Collaborative Topic Regression with Multiple Graphs Factorization for Recommendation in Social Media

Qing Zhang
Key Laboratory of Computational
Linguistics (Peking University)
Ministry of Education, China
zqicl@pku.edu.cn

Houfeng Wang[*]
Key Laboratory of Computational
Linguistics (Peking University)
Ministry of Education, China
wanghf@pku.edu.cn

Abstract

With a large amount of complex network data available from multiple data sources, how to effectively combine these available data with existing auxiliary information such as item content into the same recommendation framework for more accurately modeling user preference is an interesting and significant research topic for various recommender systems. In this paper, we propose a novel hierarchical Bayesian model to integrate multiple social network structures and content information for item recommendation. The key idea is to formulate a joint optimization framework to learn latent user and item representations, with simultaneously learned social factors and latent topic variables. The main challenge is how to exploit the shared information among multiple social graphs in a probabilistic framework. To tackle this challenge, we incorporate multiple graphs probabilistic factorization with two alternatively designed combination strategies into collaborative topic regression (CTR). Experimental results on real dataset demonstrate the effectiveness of our approach.

1 Introduction

Many real-life data have representations in the form of multiple views (Liu et al., 2013). For example, web pages usually consist of both text content and hyperlink information; images on the web have relevant tags associated with their content. In addition, it is also common that in real networks comprising multiple types of nodes are connected by multiple types of links, forming heterogeneous information networks (HIN) (Huang et al., 2012). For example, in scientific community, various types of links are formed for different types of objects, i.e., author writes paper, venue publishes paper, reader labels tag, and so on. Therefore, with a large amount of complex network data available from multiple data sources, how to effectively combine this kind of rich structure with other auxiliary information such as content information into the same recommendation framework is an interesting and significant research topic for various recommender systems. This paper aims to model multiple social graphs into a principled hierarchy Bayesian framework to improve recommending performance.

The basic idea in this paper is inspired by multi-view learning approach (Liu et al., 2013), i.e., leveraging the redundancy and consistency among distinct views (Kumar et al., 2011) to strengthen the overall performance. We extend this idea (Liu et al., 2013) originally for clustering problem to deal with rating scarcity problem when modeling user preference for recommendation. Just as in general multi-view learning, each view of objective function is assumed to be capable of correctly classifying labeled examples separately. Then, they are smoothed with respect to similarity structures in all views. Similarly, in this paper, we assume that our individual views of multiple user social relations are similar and complementary with a shared latent structure.

However, different from multi-view clustering problem, our goal is to recover a sparse rating matrix with a large number of missing user-item pairs rather than merely exploiting cluster structure with full task information. Thus, the straightforward multi-view representation of the objective (rating matrix) is

---

not available. Instead, we use side information (user social graphs) to exploit multi-view learning for improving collaborative filtering (CF). As a result, collaborative topic regression (CTR) (Wang and Blei, 2011) is employed as our basic learning framework with side information. Recently, CTR has gained considerable attention due to its well-defined mathematical framework and strong performance on user behavior prediction for various real-world applications, such as document recommendation (Li et al., 2013), tag recommendation (Wang et al., 2013), music recommendation (Purushotham et al., 2012), celebrity recommendation (Ding et al., 2013) and vote prediction (Kang and Lerman, 2013). However, all the extensions above merely focus on a single view of user or item relation. In reality, a large amount of diverse social graphs data are widely existed and particularly valuable for mutually reinforcing each other. Therefore, it should be well considered. Taking this into consideration, we extend CTR with multiple social graphs factorization for recommender systems.

The main challenge of incorporating multiple relations into CTR is how to exploit the shared information among multiple social networks and how to further deal with it to recover sparse rating matrix in a probabilistic framework. Previous efforts, purely to address the first issue for clustering problem, are usually to seek a weak consensus (Liu et al., 2013) learnd from data jointly with clustering process. Intuitively, consensus can be seen as a latent cluster structure shared by different views. Thus, it means that learning from different views should be softly regularized towards a common latent structure. However, it is not easy to directly formulate it in a probabilistic framework, because weak consensus modeling can not be separated from a joint higher task, i.e., recovering spare rating matrix, in our case.

To tackle this challenge, we propose a novel hierarchical Bayesian model with multiple social graphs factorization. We exploit two ways of modeling shared information for multi-view based recommendation. One is for heterogeneous network by directly modeling different view specific latent structures with consensus for user representation. The other is for homogeneous case, which can be used as a transformed version of heterogeneous relations. In contrast with the first strategy, we model the latter using a shared latent social structure for all views but with different user representations. Thus, we can relax strong consensus assumption in our heterogeneous case, through linear combination of each sub-latent user with maintained sharing mechanism. The multiple graphs factorization process in the proposed model can be seen as a regularization approach on each latent user for better uncovering user-item latent structures. Although, regularization technique for modeling multiple heterogeneous networks is a hot research topic, in clustering study from an algebra view (Liu et al., 2013; Kumar et al., 2011), not much is known on using it for collaborative recommendation problems in a more complex probabilistic setting.

The following sections will discuss those in details and we use the terms network and graph interchangeably throughout this paper.

