ENHANCE MATRIX FACTORIZATION BASED RECOMMENDER SYSTEM WITH SIDE INFORMATION

by

HUAN ZHAO

Department of Computer Science and Engineering
The Hong Kong University of Science and Technology
Supervised by Prof. Dik Lun Lee

June 2017, Hong Kong

ENHANCE MATRIX FACTORIZATION BASED RECOMMENDER SYSTEM WITH SIDE INFORMATION

by

HUAN ZHAO

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Matrix Factorization (MF) has become one of the most popular methods for collaborative filtering in modern recommender systems (RS). Based on the assumption that users' preferences to items are controlled by a small number of latent factors, the large user-item rating matrix can be decomposed into two smaller matrices representing user-specific and item-specific latent factors, respectively. In other words, the rating matrix is of *low rank*. Despite the success of MF-based methods, the sparsity and cold-start problems are still challenging. On the other hand, with the availability of rich side information in modern RS, like the meta-data of items, or social connections among users, a lot of previous research work are using the rich side information to improve the RS. Therefore, to address the problems facing MF, the rich side information can be very helpful when they are exploited properly. Thus, in this survey, we review existing works in the literature, which enhance the matrix factorization with the help of rich side information. Specifically, we elaborate three kinds of side information: 1) social connections among users; 2) text information; 3) geographical information. Note that there is also a Bayesian derivation for MF, called Probabilistic Matrix Factorization (PMF). All the methods in this survey are based on MF or PMF. Through this survey, we can see that the recommending performances of MF can be significantly improved with the help of side information.

0.1 Introduction

With proliferating of all kinds of online services, like news, products, musics, etc., the information overload problem becomes more and more severe, which means it is more and more difficult for us to obtain the needed information. To tackle this problem, recommender systems (RS) have become an indispensable tool for people to get interesting items more quickly and conveniently. The core idea underlying RS is to automatically provide people with items based on their preferences, which can be inferred from their past behaviors. For example, Netflix can recommend movies to users according to those they liked before, and Yelp can recommend restaurants to users based on the ones users have visited recently. Moreover, RS can create tremendous business values. On Netflix, 2/3 of the movies watched are recommended and 35% of sales on Amazon are from recommendations.[1]

In fact, recommender systems are everywhere nowadays, for example, news, books, products, music, videos, restaurants, jobs and even dating partners can be obtained based on the recommender engines, which try to provide us with those aligning with our preferences. Technically speaking, recommender systems seek to predict users' preferences to items according to users' behaviors in the past, thus providing users with what they like. Typically, users' behaviors can be categorized into two kinds. One is explicit, like the rating given to a movie by user while the other is implicit, like the webpages visited or books read.

Among various recommender methods, collaborative filtering (CF) has been one of the most popular ones, which tries to predict users' ratings (or preferences) on unseen items based on similar users or items. In the literature, the approaches of CF usually fall into two categories: memory-based and model-based ones. Besides CF, other recommendation methods include content-based, demographic-based, etc. However, in this survey, we mainly focus on the CF-based methods, specifically the matrix factorization (MF) based one. Due to its scalability and good prediction performance, MF has been the most popular CF method in the past ten years. However, they still suffer from the sparsity and cold start problems. In the literature, a lot of work have been done to address these problems, one direction of which is to use the rich side information to enhance matrix factorization. In the following sections, we will introduce in detail the CF-based methods, the problems facing it, and enhanced MF-based methods with rich side information.

The remaining of this survey is organized as follows: we introduce collaborative filtering as well as matrix factorization in Section 0.2, where the evaluation metrics and problems of MF are also introduced. Then we elaborate in Sections 0.3, 0.4, and 0.5 the methods to incorporate separately three kinds of side information, i.e. social connections, text and geographical information. Further, ensemble of these three kinds of side information are introduced in Section 0.6. Finally, we summarize this survey in Section 0.7.

[1] http://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers

0.2 Collaborative Filtering

In this section, we elaborate the most popular recommendation approach: collaborative filtering. In reality, there tends to be a number of users who give ratings to a number of items, and then we need to predict the ratings of users giving to unseen items. Formally, the problem can be defined as: Given a set of users $U = \{u_1, u_2, ..., u_m\}$ and a set of items $V = \{v_1, v_2, ..., v_n\}$ and a rating matrix $\mathbf{R} \in \mathbb{R}^{m \times n}$ with some entries missing. If a user u_i rates an item v_j, then $\mathbf{R}_{ij}$ represents the rate value, usually in the range $[1, 5]$, and $\mathbf{R}_{ij} = 0$ otherwise. See Table 0.2 1 for a toy example. In reality, users tend to rate only a small number of items, leading to large number of missing entries in $\mathbf{R}$. And in the recommendation literature, we always assume that the missing ratings to be zero. Then the task of collaborative filtering is to predict the missing values in the rating matrix $\mathbf{R}$ based on the observed ones.

	V_1	V_2	V_3	V_4	V_5
U_1	**1**	**3**	**4**	?	?
U_2	?	**4**	?	**5**	?
U_3	?	?	?	?	?
U_4	**2**	**2**	**3**	**4**	?
U_5	?	**4**	?	**3**	?

Table 1: User-Item Matrix

0.2.1 Two Types of Approaches

To solve the problem above there are two kinds of approaches: memory-based and model-based collaborative filtering.

Memory-based For the memory-based ones, they always relate to neighbor-based ones. Usually they first compute a set of similar entities, users or items, and then predict the ratings of users to items based on these similar entities. Then there are two types of methods: users-based and item-based collaborative filtering.

For user-based CF, the core idea can be explained as two steps:

1. For a given user u_i, select those who share similar rating patterns with him/her.
2. Use the ratings of the similar users in step 1 to compute the predictions of the user u_i.

For item-based CF, the core idea can be explained as two steps:

1. Build an item-item similarity matrix based on the ratings each item obtained.
2. To predict the rating of a user u_i giving to an item v_j, compute the rating based on the similarities between v_j and those rated by u_i in the rating matrix.

Despite the simplicity of the memory-based CF methods, they also suffer the problems sparsity because in reality there tends to be a small number of users who give ratings to a small number of items, leading to the sparseness of the rating matrix. To address the sparsity problem, model-based methods are proposed, among which the matrix factorization based ones stand out for its good prediction accuracy and scalability.

Model-based Model-based methods, on the other hand, fit a parametric model to the training data that can later be used to predict unseen ratings and issue recommendations. Model-based methods include cluster-based CF [36, 40, 28], Bayesian classifiers [25, 26], and regression-based methods [37]. The slope-one method [13] fits a linear model to the rating matrix, achieving fast computation and reasonable accuracy. A recent class of successful CF models are based on low-rank matrix factorization. The regularized SVD method [3, 29] factorizes the rating matrix into a product of two low rank matrices (user-specific and item-specific latent features) that are used to estimate the missing entries. An alternative method is Non-negative Matrix Factorization (NMF) [12] that differs in that it constrain the low rank matrices forming the factorization to have non-negative entries. Recent variations are Probabilistic Matrix Factorization (PMF) [27], Bayesian PMF [32], Non-linear PMF [11], Maximum Margin Matrix Factorization (MMMF) [6, 34, 31], and Nonlinear Principal Component Analysis (NPCA) [43].

0.2.2 Matrix Factorization

Due to its good performance and scalability, Matrix Factorization (MF) has become the most popular method for CF. MF is based on the assumption that users' preferences to items are controlled by a small number of latent factors. Thus, the large user-item rating matrix can be decomposed into two smaller matrices representing user-specific and item-specific latent factors, respectively. In other words, the rating matrix is of *low rank*.

