Noname manuscript No.
(will be inserted by the editor)

Fast Local Weighted Matrix Factorization for Implicit Feedback

Keqiang Wang · Hongwei Peng · Yuanyuan Jin · Chaofeng Sha · Xiaoling Wang

Received: date / Accepted: date

Abstract Item recommendation helps people to discover their potentially interested items among large numbers of items. One most common application is to recommend items on implicit feedback datasets (e.g. listening history, watching history or visiting history). In this paper, we assume that the implicit feedback matrix has *local* property, where the original matrix is not globally low-rank but some sub-matrices are low-rank. In this paper, we propose Local Weighted Matrix Factorization for implicit feedback (LWMF) by employing the kernel function to intensify *local* property and the weight function to model user preferences. The problem of sparsity can also be relieved by sub-matrix factorization in LWMF, since the density of sub-matrices is much higher than the original matrix. We propose a heuristic method to select sub-matrices which approximate the original matrix well. The greedy algorithm has approximation guarantee of factor $1 - \frac{1}{e}$ to get a near-optimal solution. The experimental results on two real datasets show that the recommendation precision and recall of LWMF are both improved about 30% comparing with the best case of WMF.

Keywords Recommendation Systems; Local Matrix Factorization; Implicit Feedback; Weighted Matrix Factorization; Item Recommender; Submodular

Keqiang Wang✉
Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, China
E-mail: sei.wkq2008@gmail.com

Hongwei Peng
Shanghai Key Laboratory of Trustworthy Computing,East China Normal University, China
E-mail: penghognwei_phw@163.com

Yuanyuan Jin
Shanghai Key Laboratory of Trustworthy Computing,East China Normal University, China
E-mail: 2547124038@qq.com

Chaofeng Sha
Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, China
E-mail: cfsha@fudan.edu.cn

Xiaoling Wang
Shanghai Key Laboratory of Trustworthy Computing,East China Normal University, China
E-mail: xlwang@sei.ecnu.edu.cn

1 Introduction

MF [4] projects users and items into a latent low-dimensional space. Further, the missing entries in the original matrix can be recovered using the dot product between user and item latent vectors. Recently, LLORMA [7] has been shown to be more effective than the traditional MF. The original matrix is divided into several smaller sub matrices, in which we can exploit local structures for better low-rank approximation. In each sub matrix, the standard MF technique is applied to generate sub matrix-specific latent vectors for both users and items.

The above techniques can achieve good performance in rating prediction when high quality explicit feedback is available. For example, ratings are explicit feedbacks which indicate users' preference. However, explicit feedbacks are not easy to get and they do not consider the item recommendation. Compared with the explicit feedback, the implicit feedbacks are more common and larger. User discovers the item if her behaviors are implicit feedbacks, such as listened, watched or visited the item. Otherwise, user is unaware of the item. Different from the explicit feedback, the numerical value to describe implicit feedback is non-negative and very likely to be noisy [10].

Therefore, we consider doing item prediction based on implicit feedback datasets. Specifically, we also assume that the implicit feedback matrix is not globally low-rank but some sub-matrices are low-rank. Instead of decomposing the original matrix, we decompose the sub-matrix intuitively. We propose Local Weighted Matrix Factorization (LWMF), integrating LLORMA [7] with WMF [10] in recommending by employing the kernel function to intensify local property and the weight function to intensify modeling user preference. The problem of sparsity can also be relieved by sub-matrix factorization in LWMF, since the density of sub-matrices is much higher than the original matrix. Two key issues of such a sub matrix-ensemble method are (1) how to generate the sub matrices and (2) how to set the ensemble weights for sub matrices. For the first problem,we propose a heuristic method DCGASC to select submatrices which approximate the original matrix well. For the second problem, we adopt the kernel function to model local property and explore user preferences by the weight function.

The main contributions can be summarized as follows:

- We propose LWMF which integrates LLORMA with WMF to recommend items on implicit feedback datasets. LWMF utilizes the local property to model the matrix by dividing the original matrix into submatrices and relieves the sparsity problem.
- Based on kernel function, we propose DCGASC (Discounted Cumulative Gain Anchor Point Set Cover) to select the sub-matrices in order to approximate the original matrix better. At the same time, we conduct the theoretical submodularity analysis of the DCGASC objective function.
- Based on item recommendation problem, we further propose a variant method user-based LWMF, which is more reasonable for item recommendation and get better performance.
- Extensive experiments on real datasets are conducted to compare LWMF with state-of-the-art WMF algorithm. The experimental results demonstrate the effectiveness of our proposed solutions.

The rest of the paper is organized as follows. Section 2 reviews related work and Section 3 presents some preliminaries about MF (Matrix factorization), WMF and LLORMA. Then we describe LWMF in Section 4 including the heuristic method DCGASC to select submatrices and the learning algorithm of local latent vectors. Experimental evaluations using real datasets are given in Section 5. Conclusion and future work are followed in Section 6.

2 Related Work

One of the most traditional and popular way for recommender systems is KNN [1]. Item-based KNN uses the similarity techniques (e.g., cosine similarity, Jaccard similarity and Pearson correlation) between items to recommend the similar items. Then, MF [2–4] methods play an important role in model-based CF methods, which aim to learn latent factors on user-item matrix. MF usually gets better performance than KNN-based methods especially on rating prediction. Recently, several studies focus on using the ensemble of submatrices for better low-rank approximation, including DFC [5], LLORMA[7,8], ACCAMS [9] and WEMAREC [26]. These methods partition the original matrix into several smaller submatrices, and a local MF is applied to each submatrix individually. The final predictions are obtained using the ensemble of multiple local MFs. Typically, clustering-based techniques with heuristic adaptations are used for submatrix generation. We give a brief review of these studies. Mackey et al. [5] introduces a Divide-Factor-Combine (DFC) framework, in which the expensive task of matrix factorization is randomly divided into smaller subproblems. LLORMA [7, 8] uses a non-parametric kernel smoothing method to search nearest neighbors; WEMAREC [26]. However, such methods focus on explicit feedback datasets while most of the feedbacks are implicit, such as listening times, click times and check-ins. The explicit feedbacks are not always available while implicit feedbacks are large and common. So Hu et al. [10] and Pan et al. [11, 12] propose Weighted Matrix Factorization (WMF) to model implicit feedback with Alternative Least Square (ALS). For details, Hu et al. [10] is a whole-data based learning approach setting a uniform weight to missing entries; i.e., giving all zero entries the same weight. Pan et al. [11,12] is a sample-based approach which samples negative instances from missing data and adopts nonuniform weighting.

To improve the efficiency of WMF, several approaches have been proposed. Pilaszy et al.[27] design an approximate solution to ALS presenting novel and fast ALS variants both for the implicit and explicit feedback datasets. Recently, Devooght et al.[28] propose the Randomized block Coordinate Descent (RCD) learner which is a dynamic framework and reduces the complexity. Further,He et al.[24] design an algorithm based on the element-wise Alternating Least Squares (eALS) technique to optimize a MF model with variably-weighted missing data. Other related work on implicit feedback datasets are ranking methods, such as BPR [13] and Pairwise Learning [14]. With the explosion of size of the training data, the ranking methods need use some efficient sampling techniques to reduce complexity. Finally, for BPR framework, there are a lot of special scenarios, such as recommending music [15], News [16], TV show [17] and POI [18,19], utilizing the additional information (e.g., POI recommender considers the geographical information) to improve prediction performance.