2 Preliminaries

In this section, we briefly review collaborative topic regression (CTR) (Wang and Blei, 2011), as the foundation of our proposed model. Figure 1 (left) shows the graphical representation of CTR, which combines the merits of traditional collaborative filtering and probabilistic topic modeling. Specifically, the key mechanism of CTR is that using topic vectors learned from LDA (Blei et al., 2003) jointly controls the prior distribution of latent items in original matrix factorization process of CF. The generative process is described as follows:

- For each user i,
    - draw user latent vector $u_i \sim N(0, \lambda_u^{-1} I)$, multivariate Gauss distribution with zero mean.
- For each item j,
    - draw topic proportions $\theta_j \sim Dirichlet(\alpha)$, Dirichlet distribution.
    - draw item latent offset vector $\epsilon_j \sim N(0, \lambda_v^{-1} I)$, and set the item latent vector as $v_j = \epsilon_j + \theta_j$.
    - For each word w_{jn}
        * draw topic assignment $z_{jn} \sim Mult(\theta)$, Multinomial distribution.
        * draw word $w_{jn} \sim Mult(\beta_{z_{jn}})$, Multinomial distribution.
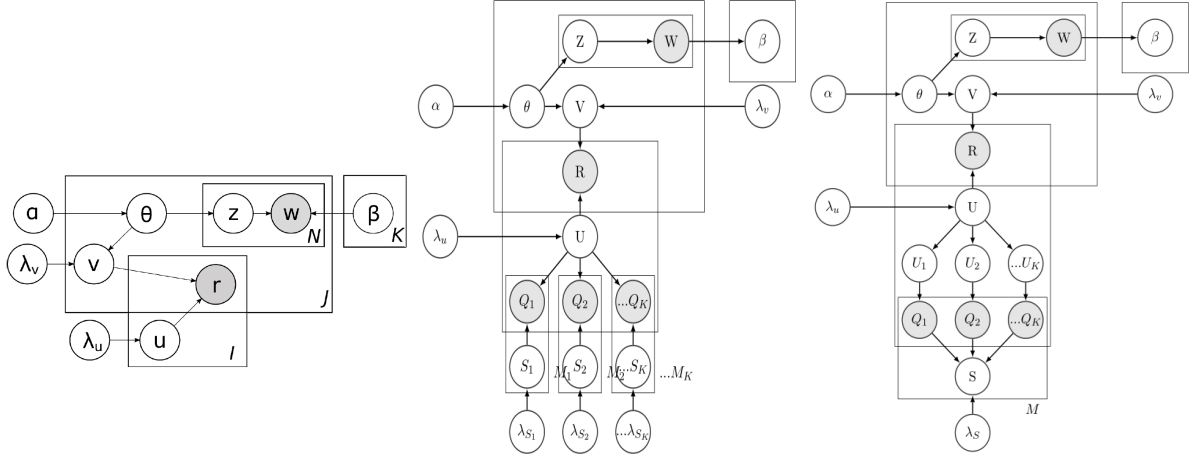- For each user-item pair (i, j),

Figure 1: CTR (left), heterogeneous CTR-MGF (middle), homogeneous CTR-MGF (right).

- draw the response $r_{ij} \sim N(u_i^T v_j, c_{ij}^{-1})$, univariate Gauss distribution, where c_{ij} is a confidence parameter for rating r_{ij}, $a > b$. $c_{ij} = a$ (higher confidence), if $r_{ij} = 1$, and $c_{ij} = b$, if $r_{ij} = 0$.

However, CTR does not take the complex social network information, which is available and crucial in many real-world applications, into consideration.

3   CTR-MGF: Collaborative Topic Regression with Multiple Graphs Factorization

In this section, we discuss our proposed method, called CTR with multiple graphs factorization (CTR-MGF). Our model is a generalized hierarchical Bayesian model which jointly learns latent user, item and multiple latent social factor spaces. Different from previous approaches, our method extends CTR to multiple complex networks setting instead of one particular type of relation for user or item. Moreover, we consider two real general contexts in various practical applications. One is the context of heterogeneous networks. The other is the context of homogeneous networks. It is noted that, for relative simplicity, in this paper we only consider user oriented complex network. The graphical representation of our models in Figure 1 (middle and right) takes $K = 3$ networks as illustration, which can be arbitrary in our derivation. It is also easy to see that Purushotham et al. (2012) is a special case of our proposed model, which is not equipped with graph sharing mechanism.

3.1   Model Notations

Each social matrix Q corresponds to a social network structure $G = \{V, E\}$, where users and their social relations are represented as vertex set V and edge set E in network structure G, respectively. The element q_{im} in Q denotes the binary relation between user 'i' and graph specific feature 'm' in heterogeneous network or the relation between two users 'i' and 'm' in homogeneous network.

3.2   CTR-MGF for Heterogeneous Networks

Heterogeneous network is formed by multiple types of nodes being connected by multiple types of links. The key characteristic of heterogeneous network is that the sizes of feature dimensions are different among multiple social graphs. For example, in a social music sharing system such as LastFM, each user has multiple heterogeneous relations associated with the interested music, i.e., user-artist, user-tag, and so on. Our model can handle all these relations in the proposed framework, CTR-MGF. Specifically, the generative process of CTR-MGF for heterogeneous networks is listed as follows:

- For each item j,
  - draw topic proportions $\theta_j \sim Dirichlet(\alpha)$, Dirichlet distribution.
  - draw item latent offset $\epsilon_j \sim N(0, \lambda_v^{-1}I)$, multivariate Gauss distribution and set the item latent vector as $v_j = \epsilon_j + \theta_j$.

235

- For each word w_{jn}
  * draw topic assignment $z_{jn} \sim Mult(\theta)$.
  * draw word $w_{jn} \sim Mult(\beta_{z_{jn}})$.
- For each heterogeneous social graph k,
  - For each social graph specific feature m
    * draw graph factor-specific latent feature vector $s_m^k \sim N(0, \lambda_{s_{km}}^{-1} I)$.
- For each user i,
  - draw the shared latent user vector among multiple social graphs $u_i \sim N(0, \lambda_{u_i}^{-1} I)$.
  - For each heterogeneous social graph k
    * For each social graph specific feature m
      · draw graph specific user heterogeneous relation pair $q_{im}^k \sim N(u_i^{\mathrm{T}} s_m^k, c_{k,q_{im}}^{-1})$.
- For each user-item pair (i, j),
  - draw the response $r_{ij} \sim N(u_i^{\mathrm{T}} v_j, c_{ij}^{-1})$.

