was supported by the following grants: National Natural Science Foundation of China 61772321, Natural Science Foundation of Shandong Province ZR2016FP07, the Innovation Foundation of Science and Technology Development Center of Ministry of Education and CERNET, Innovation Foundation of Science and Technology Development Center of Ministry of Education and New H3C Group (2017A15047), and Innovation Project NGII20170508. The funders had

Abstract

Collaborative filtering (CF) is a common recommendation mechanism that relies on user-item ratings. However, the intrinsic sparsity of user-item rating data can be problematic in many domains and settings, limiting the ability to generate accurate predictions and effective recommendations. At present, most algorithms use two-valued trust relationship of social network to improve recommendation quality but fail to take into account the difference of trust intensity of each friend and user's comment information. To this end, the recommendation system within a social network adopts topical attention and probabilistic matrix factorization (STAPMF) is proposed. We combine the trust information in social networks and the topical information from review documents by proposing a novel algorithm combining probabilistic matrix factorization and attention-based recurrent neural networks to extract item underlying feature vectors, user's personal potential feature vectors, and user's social hidden feature vectors, which represent the features extracted from the user's trusted network. Using real-world datasets, we show a significant improvement in recommendation performance comparing with the prevailing state-of-the-art algorithms for social network-based recommendation.

Introduction

In daily life, with the continuous development of network information, information overload has become a serious challenge in an environment where users are overwhelmed, therefore, develop effective programs to help users locate information about their interests is coming to a creative and very important task that attracts the attention of research and application fields. For this purpose, recommended system (RS) are one of the important means of solving this problem. They help customers find what they are looking for and have been proven to drive sales and customer loyalty [1]. Collaborative filtering (CF) [2] is a common recommendation approach that has been adopted by many e-commerce sites, from Amazon [3] to Twitter [4] and YouTube [5], and is based on using neighbors to analyze the user's interests by collecting the past behavior data of a large number of users, and identifying the neighbor users who have the same interests as the target user to forecast the interest or degree of the intended user for a

no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

certain event. Recently, collaborative filtering algorithms [6–8] tend to suffer from the serious problem of the natural sparsity of the user-item ranking data, which dues to each user only having rated a small segment of the available items.

To solve the problem of sparse data, various scholars have proposed effective solutions, including the introduction of auxiliary data, integrated cross-domain information or more hidden rules in mining data [9]. For example, item features have been utilized to improve the recommendation algorithm to deal with the problem of data sparsity. Considering a combination of data from different information sources to obtain cross-domain recommendation results, implicit feedback information in the data is mined making the most of information provided by the user or the combination of social relationship information, etc.

The recommendation algorithm combined with social information is an effective solution method to solve the disadvantage of data sparsity. In the social network, the direct trust network [10] can be extracted from each user's friend lists, which can be available and is auxiliary data, solving the cold start disadvantage and enhancing the results of the recommendation. On one hand, in life, people usually consult with their friends before being making a decision and are affected by their acquaintances; for another, people also tend to be friends with similar people. Traditionally, we have only used the similarity of rating record calculations to measure the relationship between different users. Taking social networks into consideration, we can combine trust information and score records to prioritize the relationship between users to improve the quality of the recommendations. Although some previous studies have focused on auxiliary data, most of them use trust information directly and assume that if a user trusts his friend, he will like the same items as the people trusts [11–12]. In fact, users may not like these items, despite the fact that they are liked by people they trust, and therefore, the trusted person cannot influence target users in a certain probability.

More recently, some methods have taken text information data into account in addition to ratings [13–18]. Having investigated carefully, we have observed that the text information in a large proportion of recommended tasks can usually be divided into two categories: item specification [19–21] and user review [22–23]. An item specification called text information used to describe an item's properties or attributes. For example, in an article recommendation such as a paper on cellulite, it refers to the title and summary of the paper. In Amazon recommendations and other item recommendations, the recommendation refers to item specifications and technical specifications. The other type is user reviews written by users, explaining why they favor or detest products according to their experiences. However, although both types of textual data have been found helpful to recommendation tasks, there are some intrinsic defects for them.

The matrix factorization technique [24] is one of the most widely used methods for solving the data sparsity and imbalance problems. Based on the latent factor model of matrix factorization, the commonly used matrix factorization methods mainly include normalized SVD, whose interaction information is mapped into space. Probability matrix decomposition (PMF) is a low-dimensional matrix approximate decomposition model, usually assume that the user's interest is only affected by a few factors.

In summary, we comprehensively make full use of auxiliary information and propose an approach for a recommendation system in a social network with topical attention and probabilistic matrix factorization (STAPMF), which takes into account the influence of the trusted person and the target user's own comments and predicts the user's rating for items. The main dedications of framework are summarized as follows:

We introduce a novel recommendation model which is called STAPMF, which fuses attention-based recurrent neural networks to extract topical information from review documents and utilizes the trust information in social networks to learn feature vector.

We improved the MF model with social trust network and attention-based recurrent neural networks, which learn an adaptive factor that varies between users to weigh the impact of individuals and societies on decision-making. We believe that a definite person is not easily influenced by a trusted person, and vice versa.

The rest of the study is organized as follows. Section 2 reviews the related work. The model proposed in this study is presented in section 3. Section 4 shows the experiment results. Finally, the conclusion and future work are detailed in section 5.