Our method employs the kernel function to intensify *local* property and the weight function to explore user preferences. As for parameter learning,we adopt e-ALS skillfully to learn the latent factors.

3 Preliminary

In this section, we present some preliminaries about basic MF, Weighted MF for implicit datasets and local matrix factorization method LLORMA. A glossary of notations used in the paper are list- ed in Table 1.

Table 1 Notations used in the paper.

Symbols	Descriptions
N, M	number of rows (users) and columns (items)
R	data matrix ($\in \mathbb{R}^{N\times M}$) (with missing values)
W	the weight matrix of R
R^h	the h-th sub-matrix ($\subset R$)
T^h	the weight matrix of sub-matrix R^h
$a_j = (u_j, m_j)$	the user-item pair (u_j, m_j)
$E(a_i, a_j)$	the kernel value between two points
P_u^h	the local latent vector ($\in \mathbb{R}^K$) for the user u *w.r.t.* the h-th sub-matrix
Q_m^h	the local latent vector ($\in \mathbb{R}^K$) for the item m *w.r.t.* the h-th sub-matrix
K	the number ($\ll \min(N, M)$) of dimensions for local latent vectors
H	the number of sub-matrices

3.1 Matrix Factorization

MF is a dimensionality reduction technique, which has been widely used in recommendation system especially for the rating prediction [3,4]. Due to their attractive accuracy and scalability, MF plays a vital role in recent recommendation system competitions, such as *Netflix Prize*[1], *KDD Cup 2011 Recommending Music Items*[2], *Alibaba Big Data Competitions*[3] and so on. Given a sparse matrix $R \in \mathbb{R}^{N\times M}$ with indicator matrix I, and latent factor number $F \ll \min\{N, M\}$. The aim of MF is:

$$\min_{P,Q} \sum_{u\in U} \sum_{m\in V} I_{um}(R_{um} - \hat{R}_{um})^2 \tag{1}$$

where U is the user set and V is the item set. In order to avoid over-fitting, regularization terms are usually added to the objective function to modify the squared error. So the task is to minimize $\sum_{u=1}^{N}\sum_{m=1}^{M} I_{um}(R_{um} - \hat{R}_{um})^2 + \lambda\|P\|_F^2 + \lambda\|Q\|_F^2$. The parameter λ is used to control the magnitudes of the latent feature matrices, P and Q. Stochastic gradient descent is often used to learn the parameters [4].

3.2 Weighted Matrix Factorization

[10] argues that original MF is always used on explicit feedback datasets, especially for rating prediction. So Hu et al. [10] and Pan et al. [11,12] propose Weighted Matrix Factorization (WMF) to handle the cases with implicit feedback. Recently, WMF has been widely used in TV show, music and POI (Point-of-Interests) recommendation. To utilize the undiscovered items and to distinguish between discovered and undiscovered items, a weight is added to the MF:

$$W_{um} = 1 + log(1 + R_{um} \times 10^{\varepsilon}) \tag{2}$$

where the constant ε is used to control the rate of increment. Considering the weights of implicit feedback, the optimization function is reformulated as follows:

$$\min_{P,Q} \sum_{u\in U} \sum_{m\in V} W_{um}(C_{um} - P_u Q_m^\top)^2 + \lambda_P\|P\|_F^2 + \lambda_Q\|Q\|_F^2 \tag{3}$$

where each entry C_{um} in the 0/1 matrix C indicates whether the user u has discovered the item m, which can be defined as $C_{um} = \begin{cases} 1 & R_{um} > 0 \\ 0 & R_{um} = 0 \end{cases}$.

3.3 Low-Rank Matrix Approximation

LLORMA [7,8] is under the assumption of locally low rank instead of globally low rank. That is, limited to certain types of similar users and items, the entire rating matrix R is not low-rank but a sub-matrix R_s is low-rank. It is to say that the entire matrix R is composed by a set of low-rank sub-matrices $\mathcal{R}_s = \{R^1, R^2, ..., R^H\}$ with weight matrix set $\mathcal{T} = \{T^1, T^2, ..., T^H\}$

[1] `http://www.netflixprize.com/`

[2] `http://www.kdd.org/kdd2011/kddcup.shtml`

[3] `https://102.alibaba.com/competition/addDiscovery/index.htm`

of sub-matrices, where T^h_{um} indicates the sub-matrix weight of R^h_{um} in R^h:

$$R_{um} = \frac{1}{Z_{um}} \sum_{h=1}^{H} T^h_{u_h m_h} R^h_{u_h m_h} \tag{4}$$

where $Z_{um} = \sum_{h=1}^{H} T^h_{u_h m_h}$. LLORMA uses the MF introduced in Section 3.1 to approximate the sub-matrix R^h. If the matrix has *local* property, we can achieve good accuracy in predicting ratings.

4 Local Weighted Matrix Factorization

In this section, we introduce our proposed method LWMF, and further propose a heuristic method to select sub-matrices. Finally, we adopt fast element-wise ALS to learn the local latent vectors.

4.1 Our Proposed Model

Following the LLORMA, we first select sub-matrices from the original matrix, then each sub-matrix is decomposed by WMF methods as shown in Figure 1. We propose LWMF which integrates LLORMA with WMF to recommend top-N items on implicit datasets. We estimate each sub-matrix R^h by WMF in Section 3.2 as follows:

$$\min_{P_h, Q_h} \sum_{u \in U^h} \sum_{m \in V^h} T^h_{um} W_{um} (R^h_{um} - {P^h_u}^\top Q^h_m)^2 + \lambda^h_P \|P^h\|^2_F + \lambda^h_Q \|Q^h\|^2_F \tag{5}$$

So the original Matrix R can be approximated by the set of approximated sub-matrices $\hat{\mathcal{R}}_s = \{\hat{R}^1, \hat{R}^2, ..., \hat{R}^H\}$:

$$R_{um} \approx \frac{1}{Z_{um}} \sum_{h=1}^{H} T^h_{u_h m_h} {P^h}^\top Q^h \tag{6}$$

where $\mathbf{Z}_{um} = \sum_{k=1}^{K} \mathbf{L}^{(k)}_{um}$ is the normalizer and $\mathbf{L}^{(k)}_{um}$ indicates the weight for the entry $\mathbf{R}^{(k)}_{um}$ in the submatrix $\mathbf{R}^{(k)}$. Two key issues of such a submatrix-ensemble method are (1) how to generate the submatices and (2) how to set the ensemble weights for submatrices. Following LLORMA, we use the Epanechnikov kernel to calculate the relationship between two anchor pairs $a_h = (u_h, m_h)$ and $a_j = (u_j, m_j)$. It is computed as the product of user Epanechnikov kernel ($E_b(u_h, u_h)$) and item Epanechnikov kernel ($E_b(m_j, m_j)$) as follows:

$$E(a_h, a_j) = E_b(u_h, u_j) \times E_b(m_h, m_j) \tag{7}$$

where

$$E_b(u_h, u_j) \propto (1 - d(u_h, u_j)^2)\, \mathbf{1}_{\{d(u_h,u_j) \le b\}}$$
$$E_b(m_h, m_j) \propto (1 - d(m_h, m_j)^2)\, \mathbf{1}_{\{d(m_h,m_j) \le b\}}$$

where b are the bandwidth parameters of kernel. Distance between two users or two items is the distance between two row vectors (for user kernel) or column vectors (for item kernel). The initial user latent factor and item latent factor are learned by WMF. Accordingly, the distance between users u_h and u_j is $d(u_h, u_j) = arccos(\frac{P_{u_h} \cdot P_{u_j}}{\|P_{u_h}\| \cdot \|P_{u_j}\|})$, where P_{u_h}, P_{u_j} are the h-th and j-th rows of P. The distance between items is computed in the same way. So we set the weight $T^h_{u_j m_j} = E(a_h, a_j)$ of user-item pair (u_j, m_j) for sub-matrix R^h, the sub-matrix regularization $\lambda^h_P = \lambda_P E_b(u_h, u_j)$ and $\lambda^h_Q = \lambda_Q E_b(m_h, m_j)$.