In the above generative process, the joint likelihood of data, i.e., $R, Q_{k=1,...,K}, W$, and the latent factors $U, V, S_{k=1,...,K}$ under the full model is:

$$
\begin{aligned}
&p(R, U, V, S_{k=1,...,K}, Q_{k=1,...,K}, W, \theta | \lambda_\bullet) \\
&= p(R|U,V) \cdot p(W, \theta|\beta) \cdot p(U|\lambda_U) \cdot p(V|\lambda_V) \cdot \prod_k^K p(Q_k|U, S_k, \lambda_{Q_k}) \cdot \prod_k^K p(S_k|\lambda_{S_k})
\end{aligned}
\tag{1}
$$

For learning the parameters, we develop an EM-style algorithm similar to CTR. In our model, finding the MAP is equivalent to maximizing the following log likelihood obtained by substituting univariate and multivariate Gaussian pdfs in Eq. 1:

$$
\begin{aligned}
L = &\sum_j \sum_n log\left(\sum_z \theta_{jz}\beta_{z_{jn}}\right) - \sum_{k=1}^K \frac{\lambda_{S_k}}{2} \sum_m (s_m^k)^{\mathrm{T}} s_m^k - \sum_{k=1}^K \sum_i \sum_m \frac{c_{k,q_{im}}}{2}(q_{im}^k - u_i^{\mathrm{T}} s_m^k)^2 \\
&- \frac{\lambda_U}{2} \sum_i u_i^{\mathrm{T}} u_i - \frac{\lambda_V}{2} \sum_j (v_j - \theta_j)^{\mathrm{T}}(v_j - \theta_j) - \sum_i \sum_j \frac{c_{ij}}{2}(r_{ij} - u_i^{\mathrm{T}} v_j)^2
\end{aligned}
\tag{2}
$$

We employ coordinate ascent (CA) approach alternatively optimizing latent factor variables u_i, v_j, $s_m^{k=1,2,...,K}$ and the simplex variables θ_j as topic proportions. Specifically, the following update rules in CA are obtained by setting the derivative of L with respect to u_i, v_j, and $s_m^{k=1,2,...,K}$ to zero.

$$
u_i = \left(\lambda_U I + V^{\mathrm{T}} D_{c_i} V + \sum_k^K S_k^{\mathrm{T}} D_{q_i}^k S_k\right)^{-1} \cdot \left(V^{\mathrm{T}} D_{c_i} R_i + \sum_k^K S_k^{\mathrm{T}} D_{q_i}^k Q_i^k\right)
\tag{3}
$$

$$
v_j = \left(\lambda_V I + U^{\mathrm{T}} D_{c_j} U\right)^{-1} \cdot \left(\lambda_V \theta_j + U^{\mathrm{T}} D_{c_j} R_j\right)
\tag{4}
$$

$$
s_m^{k=1,2,...,K} = \left(\lambda_{S_k} I + U^{\mathrm{T}} D_{q_m}^k U\right)^{-1} \cdot \left(U^{\mathrm{T}} D_{q_m}^k Q_m^k\right)
\tag{5}
$$

where K is the total number of graphs. I is an identity matrix of the same dimension as that of latent space. U and V are matrices with rows as latent users and latent items, respectively. S_k is a matrix with rows as social factor-specific latent feature vectors for graph k. R_i is a column vector with values $[r_{i1},...,r_{iJ}]^{\mathrm{T}}$. Similarly, $R_j = [r_{1j},...,r_{Ij}]^{\mathrm{T}}$. For graph k, $Q_i^k = [q_{i1}^k,...,q_{iM}^k]^{\mathrm{T}}$ and $Q_m^k = [q_{1m}^k,...,q_{Im}^k]^{\mathrm{T}}$ respectively. Likewise, $D_{q_i}^k$, and $D_{q_m}^k$ are similarly defined with diagonal elements c_{k,q_i} and $c_{k,q.m}$, respectively. D_{c_i} is a diagonal matrix with values $diag(c_{i1},...,c_{iJ})$. $D_{c_j} = diag(c_{1j},...,c_{Ij})$. In addition, c_{ij} and $c_{k,q_{im}}$ are also seen as the confidence parameters for r_{ij} and q_{im}^k, respectively. The high confidence value a is set to the observed interactive pairs and the low confidence value b is set to the unobserved interactive pairs, where $a > b > 0$.

For our brevity, the remaining update rules for θ and β, can be obtained using the same way as described in CTR (Wang and Blei, 2011). Please see that for details.

It is worth noting that through our assumption and the derivation above, we have theoretically proved that our modeling in this case is equivalent to first concatenating features of different views together and then applying Purushotham et al. (2012) for recommendation.

3.3 CTR-MGF for Homogeneous Networks

In this section, we further extend the basic CTR to the context of homogeneous networks. In fact, any user specific homogeneous networks can be obtained through transforming corresponding heterogeneous networks. For example, in LastFM, we can construct two user-user homogeneous networks by computing the similarities of user-tag and user-artist from original heterogeneous networks. The goal of this transformation is to further exploit weak consensus modeling scheme based on Section 3.2. Different from the graph sharing mechanism presented in last section, we relax the restriction that all users have the same representation. Specifically, we assume each latent user has multiple sub-graph specific representations.

However, it is nontrivial to model the relaxed assumption directly from original perspective. To achieve this more weaker sharing mechanism, we are towards its transformed perspective, i.e., sacrificing heterogeneous characteristic, because we need to exploit shared information from latent graph specific feature perspective. Thus, we require equal dimensions of different graphs, which motivates us to investigate the homogeneous case.