Related work

In this section, we will introduce the literature related to our work from the following three aspects.

Probability matrix factorization

The motivation of the probability matrix factorization algorithm [25] is tantamount to adding the probability based on the matrix factorization model. Specifically, assuming that there are M users and an N item scoring matrix $R \in \mathbb{R}^{N \times M}$, our goal is to reconstruct the scoring matrix by finding the potential model of the user and the product. In the collaborative filtering algorithm, the probabilistic matrix factorization model is used to study the eigenvectors of the users and the items, and then, the feature vectors are used to predict the score. We know that the conditional probability of the existing scoring data is:

$$p(R|U, V, \sigma_R^2) = \prod_{i=1}^N \prod_{j=1}^M [N(R_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}} \quad (1)$$

where $N(x|\mu, \sigma_R^2)$ is the normal distribution with variance σ^2 and mean μ , and I_{ij} is the indicator function that is equal to 1 if u has rated i ; otherwise, it is equal to 0. The user's actual score matrix R and the prediction score have a Gaussian distribution with a mean value of zero. Assume that the user eigenvector matrix $U \in \mathbb{R}^{N \times M}$ and the commodity eigenvector matrix $V \in \mathbb{R}^{N \times M}$ obey the Gaussian prior distribution with a mean of 0 and finally assume that the user score is independent of the same distribution, subject to spherical Gaussian prior distribution:

$$p(U|\sigma_U^2) = \prod_{i=1}^N N(U_i|0, \sigma_U^2 I), p(V|\sigma_V^2) = \prod_{j=1}^M N(V_j|0, \sigma_V^2 I) \quad (2)$$

Then, we use the Bayesian derivation of the user and the product of the implicit eigenvector posterior probability to perform the maximum likelihood estimate:

$$p(U, V|R, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto p(R|U, V, \sigma_R^2) p(U|\sigma_U^2) p(V|\sigma_V^2) = \prod_{i=1}^N \prod_{j=1}^M [N(R_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}} \times \prod_{j=1}^M N(V_j|0, \sigma_V^2 I) \times \prod_{i=1}^N N(U_i|0, \sigma_U^2 I) \quad (3)$$

Fig 1 shows the graph of method. To prevent overfitting, the actual score is usually mapped to the $[0, 1]$ interval by $f(x) = (x-1)/(R_{max}-1)$, and the prediction score is mapped to the $[0, 1]$

Fig 1. Probabilistic matrix factorization diagram model.

<https://doi.org/10.1371/journal.pone.0223967.g001>

interval by $g(x) = 1/(1+\exp(x))$. The final objective function is as follows:

$$L = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_U}{2} \sum_{i=1}^N \|U_i\|^2 + \frac{\lambda_V}{2} \sum_{j=1}^M \|V_j\|^2 \quad (4)$$

Deep neural networks in natural language processing

Recently, the application of NLP has gradually become another popular topic in the research of deep learning. In 2013, with the rise of word vector word2vec [26], many studies on the distributed feature correlation of various words emerged. At the beginning of 2014, researchers used different DNN models, such as convolution networks, cyclic networks and recursive networks, which made great progress in traditional NLP applications, including part of speech tagging, effective analysis, and syntactic analysis. [27] proposed employing an attention mechanism to achieve the most advanced results in machine translation [28] by using the neural language model.

The depth of learning in the research field of natural language processing tasks has been a great success, and thus, in the field of recommender systems, it has also attracted broad interest [29–30]. For example, [31] proposed to utilize convolution neural network embedding matrix factorization to learn features. Bahdanau et al [32] proposed an important concept of neurological attention. It is a weighted summation technique, but it can automatically analyze which parts are more important, such as words in an image or words in a sentence. With this

attention mechanism, we can extract crucial words and sentences in the comment text and provide semantic explanations for the recommendations. Similarly, [33] utilized attention-based convolution neural networks to model review documents and obtained the most advanced results in the evaluation and prediction tasks. Wang et al. [34] proposed the DAMD model, using attention model to merge multiple prediction models for article recommendation.

Social networks

The booming development of social network enables us to get user-generated content information from the Internet, such as social relations, tags, comments, etc. The recommended content also develops in a more diversified direction, including items, friends.

Social network recommendation not only considers the relationship between users and items, but also adds social information between users, as shown in Fig 2A. The scoring prediction process integrates users' scoring information of items shown in Fig 2B. and users' trust, making the final scoring result more accurate.

Proposed method

In this part, we will detail our proposed STAPMF model. We first describe our attention-based recurrent neural network architecture for document modeling, followed by a method for extracting text features from users and item review documents. We then introduce user social latent features from social networks, and we extend the traditional probability matrix probabilistic model by combining textual regularization. Finally, the parameters are optimized.

Attention-based RNN for document modeling

We use a double-sided RNN with an attention mechanism to learn the features in the comment document. The model consists of four main components: (1) a word embedding layer (2) a sequence coding layer (3) a topical attention layer, and (4) a feature projection layer.