Based on the assumption that a user's rating behavior is governed by a small number of latent factors, the rating matrix $\mathbf{R} \in \mathbb{R}^{m \times n}$ can be approximated by the product of l-rank factors,

$$\hat{\mathbf{R}} \approx \mathbf{U}\mathbf{V}^{\top}, \tag{0.1}$$

where $\mathbf{U} \in \mathbb{R}^{m \times l}$ and $\mathbf{V} \in \mathbb{R}^{n \times l}$ with $l \ll min(m, n)$, representing, respectively, the latent features of users' preferences and items. The approximating process can be completed by solving the following optimizing problem:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \sum_{(i,j) \in \Omega} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^{\top})^2, \tag{0.2}$$

where $\mathbf{U}_i$ and $\mathbf{V}_j$ are the ith and jth rows of $\mathbf{U}$ and $\mathbf{V}$, representing the latent feature vectors of the user u_i and the item v_j, respectively. $||\cdot||_F$ denotes the Frobenius norm, and Ω is the index set of the observations in $\mathbf{R}$, and $\mathbf{R}_{ij}$ is the observed rating of the user u_i to the item v_j. In order to avoid overfitting, two regularization terms, $\frac{\lambda}{2}(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2)$, are added to Equation (0.2). In the literature, this method is also termed Regularized SVD (RegSVD) [29]. The complete formula can be rewritten as:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \sum_{(i,j) \in \Omega} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^{\top})^2 + \frac{\lambda}{2} \left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2 \right). \tag{0.3}$$

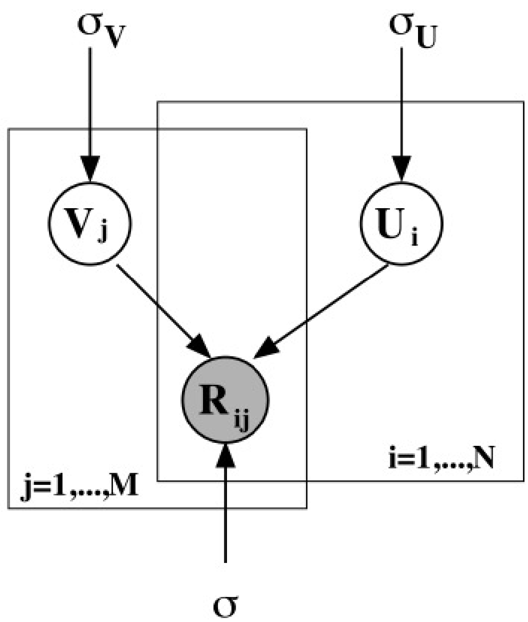


Figure 1: The graphical representation for Probabilistic Matrix Factorization.

In [27], Ruslan et.al. give a Bayesian generative model for matrix factorization, as shown in Figure 1. In Figure 1, suppose we have M items, N users, and integer rating values from 1 to K. Let R_ij represent the rating of user i for item j, $U \in \mathbb{R}^{N\times D}$ and $V \in \mathbb{R}^{M\times D}$ be latent user and item feature matrices, with row vectors $\mathbf{U}_i$ and $\mathbf{V}_j$ representing user-specific and item-specific latent feature vectors respectively. Then in Figure 1, they define the conditional distribution over the observed ratings as

$$p(R|U,V,\sigma^2) = \prod_{i=1}^{N}\prod_{j=1}^{M}\Big[\mathcal{N}(R_{ij}|U_iV_j^T,\sigma^2)\Big], \tag{0.4}$$

where $\mathcal{N}(R_{ij}|U_iV_j^T,\sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2, and I_{ij} is the indicator function with $I_{ij}=1$ if user i rated item j and 0 otherwise. Besides, zero-mean spherical Gaussian priors are placed on user and item latent feature vectors, then the log of the posterior distribution over the user and item features is given by:

$$\begin{aligned} ln(p(U,V|R,\sigma^2,\sigma_U^2,\sigma_V^2)) = &- \frac{1}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij}(R_{ij}-U_iV_j^{\top})^2 - \frac{1}{2\sigma_U^2}||U||_F^2 - \frac{1}{2\sigma_V^2}||V||_F^2 \\ &- \frac{1}{2}\bigg((\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}) \ln\sigma^2 + NDln\sigma_U^2 + MD\ln\sigma_V^2 \bigg) + C, \end{aligned} \tag{0.5}$$

where C is a constant that does not depend on the parameters. Then maximizing the log of posterior distribution over the user and item features is equivalent to minimizing the sum-of-squared-erros with regularization in Eq. (0.3).

0.2.3 Evaluation Metrics

To test the effectiveness of matrix factorization, we need to evaluate the accuracy of rating prediction in the user-item rating matrix. In the literature, there are two popular metrics, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), which are defined as:

$$\text{MAE} = \frac{\sum_{(i,j)\in\bar{\Omega}} |\mathbf{R}_{ij} - \widehat{\mathbf{R}_{ij}}|}{|\bar{\Omega}|},$$

$$\text{RMSE} = \sqrt{\frac{\sum_{(i,j)\in\bar{\Omega}} (\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij})^2}{|\bar{\Omega}|}},$$

where $\bar{\Omega}$ is the set of all user-item pairs (i, j) in the test set, and $\mathbf{R}_{ij}$ is the corresponding rating predicted by algorithms. In practice, the lower values of RMSE or MAE, the better performance the model obtains.

0.2.4 Challenges

Due to the good performance of matrix factorization, there are two problem facing it. The first one is the sparsity of the rating matrix. In reality, the sparseness of the rating matrix tend to be greater than 99%. It will be difficult for a further better performance when the rating matrix is so sparse. Secondly, there are a lot of users who give only a small number of ratings, say less than 5, thus it will be impossible to compute a accurate latent vectors with such a small number of ratings. To alleviate these two problems, auxiliary information can be used. For example, if we can obtain the reviews written by users, we can make use of text analysis tools to uncover the specific interests of users, leading to a better recommender system. Another example is that users tend to turn to their close friends for recommendation, thus we can recommend the items to cold start users, which are liked by their close friends. In the literature, they are also called side information, with which we can address these two problems facing matrix factorization. In the following sections, we will review previous work incorporating three kinds of side information separately and then some work incorporating them together.

0.3 Social Enhanced MF

With the developments of online social networks, like Facebook, Twitter, more and more research have been done to enhance the recommender system with the help of online social connections. The key part of RS is to compute users' preferences, and our preferences are inherently influenced by our friends. According to the social homphily theory [24] that people with similar interests tend to be connected as friends, thus it will be very useful to take into consideration users' social connections when modeling users' preferences. Based on this, a serials of work in the literature are proposed to incorporate social connections into the matrix factorization framework, which helps improve the recommendation performances.

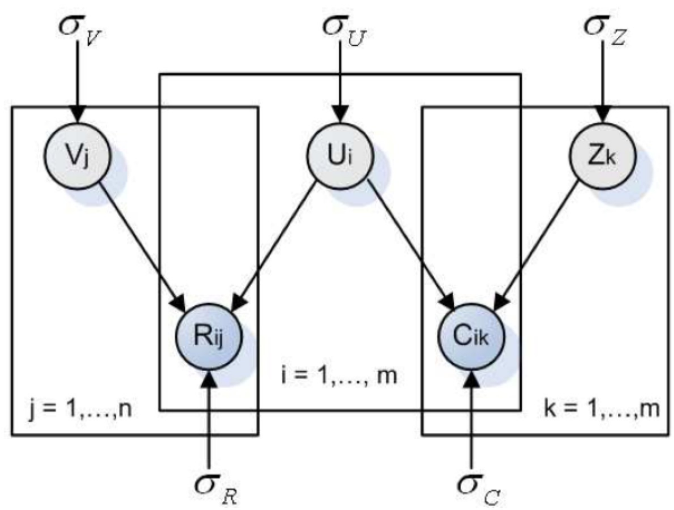


Figure 2: The graphical representation for the co-factorization framework.