The key differences between our model in this section and that in last section are the strategies of latent user modeling and its social factor modeling. More specifically, we model each latent user as a linear combination of all sub-latent users associated with multiple homogeneous networks. All these sub-latent homogeneous users are associated with a shared social factor feature space. Thus, the shared information among multiple graphs can be exploited and it is more flexible to adjust the contribution of each sub-latent user to the final latent user representation. The generative process of CTR-MGF for homogeneous networks is listed as follows:

- For each item j,
  - draw topic proportions $\theta_j \sim Dirichlet(\alpha)$, Dirichlet distribution.
  - draw item latent offset $\epsilon_j \sim N(0, \lambda_v^{-1}I)$, multivariate Gauss distribution and set the item latent vector as $v_j = \epsilon_j + \theta_j$.
  - For each word w_{jn}
    * draw topic assignment $z_{jn} \sim Mult(\theta)$.
    * draw word $w_{jn} \sim Mult(\beta_{z_{jn}})$.
- For each social graph specific feature m, regarding to all related homogeneous social graphs
  - draw a shared factor-specific latent feature vector across multiple graphs $s_m \sim N(0, \lambda_{s_m}^{-1}I)$.
- For each user i,
  - For each homogeneous social graph k
    * draw a social graph specific latent user $u_i^k \sim N(0, (\lambda_{u_i}^k)^{-1}I)$.
    * For each social graph specific feature m
      · draw graph specific user homogeneous relation pair $q_{im}^k \sim N((u_i^k)^{\mathrm{T}}s_m, c_{k,q_{im}}^{-1})$.
  - draw a final latent user $u_i \sim N(\sum_{k=1}^{K} T_k u_i^k, \lambda_{u_i}^{-1}I)$.
- For each user-item pair (i, j),
  - draw the response $r_{ij} \sim N(u_i^{\mathrm{T}}v_j, c_{ij}^{-1})$.

In the above generative process, the joint likelihood of data, i.e. R, $Q_{k=1,...,K}$ and W, and the latent factors $U, U_{k=1,...,K}, V$ and S under the full model is:

$$p(R, U, V, S, U_{k=1,...,K}, Q_{k=1,...,K}, W, \theta | \lambda_\bullet) = p(R|U,V) \cdot p(W, \theta|\beta) \cdot$$

$$(\prod_{k=1}^{K} p(Q_k|U_k, S, \lambda_{Q_k})) \cdot p(S|\lambda_S) \cdot (\prod_{k=1}^{K} p(U_k|\lambda_{U_k})) \cdot p(V|\lambda_V) \cdot p(U|\lambda_U) \cdot p(U|U_{k=1,...,K}, \lambda_C) \tag{6}$$

Similarly to last section, we develop an EM-style algorithm to find the MAP solutions, which is equivalent to maximizing the following log likelihood by substituting univariate and multivariate Gaussian pdfs in Eq. 6:

$$L = \sum_j \sum_n log(\sum_z \theta_{jz}\beta_{z_{jn}}) - \sum_{k=1}^K \frac{\lambda_{U_k}}{2} \sum_i (u_i^k)^{\mathrm{T}} u_i^k - \frac{\lambda_S}{2} \sum_m s_m^{\mathrm{T}} s_m$$

$$- \sum_{k=1}^K \sum_i \sum_m \frac{c_{k,q_{im}}}{2}(q_{im}^k - (u_i^k)^{\mathrm{T}} s_m)^2 - \frac{\lambda_V}{2} \sum_j (v_j - \theta_j)^{\mathrm{T}}(v_j - \theta_j) \tag{7}$$

$$- \sum_i \sum_j \frac{c_{ij}}{2}(r_{ij} - u_i^{\mathrm{T}} v_j)^2 - \frac{\lambda_C}{2} \sum_i (u_i - \sum_{k=1}^K T_k u_i^k)^{\mathrm{T}}(u_i - \sum_{k=1}^K T_k u_i^k) - \frac{\lambda_U}{2} \sum_i u_i^{\mathrm{T}} u_i$$

We employ coordinate ascent (CA) approach as previous section alternatively optimizing latent factor variables and simplex variables as topic proportions. Then we acquire the update rules by setting the derivative of L with respect to the following variables to zero.

$$u_i^{k=1,2,...,K} = (\lambda_{U_k}I + \lambda_C T_k^2 I + S^{\mathrm{T}} D_{q_i}^k S)^{-1} \cdot (\lambda_C u_i T_k - (\sum_{p \neq k}^K T_p u_i^p)\lambda_C T_k + S^{\mathrm{T}} D_{q_i}^k Q_i^k) \tag{8}$$

$$u_i = (\lambda_U I + \lambda_C I + V^{\mathrm{T}} D_{c_i} V)^{-1} \cdot (V^{\mathrm{T}} D_{c_i} R_i + \lambda_C \sum_k^K T_k u_i^k) \tag{9}$$

$$v_j = (\lambda_V I + U^{\mathrm{T}} D_{c_j} U)^{-1} \cdot (U^{\mathrm{T}} D_{c_j} R_j + \lambda_V \theta_j) \tag{10}$$

$$s_m = (\lambda_S I + \sum_{k=1}^K U_k^{\mathrm{T}} D_{q_m}^k U_k)^{-1} \cdot (\sum_{k=1}^K U_k^{\mathrm{T}} D_{q_m}^k Q_m^k) \tag{11}$$

where K is the total number of graphs. I is an identity matrix of the same dimension as that of latent space. U and V are matrices with rows as latent users and latent items, respectively. S is a matrix with rows as the shared social factor-specific latent feature vectors for all graphs. T is the graph selection weight, $\sum_{k=1}^K T_k = 1, T_k >= 0$. $R_i = [r_{i1}, \ldots, r_{iJ}]^{\mathrm{T}}$ and $R_j = [r_{1j}, \ldots, r_{Ij}]^{\mathrm{T}}$. For graph k, $Q_i^k = [q_{i1}^k, \ldots, q_{iM}^k]^{\mathrm{T}}$, $Q_m^k = [q_{1m}^k, \ldots, q_{Im}^k]^{\mathrm{T}}$ and U_k is a matrix with rows as the social graph k specific latent user vectors. Likewise, $D_{q_i}^k$, and $D_{q_m}^k$ are similarly defined with diagonal elements $c_{k,q_{i.}}$ and $c_{k,q_{.m}}$, respectively. $D_{c_i} = diag(c_{i1}, \ldots, c_{iJ})$ and $D_{c_j} = diag(c_{1j}, \ldots, c_{Ij})$. In addition, c_{ij} and $c_{k,q_{im}}$ are also seen as the confidence parameters for r_{ij} and q_{im}^k, respectively. The high confidence value a is set to the observed interactive pairs and the low confidence value b is set to the unobserved interactive pairs, where $a > b > 0$.