Word embedding layer. We initialize the word embedding layer with the pretrained word vector obtained from word2vec [35] and adjust it using backpropagation. The word

A User-item diagram representation

B User-item-rating matrix

Fig 2. The expression of trust social network recommendations. A. User-item diagram representation. B. User-item-rating matrix.
<https://doi.org/10.1371/journal.pone.0223967.g002>

embedding layer takes a series of words $(d_1, d_2, d_3 \dots d_N)$ as input and maps each word to its respective k -dimensional vector representation $x_i \in R_k$.

Sequence encoding layer. Context annotations with the input sequence are supplied by the sequence encoding layer. [36–37] proposed the bidirectional gated recurrent unit (GRU) architecture, which is a popular variant of the normal recurrent hidden unit. By combining a forget gate with an input gate to synthesize a single update gate, and mixing cell and hidden states, each recurrent unit is capable of encapsulating sequential dependencies across different time scales.

Formally, a GRU computes its activation at time step t as the linear interpolation between the antecedently activation h_{t-1} and the candidate act $\bar{h}_t$.

$$h_j^k = (1 - z_j^t)h_j^{t-1} + z_j^t\bar{h}_j^t \quad (5)$$

$$net_z^t = x^t W^z + h^{t-1} Y^z \quad (6)$$

$$z_j^t = \sigma(net_z^t)^j \quad (7)$$

$$net_h^t = x^t W + (r^t \odot h^{t-1}) Y \quad (8)$$

$$\bar{h}_j^t = \tanh(net_h^t)^j \quad (9)$$

$$net_r^t = x^t W^r + h^{t-1} Y^r \quad (10)$$

$$r_j^t = \sigma(net_r^t)^j \quad (11)$$

$$net_y^t = h^t W^y \quad (12)$$

$$y_j^t = \sigma(net_y^t)^j \quad (13)$$

where $\odot$ is the Hadamard product, which is the product of the corresponding element; the subscript indicates the index of the node, and t indicates the time. $W^y \in R^{hd \times yd}$ represents the parameter matrix of the hidden layer to the output layer and hd, yd is the amount of nodes in the hidden layer and the output layer, respectively. $W^z \in R^{xd \times hd}$, $Y^z \in R^{hd \times hd}$, respectively, represent the input matrix and the connection matrix of the hidden layer to the update gate z at the previous moment. Xd is the dimension of the input data. $W^r \in R^{xd \times hd}$, $Y^r \in R^{hd \times hd}$ are the input matrix and the connection matrix of the hidden layer to the reset gate r at the previous moment. $W \in R^{xd \times hd}$, $Y \in R^{hd \times hd}$ are the connection matrix of the input and the hidden layer to the selected state at the previous moment.

We want the annotations for each word to not only summarize the previous words but also summarize the words below. Hence, we employ a bidirectional GRU consisting of forward and backward GRUs. The activation functions of the forward GRU and the backward GRU at time step t are expressed by $\vec{h}_j^t$ and $\bar{h}_j^t$, respectively. At each time step, we connect forward activation and backward activation to obtain the final comment. That is $\vec{h}_j^t = [\vec{h}_j^t, \bar{h}_j^t]$.

Topical attention layer. We assume that not all parts of the document are related to a particular topic. Therefore, we introduce attention mechanisms to capture the relative importance of different words. Suppose that each dimension of the reviews final representation vector is

considered a topic related to the user's characteristics or the items characteristics, and the value of each dimension represents the strength of the topic. For a specific topic, each word in the review contributes differently to the topic, so we propose an attention weighting method, that is, each word learns an attention for each topic, and then the representation vector of the topic is weighted.

Consider, for example, the k -th attention module. Given the sequence of word annotations $(h_1, h_2, h_3 \dots h_n)$, the attention module first transforms each word annotation through a single layer perceptron with the tanh activation function:

$$q_t^k = \tanh(W_s^k h_t^s + b_s^k) \quad (14)$$

Then, the attention module compares the similarity between the context vector z_k and the transformed annotation by calculating the dot product and assigns a weighted score to each annotation using the softmax function:

$$a_t^k = \frac{z_k \cdot q_t^k}{\sum_{t=1}^n z_k \cdot q_t^k} \quad (15)$$

Then, the attention module weights the representative vector of the topic:

$$\bar{h}_k = \sum_{t=1}^n a_t^k h_t \quad (16)$$

Feature Projection Layer. After the representative vector of the topic is obtained, the projection layer obtains a value indicating the topic in the vector.

$$d_k = \tanh(W_c^k \bar{h}_k + b_c^k) \quad (17)$$

The representation vector of review is $d = [d_1, d_2 \dots dk]$, the representation vector of user reviews and item reviews is the average of the review representation vectors.

Extracting textual features from review documents

First, we define the comment text of user i as $d_{u,i}$, which is all of user's comments. Similarly, the relevant comment text of item j is $d_{v,j}$. The recursive neural network based on user comment document and item comment document is called user attention network and item attention network respectively. Considering the review document, we first use an attention-based recurrent neural network to generate potential document representations for each individual comment and then average them as text features extracted from the review document:

$$U_i = UAN(W, X_i) + \epsilon_i, \epsilon_i \sim N(0, \sigma_u^2 I) \quad (18)$$

$$V_j = IAN(Z, Y_j) + \epsilon_j, \epsilon_j \sim N(0, \sigma_v^2 I) \quad (19)$$

where UAN refers to the user attention network, the function denotes the textual features for user i generated by feeding the user review document $D_{u,i}$ into the user attention network with parameters W . where the function denotes the textual features for item j generated by feeding the item review document $D_{v,j}$ into the item attention network with parameters Z . IAN refers to the item attention network.