And each anchor point stands for a sub-matrix. Selecting the sub-matrix set $\mathcal{R}_s$ is in fact to select a set of anchor points. The details of selecting anchor point set is discussed in next section.

4.2 Anchor Point Set Selection

Intuitively, the sub-matrix set $\mathcal{R}_s = \{R^1, R^2, ..., R^H\}$ should cover the original matrix R, that is $R = \cup_{R^h \in \mathcal{R}_s} R^h$, so these sub-matrix sets $\mathcal{R}_s$ can approximate the original matrix R better than the set that does not cover. So the anchor points selection problem can be reduced to the set cover problem.

4.2.1 Anchor Point Set Cover (ASC)

Given the candidate anchor point set $A = \{a_1, a_2, ..., a_G\}$ (all the none-zero user-item pairs) while every candidate point a_i can cover itself several other candidate points denoted by $A^i = \{a_i, a_{i1}, a_{i2}, ..., a_{iD}\} \subset A$, we propose the naive anchor points cover method, called Anchor Point Set Cover: returns an anchor point set $S \subset A$ such that

$$\max J(S) = |\cup_{i \in S} A^i| \quad s.t. |S| = K \tag{8}$$

Here, we use all pairs (u, m) in training data as the candidate anchor points. Obviously, the ASC problem is submodular and monotone [25]. So the greedy algorithm has $1 - \frac{1}{e}$ approximation of optimization.

4.2.2 Discounted Cumulative Gain Anchor Point Set Cover (DCGASC)

However, set cover problem only need to cover a point only once while covering all training data only once is not enough. Although performance is improved by increasing cover times, the gain is discounted, which is similar to the situation in ranking quality measures NDCG (Normalized Discounted Cumulative

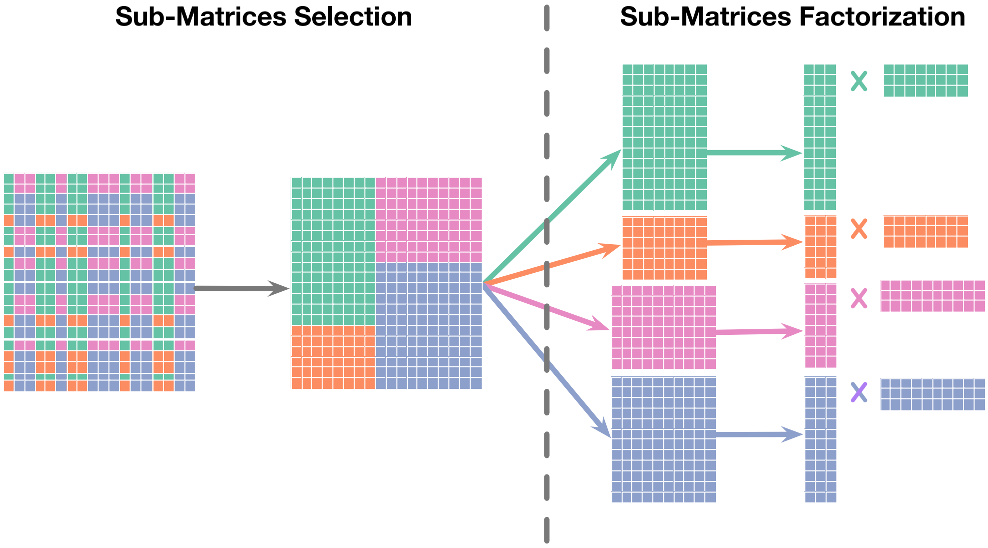


Fig. 1 Local Matrix Factorization

Gain) [22] and ERR (Expected Reciprocal Rank) [21] in IR(Information Retrial). The premise of NDCG and ERR is that highly relevant documents appearing lower in a search result list should be penalized as the graded relevance value is reduced proportional to the position of the result. Learning from this discounted approach, we propose a heuristic method to model this situation, called Discounted Cumulative Gain Anchor Point Set Cover (DCGASC): returns an anchor point order list $S = \{a_{j1}, a_{j2}, ..., a_{jH}\} \subset A$ such that

$$\begin{aligned} &\max J(S) \\ &= \sum_{h=1}^{H} \sum_{a_l \in A^{jh}} \alpha^{o_{lh}-1}(1 - max_{h' \in \{1,...,h-1\}} E_b(a_{jh}, a_{jh'})) \\ &\qquad s.t. |S| = H \end{aligned} \tag{9}$$

where o_{lh} denotes the covered times of a_l by itself or other selected anchor points in $\{A^{j1}, A^{j2}, ..., A^{jh}\}$. $alpha \in (0, 1)$ is the discount parameter. When point a_j has been covered by a anchor point before, the covered gain will be reduced next time. When $\alpha = 0$, this problem reduces to the set cover problem. And when $\alpha = 1$, it just gets the anchor point which covers the other points at most every time. The $(1-max_{h' \in \{1,...,h-1\}} E_b(a_{jh}, a_{jh'}))$ term means that it tends to select the point which is far from the selected anchor points. Below we prove that $f(\cdot)$ is submodular and monotone.

Theorem 1 *DCGASC function is submodular and also monotone nondecreasing.*

Proof Let $S \subseteq V \subseteq A$, $H = |S| + 1$, $X = |V| + 1 \geq H$ and $A^i \in A \backslash V$. We have that

$$\begin{aligned} &J(S \cup \{A^i\}) - J(S) \\ &= \sum_{h=1}^{H} \sum_{a_l \in A^{jh}} \alpha^{o_{lh}-1}(1 - max_{h' \in \{1,...,h-1\}} E_b(a_{jh}, a_{jh'})) \\ &\quad \sum_{h=1}^{H-1} \sum_{a_l \in A^{jh}} \alpha^{o_{lh}-1}(1 - max_{h' \in \{1,...,h-1\}} E_b(a_{jh}, a_{jh'})) \\ &= \sum_{a_l \in A^i} \alpha^{o_{lH}-1}(1 - max_{h' \in \{1,...,H-1\}} E_b(a_{jH}, a_{jh'})) \geq 0 \end{aligned} \tag{10}$$

and

$$\begin{aligned} &J(S \cup \{A^i\}) - J(S) - (J(V \cup \{A^i\}) - J(V)) = \\ &\quad = \sum_{a_l \in A^i} \alpha^{o_{lH}-1}(1 - max_{h' \in \{1,...,H-1\}} E_b(a_{jH}, a_{jh'})) \\ &\quad - \sum_{a_l \in A^i} \alpha^{o_{lX}-1}(1 - max_{h' \in \{1,...,X-1\}} E_b(a_{jX}, a_{jh'})) \end{aligned} \tag{11}$$

Because the number of anchor points covered satisfies that $o_{lH} \leqslant o_{lX}$, discount parameter $\alpha \in [0, 1]$ and $max_{h' \in \{1,...,H-1\}} E_b(a_{jH}, a_{jh'}) \leq \max_{h' \in \{1,...,X-1\}} E_b(a_{jX}, a_{jh'})$, we know that $f(S \cup \{A^i\}) - f(S) - (f(V \cup \{A^i\}) - f(V)) \geq 0$. Therefore, it is proved that the DCGASC function is monotone and submodular.