For our brevity, the remaining update rules for θ and β, can be obtained using the same way as described in CTR (Wang and Blei, 2011). Please see that for details.

3.4 Prediction

Using the learned parameters above, we can make in-matrix and out-of-matrix predictions defined in Wang and Blei (2011). For in-matrix prediction, it refers to the case where those items have been rated by at least one user in the system. To compute predicted rating, we use

$$r_{ij}^* \approx (u_i^*)^T v_j^*. \tag{12}$$

For out-of-matrix prediction, it refers to the case where those items have never been rated by any user in the system. To compute predicted rating, we use

$$r_{ij}^* \approx (u_i^*)^T \theta_j^*, \tag{13}$$

where the corresponding θ_j^* is defined as topic proportion in Section 3.2 and 3.3.

3.5 Computational Issue

To reduce computational costs when updating u_i, v_j and other variables with similar structure in update rule, we adopt the same strategy of matrix operation shown in Hu et al. (2008). Specifically, directly computing $V^T D_{c_i} V$ and $U^T D_{c_j} U$ requires time $O(L^2 J)$ and $O(L^2 I)$ for each user and item, where J and I are the total number of items and users respectively, L is the dimension of latent representation space. Instead, we rewrite $U^T D_{c_j} U = U^T (D_{c_j} - bI)U + bU^T U$. Then, $bU^T U$ can be pre-computed and $D_{c_j} - bI$ has only I_r non-zeros elements, where I_r refers to the number of users who rated item j and empirically $I_r \ll I$. For other similar structures, i.e., $V^T D_{c_i} V$, $S^T D_{q_i}^k S$, and so on, they are similar. Therefore, we can significantly speed up computation by this sparsity property.

4 Experiments

4.1 Data

We evaluate our proposed method on real life dataset [1] from LastFm. LastFm[2] is an online music catalogue, powered by social music discovery service for personalized recommendation. This dataset (Cantador et al., 2011) is challenging. Though it contains 92,834 pairs of observed ratings with 1892 users and 17,632 items, the sparseness is quite low, i.e., merely 0.2783% , which is much lower than that of the well-known Movielens dataset with the sparseness 4.25%. On average, each user has 44.21 items in the play list, ranging from 0 to 50, and each item appears in 4.95 users libraries, ranging from 0 to 611. For each item, the tag information is used as bag-of-word representation. After text processing, 11,946 distinct words are remained in the corpus. In addition, we further remove noisy users which have no items. We also construct two additional social graphs for our experiments. One is user-tag network extracted from user-tag-item relations in original dataset. The other is user-user network through transforming the constructed user-tag network. The relation in all graphs is binary, i.e., the available denoted as 1 and the unavailable denoted as 0.

Table 1: Original dataset description

Dataset	users	items	tags	user-user relations	user-tags-items	user-items relations
LastFm	1892	17632	11946	25434	186479	92834

4.2 Metrics

Two metrics for evaluating the recommendation performance are employed, i.e., Recall and NDCG. Measure for plain relevance:

$$Recall@k = \frac{\sharp relevance}{k}, \tag{14}$$

where $\sharp relevance$ denotes the total relevant papers in returned top-k result. Measure for ranking-based relevance:

$$NDCG@k = \frac{\sum_{i=1}^{k} \frac{2^{rel_i} - 1}{log_2(1+i)}}{IDCG}, \tag{15}$$

where rel_i denotes the relevant degree which is binary in our task and $IDCG$ is the optimal score computed using the same form in numerator but with optimal ranking known in advance.

4.3 Experimental Design

In this paper, we expect the proposed model 'Our-Homo' in Section 3.2 and 'Our-Heter' in Section 3.3 can jointly provide a general and systematic solution to handling the following cases of using multiple graphs for recommendation:

- **Case 1**:(Heterogeneous networks with noise) Network data or the extraction process is usually imprecise or noisy in practice. Transform it into homogeneous case and then use 'Our-Homo'.

---

[1]Data available at http://grouplens.org/datasets/hetrec-2011/
[2]http://www.last.fm/

- **Case 2**:(Homogeneous networks) 'Our-Homo' can be directly employed as the tool for case 1.

- **Case 3**:(Heterogeneous networks with high quality) 'Our-Heter' might be directly employed. It is not needed to be further transformed into Homogeneous case.

The detail experiments in the following sections are presented to justify the effectiveness of our methods for the three cases above.

4.4   Experiments for Case 1 and Case 2

In this section, we mainly focus on the most complex and common case 1 with case 2 in practice.

4.4.1   Baselines

We compare our proposed two models, the model in Section 3.2 denoted as Our-Heter and the model in Section 3.3 denoted as Our-Homo, with some state-of-the-art algorithms.

- **CTR**: This method, described in Wang and Blei (2011), combines both item content information and user-item ratings for CF.
- **PMF**: This method, described in Salakhutdinov and Mnih (2007), is a well-known matrix factorization method for CF, only using interactive rating information.
- **SMF-1**: This method, described in Purushotham et al. (2012), exploits single user's social network structure combined with item's content information for CF. SMF-1 denotes using our extracted user-tag relation.
- **SMF-2**: The same SMF method, described in (Purushotham et al., 2012). SMF-2 denotes using original user-user relation.
- **Our-Heter**: Our model for heterogeneous networks, proposed in Section 3.2, uses our extracted user-tag network and original user-user network.
- **Our-Homo**: Our model for homogeneous networks, proposed in Section 3.3, uses two homogeneous networks, i.e., 1) the transformed user-user network through our extracted user-tag relation, and 2) original user-user network.

