Interval-valued Matrix Factorization with Applications

Zhiyong Shen^{1,3}, Liang Du^{2,1}, Xukun Shen³, Yidong Shen²

¹ Hewlett Packard Labs China, zhiyongs@hp.com

² State Key Laboratory of Computer Science, China,{duliang,ydshen}@ios.ac.cn

³State Key Laboratory of Virtual Reality Technology and system, China, xkshen@vrlab.buaa.edu.cn

Abstract—In this paper, we propose the Interval-valued Matrix Factorization (IMF) framework. Matrix Factorization (MF) is a fundamental building block of data mining. MF techniques, such as Nonnegative Matrix Factorization (NMF) and Probabilistic Matrix Factorization (PMF), are widely used in applications of data mining. For example, NMF has shown its advantage in Face Analysis (FA) while PMF has been successfully applied to Collaborative Filtering (CF). In this paper, we analyze the data approximation in FA as well as CF applications and construct interval-valued matrices to capture these approximation phenomenons. We adapt basic NMF and PMF models to the interval-valued matrices and propose Interval-valued NMF (I-NMF) as well as Intervalvalued PMF (I-PMF). We conduct extensive experiments to show that proposed I-NMF and I-PMF significantly outperform their single-valued counterparts in FA and CF applications.

Keywords-Matrix factorization, uncertainty

I. Introduction

Exploring data approximation has attracted much attention in uncertain data mining [1] and privacy preserving data mining [2]. Data approximation might be caused by limitations of measuring, delayed data update or intensional data perturbation. When traditional data mining techniques are employed, the consideration of data approximation may improve the quality of results. Thus, various data mining techniques, such as clustering, classification, association mining have been adapted to handling data approximation. In this paper, we devote to inject data approximation into Matrix Factorization (MF) techniques. MF, also known as matrix decomposition, underlies many data mining techniques including clustering, dimensionality reduction and missing data prediction etc.. It decomposes an input data matrix into a number of low-rank factor matrices, which leads to a more compact linear approximation for the original data matrix.

Variations MF have been extensively studied in literatures. In this paper, we pay special attention to *Nonnegative Matrix Factorization* (NMF) [3], [4] and *Probabilistic Matrix Factorization* (PMF) [5]. Each of these MF techniques is suited for a particular class of applications. For example, NMF has shown its advantage in Face Analysis (FA) [4]. In FA applications, each face is represented by a feature vector. NMF factorizes the matrix of multiple face feature vectors into factor matrices and then achieve a more compact representation of the original face data. On the other hand,

PMF has been successfully applied to Collaborative Filtering (CF) [6]. CF is one of the most successful techniques for automatic recommendation systems which need only an observed rating matrix as input. PMF decomposes the sparse rating matrix into user profile matrix and item profile matrix, and then makes predictions for the unknown entries. However, traditional NMF and PMF ignore the following data approximation phenomenons in FA and CF.

Alignment approximation in FA: The faces need to be rotated and aligned to make sure that same columns in the data matrix are corresponding to the same positions in faces. Such alignment is hardly to be perfect in practice, i.e. there is approximation with the alignment in FA applications (see Section II-A for details).

Rating approximation in CF: When a user rates an item in a real-life rating system she/he usually selects a discretized *rating value* which is close to the ideal numerical *preference value* (the exact preference degree). Thus, the rating matrix does contain approximations to some degree (see Section II-B for details).

Interval bounds are better than single-valued variables to describe the above phenomenons of approximation. Many application areas have taken advantage of interval-valued data analysis (see for instance [7]), such as object tracking, market analysis, quantitative economics and so on. In traditional MF techniques, input data matrices might be real values, non-negative values or binary values etc., all of which are single-valued. In this paper, we introduce a new type of data matrix – *interval-valued* matrix to MF, which captures approximation in the observed data matrix. Then, we propose a novel MF framework — *Interval-valued Matrix* Factorization (IMF) to decompose such matrices. Under the IMF framework, we inject data approximation into NMF and PMF and extend them to interval-valued NMF (I-NMF for short) and interval-valued PMF (I-PMF for short). Therefore, our work is a marriage between interval-valued data analysis [7] and MF, and our contributions on both sides of research area are summarized as follows