We suppose that the user and the item underlying factors are highly correlated with text features extracted from the review document. Therefore, we can get the prior probability

distribution of users and the potential factor vector of the items.

$$p(U|W, X, \sigma_U^2) = \prod_i^N N(U_i|UAN(W, X_i), \sigma_U^2 I) \quad (20)$$

$$p(V|Z, Y, \sigma_V^2) = \prod_j^M N(V_j|IAN(Z, Y_j), \sigma_V^2 I) \quad (21)$$

where U_i and V_j are the textual features extracted from the comment texts of user i and item j respectively.

User social latent feature

For the extraction of user characteristic vectors, this part takes the adjacency matrix called the trust matrix $T = [T_{il}]_{n \times m}$. From the perspective of reality, there exists an asymmetric matrix S , because the user relationship network is oriented, and the target user trusts the user does not mean that the user will trust the target user. For each pair of users, the trust value is 1, otherwise 0. The individual potential characteristics of user I are represented by U_i , while the social potential characteristics are represented by $S \in R^{m \times l}$. The potential feature vectors of each society obey the gaussian prior distribution with a mean value of 0, and the specific form is shown as follows

$$p(S|\sigma_s^2) = \prod_{i=1}^N N(S_i|0, \sigma_s^2 I) \quad (22)$$

By directly trusting the feature vector, the average feature vector U_i of target user i can be obtained. It can be represented as:

$$U = \frac{\sum_{l=1}^n T_{il} U_l}{\sum_{l=1}^n T_{il}} \quad (23)$$

At the same time, the trust matrix is normalized so that $\sum_{l=1}^n T_{il} = 1$. Now, we can rewrite the formula as:

$$\bar{U}_i = \sum_{l=1}^n T_{il} U_l \quad (24)$$

As is known, we can be influenced by trusted people, but not be exactly like them, but the characteristics of the people we trust are not completely transferred to the target users. So, T and U are not the same. Now, we calculate the conditional distribution of T given by the potential characteristics of the individual and the trust relationship as follows:

$$\begin{aligned} p(S|U, T, \sigma_s^2, \sigma_T^2) \\ \propto P(U, T|S, \sigma_T^2) \times P(S|\sigma_s^2) \\ = \prod_{i=1}^N \prod_{j=1}^M N(\sum_{l=1}^N S_{il} U_l | S_i, \sigma_T^2 I) \times \prod_{i=1}^N N(S_i | 0, \sigma_s^2 I) \end{aligned} \quad (25)$$

Now, we have an item with two different prediction scores based on individual characteristics and potential social characteristics of the target users. One needs to consider only the personal characteristics of the users regardless of the user's features users from trusted out of the man to predict. In real life, some people decide not to consider the opinions of people they

Fig 3. STAPMF model.

<https://doi.org/10.1371/journal.pone.0223967.g003>

trust when choosing an item. Even a trusted person can't easily change his decision. Others are easily influenced by trusted people and their opinions are easily changed. Therefore, the same user is affected by the potential characteristics of different societies, and only from the perspective of user trust is not comprehensive enough.

Adaptive weighting of each user's personal and social underlying characteristics. As shown in Fig 3, $\alpha \in R_{m \times k}$ is the influence factor, and $1 - \alpha_i$ is the influence degree of users. And assume that it obeys the gaussian prior distribution of one half mean

$$P(\alpha | \sigma_\alpha^2) = \prod_{i=1}^N N(\alpha_i | 1/2, \sigma_\alpha^2) \quad (26)$$

Optimization methodology

The user and the item potential factor vector U , the user social potential feature S , the item potential factor vector V and the firm factor α are synthesized to make better recommendations on the social network. According to the formulas (25) and (26), then, through Bayesian reasoning, the formula for the posterior distribution of the potential factors of the user and the product is given as follows:

$$\begin{aligned} & P(U, S, V, Z, W, \alpha | R, X, Y, \sigma_R^2, \sigma_U^2, \sigma_S^2, \sigma_V^2, \sigma_\alpha^2, \sigma_Z^2, \sigma_W^2) \\ & \propto P(R | U, S, V, \alpha, \sigma_R^2) \times P(S | U, S, \sigma_S^2) \\ & \times P(U | W, X, \sigma_U^2) \times P(V | Z, Y, \sigma_V^2) \\ & \times P(S | \sigma_S^2) \times P(W | \sigma_W^2) \times P(Z | \sigma_Z^2) P(\alpha | \sigma_\alpha^2) \\ & = \frac{1}{2} \prod_{i=1}^N \prod_{j=1}^M \left[I_{ij} (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S) i V_j^T) \right]^2 \\ & \times \prod_{i=1}^N N(\sum_{k=1}^N S_{ik} U_k | S_i, \sigma_S^2 I) \times \prod_{i=1}^N N(U_i | UAN(W, X_i, \sigma_U^2 I)) \\ & \times \prod_{i=1}^N N(S_i | 0, \sigma_S^2 I) \times \prod_{j=1}^M N(V_j | LAN(Z, Y_j, \sigma_V^2 I)) \times \prod_{i=1}^N N(\alpha_i | 1/2, \sigma_\alpha^2) \end{aligned} \quad (27)$$