4.4.2   Settings

For a fair comparison, we use the similar settings as prior work in Purushotham et al. (2012). Specifically, to well judge the influence of multiple social network structures, we fix the effects of content information to the same level that is optimal in SMF, $\lambda_v = 0.1$. We randomly split the dataset into two parts, training (90%) and test datasets (10%), with constraint that users in test dataset have more than half of the average number of rated items, i.e., 20. This expands the range of performance analysis for our evaluation compared with Purushotham et al. (2012). The optimal parameters are obtained on a small held-out dataset. For PMF, we set $\lambda_v = 100, \lambda_u = 0.01$. For all CTR-based methods, we set $a = 1, b = 0.01, \lambda_v = 0.1$. Specifically, for CTR, we set $\lambda_u = 0.01$. For SMF-1 and SMF-2, we set $\lambda_u = 0.01$. For Our-Homo, we set $\lambda_u = 0.01, \lambda_{u1} = \lambda_{u2} = \lambda_s = 100, \lambda_c = 0.01$. For Our-Heter, we set $\lambda_u = 0.01, \lambda_{s1} = \lambda_{s2} = 100$. The remaining paramters are varied for experiment analysis.

It is noted that the task of out-of-matrix prediction is originally designed for evaluating item content modeling in CTR rather than user social graphs as in CTR-smf. Thus, we followed the same setting in baseline CTR-smf (Purushotham et al., 2012), not considering this task.

4.4.3   Performance Comparison with State-of-the-Art Methods

Figure 2 shows the recall and NDCG results of all the methods when the number of latent factor is fixed to 200 (optimal for the baselines). The proposed model 'Our-Homo' consistently outperforms the baselines and 'Our-Heter' model under both recall and NDCG measures. This finding demonstrates that (1) using multiple graphs for CTR is a necessary for improving recommendation performance from both ranking and plain accuracy perspectives. (2) strong consensus for modeling shared information undermines the performance for multiple graphs factorization as designed in Our-Heter. (3) For heterogeneous case, we address that through simply transforming the heterogeneous network to homogeneous one and then use
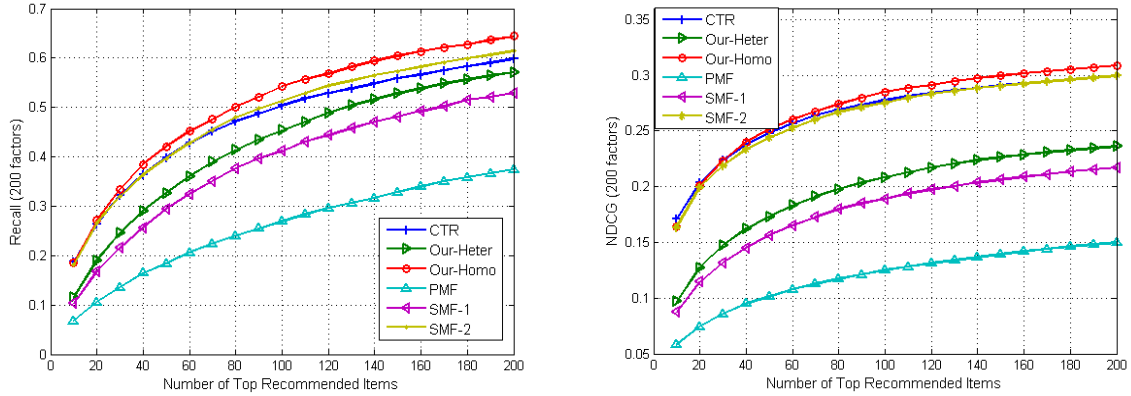
Figure 2: Our model comparison with the state-of-the-art methods, for Recall and NDCG.
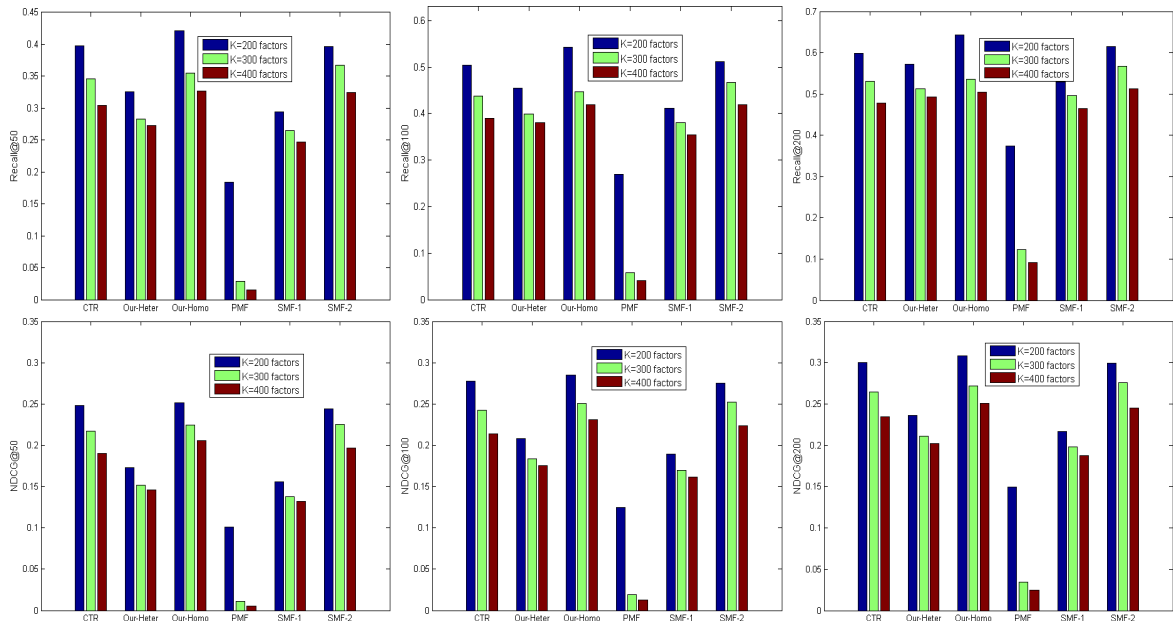

Figure 3: Performance comparison for different latent factors (K=200,300,400) @ top (50,100,200).