Due to the monotonicity and submodularity of DCGASC function, the greedy algorithm 1 can provide a theoretical approximation guarantee of factor $1 - \frac{1}{e}$ as

Algorithm 1: `DCGASC Greedy Algorithm`

Input : Set of anchors A, anchor number H, DCGASC function f and sets A_j covered by each anchor a_j
Output: An anchor point order list $S \subseteq A$ with $|S| = H$

$S \leftarrow \{\arg\max_{a_j \in A} |A_j|\}$;
while $|S| < H$ **do**
  $a_j \leftarrow \arg\max_{a'_j \in A \setminus S} f(S \cup \{a'_j\}) - f(S)$
  $S \leftarrow S \cup \{a_j\}$
end
return S

described in [23]. Algorithm. 1 shows the greedy algorithm: it first obtains the anchor point which cover other points at most, then uses Eq. 11 to get the following anchor points in turn.

4.3 Learning Algorithm

Alternating Least Square (ALS) is a popular approach to optimize Weighted Matrix Factorization [10]. [24] proposed a fast element-wise ALS learning algorithm which optimizes each coordinate of the latent vector with the other fixed and sped up learning by avoiding the massive repeated computations introduced by the weighted missing data. In this paper, we use the element-wise ALS learning algorithm to learn the sub-matrix latent vectors. More Specifically, the latent factors of user u is updated based on

$$P_{uk}^h = \frac{\sum_{m \in V^h} (R_{um} - \hat{R}_{um,k}^h) T_{um}^h W_{um} Q_{mk}^h}{\sum_{m \in V^h} T_{um}^h W_{um} Q_{mk}^h Q_{mk}^h + \lambda_P^h} \tag{12}$$

where $\hat{R}_{um,k}^h = \hat{R}_{um}^h - P_{uk}^h Q_{mk}^h$, *i.e.*, the prediction without the component of latent factor k . The sub-matrix weight T_{um}^h is the only difference in Eq. 12 with the original WMF which may lead to high running time. Fortunately, due to $T_{um}^h = E_b(u_h, u) \times E_b(m_h, m)$ and $\lambda_P^h = \lambda_P E_b(u_h, u)$, we also can speed up learning by memorizing the massive repeated computations. Firstly, $E_b(u_h, u)$ is both in the numerator and denominator so it can be canceled. Then, we focus on the numerator:

$$\begin{aligned} &\sum_{m \in V^h} (R_{um} - \hat{R}_{um,k}^h) E_b(m_h, m) W_{um} Q_{mk}^h \\ &\quad = \sum_{m \in V_u^h} [W_{um} R_{um} - (W_{um} - 1) \hat{R}_{um,k}^h] E_b(m_h, m) Q_{mk}^h \\ &\qquad - \sum_{m \in V^h} E_b(m_h, m) \hat{R}_{um,k}^h Q_{mk}^h \end{aligned} \tag{13}$$

where V_u^h means the item set of user u in the h-th sub-matrix. Because $E_b(m_h, m)$ is the same for different user u, the cache method can also be utilized here. The $\sum_{m \in V^h} E_b(m_h, m) \hat{R}_{um,k}^h Q_{mk}^h$ term can be speed up:

$$\begin{aligned} &\sum_{m \in V^h} E_b(m_h, m) \hat{R}_{um,k}^h Q_{mk}^h \\ &\qquad = \sum_{f \neq k} P_{uf} \sum_{m \in V^h} E_b(m_h, m) Q_{mk} Q_{mf} \end{aligned} \tag{14}$$

So the $\sum_{m \in V^h} E_b(m_h, m) Q_{mk} Q_{mf}$ can be pre-computed and used in updating the latent vectors for all users. Similarly, the same cache method can be used in the calculate of denominator. We define the S^Q as $S^Q = \sum_{m \in V^h} E_b(m_h, m) Q_m Q_m^\top$, so Eq. 12 can be calculated as:

$$\begin{aligned} P_{uk}^h = \{ &\sum_{m \in V_u^h} [W_{um} R_{um} - (W_{um} - 1) \hat{R}_{um,k}^h] E_b(m_h, m) Q_{mk}^h \\ &- \sum_{f \neq k} P_{uf} S_{fk}^Q \} / \{ \sum_{m \in V_u^h} E_b(m_h, m)(W_{um} - 1) Q_{mk}^h Q_{mk}^h \\ &+ S_{kk}^Q + \lambda_P \} \end{aligned} \tag{15}$$

where S_{fk}^Q is the (f, k)-th element of the S^Q. Similarly, we define the S^P as $S^P = \sum_{u \in U^h} E_b(u_h, u) P_u P_u^\top$ and the update of item latent vectors is:

$$\begin{aligned} Q_{mk}^h = \{ &\sum_{u \in U_m^h} [W_{um} R_{um} - (W_{um} - 1) \hat{R}_{um,k}^h] E_b(u_h, u) P_{uk}^h \\ &- \sum_{f \neq k} P_{mf} S_{fk}^P \} / \{ \sum_{u \in V_m^h} E_b(u_h, u)(W_{um} - 1) P_{uk}^h P_{uk}^h \\ &+ S_{kk}^P + \lambda_Q \} \end{aligned} \tag{16}$$

So with the local sub-matrix weights, one iteration takes $O(NK^2 + MK^2 + |R|K)$ time as the same as the fast element-wise ALS [24].

Algorithm. 2 summarizes the process of learning local weighted latent vectors. First we use the fast element-wise ALS [23] to learn the global latent vectors (Line 1). Then we obtain the anchor set by Algorithm. 1. At last, we adopt the fast element-wise ALS to learning every sub-matrix latent vectors.

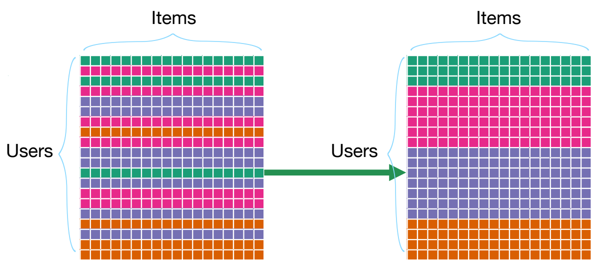


Fig. 2 User-based Local Matrix Factorization

Algorithm 2: `LWMF Learning Algorithm`

Input : R, anchor number H, DCGASC function J and sets A^i covered by each anchor a_i,W,λ_P and λ_Q, the number K of dimensions for latent vectors
Output: Latent featue matrix sets $\mathbf{P} = \{P^1, P^2, ..., P^H\}$ and $\mathbf{Q} = \{Q^1, Q^2, ..., Q^H\}$