In this survey, we focus on enhanced matrix factorization with various side information, among which social connections stands out. Specifically speaking, we can classify the previous works in the literature into four aspects in terms of social connections: 1) friendship; 2) trust; 3) social circles; 4) social influences. They represent the degree to which we can be influenced by those we are connected online. Friendship means Facebook-style connections, i.e. we need to know each other and the connection is bidirectional. While for the trust in the social networks, it tends to represent how much a person like or trust others, and the connection is Twitter-style, i.e. unidirectional. For the third one, we know that people tend to form groups because of similar interests, like the football club or reading group. This can be an explicit indicator of similar interests. Finally, social influences corresponds to a fact that renowned people can influence people not only their friends but also friends of friends. It means their influences can propagate through the social networks. An example is that nowadays companies are willing to pay a lot of money to a person with large number of followers in Twitter to help to promote their products, say writing a tweet related to their products. These four kinds of social information can play different roles when designing a modern recommender system. In the remaining part of this section, we will review works in the literature to process them. Note that, our focus is the matrix factorization framework, thus we only introduce those incorporating social connections into RegSVD [29] or PMF [27].

First, in terms of friendships, most of the existing work are based on social homophily and assume that our preferences can be affected by our friends. Thus, the latent vectors obtained from the matrix factorization need to be adjusted to adapt to this influences. In [21], Ma et.al. proposed a co-factorization framework to factorize the rating matrix and social matrix together by assuming the user-specific latent features share in these two matrices. Specifically, in Figure 2, we can see that a latent variable $\mathbf{Z}$ is added to decide the observed social connections together with the shared variable $\mathbf{U}$. $\mathbf{C}$ is the observed social connections with $\mathbf{C}_{ik} = 1$ if users u_i and u_k are friends and $\mathbf{C}_{ik} = 0$ otherwise. Similar to PMF, gaussian prior is placed to $\mathbf{Z}$ as well. Then by maximizing the log of posterior distribution, which is equivalent to minimizing the following:

$$\mathcal{L}(\mathbf{R}, \mathbf{C}, \mathbf{U}, \mathbf{V}, \mathbf{Z}) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{I}_{ij}^{R}(\mathbf{R}_{ij} - \mathbf{U}_i\mathbf{V}_j^\top)^2 + \frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{k=1}^{m}\mathbf{I}_{ik}^{C}(\mathbf{C}_{ik} - \mathbf{U}_i\mathbf{Z}_k^\top)^2 \\ + \frac{\lambda_U}{2}||\mathbf{U}||_F^2 + \frac{\lambda_V}{2}||\mathbf{V}||_F^2 + \frac{\lambda_Z}{2}||\mathbf{Z}||_F^2. \tag{0.6}$$

From Figure 2 and Eq. (0.6), we can see that the key idea is that it assumes a low-rank structure of the social matrix, and propose a joint model to factorize the rating matrix and social matrix together, thus it is called co-factorization framework. Further, Ma et.al. [22] propose a regularization framework to incorporate the friendships, which assumes that a user's should be closer to his or her friends in the latent space. It means when factorizing the rating matrix, the latent vectors of users should be adjusted according to their friends. Specifically, two regularization terms are designed as follows:

$$\sum_{i}\sum_{u_k \in N(u_i)} ||\mathbf{U}_i - \mathbf{S}_{ik}\mathbf{U}_k||^2, \tag{0.7}$$

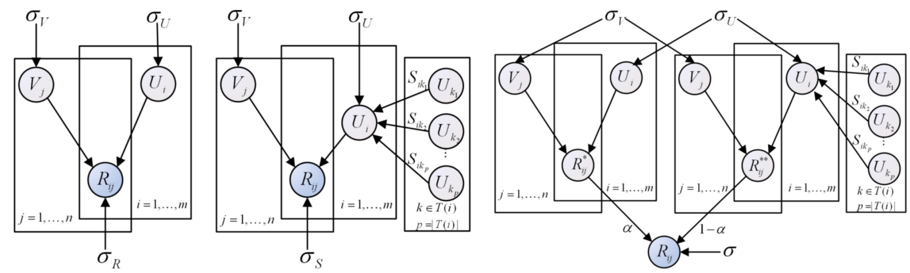
$$\sum_{i} ||\mathbf{U}_i - \sum_{u_k \in N(u_i)} \mathbf{S}_{ik}\mathbf{U}_k||^2, \tag{0.8}$$

where $N(u_i)$ is the set representing all the friends of the user u_i. $\mathbf{U}_i$ and $\mathbf{U}_k$ represent the latent vectors of user u_i and user u_k, respectively. $\mathbf{S}_{ik}$ represents the similarity between u_i and u_k computed based on the rating or social matrix. Then for Eq. (0.7), it requires the more similar two users, the closer their latent vectors should be in the factorization process. And for Eq. (0.8), it requires that the latent vectors of a user u_i should be closer to the average of his or her friends'. By adding the regularization terms into the matrix factorization framework, we then have:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\sum_{(i,j)\in\Omega}(\mathbf{R}_{ij} - \mathbf{U}_i\mathbf{V}_j^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\right) + \frac{\beta}{2}\Phi(\mathbf{U}), \tag{0.9}$$

where $\Phi(\mathbf{U})$ is Eq. (0.8) or (0.7). Note that both of Eq. (0.6) and (0.9) can be solved by gradient descent methods. These two are the typical ways to incorporate friendships into the matrix factorization framework, and both of them have performed experiments on Facebook-style datasets to demonstrate the efficacy of the models comparing the RegSVD and PMF.

Secondly, besides the bidirectional social relations, the unidirectional ones, i.e. twitter-style, are also very common in the social networks nowadays. The difference is that bidirectional connections tend to be built between two people know each other, while for unidirectional ones, there is no this constrain. For example, we can choose anyone to follow in twitter as long as we want to. Friends can follow each other, but usually we can follow those who are famous but don't know us. Thus it can be regarded as an explicit indicator of their interests in, or trust to the followees. A typical example is that a football fan can follow Messi, a famous football player, on Twitter to keep in touch with all his updates. Therefore, comparing to the friendships, it is also very important to take this unidirectional relations into consideration when designing recommender system. We call this trust-based recommender system. In [19], Ma et.al. first introduce the trust



(a) Factorization of User-Item Matrix (b) Recommendations by Trusted Friends (c) Recommendations with Social Trust Ensemble

Figure 3: Recommendation with Social Trust Ensemble (RSTE)

relations into the matrix factorization framework by proposing a social trust ensemble method. They assume that users' preferences can be influenced by those he trust and their decisions tend to be the balance between his or her own tastes(See Figure 3(a)) and those he or she trusts. (See Figure 3(b)). Therefore, an ensemble method is designed to capture these two factors as shown in Figure 3(c). Note that Ma et.al. design their framework based on PMF and the log of posterior can be written as follows:

$$\ln p(U,V|R,S,\sigma^2,\sigma_U^2,\sigma_V^2) = -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}(R_{ij} - g\big(\alpha U_i V_j^\top + (1-\alpha)\sum_{k\in N(i)} S_{ik}U_k V_j^\top\big))^2 - \frac{1}{2\sigma_U^2}\sum_{i=1}^{m} U_i U_i^\top - \frac{1}{2\sigma_U^2}\sum_{i=1}^{m} V_i^\top V_i + C, \tag{0.10}$$

where C is constant that does not depend on the parameters. Maximizing the log-posterior over two latent features is equivalent to minimizing the following sum-of-squared-errors objective functions with regularization terms:

$$\mathcal{L}(\mathbf{R},\mathbf{S},\mathbf{U},\mathbf{V}) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{I}_{ij}^{\mathbf{R}}(\mathbf{R}_{ij} - g\big(\alpha\mathbf{U}_i\mathbf{V}_j^\top + (1-\alpha)\sum_{k\in N(i)}\mathbf{S}_{ik}\mathbf{U}_k\mathbf{V}_j^\top\big))^2 + \frac{\lambda_U}{2}||\mathbf{U}||_F^2 + \frac{\lambda_V}{2}||\mathbf{V}||_F^2. \tag{0.11}$$

where α control the balances between users' own preferences and the influences from those they trust. Similarly, Eq. (0.11) can be solved by gradient descent. Besides trust, Ma et.al. also explore the distrust connections in the matrix factorization [20]. Based on the assumption that users' latent vectors should be closer to those they trust and farther to those they distrust, a similar regularization framework to Eq. (0.9) is designed as follows:

$$\min_{\mathbf{U},\mathbf{V}}\frac{1}{2}\sum_{(i,j)\in\Omega}(\mathbf{R}_{ij} - \mathbf{U}_i\mathbf{V}_j^\top)^2 + \frac{\lambda}{2}\big(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\big) - \frac{\beta}{2}\sum_{i}\sum_{u_k\in D(u_i)}\mathbf{T}_{ik}||\mathbf{U}_i - \mathbf{U}_k||_F^2, \tag{0.12}$$

where $D(u_i)$ is the set of users u_i distrusts and T_{ik} is the degree of trust. Similarly, trust