'Our-Homo'. This is natural but the opposite is hard. Thus, our solution 'Our-Homo' for modeling weak consensus is effective for both homogeneous and heterogeneous cases.

In addition, we can see that CTR-smf (Purushotham et al., 2012) is sensitive to the quality of graph (SMF-1 with low quality and SMF-2 with high quality as shown in Figure 2). In contrast, we can use the low quality noisy graph (SMF-1) to improve the overall performance by this transformation process. In fact, why Our-Heter does not perform well is mainly due to the noisy graph-1. 'Transformation' can be seen as a 'denoising' process.

4.4.4 Performance Analysis with Different Latent Factors

Figure 3 shows the results of the compared algorithms, with different number of latent factors for varied top recommended item. It shows that $K = 200$ factors is optimal for all baselines compared with other choices of the number of latent factors. This justifies our fire choice of 200 latent factors reported in Figure 3 (Other factor choices are omitted here due to page constraint, which is not optimal for our baselines) and suggests that the choice of latent factor number is crucial for all algorithms especially for PMF. In contrast, the proposed 'Our-Homo' is more stable compared with PMF and outperforms the other baselines in an overall performance as reported in Figure 2.
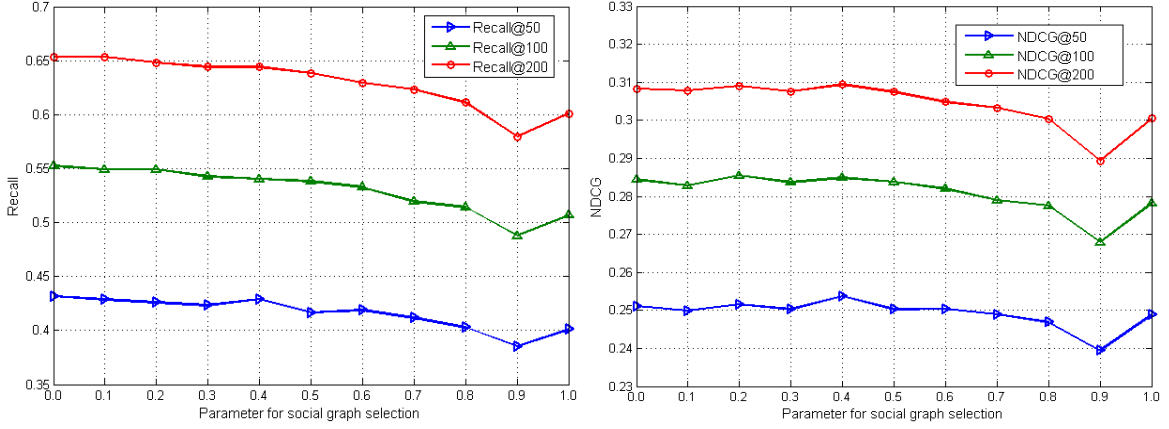
Figure 4: Parameter analysis of graph selection for our weak consensus modeling, i.e., Our-Homo.

4.4.5 Impact of the Parameter for Graph Selection

Next, we examine how our algorithm 'Our-Homo' is influenced by the graph-selection weights. In Figure 4, the horizontal axis shows the graph proportion weight T_1 for graph-1 (user-tag network). In our case, two social graphs are considered. The weight for second graph is $T_2 = 1 - T_1$, which is not shown in Figure 4. Figure 4 clearly proves the effectiveness of our weak consensus modeling. Specifically, we can see that 0.0 seemingly means that the first graph is not selected due to the weight of the first graph $T_1 = 0.0$ and that of second graph $T_2 = 1$ (fully selected). However, due to our weak consensus modeling scheme, though the graph-1 weight is 0.0, it does not mean that the first graph is removed. In fact, the effect of graph-1 is also active through the shared variable 's' in Figure 1 (right). This further can be investigated from the Equation 9. Apparently, the case of 0.0 weight for graph-1, u is only relevant to u_1, but from Equation 8, we can see that u_1 is also influenced by u_2 via the shared social graph factors s. Thus, this also explains why 0.0 weight for graph-1 is not equal to the result of SMF-2 as shown in Figure 2, which only uses graph-2, original user-user graph in SMF-2.

In addition, the 'valley' in Figure 4 might be explained that in the extreme cases (0.0 and 1.0), the denoising effect of weak consensus is slightly strengthened because only one specific graph (higher quality smf-2 or lower quality smf-1 shown in Fig.2) is directly associated with final latent user combined with shared variable. Therefore, the extreme case (1.0) is towards a relative higher performance.

4.5 Experiments for Case 3

Though the proposed 'Our-Homo' is more effective than CTR and CTR-smf, it does not mean that the proposed another method 'Our-Heter' is useless. In this section, we show how the case 3 will be justified.

4.5.1 Baselines

- **Our-Heter(N)**: Our model for heterogeneous networks, proposed in Section 3.2, uses our modified high quality user-tag network as described in Section 4.5.2 and original user-user social network as SMF-2 in Section 4.4.1.
- **SMF-1(N)**: The same SMF method with single social network, described in Purushotham et al. (2012). SMF-1 (N) denotes using our modified high quality user-tag network as described in Section 4.5.2.
- **Our-Homo(O)**: The result of this method is the same as that reported in Section 4.4.3.
- **CTR(O)**: The result of this method (Wang and Blei, 2011) is the same as that reported in Section 4.4.3.