Keep the super parameter σ^2 , σ_U^2 , σ_V^2 , σ_W^2 constant and maximize formula (27) to minimize the squared function:

$$\begin{aligned} L = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T)^2 \\ & + \frac{1}{2} \lambda_T \sum_{i=1}^N ((\sum_{k=1}^n T_{ik} U_k - S_i) (\sum_{k=1}^n T_{ik} U_k - S_i)^T) \\ & + \frac{1}{2} \lambda_U \sum_{i=1}^N \|U_i - UAN(W, X_i, \sigma_U^2 I)\|_F^2 + \frac{\lambda_U}{2} \sum_{i=1}^N S_i S_i^T \\ & + \frac{\lambda_V}{2} \sum_{j=1}^M \|V_j - IAN(Z, Y_j, \sigma_V^2 I)\|_F^2 + \frac{\lambda_Z}{2} \sum_{i=1}^N \alpha_i^2 + \frac{\lambda_W}{2} \sum_{a=1}^{|W|} \|W_a\|^2 + \frac{\lambda_Z}{2} \sum_{b=1}^{|Z|} \|Z_b\|^2 \end{aligned} \quad (28)$$

where $\lambda_U = \sigma^2 / \sigma_U^2$, $\lambda_V = \sigma^2 / \sigma_V^2$, $\lambda_S = \sigma^2 / \sigma_S^2$, $\lambda_\alpha = \sigma^2 / \sigma_\alpha^2$ and $\|\bullet\|_F$ denote the Frobenius norm, and the method of gradient descent is adopted to estimate parameters, and the formula is as follows:

$$\begin{aligned} \frac{\partial L}{\partial U_i} = & - \sum_{j=1}^M \alpha_i I_{ij}^R (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) \\ & \times g'((\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) V_j + \lambda_T \sum_{k=1}^n S_{ik} + \lambda_u U_i \end{aligned} \quad (29)$$

$$\begin{aligned} \frac{\partial L}{\partial S_i} = & - \sum_{j=1}^M (1 - \alpha_i) I_{ij}^R (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) \\ & \times g'((\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) V_j + \lambda_S \sum_{k=1}^n (S_{ik} - S_k) + \lambda_s S_i \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{\partial L}{\partial V_j} = & - \sum_{i=1}^N I_{ij}^R (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) \\ & \times g'((\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) (\alpha_i U_i + (1 - \alpha_i) S_i) + \lambda_v V_j \end{aligned} \quad (31)$$

$$\begin{aligned} \frac{\partial L}{\partial \alpha_i} = & - \sum_{j=1}^M I_{ij}^R (R_{ij} - g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) \\ & \times g'((\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T) (U_i - S_i) V_j^T + \lambda_\alpha \alpha_i \end{aligned} \quad (32)$$

When U and V are assumed to be constant, the goal of the network for users and products is to adjust its internal weights W and Z to make the extracted text features close to the target's U and V .

$$L_W(X_i, U_i) = \|U_i - UAN(W, X_i, \sigma_U^2 I)\|_F^2 \quad (33)$$

$$L_Z(Y_j, Z) = \|V_j - IAN(Z, Y_j, \sigma_V^2 I)\|_F^2 \quad (34)$$

While U and V are fitted with alternating least squares, W_U and W_V are optimized with

mini-batch gradient descent.

$$W := W - \eta \nabla E_p(W), Z := Z - \eta \nabla E_p(Z) \quad (35)$$

In the optimization process, U , V , T , α , W_U , and W_V , are alternately updated until the results converge, and the user can predict the unknown score of the project through the optimization process.

$$R_{i,j} \approx E[R_{i,j} | g(\alpha_i U_i + (1 - \alpha_i) S_i) V_j^T, \sigma^2] \quad (36)$$

Experiment

Datasets and evaluation protocol

To verify the validity of our model in rating prediction, we used two real Epinions and Ciao data sets, which contain both commentary and online social data collected and published by Tang et al. [38] in 2011. Epinions and Ciao are two well-known mass consumer reviews websites with major markets in the US and Europe. On both sites, any user can rate and comment on all items after signing up, as well as browse other users' ratings and reviews to help them make more favorable decisions. At the same time, users can establish a friendship relationship with trusted users to establish a social network. First of all, we randomly selected 80% of the data from the data set as training data, and the rest data were evenly divided into two parts, namely verification data and test data. The size of used datasets is presented in Table 1.

The evaluation index in the experiment is the root mean square error (RMSE). The smaller RMSE value, the better performance of the recommended algorithm. RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_{i,j}^{N,M} (R_{i,j} - \hat{R}_{i,j})^2}{R_C}} \quad (37)$$

where R_C is the test dataset, $R_{i,j}$ is the real value, and $\hat{R}_{i,j}$ is the predicted value.

In order to evaluate the recommendation accuracy of our algorithm, we compared the following algorithms.

The probability matrix factorization (PMF) [30] is a classical rating prediction model, but it uses rating information in the collaborative filtering recommendation, ignoring the impact of the relationship between the users and the text information.