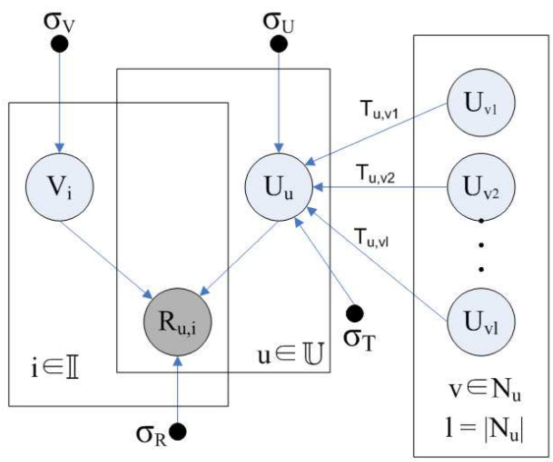


Figure 4: The graphical representation for the SocialMF model.

can also be incorporated as regularization terms as follows:

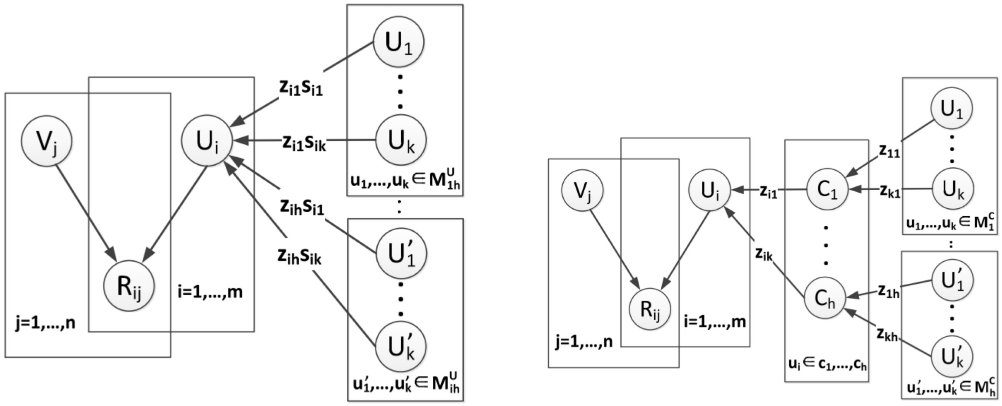
$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \sum_{(i,j)\in\Omega} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\right) + \frac{\beta}{2} \sum_i \sum_{u_k \in T(u_i)} \mathbf{S}_{ik} ||\mathbf{U}_i - \mathbf{U}_k||_F^2, \tag{0.13}$$

where $T(u_i)$ is the set of users u_i distrusts and S_{ik} is the degree of trust. The experimental results show that the distrust connections is at least as important as the trust ones. In [7, 8], Jamali et.al. further proposed a model, called SocialMFa, to make use of trust propagation based on the matrix factorization, which assumed that the users' latent features are dependent on those he trust, as shown in Figure 4. Here Jamali et.al. built the SocialMF based on the PMF framework, then we can also derive the log posterior of the model in Figure 4. Similarly, maximizing the log-posterior is equivalent to minimizing the following objective function:

$$\min_{\mathbf{U},\mathbf{V}} \frac{1}{2} \sum_{(i,j)\in\Omega} (\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\right) + \frac{\beta}{2} \sum_i \sum_{u_k \in N(u_i)} ||\mathbf{U}_i - \mathbf{S}_{ik}\mathbf{U}_k||_F^2, \tag{0.14}$$

where $N(u_i)$ is the set of users a user u_i trusts, and $\mathbf{S}_{ik}$ denotes the trust value of the user u_i to u_k.

Thirdly, apart from the friendship and trust, circles, i.e. social groups or communities, can also be indicators of users' preferences. It's very common that people participate in different online groups because of different interests, like reading group or football clubs. In [42], Yang et.al. proposed the circle-based recommender system, which is based on an assumption that people tend to trust different people regarding different domains. This is a more fine-grained work to incorporate social trust into the matrix factorization. To achieve this, they design algorithms to infer the category-specific circles and three methods to infer the trust value in the circles. And they train a category-specific matrix factorization with category-specific trust regularization terms. The objective function is



(a) MFC. (b) MFC++.

Figure 5: The graphical representation for the overlapping community regularized MF.

as follows:

$$\min_{\mathbf{U}^{(c)},\mathbf{V}^{(c)}} \frac{1}{2} \sum_{(i,j)\in\Omega} (\mathbf{R}_{ij}^{(c)} - \mathbf{U}_i^{(c)}(\mathbf{V}_j^{(c)})^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}^{(c)}||_F^2 + ||\mathbf{V}^{(c)}||_F^2\right)$$
$$+ \frac{\beta}{2}\sum_i \sum_{u_k\in N(u_i)} ||\mathbf{U}_i^{(c)} - \mathbf{S}_{ik}^{(c)}\mathbf{U}_k^{(c)}||^2, \qquad (0.15)$$

where the superscript c represents the variables in the category c, i.e. $\mathbf{U}^{(c)}$ denotes the latent features of users in the category c, and $\mathbf{S}_{ik}^{(c)}$ denotes the trust value between the user u_i and u_k in the category c. Note that this framework is similar to SocialMF [7] but it deals with the social trust in a more fine-grained way. The experimental results show the circle-based MF outperforms SocialMF, which demonstrates the usefulness of circles. A similar work is also proposed in [45], which first use the Latent Dirichlet Allocation [4] to discovery the communities and apply the matrix factorization to each community. These two work try to factorize the rating matrix based on the social circles, while in [14], Li et.al. proposed to use the communities to regularized the latent vectors of users in the factorization process. Li also built his model based on the fact that users tend to join in different groups out of different interests and assumes that users' preferences are influenced by those who are in the same communities with them. As shown in Figure 5(a) and (b), these are the two proposed methods to incorporate the overlapping communities to PMF. In Figure 5(a), the assumption is that the latent vectors of a user is influenced by the people in all the communities he or she belongs to. The larger interest a user is in one community, the larger influences people in that community will exert on him or her. In Figure 5(b), latent vector C representing the community c is used and the users' latent vectors are influenced by all the latent vectors of the communities he or she belongs to.

The objective function for these two methods are in the following:

$$\min_{\mathbf{U}^{(c)},\mathbf{V}^{(c)}} \frac{1}{2} \sum_{(i,j)\in\Omega} (\mathbf{R}_{ij}^{(c)} - \mathbf{U}_i^{(c)}(\mathbf{V}_j^{(c)})^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}^{(c)}||_F^2 + ||\mathbf{V}^{(c)}||_F^2\right)$$
$$+ \frac{\lambda_Z}{2} \sum_i \sum_{h=1}^{l} \mathbf{I}_{ih}^Z \mathbf{Z}_{ih} \sum_{u_k \in M_{ih}^U} \mathbf{S}_{ik} ||\mathbf{U}_i - \mathbf{U}_k||_F^2, \quad (0.16)$$

and

$$\min_{\mathbf{U}^{(c)},\mathbf{V}^{(c)}} \frac{1}{2} \sum_{(i,j)\in\Omega} (\mathbf{R}_{ij}^{(c)} - \mathbf{U}_i^{(c)}(\mathbf{V}_j^{(c)})^\top)^2 + \frac{\lambda}{2}\left(||\mathbf{U}^{(c)}||_F^2 + ||\mathbf{V}^{(c)}||_F^2\right)$$
$$+ \frac{\lambda_Z}{2} \sum_i \sum_{h=1}^{l} \mathbf{I}_{ih}^Z \mathbf{Z}_{ih} \left|\left| \mathbf{U}_i - \frac{\sum_{u_k \in M_h^C} \mathbf{Z}_{kh} \mathbf{U}_k}{\sum_{u_k \in M_h^C} \mathbf{Z}_{kh}} \right|\right|_F^2, \quad (0.17)$$

where M_{ih}^U contains the users in the same community c_h as user u_i, and $\mathbf{Z}_{ih}$ represents the interest of user u_i to the community c_h, which can be computed based on overlapping community detection methods, like the BIGCLAM [41].