4.5.2 Settings

We want to investigate whether Our-Heter will outperform Our-Homo in the case where the heterogeneous networks with less noise are available, compared with previous results in Figure 2. The settings
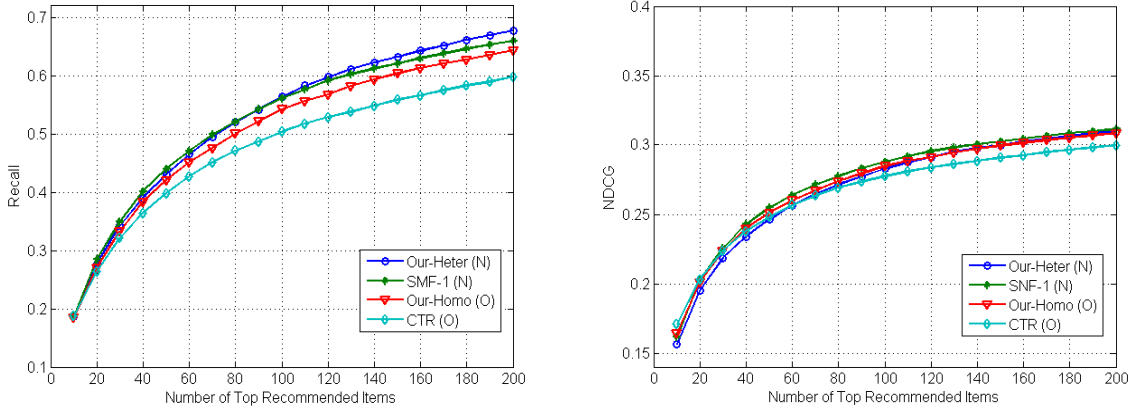
Figure 5: Our model comparison with the state-of-the-art methods (case 3), for Recall and NDCG.

in case 3 are the same as that in Section 4.4.2 except for the refined user-tag graph. In this case, we construct a less noisy user-tag network by selecting top-10% tags according to $tf * idf$ value. The optimal latent factor number is set to 100 for Our-Heter (N) through a small held-out dataset. The remaining parameters are kept as the same values in Section 4.4.2. For notations in baselines Section, 'O' denotes old setting and 'N' denotes new setting with updated user-tag network presented in this section.

4.5.3 Performance Comparison with State-of-the-Art Methods

Figure 5 shows Our-Heter (N) can achieve improved performance compared with baselines without transformation, for the case where high quality graphs are available. Specifically, for recall measure, Our-Heter (N) produces the best result with the increasing number of top recommended items. In addition, we observe that modeling multiple graphs is necessary to further improve recommending performance, while multiple high quality heterogeneous graphs are available.

For NDCG measure, Our-Heter (N) is comparable to our baselines. Since recall measure is only considered for several reasons in previous work (Wang and Blei, 2011; Purushotham et al., 2012), NDCG is introduced as a plus compared with primarily focused recall. Therefore, Our-Heter (N) is also competitive in overall performance in case 3.

In fact, as discussed in Kang and Lerman (2013), CTR-smf (Purushotham et al., 2012) is not always superior to CTR (Wang and Blei, 2011) and vice versa due to different contexts. Likewise, our model is under the multi-view assumption as discussed in Section 1 that should be checked in practice.

5 Conclusions

In this paper, we propose a general recommendation framework with multiple data sources based on CTR. It is a principled hierarchy Bayesian framework with multiple social graphs factorization for recommender systems. In this framework, two ways of consensus modeling are exploited. Specifically, the proposed models Our-Homo and Our-Heter can jointly provide a general and systematic solution to handling three real cases of using multiple graphs with item content information for recommendation: case 1) Heterogeneous networks with noise; case 2) Homogeneous networks; case 3) Heterogeneous networks with high quality. Experimental results on real dataset demonstrate the effectiveness of our approach. While this framework is used for modeling multiple user social graphs, it can be easily extended to exploiting other side information such as multiple complex relations for items in various applications.

Acknowledgements

References

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2011. 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proceedings of the 5th ACM Conference on Recommender Systems*, RecSys 2011, New York, NY, USA. ACM.

Xuetao Ding, Xiaoming Jin, Yujia Li, and Lianghao Li. 2013. Celebrity recommendation with collaborative social topic regression. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI*, pages 2612–2618. AAAI Press.

Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Eighth IEEE International Conference on Data Mining, ICDM*, pages 263–272. IEEE.

Hongzhao Huang, Arkaitz Zubiaga, Heng Ji, Hongbo Deng, Dong Wang, Hieu Khac Le, Tarek F Abdelzaher, Jiawei Han, Alice Leung, John Hancock, et al. 2012. Tweet ranking based on heterogeneous networks. In *Proceedings of International Conference on Computational Linguistics, COLING*, pages 1239–1256.

Jeon-Hyung Kang and Kristina Lerman. 2013. La-ctr: A limited attention collaborative topic regression for social media. In *Proceedings of AAAI Conference on Artificial Intelligence, AAAI*, pages 119–125.

Abhishek Kumar, Piyush Rai, and Hal Daumé III. 2011. Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems, NIPS*, pages 1413–1421.

Yingming Li, Ming Yang, and Zhongfei Mark Zhang. 2013. Scientific articles recommendation. In *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management, CIKM*, pages 1147–1156. ACM.

Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of SIAM Data Mining Conference, SDM*. SIAM.

Sanjay Purushotham, Yan Liu, and C-c J Kuo. 2012. Collaborative topic regression with social matrix factorization for recommendation systems. In *Proceedings of the 29th International Conference on Machine Learning, ICML*, pages 759–766.

Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems, NIPS*, pages 1257–1264.

Chong Wang and David M Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD*, pages 448–456. ACM.

Hao Wang, Binyi Chen, and Wu-Jun Li. 2013. Collaborative topic regression with social regularization for tag recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, IJCAI*, pages 2719–2725. AAAI Press.