The SoRec[39] algorithm adopts matrix decomposition method and assumes that recommendation system and social network share the same user implicit space. User implicit space is a matrix composed of multiple vectors, each of which is the user's implicit eigenvector

SoReg[40] was proposed by Ma et al. and uses the similarity of ratings as a measure of trust between users, and the main mode of communication in SocialMF is improved.

RSTE[41] proposes a matrix decomposition framework with social regularization and explains the differences between social recommendation system and trust perception recommendation system

TrustMF[42] was proposed Yao et al. to separately model the situation when the user is a trustee and a trusted person and integrates the users to obtain the recommended result.

Table 1. Data statistics on two real-world datasets.

Datasets	#users	#items	#ratings	#trust
Epinions	22166	296277	922267	355754
Ciao	7375	106797	284086	111781

<https://doi.org/10.1371/journal.pone.0223967.t001>

Recommendation based on social relations (SocialMF)[43] proposed integrating the user's social relations into the recommendation system, but this model project feature vector does not take into account the additional information of the item product.

Collaborative Topic Regression (CTR) [44] learns interpretable latent structures from user-generated content so that probabilistic topic modeling can be integrated into collaborative filtering.

Social relationships based on marginalized stacked Denoising Autoencoders(SDAE) [45] proposed a deep learning method to learn user preferences and the social influence of friends simultaneously when generating recommendations.

Experimental results and analysis

A set of experiments considered the influence of the characteristic dimension of experimental results; for the other parameters, we adjust the parameters in advance for each method, and the optimum value is used for all experiments. The table indicates that the potential feature dimension K of all algorithms is the RMSE value under different value conditions, and the dimension of the hidden feature vector K is set to 5, 10, and 20, the results are shown in Table 2.

The following conclusions are obtained by comparison: Our method STAPMF performs best in all cases. The social recommendation algorithms (such as SocialMF and SoReg) are superior to RMSE in the probability matrix factorization method PMF, which only relies on the user rating. This shows that making full use of the implicit user preference information in social relationships can effectively improve the precision of the recommendation algorithm. Collaborative topic regression learns interpretable potential feature vectors from user-generated content, which also suggests that leveraging user reviews and product information can help to improve recommendation algorithms. The algorithm shows good recommendation ability in Epinions and Ciao data sets, and verifies the reliability of this algorithm, and has no obvious error for specific data sets.

In this paper, STAPMF has a certain degree of performance degradation when the dimension is increased, but it still has obvious competitive advantages compared with other algorithms. Since the algorithm in the paper models each user from two dimensions, including (rating, truster), the dimension of each user is actually twice that of a single vector dimension, which enhances the fitting ability of the algorithm. It also increases the risk of overfitting. Table 1 also shows that the performance at dimension 5 is better than the other dimensions. Therefore, this algorithm is more suitable for adopting a lower vector dimension.

Table 2. RMSE comparisons for different K .

Algorithm	K = 5		K = 10		K = 20	
	Epinions	Ciao	Epinions	Ciao	Epinions	Ciao
PMF	1.1347	1.0483	1.1491	1.0438	1.1581	1.0455
SoRec	1.0758	1.0308	1.1047	1.0645	1.2260	1.1620
SoReg	1.0674	0.9924	1.1109	1.0011	1.2399	1.0970
RSTE	1.0915	1.0471	1.1387	1.0300	1.2291	1.1656
SocialMF	1.0809	1.0087	1.0904	1.0700	1.1932	1.0060
TrustMF	1.0722	0.9794	1.1208	0.9823	1.2893	0.9812
CTR	1.0644	0.9543	1.0576	0.9662	1.0916	0.9678
SDAE	1.0457	0.9321	1.0501	0.9525	1.0732	0.9603
STAPMF	1.0301	0.9131	1.0416	0.9341	1.0478	0.9517

<https://doi.org/10.1371/journal.pone.0223967.t002>

Fig 4. Validation RMSE as a result of varying d_w, d_z .

<https://doi.org/10.1371/journal.pone.0223967.g004>

We will explore the different settings of hyperparameters that affect the performance of our proposed model. The hyperparameters in the experiment include the dimension d_w, d_z of word embedding, the state dimension of the sequence encoder d_x, d_y , the dimension in the attention mechanism module d_A . On the Ciao data set, λ_u is set to 0.1, 1, 10, 30, 40, 100, respectively, and λ_v is set to 1, 10, 30 respectively. The parameters λ_u set on the Epinions data set are 0.01, 0.1, 1, 10 respectively, λ_v are 10, 30, 100 respectively.

According to the experimental results, as shown in Figs 4 and 5, we set the embedding dimension to 256, the sequence encoder state dimension to 128, and the dimension of the conversion annotation in the attention mechanism module to 128. We observe λ_u, λ_v . The effect of the recommendation is significant, and the effect is shown in Fig 6. In these two data sets, when λ_u and λ_v take the value of 100, the effect is the best. When the value of λ_u increases continuously, the user vector is not updated. The user implicit feature is projected into the project implicit feature vector space mainly through Describe the characteristics obtained by the file.

Fig 5. Validation RMSE as a result of varying d_x, d_y .

<https://doi.org/10.1371/journal.pone.0223967.g005>

Fig 6. Validation RMSE as a result of varying λ_u , λ_v .