Use fast element-wise ALS [24] to learn the whole latent vectors P and Q;
Use algorithm. 1 to get anchor set $S = \{a_{j1}, a_{j2}, ..., a_{jH}\}$;
for $h \longleftarrow 1$ *to* H **do**
  for $u \in U$ **do** calculate $E_b(u, u_h)$;
  for $m \in V$ **do** calculate $E_b(m, m_h)$;
  //update local user latent vectors $S^Q = \sum_{m \in V^h} E_b(m_h, m) Q_m Q_m^\top$;
  for $u \in U$ **do**
    for $m \in V_u^h$ **do** $\hat{r}_{um}^h \longleftarrow P_u^{h\top} Q_m^h$
    for $k \longleftarrow 1$ *to* K **do**
      for $m \in V_u^h$ **do** $\hat{r}_{um,k}^h \longleftarrow \hat{r}_{ui} - P_{uk} Q_{mk}$
      calculate P_{uk} using Eq. 15;
      for $m \in V_u^h$ **do** $\hat{r}_{um,k}^h \longleftarrow \hat{r}_{ui} + P_{uk} Q_{mk}$
    end
  end
  //update local item latent vectors $S^P = \sum_{u \in U^h} E_b(m = u_h, u) P_u P_u^\top$;
  for $m \in V$ **do**
    for $u \in U_m^h$ **do** $\hat{r}_{um}^h \longleftarrow P_u^{h\top} Q_m^h$
    for $k \longleftarrow 1$ *to* K **do**
      for $u \in U_m^h$ **do** $\hat{r}_{um,k}^h \longleftarrow \hat{r}_{ui} - P_{uk} Q_{mk}$
      calculate Q_{mk} using Eq. 16;
      for $u \in U_m^h$ **do** $\hat{r}_{um,k}^h \longleftarrow \hat{r}_{ui} + P_{uk} Q_{mk}$
    end
  end
end
return $\mathbf{P}, \mathbf{Q}$

4.4 User-based Local Weighted Matrix Factorization

The above method LWMF uses the selected sub-matrices to model the *local* property and ignore global feature. Especially for the item recommendation problem, we should recommend items for a user from all the items. So we propose a variant method, called User-based Local Weighted Matrix Factorization, which only considers users to select the anchor points and puts all items into the sub-matrix. Given the user set $U = \{u_1, u_2, ..., u_N\}$ (all users) while every user u_i can cover itself several other users denoted by $U^i = \{u_i, u_{i1}, u_{i2}, ..., u_{iD}\} \subset U$, we need to maximize the following function:

$$\max J(S) = \sum_{h=1}^{H} \sum_{u_l \in U^{jh}} \alpha^{o_{lh}-1} (1 - max_{h' \in \{1,...,h-1\}} E_b(u_{jh}, u_{jh'})) \quad s.t. |S| = H \tag{17}$$

Obviously, this user-based DCGASC function is also submodular and also monotone nondecreasing. Fig. 2 shows the user-based LWMF to select the sub-matrices. Because we do not need to consider the items, it is much faster to the user anchor point set. Moreover, user-based LWMF is more reasonable for item recommendation problem. As a direct comparison of user-based LWMF, we also implement item-based LWMF which only considers items to select the anchor points and lets all users into the sub-matrix.

5 Experiments

In this section, we evaluate the method proposed in this paper using real datasets. We first introduce the datasets and experimental settings. Then we compare our method with WMF under specific parameter settings. We also compare results with different anchor numbers and two anchor points selection methods.

5.1 Dataset

We choose two real-world datasets from [**?**]. One is the Foursquare check-in data made in Singapore between Aug.2010 and Jul.2011, and another is the Gowalla check-in data made in California and Nevada between Feb.2009 and Oct.2010. Both are popular online LBSNs datasets.

The Foursquare check-in data comprises 194,108 check-ins made by 2,312 users at 5,596 POIs, and the density is 1.49×10^{-2}. The Gowalla check-in data comprises 456,988 check-ins made by 10,162 users at 24,250 POIs, and the density is 1.86×10^{-3}. Two datasets are very sparse.

More details about two datasets are showed in the Table 3. We randomly select 80% of each user's visiting locations as the training set and the rest 20% as the testing set.

5.2 Setting

Next, we show the parameter values. The regularization λ is set to 10 and the performance of recommendation is not sensitive to this parameter. The weight parameter ε for Fousquare is set to 2 and for Gowalla is set

Table 2 Precision and Recall Comparison on Foursquare and Gowalla, where column 'Improve' indicates the relative improvements that our approach LWMF achieves relative to the basic WMF results

ALL	Metrics	MP	KNN$_u$	KNN$_m$	WMF	LWMF$_{both}$	LWMF$_m$	LWMF$_u$	Improve
Foursquare	Precision	0.0615	0.0741	0.0698	0.0792	0.0823	**0.0869**	0.0852	9.80%
d=5	Recall	0.0680	0.8212	0.7975	0.0905	0.0952	0.0962	**0.0999**	10.34%
d=10	Precision	0.0615	0.0741	0.0698	0.0847	0.0847	0.0878	**0.0898**	6.03%
	Recall	0.0680	0.8212	0.7975	0.0993	0.0995	0.0990	**0.1047**	5.44%
d=20	Precision	0.0615	0.0741	0.0698	0.0844	0.0832	0.0893	**0.0915**	8.39%
	Recall	0.0680	0.8212	0.7975	0.0980	0.0982	0.1021	**0.1067**	8.85%
d=40	Precision	0.0615	0.0741	0.0698	0.0741	0.0828	**0.0907**	**0.0902**	22.45%
	Recall	0.0680	0.8212	0.7975	0.0922	0.0945	0.1028	**0.1054**	14.27%
Gowalla	Precision	0.0203	0.0552	0.0587	0.0321	0.0489	0.0478	0.0445	52.56%
d=5	Recall	0.0460	**0.1055**	0.1014	0.0664	0.0923	0.0884	0.0881	39.05%
d=10	Precision	0.0203	0.0552	**0.0587**	0.0385	0.0528	0.0526	0.0504	37.10%
	Recall	0.0460	**0.1055**	0.1014	0.0779	0.0990	0.0936	0.0989	27.01%
d=20	Precision	0.0203	0.0552	**0.0587**	0.0442	0.0558	0.0565	**0.0581**	31.44%
	Recall	0.0460	0.1055	0.1014	0.0871	0.1035	0.1006	**0.1110**	27.41%
d=40	Precision	0.0203	0.0552	0.0587	0.0485	0.0578	0.0584	**0.0623**	28.36%
	Recall	0.0460	0.1055	0.1014	0.0953	0.1067	0.1034	**0.1191**	25.04%

Table 3 Detail information of Gowalla and Foursquare

	Foursquare	Gowalla
#user	2,321	10,162
#locations	5,596	24,238
#check-ins	194,108	456,967
avg. #users per loc.	34.69	18.85
avg. #loc. per user	83.63	44.97
max #users per loc.	695	2,195
max #loc. per user	311	1,113

to 3. We set the bandwidth parameter in Epanechanikov kernel as $b_u = b_m = 0.8$. The discount α of DCGASC is set to 0.4. We select 100 anchor points for both datasets. In the experiments, we observe that if the number of anchor points is larger, the performance is better. But the training time increases accordingly.