- We analyze the alignment approximation in FR as well as the rating approximation in CF, and formalize them with interval-valued matrices (Section II).
- We propose the IMF framework, under which we extend two representative basic MF techniques NMF



and PMF to I-NMF and I-PMF which are capable handling interval-valued matrices (Section IV). We conduct extensive experiments to show that the proposed I-NMF and I-PMF significantly outperfetheir traditional single-valued counterparts in FA CF applications (Section V).

II. INTERVAL-VALUED MATRIX AND DATA APPROXIMATION

In this section we formalize the approximation in CF and FR problems with interval-valued matrices. First of all, give formal definitions of interval-valued matrix.

Let $X \in \mathbb{R}^{n \times d}$ denote the input data matrix, v entries denoted as X_{ij} . Let I(X) denote the interval-val matrix corresponding to X, and we have the following equivalent representations for I(X).

Definition 1 (Center-radius representation). We denote interval with center X_{ij} and radius δ_{ij} as

$$I(X_{ij}) = \langle X_{ij}, \delta_{ij} \rangle \tag{1}$$

For entire matrices, we have $I(X) = \langle X, \delta \rangle$.

Definition 2 (Min-max representation). We denote the interval bounds as $X_{ij}^{\text{low}} = X_{ij} - \delta_{ij}$ and $X_{ij}^{\text{up}} = X_{ij} + \delta_{ij}$.

$$I(X_{ij}) = [X_{ij}^{\text{low}}, X_{ij}^{\text{up}}]$$
(2)

For entire matrices, we have $I(X) = [X^{low}, X^{up}]$.

matrices rather than interval-valued ones. In the following subsections we'll give the empirical method to construct \boldsymbol{X}) based on \boldsymbol{X} . The above definitions have already been adopted in interval-valued data analysis [8]. In our work, we'll use the center-radius representation (Definition 1) to formalize the rating approximation in CF and alignment approximation in FR and then construct interval-valued matrices. The min-max representation (Definition 2) will be used as input for the proposed IMF models introduced in Section IV.

In practice, we might only observe single-valued data

A. Alignment Approximation in FR

In many FA techniques, we need to align the faces image such that, ideally, the pixels with the same coordinates correspond to the identical positions of a face. In Figure 1, we take the position of the nose tip as an example to show that the alignment is not perfect. Although the same position of a face is not exactly aligned, they should be near to each other. Take the first row as examples, the pixel with coordinate (33,35) may corresponds to the face position coordinated by (41,34) in the second image, or (33,40) in the third and so on. Formally, the value of a pixel with coordinates $(x,y),x\in\{1,...,d_x\},y\in\{1,...,d_y\}$, might correspond to a pixel with coordinates $(x+\Delta x,y+\Delta y),$ $\leq \Delta x, \Delta y \leq r$. In MF, the i'th face is represented by 0

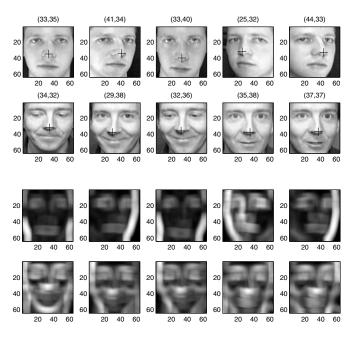


Figure 2. An example of δ matrix corresponding to faces in Figure 1.

a vector X_i . with dimensionality $d = d_x \times d_y$. We use $(x^{(i,j)}, y^{(i,j)})$ to denote the coordinates of pixels in the i'th image which corresponds to the j'th element in vector X_i , namely X_{ij} . Then, we define the following set of the entries in X for each X_{ij}

$$S_{ij}^{\text{FA}(r)} = \{ X_{ij'} | |x^{(i,j')} - x^{(i,j)}| \le r \land |y^{(i,j')} - y^{(i,j)}| \le r \}$$
(3)