<https://doi.org/10.1371/journal.pone.0223967.g006>

In our STAPMF approach, the parameter λ_T is the degree to which the trusted regular term influences the objective function (25). However, the maximum value of the value will lead to the social trust information to dominate the prediction model, which may limit the accuracy of the prediction. In this part, we use different sizes of training data and analyze the influence of different parameter values on the algorithm results. The influence of λ_T value between 0 and 1 on prediction accuracy is studied. For the other parameters, we use the grid search method to figure out the best combination of values. And the default is $k = 5$ and $k = 10$, and $\lambda_u = \lambda_v = 0.1$. Figs 7 and 8 compare the RMSE values of the model are different. As presented in the results, as the RMSE value increases, the RMSE value begins to decrease. When a threshold value is exceeded, the RMSE value increases again.

Another challenge in the recommendation system is cold start. In order to verify the efficiency of our method in cold start, we conducted experiments on two data sets and compared

Fig 7. Parameter analysis of Ciao.

<https://doi.org/10.1371/journal.pone.0223967.g007>

Fig 8. Parameter analysis of Epinions.

<https://doi.org/10.1371/journal.pone.0223967.g008>

Fig 9. Comparison results on the different datasets.

<https://doi.org/10.1371/journal.pone.0223967.g009>

them with other methods. Fig 9 shows the RMSE results for the three datasets used only for cold-start users. It can be seen that compared with other algorithms, the proposed STAPMF algorithm can provide the most precise rating forecast for cold start users. The main reason is not only considering the influence of trust network, but also the individual features of each user are adaptive to balance item and social features.

Conclusions

Although the recommended algorithm achieved great success in a large number of practical applications, the recommendation system is still subject to issues such as data sparsity, imbalance and cold start, and the user's decision on the Internet item is influenced by its own characteristics and the recommendation of a trusted friend. A local attention probability matrix factorization (STAPMF) method based on social network is proposed. Our STAPMF method includes three stages of learning. Firstly, we exploit a topical attention model to learn the latent

feature vectors of users and items. Secondly, we take into consideration the influence of the community impact in the trusted social network. Finally, the proposed objective function is minimized with the users' characteristics and the user social latent feature vectors. Unlike the most existing models, our model leverages the user review information and the impact of the trust networks and sets the impact factors. Our experimental analysis of the two datasets, Ciao and Epinions, shows that our method outperforms traditional recommendation algorithms and recommendation algorithms based on social networks. In addition, the proposed algorithm performs better than the existing algorithms in handling cold start.

The trust relationship of users in the social trust network is constant while the trust relationship can change with time. In addition, ratings and comments from users are also time sensitive, and outdated comments may become noise information for recommendations. Therefore, the time sensitivity can be incorporated into the trust social network in future works.

Supporting information

S1 File. The proposed STAPMF model and experimental results.
(DOC)

S1 Dataset. Data set used in the manuscript.
(ZIP)

Acknowledgments

This work was supported by the following grants: National Natural Science Foundation of China 61772321, the Innovation Foundation of Science and Technology Development Center of Ministry of Education and CERNET Innovation Project NGII20170508.

Author Contributions

Data curation: Weiwei Zhang.

Formal analysis: Weiwei Zhang.

Investigation: Daomeng Xu, Lu Jiang.

Project administration: Fangai Liu.

Writing – original draft: Weiwei Zhang.

Writing – review & editing: Weiwei Zhang, Fangai Liu, Daomeng Xu, Lu Jiang.

References

1. Peng M, Zeng G, Sun Z, Huang J, Wang H, & Tian G. Personalized app recommendation based on app permissions. *World Wide Web*.2018; 21(1):89–104.
2. Sarwar B., Karypis G., Konstan J., & Riedl J. Item-based collaborative filtering recommendation algorithms. *WWW*. 2001; 1: 285–295.
3. Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*. 2003;(1): 76–80.
4. Elmongui, H. G., Mansour, R., Morsy, H., Khater, S., El-Sharkasy, A., & Ibrahim, R. TRUPI: Twitter recommendation based on users' personal interests, *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, Cham.2015: 272–284.
5. Davidson, J., Livingston, B., Sampath, D., Liebald, B., Liu, J., & Nandy, P., et al. The YouTube video recommendation system, *Proceedings of the fourth ACM conference on Recommender systems*. ACM.2010: 293–296.