We employ the Precision@N and Recall@N to measure the performance. For a user u, we set $\mathcal{I}^P(u)$ as the predicted item list and $\mathcal{I}^T(u)$ as the true list in the testing dataset. So the Precision@N and Recall@N are:

$$Precision@N = \frac{1}{|U|}\sum_{u\in U}\frac{|\mathcal{I}^P(u)\bigcap\mathcal{I}^T(u)|}{N}$$

$$Recall@N = \frac{1}{|U|}\sum_{u\in U}\frac{|\mathcal{I}^P(u)\bigcap\mathcal{I}^T(u)|}{|\mathcal{I}^T(u)|}$$

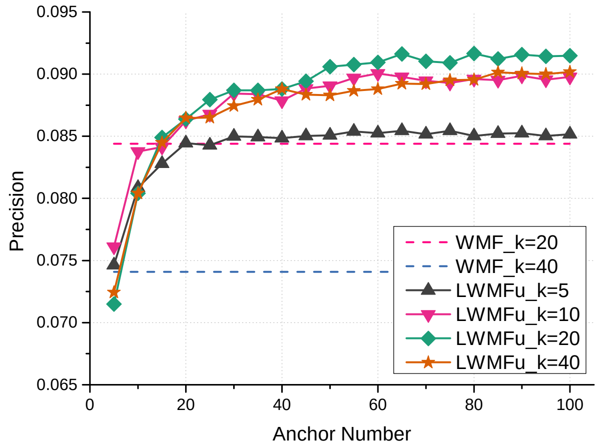
where $|\mathcal{I}^P(u)| = N$. In our base experiments, we choose top 10 as evaluation metrics.

We compare seven methods for implicit feedback datasets:

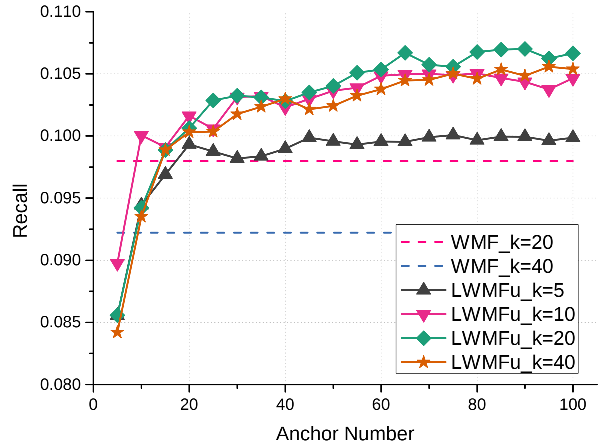
– MostPopular: This is the most basic method, which recommends the most popular items to the target user.
– KNN$_u$: This is user-based CF method ,where user-user similarity is calculated based on the training data.
– KNN$_m$: This method is similar with UserKNN, and the difference is that ItemKNN calculates item-item similarity based on the training data.
– WMF: This is the state-of-the-art method which is a whole-data based learning approach setting a uniform weight to missing entries [10,24].
– LWMF$_{both}$:This is our proposed method that employs the kernel function to intensify *local* property and the weight function to explore user preferences.
– LWMF$_u$: A variant method of LWMF$_{both}$ which only considers users to select the anchor points and puts all items into the sub-matrix.
– LWMF$_m$: A variant method of LWMF$_{both}$ which only considers items to select the anchor points and puts all users into the sub-matrix.

Then we compare two anchor points selection methods to study the performance of LWMF:

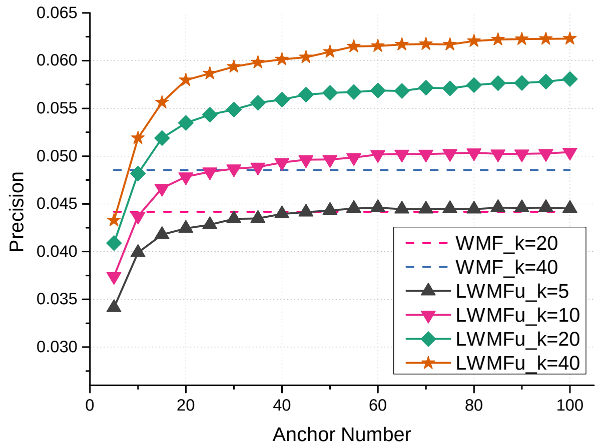
– Random: Sampling anchor points uniformly from training dataset as paper [7] does.



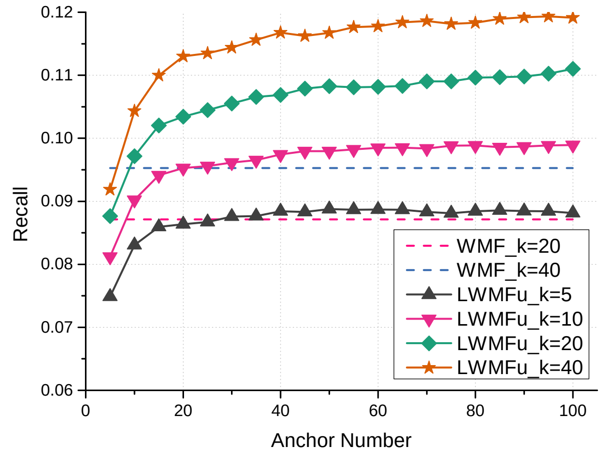
(a) Precsion on Foursquare



(b) Recall on Foursquare



(c) Precsion on Gowalla



(d) Recall on Gowalla

Fig. 3 Comparison with different number of anchor points

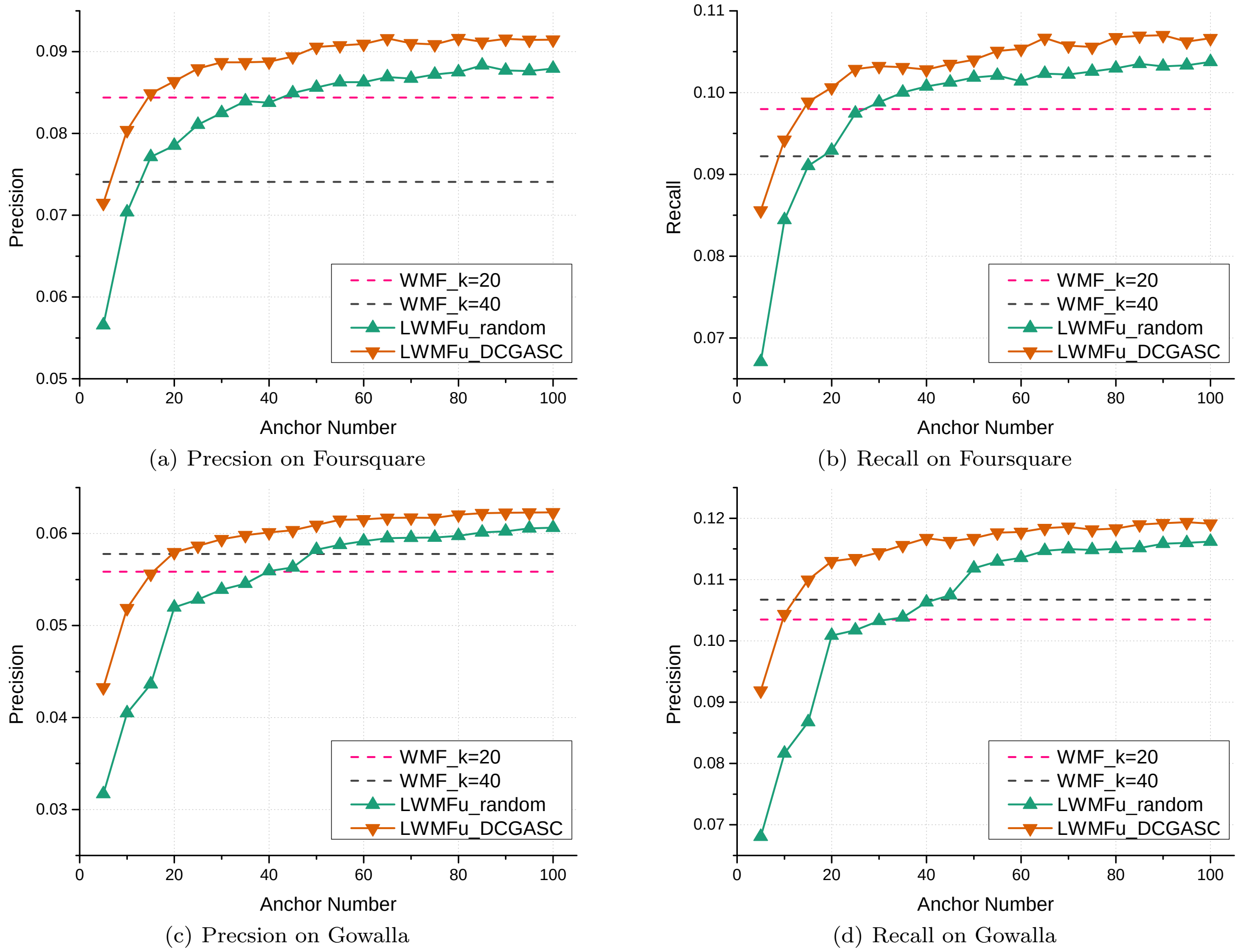
– Discounted Cumulative Gain Anchor Set Cover (DCGASC):Discounting cumulative gain of covering the points which is also submodular and monotone.