The elements in $\mathcal{S}^{\mathrm{FA}(r)}_{ij}$ correspond to pixels around $(x^{(i,j)},y^{(i,j)})$ in a range r. Intuitively, X_{ij} may corresponds to a value in the interval of $[\min(\mathcal{S}^{\mathrm{FA}(r)}_{ij}),\max(\mathcal{S}^{\mathrm{FA}(r)}_{ij})]$, which coincides the min-max definition (Definition 2). However, min-max statistics are not robust in practice and alternatively, we construct $I(X_{ij})$ based on the standard deviation to capture the variation in $\mathcal{S}^{\mathrm{FA}(r)}_{ij}$. According to Definition 1, we set X_{ij} as the center of $I(X_{ij})$ and calculate the radius via

$$\delta_{ij}^{\text{FA}(r)} := \alpha \cdot \text{std}(\mathcal{S}_{ij}^{\text{FR}(r)})$$
 (4)

where $\alpha \in \mathbb{R}^+$ is a multiplicative scale coefficient. Based on Defintion 2, it's easy to calculate the bounds of interval-valued input for I-NMF according to min-max representation (Definition 2). Examples of the $\delta_i^{\mathrm{FR}(r)}$ corresponding to the faces in Figure 1 are shown in Figure 2, where lighter gray level represents larger radius. In Figure 2, we can see positions such as eyes or nose have larger radius. These positions are more sensitive to alignment error that may hurt the performance of single-valued techniques. With a relatively large radius, the interval-valued techniques may be more tolerant to such alignment errors.

Table I EXAMPLES OF SINGLE-VALUED AND INTERVAL-VALUED RATING MATRICES FOR CF

(a) A single-valued rating matrix: X

	m_1	m_2	m_3	m_4	m_5
u_1		1	4		5
u_2	3		1	2	
u_3		1			4
u_4	5				
u_5		1	4	2	
u_6		3		2	5

(b) A interval-valued rating matrix: I(X)

	m_1	m_2	m_3	m_4	m_5
u_1		[0.6, 1.4]	[3.5,4.5]		[4.8,5.2]
u_2	[2.8,3.2]		[0.5, 1.5]	[1.5,2.5]	
u_3		[0.7, 1.3]			[3.5,4.5]
u_4	[4.5,5.5]				
u_5		[0.4, 1.6]	[3.7,4.3]	[1.8,2.2]	
u_6		[2.7, 3.3]		[1.4, 2.6]	[4.2, 5.8]

B. Rating Approximation in CF

In CF, the rating degree is actually an approximate to its actual preference degree of a user u over an item. For example, a web site allows users to rate items from one star to five stars. User u may think the two items a and b are beyond two stars while not worth four stars, and he may prefer a to b. Suppose the continuous preference degrees of user u on a and b are 3.4 and 2.8, respectively. However, due to the constraint of the rating system, u can only rate both a and b as three stars, and the difference between a and b disappears. It also indicates that the rating degree actually represents a continuous interval, which may include the ideal preference degree. Intuitively, the rating degree X_{ij} is affected by both the i'th user and j'th item. Therefore, we define the observations relevant to X_{ij} with the set as follows:

$$S_{ij}^{\text{CF}} = \{X_{i'j'} | (i' = i \lor j' = j) \land (i', j') \in (\mathbf{i}, \mathbf{j})\}$$
 (5)

 $\mathcal{S}^{\mathrm{CF}}_{ij}$ is actually constructed by the observed rating degrees in the *i*-th row and *j*-th column of the rating matrix in CF. Again, we calculate the radius $\delta^{\mathrm{CF}}_{ij}$ for each observed rating degree X_{ij} according to Definition 1 based on the standard deviation of the ratings in $\mathcal{S}^{\mathrm{CF}}_{ij}$:

$$\delta_{ij}^{\text{CF}} := \alpha \cdot \text{std}(\mathcal{S}_{ij}^{\text{CF}})$$
 (6)

where $\alpha \in \mathbb{R}^+$ is again a multiplicative scale coefficient. Intuitively, a user's ratings on different items (or the ratings of a item from different users) vary greatly, we should assign a big value of interval radius to this entry. Then, it's easy to calculate the bounds of interval-valued input for I-PMF according to min-max representation (Definition 2). A example of interval-valued rating matrix with its corresponding single-valued matrix in min-max representation are shown in Table II-A

III. MATRIX FACTORIZATION WITH APPLICATIONS

In this section we briefly discuss the MF techniques with their applications. We devote special attention the NMF and PMF techniques since they serve to be the single-valued counterparts of the proposed IMF models.