In all the above methods, they mainly focus on users' local friends or those in the same communities. An important kind of social connection is ignored, which is that everyone in the social networks can be influenced indirectly by those with high reputation. Based on this fact, Tang et.al. [35] proposed a joint model to capture both the influences from local friends as well as those with high reputation. To capture the local friendship, they use the following equation:

$$\min \sum_{i=1}^{m} \sum_{u_k \in N(u_i)} \left(\mathbf{S}_{ik} - \mathbf{U}_i \mathbf{H} \mathbf{U}_k^\top\right), \quad (0.18)$$

where $\mathbf{S}_{ik}$ denotes the similarity between user u_i and u_k based on their ratings. $\mathbf{U}_i$ and $\mathbf{U}_k$ represent the feature vector of user u_i and u_k, respectively. $\mathbf{H} \in \mathbb{R}^{K\times K}$ is the matrix to capture the user preference correlation. A larger $\mathbf{S}_{ik}$ denotes a strong connection between u_i and u_k, thus their latent vectors are tightly correlated via $\mathbf{H}$. Eq. (0.18) can be rewritten in matrix form as follows:

$$\min ||\mathbf{T} \odot (\mathbf{S} - \mathbf{U}\mathbf{H}\mathbf{U}^\top)||_F^2, \quad (0.19)$$

where $\odot$ is the Hadamard product, i.e. element-wise multiplication. $\mathbf{T}_{ij} = 1$ if u_i and u_j are friends and 0 otherwise. And for the global influences, they first perform PageRank to rank users according their PageRank values, and then a reputation score w_i for user u_i is defined as:

$$w_i = f(i) = \frac{1}{1 + \log(r_i)}, \quad (0.20)$$

where r_i the rank of users, i.e. $r_i = 1$ denotes that u_i has the highest reputation according to the PageRank value. Then they use the following equation to capture the global social context:

$$\min \sum_{\langle u_i, v_j\rangle \in \mathcal{O}} w_i(\mathbf{R}_{ij} - \mathbf{U}_i \mathbf{V}_j^\top), \quad (0.21)$$

where it means that the higher reputation of a user u_i, the more important of his/her ratings, then it will force $\mathbf{U}_i\mathbf{V}_j^\top$ to tightly fit the rating $\mathbf{R}_{ij}$. It can also be rewritten as

$$\min ||\mathbf{W} \odot (\mathbf{R} - \mathbf{U}\mathbf{V}^\top)||_F^2. \tag{0.22}$$

Then taking Eq. (0.19) and (0.22) together, we have the objective function as follows:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{H}} ||\mathbf{W}\odot(\mathbf{R}-\mathbf{U}\mathbf{V}^\top)||_F^2+\alpha||\mathbf{T}\odot(\mathbf{S}-\mathbf{U}\mathbf{H}\mathbf{U}^\top)||_F^2+\lambda\left(||\mathbf{U}||_F^2+||\mathbf{V}||_F^2+||\mathbf{H}||_F^2\right). \tag{0.23}$$

Eq. (0.23) can also be solved by the gradient descent method.

In summary, all the existing work in the literature try to operate the latent vectors of users in the matrix factorization framework, no matter through the regularization, like [22], or co-factorization [21], or ensemble [19]. Based on the social homophily theory that users with similar interests tend to be connected as social friends, the social information is then incorporated to enhance the MF methods. In the following section, we show another important side information, the text, which can also be used to operate on the latent vectors of items.

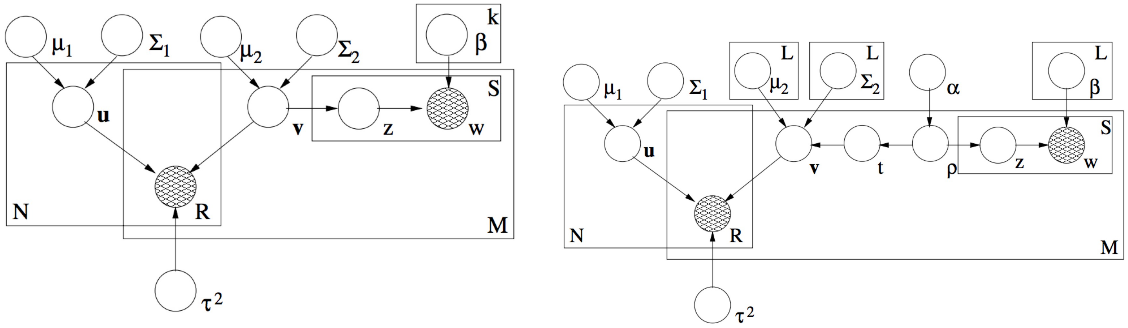
0.4 Text Enhanced MF

In modern recommender system, text information have become everywhere, like the descriptions of products, or reviews written by the users, or the items are text themselves, i.e. news or books. Then it will be very useful when the text information is modeled properly because it can provide informative indicators for users' interests through texts. In the literature, Latent Dirichlet Allocation (LDA) [4] has become a state-of-art text analysis tool, thus it is natural for researchers to integrate the LDA with MF-methods. Moreover, both of them are latent-feature based models, thus in the literature, they tend to be connected through the latent variables.

In [33], Shan et.al. proposed a parametric probabilistic matrix factorization framework (PPMF) to generalize the PMF [27] and BPMF[32] with different priors, and then they further explore to model the meta-data of items under PPMF with the help of correlated topic models (CTM) [10] and LDA. As shown in Figure 6(a) and (b), CTM and LDA are combine with PPMF and mixture probabilistic matrix factorization (MPMF). Note that the difference between PPMF and MPMF lies in that $\mathbf{u}$ and $\mathbf{v}$ are generated from a single Gaussian distribution in PPMF while MPMF further generalize PPMF by allowing $\mathbf{u}$ and $\mathbf{v}$ be generated from a mixture of Gaussians. For the detail of PPMF and MPMF, we refer the reader to [33].

The main idea in CTM-PPMF is as follows: For each itemm, $\mathbf{v}$ not only serves as PPMFs latent feature vector for its ratings, but also serves as CTMs membership vector over topics (after logistic transformation) for its side information. Therefore the common $\mathbf{v}$ for both the ratings and side information of movie becomes the glue to combine PPMF and CTM together.

The main idea in LDA-MPMF is as follows: For MPMF on the rating matrix, each item has a Dirichlet-discrete prior $\alpha \to \rho_{\mathbf{j}}$. Meanwhile, if we use LDA on side information,



(a) CTM-PPMF (b) LDA-MPMF

Figure 6: The graphical representations in [33].

each movie also needs a Dirichlet-discrete prior to generate the topics for the words in side information. Therefore, letting MPMF and LDA share the Dirichlet discrete prior, we can combine MPMF and LDA together.

Experimental results show the improvements of the prediction accuracy, which demonstrates the efficacy of combining PMF with topic models.