So LWMF can be expanded into two sub-methods LWMF_Random, LWMF_DCGASC. By defualt, LWMF means LWMF_DCGASC. Each method is conducted 5 times independently. Therefore, the average score indicates the performance of the recommendation methods.

In this section, we discuss the experimental results on Foursquare and Gowalla datasets.

5.2.1 Recommendation Methods Comparison

Table 2 lists the precision and recall of seven methods mentioned above on Foursquare and Gowalla datasets. It shows the same result as [7] that LORMA outperforms SVD, and LWMF always outperforms WMF.The performances of WMF and LWMF are increasing with the increase of K. However, on Foursquare , when K gets to 40, the performances both fall, which indicates that the value of K has resulted in overfitting. So we choose K to be 20. On the other hand, the experiments based on Gowalla dataset show that the value of K is bigger than 40 when the performance is best. It is obvious that performances of all LWMF methods are better than WMF methods in all dimensions. Especially on Gowalla dataset, the precision and recall of LWMF are more than 25 percent better than WMF. Specifically when K equals to 5, the precision of LWMF$_{both}$ are 52.56 percent better than WMF. More obvious improvements on Foursquare and Gowalla is due to the local property. For example, there are some business districts in a city and business POIs are geographically close to each other within each business district. Additionally as for our three approaches, we can find that the differences between their performances are not very obvious. But from an overall view, LWMF$_u$ are better than the other two methods. LWMF$_u$ does the recommendation task based on users, so it can be infered that selecting points based on users are more reasonable



(a) Precsion on Foursquare
(b) Recall on Foursquare
(c) Precsion on Gowalla
(d) Recall on Gowalla

Fig. 4 Anchor Point Set Selection Methods Comparison

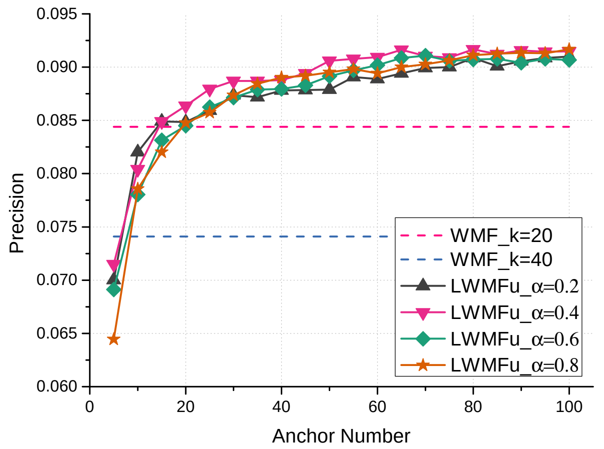
than the other two methods. We also do the comparison of three basic methods, which are MostPopular, KNN_u and KNN_m. The experimental results indicate that our methods are better than these three basic methods. Although KNN_u and KNN_m are better than LWMF when K is low on Gowalla , the performance of LWMF goes up with the increase of K and is far more better than KNN_u and KNN_m.

5.2.2 Comparison with different number of anchor points

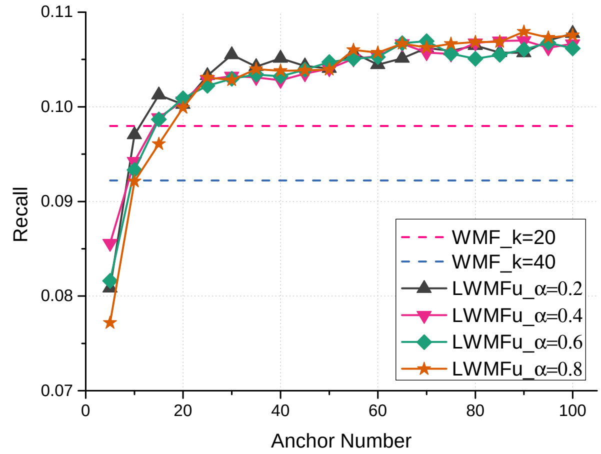
Fig. 3 shows the performance of LWMF with different anchor numbers. For both datasets, the precision and recall of both LWMF and WMF improve while K increases and LWMF performs better than WMF with $K \geq 20$. For Foursquare dataset, LWMF with $K = 20$ and anchor number $n \geq 20$ outperforms WMF with $K = 20$. While the same performance on Gowalla dataset needs $n \geq 40$ anchor points. We can see that as the number of anchor points increases, the performance gets better. When the number of anchor points gets to 50,we can get a good performance. Although the training time increases, the gap of running time of matrix factorization between LWMF and WMF is small. Because the running time of WMF is $O(K|R| + K^2N + K^2M)$ and the sub-matrices of LWMF are much smaller than the original matrix (i.e., in both datasets, each sub-matrix is about 10% of original matrix averagely). Only one sub-matrix factorization is much faster than original matrix factorization. Despite all this, LWMF costs more time on calculating the KDE between users and items and selecting anchor points.