6. Xing S., Liu F., Zhao X., & Li T. Points-of-interest recommendation based on convolution matrix factorization. *Applied intelligence*. 2018; 48(8):2458–2469.
7. Zhang, J., Lin, Z., Xiao, B., & Zhang, C. An optimized item-based collaborative filtering recommendation algorithm. In *Network Infrastructure and Digital Content*, 2009. IC-NIDC 2009. IEEE International Conference. 2009: 414–418.
8. Sa L. Collaborative filtering recommendation algorithm based on cloud model clustering of multi-indicators item evaluation, 2011 International Conference on Business Computing and Global Informatization. IEEE. 2011: 645–648.
9. Min P., Qianqian X., Hua W., Yanchun Z., & Gang T. Bayesian sparse topical coding, *IEEE Transactions on Knowledge and Data Engineering*. 2018; 31(6): 1080–1093.
10. Tang, J., Gao, H., Hu, X., & Liu, H. Exploiting homophily effect for trust prediction. In *Proceedings of the sixth ACM international conference on Web search and data mining*. 2013: 53–62.
11. Zhang J., & Curley S. P. Exploring explanation effects on consumers' trust in online recommender agents. *International Journal of Human–Computer Interaction*. 2018; 34(5): 421–432.
12. Choudhary N, Bharadwaj K K. Leveraging Trust Behaviour of Users for Group Recommender Systems in Social Networks, *Integrated Intelligent Computing, Communication and Security*. Springer, Singapore. 2019: 41–47.
13. Almahairi, A., Kastner, K., Cho, K., & Courville, A. Learning distributed representations from reviews for collaborative filtering. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 2015: 147–154.
14. Ling, G., Lyu, M. R., & King, I. Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*. 2014: 105–112.
15. Peng M., Zhu J., Wang H., Li X., Zhang Y., Zhang X., & Tian G. Mining event-oriented topics in microblog stream with unsupervised multi-view hierarchical embedding. *ACM Transactions on Knowledge Discovery from Data (TKDD)*. 2018; 12(3): 38.
16. Ren, Z., Liang, S., Li, P., Wang, S., & de Rijke, M. Social collaborative viewpoint regression with explainable recommendations. In *Proceedings of the tenth ACM international conference on web search and data mining*. 2017: 485–494.
17. Lu, Y., Dong, R., & Smyth, B. Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. 2018: 773–782.
18. Zheng, L., Noroozi, V., & Yu, P. S. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 2017: 425–434.
19. Peng, M., Xie, Q., Zhang, Y., Wang, H., Zhang, X., Huang, J., & Tian, G. Neural sparse topical coding, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018: 2332–2340.
20. Peng, M., Chen, D., Xie, Q., Zhang, Y., Wang, H., Hu, G., . . . & Zhang, Y. Topic-net conversation model, *International Conference on Web Information Systems Engineering*. Springer, Cham. 2018: 483–496.
21. Wang, H., Wang, N., & Yeung, D. Y. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2015: 1235–1244.
22. Wang, H., Xingjian, S. H. I., & Yeung, D. Y. Collaborative recurrent autoencoder: Recommend while learning to fill in the blanks. In *Advances in Neural Information Processing Systems*. 2016: 415–423.
23. Xu, Y., Lam, W., & Lin, T. Collaborative filtering incorporating review text and co-clusters of hidden user communities and item groups. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. 2014: 251–260.
24. Xu, Y., Shi, B., Tian, W., & Lam, W. A unified model for unsupervised opinion spamming detection incorporating text generality. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015: 725–731.
25. Mnih A, Salakhutdinov R R. Probabilistic matrix factorization, *Advances in neural information processing systems*. 2008: 1257–1264.
26. Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
27. Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. Convolutional sequence to sequence learning, *Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. Org*. 2017: 1243–1252.

28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. et al. Attention is all you need, *Advances in neural information processing systems*. 2017: 5998–6008.
29. Xing S., Wang Q., Zhao X., & Li T. A hierarchical attention model for rating prediction by leveraging user and product reviews. *Neurocomputing*. 2019; 332: 417–427.
30. Sainath T N, Kingsbury B, Sindhwani V, et al. Low-rank matrix factorization for deep neural network training with high-dimensional output targets, 2013 IEEE international conference on acoustics, speech and signal processing. IEEE.2013: 6655–6659.
31. Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering, *Proceedings of the 24th international conference on Machine learning*. ACM. 2007: 791–798.
32. D. Bahdanau, K. Cho, Y. Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, 2014, ArXiv:1409.0473.
33. Kim, D., Park, C., Oh, J., Lee, S., & Yu, H. Convolutional matrix factorization for document context-aware recommendation, *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM. 2016: 233–240.
34. X. Wang, L. Yu, K. Ren, G. Tao, W. Zhang, Dynamic attention deep model for article recommendation by learning human editors demonstration, in: *Proceedings of the 23th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017: 2051–2059.
35. Seo, S., Huang, J., Yang, H., & Liu, Y. Interpretable convolutional neural networks with dual local and global attention for review rating prediction. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 2017: 297–305.
36. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 2013: 3111–3119.
37. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. *arXiv preprint arXiv:1406.1078*, 2014.
38. Tang, J., Gao, H., & Liu, H. mTrust: discerning multi-faceted trust in a connected world. In *Proceedings of the fifth ACM international conference on Web search and data mining*.2012:93–102.
39. Ma, H., Yang, H., Lyu, M. R., & King, I. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*. 2008:931–940.
40. Ma, H., Zhou, D., Liu, C., Lyu, M. R., & King, I. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*.2011:287–296.
41. Ma, H., King, I., & Lyu, M. R. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*.2009:203–210.
42. Jamali, M., & Ester, M. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*.2010:135–142.
43. Yao, W., He, J., Huang, G., & Zhang, Y. Modeling dual role preferences for trust-aware recommendation. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*.2014:975–978.
44. Wang, C., & Blei, D. M. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*.2011:448–456.
45. Rafailidis, D., & Crestani, F. Recommendation with Social Relationships via Deep Learning. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*.2017:151–158.