In [1], Agarwal et.al. propose a novel model, fLDA, which specifically explores the MF with LDA by by regularizing both user and item factors simultaneously through user features and the bag of words associated with each item. The main idea of fLDA is shown in Figure 7. The experimental results not only show the efficacy of fLDA, but also

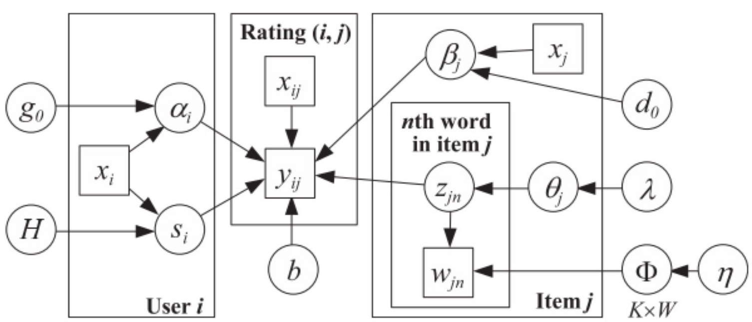


Figure 7: The graphical representation for fLDA.

bring a by-product, which is that fLDA can identify interesting topics that explain users' behaviors.

Then in [38], Wang et.al. work on scientific articles recommendation and propose a collaborative topic regression (CTR) which combines matrix factorization with topic model. Specifically, they integrate item latent vectors $\mathbf{v}_j$ in MF with item topics θ_j in LDA by adding an Gaussian offset $\epsilon_j \sim \mathcal{N}(\theta_j, \lambda_v^- 1 I_K)$, i.e. let $\mathbf{v}_j = \theta_j + \epsilon_j$. The main idea is shown in Figure 8 and we can see that it is the combination of PMF and LDA through the latent vectors of item. Due to the text property of the scientific articles, PMF and LDA can be naturally combined.

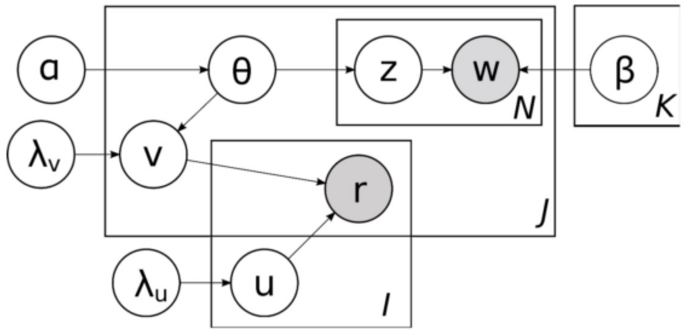
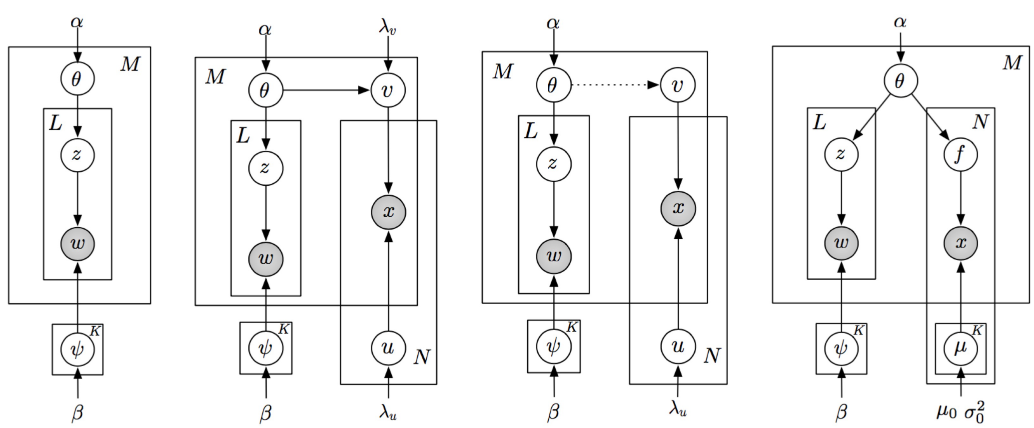


Figure 8: The graphical representation for CTR [38].

In [23], another important text information is used to combine MF with LDA. McAuley et.al. propose a novel model HFT to combine the rating dimensions, learned by MF, with latent review topics, learned by LDA, which can help uncover the specific aspects of users' tastes. Moreover, it can help identify representative reviews as well as discover genres attached to the items from the reviews. Their final model is based on the idea that the factors $\mathbf{v}_j$ should accurately model users ratings but also that the review corpus should be likely when these factors are transformed into topics. To achieve this, the define the objective of a corpus $\mathcal{T}$ (ratings and reviews) as:

$$f(\mathcal{T}|\Theta, \Phi, K, \mathbf{z}) = \sum_{r_{i,j} \in \mathcal{T}} (\hat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}) - \mu l(\mathcal{T}|\theta, \phi, z). \tag{0.24}$$

From Eq. (0.24), we can see that it is actually a joint model that optimize the error of rating prediction and likelihood of the reviews.



(a) LDA (b) CTR (c) HFT (d) RMR

Figure 9: The graphical representations for LDA [4], CTR [38], HFT [23], and RWR [16].

Then in [16], Ling et.al. extend the idea of HFT [23] and propose a novel model Rating Meets Reviews (RMR) to combine content-based recommendation with matrix factorization, which applies LDA to the review texts and aligns the topics with the rating

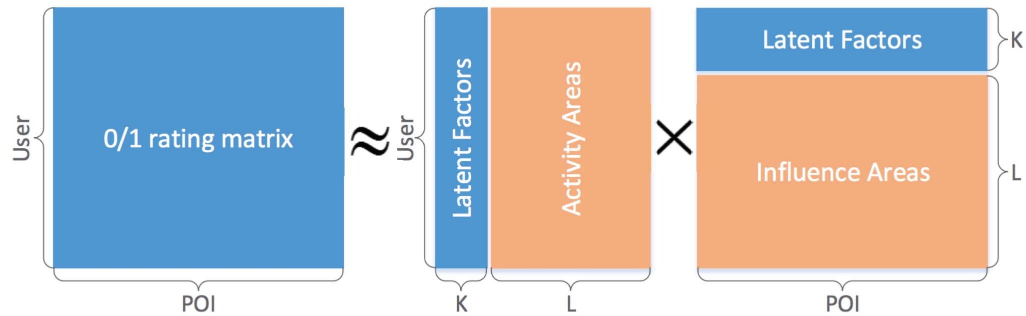


Figure 10: The illustration of GeoMF [15].

dimensions to improve the rating prediction accuracies. Ling et.al. also compare the RMR with previous HFT and CTR as shown in Figure 9. We can see that they all try to operate on the latent vectors of items from MF and connect them with the topics obtain by LDA from the texts, meta-data, content, or reviews, attached with items. But the differences lie in their methods to connect these two latent variables, say latent vectors of items from MF and topics obtained by LDA.

In summary, due to the existence of rich text informations in recommending scenarios, a lot of previous work in the literature try to enhance recommendation with the help of text analysis. Matrix factorization is one of the state-of-art methods as well as a latent model for RS, thus it is very natural to combine MF and topic model tools, like LDA, which will not only improve the prediction accuracy in MF-based collaborative filtering but also provide a possibility to explain the recommending results based on the topics from the texts.

0.5 Geographical Enhanced MF

In recent years, with the development of mobile devices, more and more location-based services come out, like Yelp or Foursquare, which users can check-in at some place in the online platforms and then these platforms can recommend places-of-interest (POI) according to the users' past check-ins. It corresponds to a new recommendation problem, i.e. the POI recommendation. Besides, due to the availability of users' locations, many traditional recommender system can also improve their recommending performance with the help of users' locations. For example, a news mobile APP can recommend news related to where the users work and live. Therefore, the location information can also be regarded as one important kind of side information. In this part, we review some existing work based on the MF to incorporate the location.