5.2.3 Anchor Point Set Selection Methods Comparison

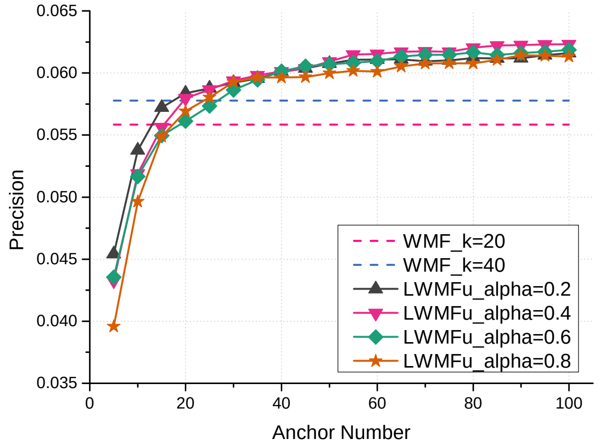
Next, we compare the performance of LWMF_u_Random and LWMF_u in Fig. 4.The discount parameter α is set 0.4. K is set to 20 for Foursquare dataset while 40 for Gowalla dataset. From Fig. 4, when the number of anchor points is small, LWMF_u performs better in preci-



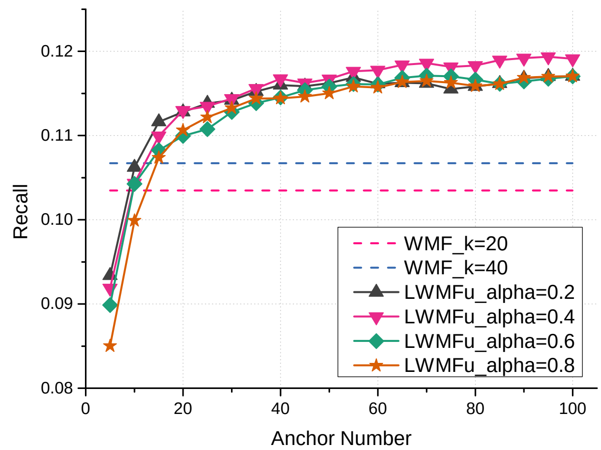
(a) Precsion on Foursquare



(b) Recall on Foursquare



(c) Precsion on Gowalla



(d) Recall on Gowalla

Fig. 5 Comparison with different discounts of anchor points

sion and recall. When the number of anchor points increases, the gap of performance among three gets less. Despite of this, LWMF$_u$ outperforms LWMF$_u$_Random on both datasets.

5.2.4 Comparison with Different Discounts for DCGASC

Finally, we study the performance of LWMF$_u$ with different discount parameters. K is set to 20 for Foursquare dataset while 40 for Gowalla dataset. For each α, we explore results obtained by varying the parameter in the range $(0, 1]$ with decimal steps. Because the results with discount parameter $\alpha \in [0.2, 0.8]$ are similar, we only plot the curves with $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$ in Fig. 5. The gap of performance with four discount parameters is small. The performance with discount parameter $\alpha = 0.4$ is better slightly. In general, the performance of LWMF is not sensitive to the discount parameter but mainly depends on the number of anchor points.

6 Conclusion and Future Work

In this paper, we propose LWMF which selects submatrices to model the user behavior better. LWMF relieves the sparsity problem by sub-matrix factorization. Moreover, we propose DCGASC to select sub-matrix set which improves the performance of LWMF. The extensive experiments on two real datasets demonstrate the effectiveness of our approach compared with state-of-the-art method WMF.

We want to study the three further directions: (1) To speed up selecting sub-matrices; (2) In this paper, we first select the sub-matrix set by selecting anchor points, then do the weighted matrix factorization for each sub-matrix. So we need two steps to optimize the objective function. We can try to find the methods to optimize the local matrix factorization in only one objective function; (3) We can further leverage other special additional information into LWMF in some special scenarios (such as, the geographical information in POI recommender).

Acknowledgements This work was supported by the NSFC grants (No. 61472141, 61370101 and 61021004), Shanghai Leading Academic Discipline Project (No. B412), and Shanghai Knowledge Service Platform Project (No. ZF1213).

References

1. Deshpande M, Karypis G. Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 143-177.
2. Paterek A. Improving regularized singular value decomposition for collaborative filtering. Proceedings of KDD cup and workshop. 2007, 2007: 5-8.
3. Mnih A, Salakhutdinov R. Probabilistic matrix factorization. Advances in neural information processing systems. 2007: 1257-1264.
4. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009 (8): 30-37.
5. Mackey L W, Jordan M I, Talwalkar A. Divide-and-conquer matrix factorization. Advances in Neural Information Processing Systems. 2011: 1134-1142.
6. Zhang Y, Zhang M, Liu Y, et al. Localized matrix factorization for recommendation based on matrix block diagonal forms. Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013: 1511-1520.
7. Lee J, Kim S, Lebanon G, et al. Local low-rank matrix approximation. Proceedings of The 30th International Conference on Machine Learning. 2013: 82-90.
8. Lee J, Bengio S, Kim S, et al. Local collaborative ranking. Proceedings of the 23rd international conference on World wide web. ACM, 2014: 85-96.
9. Beutel A, Ahmed A, Smola A J. ACCAMS: Additive Co-Clustering to Approximate Matrices Succinctly. Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015: 119-129.
10. Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on. IEEE, 2008: 263-272.
11. Pan R, Zhou Y, Cao B, et al. One-class collaborative filtering. Eighth IEEE International Conference on Data Mining. IEEE, 2008: 502-511.
12. Pan R, Scholz M. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 667-676.
13. Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2009: 452-461.
14. Rendle S, Freudenthaler C. Improving pairwise learning for item recommendation from implicit feedback. Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014: 273-282.
15. Yang D, Chen T, Zhang W, et al. Local implicit feedback mining for music recommendation. Proceedings of the sixth ACM conference on Recommender systems. ACM, 2012: 91-98.
16. Ilievski I, Roy S. Personalized news recommendation based on implicit feedback. Proceedings of the 2013 International News Recommender Systems Workshop and Challenge. ACM, 2013: 10-15.
17. Zibriczky D, Hidasi B, Petres Z, et al. Personalized recommendation of linear content on interactive TV platforms: beating the cold start and noisy implicit user feedback. UMAP Workshops. 2012.
18. Lian D, Zhao C, Xie X, et al. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 831-840.
19. Liu Y, Wei W, Sun A, et al. Exploiting geographical neighborhood characteristics for location recommendation. Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014: 739-748.
20. Cho E, Myers S A, Leskovec J. Friendship and mobility: user movement in location-based social networks. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011: 1082-1090.
21. Chapelle O, Metlzer D, Zhang Y, et al. Expected reciprocal rank for graded relevance. Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 621-630.
22. Clarke C L A, Kolla M, Cormack G V, et al. Novelty and diversity in information retrieval evaluation. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2008: 659-666.
23. Nemhauser G L, Wolsey L A, Fisher M L. An analysis of approximations for maximizing submodular set functionsI[J]. Mathematical Programming, 1978, 14(1): 265-294.
24. He, X., Zhang, H., Kan, M.Y. and Chua, T.S., 2016. Fast matrix factorization for online recommendation with implicit feedback. In Proc. of SIGIR (Vol. 16).
25. Guillory, A. and Bilmes, J., 2010. Interactive submodular set cover. arXiv preprint arXiv:1002.3345.
26. Chen C, Li D, Zhao Y, et al. WEMAREC: Accurate and Scalable Recommendation through Weighted and Ensemble Matrix Approximation[C]//Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015: 303-312.
27. GB/T 7714 Pilszy I, Zibriczky D, Tikk D. Fast als-based matrix factorization for explicit and implicit feedback datasets[C]//Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010: 71-78.
28. R. Devooght, N. Kourtellis, and A. Mantrach. Dynamic matrix factorization with priors on unknown values. In KDD 2015, pages 189-198.
29. Yuan Q, Cong G, Ma Z, et al. Time-aware point-of-interest recommendation[C]//Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2013: 363-372.