In [15], Lian et.al. we propose a novel framework GeoMF to jointly model geographical facts and matrix factorization in the POIs recommending scenarios. As shown in Figure 10, besides the latent factors in the MF framework, they design two more latent vectors appending to the latent factors by MF, which are called User Activity Areas and POI's Influences Areas. By assuming that the areas are obtained by splitting the whole world into L even grids, denoted as $L = \{g_1, g_2, ..., g_L\}$, the User Activity Areas is represented by a non-negative vector $\mathbf{x} \in \mathbb{R}_+^L$ and each entry x_l in the vector $\mathbf{x}$ indicates the possibility that this user will appear in the grid $g_l \in L$.

For the POI's Influence Areas, they consist of a collection of pairs of a grid area $l \in L$ to which the influence of this POI can be propagated, and each entry in the vector $\mathbf{y} \in \mathbb{R}_+^L$, is a non-negative real value $v_j \in \mathbb{R}_+$, indicating the quantity of influence from this POI.

Then the objective function is formulated as follows:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{X}} ||\mathbf{W} \odot (\mathbf{R} - \mathbf{U}\mathbf{V}^\top - \mathbf{X}\mathbf{Y}^\top)||_F^2 + \gamma(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2) + \lambda||\mathbf{X}||_1$$
$$\text{s.t. } \mathbf{X} \geq 0, \tag{0.25}$$

where $||\mathbf{X}||_1$ is ℓ_1 norm of matrix $\mathbf{X}$, which induces sparsity within $\mathbf{X}$, due to the fact that users tend to appear in a small number of locations, like home, work place, etc.

In [17], they propose a novel geographical probabilistic factor analysis framework which strategically takes various factors into consideration. Specifically, this framework allows to capture the geographical influences on a users check-in behavior. Also, the user mobility behaviors can be effectively exploited in the recommendation model.

Note that in most cases, the location, or geographical information, are used together with other side information, like the text or social. In the following parts, we review some existing work in the literature which model at least two of the three side information, i.e. social connections, text, geographical information. The framework is shown in Figure 11.

0.6 Ensemble of Social, Text and Geographical Information

In the previous three sections, we introduce separately the existing work on incorporating three kinds of side information into the matrix factorization: social connections, text and geographical information. Nowadays, with the popularity of mobile devices, many platforms can obtain these three kins of side information at the same time. Then it will further improve the recommending performance when we model as more as we can these side information comparing to utilizing only one kind of them.

In [30], Purushotham et.al. extend the CTR [38] to further integrate CTR with the social connections and propose a generalized hierarchical Bayesian model which jointly learns the user, item and social factor latent spaces. They user LDA [4] to capture items' content information in latent topic space, and use matrix factorization to derive latent feature space of user from his social network graph. It can be seen that CTR model and social matrix factorization (SMF) [21] are fused to obtain a consistent and compact feature representation. The graphical representation can be seen in Figure Then the objective function is also a joint optimizing objective function, which maximizes the log-posterior,

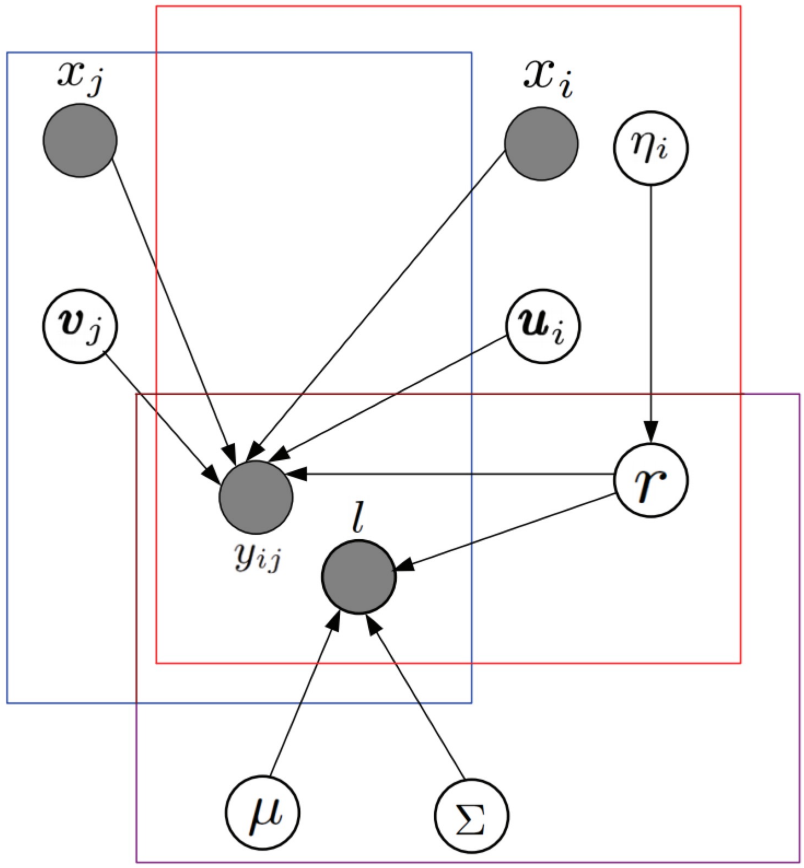


Figure 11: Proposed Model in [17].

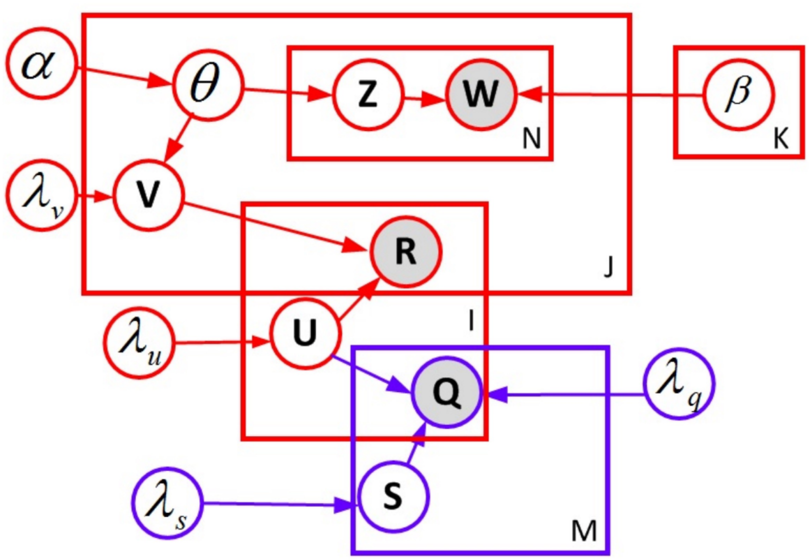


Figure 12: Proposed Model - CTR with SMF, CTR part shown in red color, Social Matrix Factorization (SMF) shown in blue color.

in the following:

$$\mathcal{L} = -\frac{\lambda_U}{2}\sum_i u_i^\top u_i - \frac{\lambda_V}{2}\sum_j (v_j-\theta_j)^T(v_j-\theta_j) + \sum_j\sum_n \log\Big(\sum_k \theta_{jk}\beta_{k,w_{j,n}}\Big)$$
$$-\sum_{ij}\frac{c_{ij}}{2}(\mathbf{R}_{ij}-u_i^Tv_j)^2 - \frac{\lambda_Q}{2}\sum_{i,m}\frac{d_{im}}{2}(q_{im}-u_i^Ts_m)^2 - \frac{\lambda_S}{2}\sum_k s_k^Ts_k, \qquad (0.26)$$

where $\lambda_U = \sigma_R^2/\sigma_U^2$, $\lambda_S = \sigma_R^2/\sigma_S^2, \lambda_Q = \sigma_R^2/\sigma_Q^2$. And to solve this optimizing problem, they use gradient ascent approach by iteratively optimizing the collaborative filtering and social network variables u_i, v_j, s_m and topic proportions θ_j.

A similar idea is proposed in [39]. However, instead of a joint model like that in [30], Wang et.al. use the regularization framework to constrain the items with a link in the CTR part as shown in Figure 13.

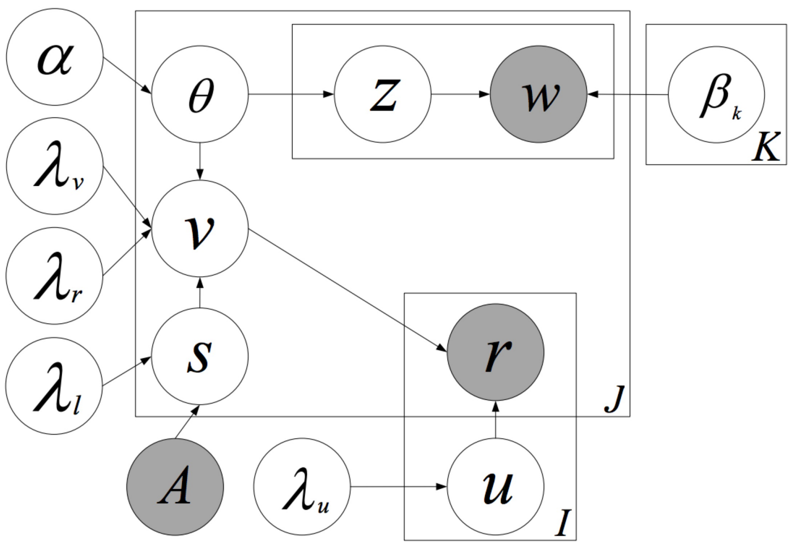


Figure 13: Graphical representation for CTR with Social Regularization (SR)(CTR-SR).

Both of the experimental results of [30] and [39] shows the effectiveness of these two extensions of CTR by incorporating the social connections.

The above two work explore the integration of social connections and text information, and in the following, several work on combining social connections and geographical information are reviewed. In [5], Chen et.al. propose a fused MF-based model to capture simultaneously the geographical influences and the social connections between users.

0.7 Summary

In this survey, we review the existing works, which try to improve the performance of matrix factorization with the help of rich side information. Specifically, we focus on three kinds of side information: social connections, text and geographical informations. Due to the fact that matrix factorization are latent factor methods, by which we obtain user-specific and item-specific latent features, or representations. The sparsity and cold-start problems impairs the accuracies of the learned latent representations of users and items, for which the side information can help.

Through all the reviewed works, we can see that all of them are operating on the latent vectors of users and items. With the availability of different kinds of side information, users' interests or items' properties can be modeled in a more fine-grained way, which adjusts the latent vectors in the factorizing process. Besides, some work incorporating more than one side information out of these three are also reviewed, and the performances are further improved. Therefore, we can know that the more side information we can incorporate, the better performance we can obtain.

REFERENCES

[1] D. Agarwal and B.-C. Chen. flda: matrix factorization through latent dirichlet allocation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 91–100. ACM, 2010.

[2] Y. Bao, H. Fang, and J. Zhang. Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In *AAAI*, volume 14, pages 2–8, 2014.

[3] D. Billsus and M. J. Pazzani. Learning collaborative information filters. In *Proceedings of the Fifteenth International Conference on Machine Learning*, volume 98, pages 46–54, 1998.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[5] C. Cheng, H. Yang, I. King, and M. R. Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Aaai*, volume 12, page 1, 2012.

[6] D. DeCoste. Collaborative prediction using ensembles of maximum margin matrix factorizations. In *Proceedings of the 23rd international conference on Machine learning*, pages 249–256. ACM, 2006.

[7] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 135–142. ACM, 2010.

[8] M. Jamali and M. Ester. A transitivity aware matrix factorization model for recommendation in social networks. In *Twenty-Second International Joint Conference on Artificial Intelligence*, 2011.

[9] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 426–434, 2008.

[10] J. D. Lafferty and D. M. Blei. Correlated topic models. In *Advances in neural information processing systems*, pages 147–154, 2006.

[11] N. D. Lawrence and R. Urtasun. Non-linear matrix factorization with gaussian processes. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 601–608. ACM, 2009.

[12] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[13] D. Lemire and A. Maclachlan. Slope one predictors for online rating-based collaborative filtering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*, pages 471–475. SIAM, 2005.

[14] H. Li, D. Wu, W. Tang, and N. Mamoulis. Overlapping community regularization for rating prediction in social recommender systems. In *ACM Conference on Recommender Systems*, pages 27–34, 2015.

[15] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 831–840. ACM, 2014.

[16] G. Ling, M. R. Lyu, and I. King. Ratings meet reviews, a combined approach to recommend. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 105–112. ACM, 2014.

[17] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM, 2013.

[18] X. Liu and K. Aberer. Soco: a social network aided context-aware recommender system. In *Proceedings of the 22nd international conference on World Wide Web*, pages 781–802. ACM, 2013.

[19] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 203–210, 2009.

[20] H. Ma, M. R. Lyu, and I. King. Learning to recommend with trust and distrust relationships. In *Proceedings of the third ACM conference on Recommender systems*, pages 189–196. ACM, 2009.

[21] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008.

[22] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

[23] J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *ACM Conference on Recommender Systems*, pages 165–172, 2013.

[24] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[25] K. Miyahara and M. Pazzani. Collaborative filtering with the simple bayesian classifier. *PRICAI 2000 Topics in Artificial Intelligence*, pages 679–689, 2000.

[26] K. Miyahara and M. J. Pazzani. Improvement of collaborative filtering with the simple bayesian classifier. *Information Processing Society of Japan*, 43(11), 2002.

[27] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *Neural Information Processing Systems*, pages 1257–1264, 2007.

[28] M. OConnor and J. Herlocker. Clustering items for collaborative filtering. In *Proceedings of the ACM SIGIR workshop on recommender systems*, volume 128. UC Berkeley, 1999.

[29] A. Paterek. Improving regularized singular value decomposition for collaborative filtering, 2007.

[30] S. Purushotham, Y. Liu, and C.-c. J. Kuo. Collaborative topic regression with social matrix factorization for recommendation systems. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 759–766, 2012.

[31] J. D. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the 22nd international conference on Machine learning*, pages 713–719. ACM, 2005.

[32] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887. ACM, 2008.

[33] H. Shan and A. Banerjee. Generalized probabilistic matrix factorizations for collaborative filtering. In *2010 IEEE 10th International Conference on Data Mining (ICDM)*, pages 1025–1030. IEEE, 2010.

[34] N. Srebro, J. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In *Advances in neural information processing systems*, pages 1329–1336, 2005.

[35] J. Tang, X. Hu, H. Gao, and H. Liu. Exploiting local and global social context for recommendation. In *IJCAI*, pages 264–269, 2013.

[36] L. H. Ungar and D. P. Foster. Clustering methods for collaborative filtering. In *AAAI workshop on recommendation systems*, volume 1, pages 114–129, 1998.

[37] S. Vucetic and Z. Obradovic. Collaborative filtering using a regression-based approach. *Knowledge and Information Systems*, 7(1):1–22, 2005.

[38] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 448–456. ACM, 2011.

[39] H. Wang, B. Chen, and W.-J. Li. Collaborative topic regression with social regularization for tag recommendation. In *IJCAI*, 2013.

[40] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Y. Yu, and Z. Chen. Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 114–121. ACM, 2005.

[41] J. Yang and J. Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *ACM International Conference on Web Search and Data Mining*, pages 587–596, 2013.

[42] X. Yang, H. Steck, and Y. Liu. Circle-based recommendation in online social networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1267–1275. ACM, 2012.

[43] K. Yu, S. Zhu, J. Lafferty, and Y. Gong. Fast nonparametric matrix factorization for large-scale collaborative filtering. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 211–218. ACM, 2009.

[44] F. Zhao, M. Xiao, and Y. Guo. Predictive collaborative filtering with side information. In *IJCAI*, pages 2385–2391, 2016.

[45] G. Zhao, M. L. Lee, W. Hsu, W. Chen, and H. Hu. Community-based user recommendation in uni-directional social networks. In *Proceedings of the 22nd ACM international conference on Information Knowledge Management*, pages 189–198. ACM, 2013.