MF is a linear approximation data representation for the original data matrix $X \in \mathbb{R}^{n \times d}$. Generally, we have

$$X \to UV$$
 (7)

where $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{k \times d}$. Each data instance X_i is approximated by a linear combination of the rows of V with weight vector U_i , the i'th row of U. Thus, we call U as weight matrix and V as basis matrix. The ranks of U and V are always much lower than the rank of X, i.e. $1 \times M$ in $1 \times M$ and $1 \times M$ as follows

$$\hat{X} \leftarrow UV$$
 (8)

Various assumptions over U and V lead to different MF models which have been widely used in data mining applications. The following two series of applications are relevant to this paper:

Parts-based representation: MF naturally represent the original data matrix X by parts. The rows in V, so-called *basis vectors*, are optimized for the linear approximation for X, and U_i . could be regard as a representation for the X_i . with lower dimensionality. NMF has been successfully applied to find addictive *parts-based* representations for face images (see for detail in Section III-A).

Missing Data Prediction: The reconstructed matrix \hat{X} is a full matrix. Therefore, when X is sparse, we can make prediction for its missing entries based on \hat{X} . For example, PMF is successfully applied to predict the missing entries of the rating matrices in CF (see for detail in Section III-B).

A. Nonnegative Matrix Factorization

NMF aims to factorize a nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{n \times d}$ with two nonnegative matrices $\boldsymbol{U} \in \mathbb{R}_+^{n \times k}$ and $\boldsymbol{V} \in \mathbb{R}_+^{k \times d}$ which minimize the following L_2 loss function

$$\mathcal{L}_{\text{NMF}} = \|X - UV\|_{\text{F}}^{2}$$
s.t. $U \ge 0, V \ge 0$ (9)

where $\|\cdot\|_F$ denotes the Frobenius norm. The estimations of U and V can be find via the *multiplicative update rules* proposed in [3], which iteratively update U and V as follows

$$U_{ij} \leftarrow U_{ij} \frac{(\boldsymbol{X}\boldsymbol{V}^T)_{ij}}{(\boldsymbol{U}\boldsymbol{V}\boldsymbol{V}^T)_{ij}}$$

$$V_{ij} \leftarrow V_{ij} \frac{(\boldsymbol{U}^T\boldsymbol{X})_{ij}}{(\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{V})_{ij}}$$
(10)

The update rules in (10) can be deduced according to Karush-Kunhn-Trucker optimal condition [9] of inequality constraint (see for detail in [10]). In [3], it is proved that the updates in (10) lead to a local minimum of (9). The non-negative constraints on U and V only allow addictive linear combination of basis vectors in V, so-called parts-based representation [4]. NMF is suited for many real world applications such as human face analysis [4]. In human face analysis, the resultant matrix U constructs a optimized representation for the original data instances. Many FA algorithms, such as face recognition, face clustering, may be directly applied on U instead of the original data matrix X.

B. Probabilistic Matrix Factorization

In CF, the PMF model [5] assume that the ratings are drawn from some Gaussian distribution.

$$p(X_{ij}|i,j,\boldsymbol{U},\boldsymbol{V},\sigma^2) = G(X_{ij}|\boldsymbol{U}_{i\cdot}\boldsymbol{V}_{\cdot j},\sigma^2)$$
(11)

For U and V, they place zero-mean spherical Gaussian priors

$$p(\boldsymbol{U}|\sigma_1^2) = \prod_i G(\boldsymbol{U}_{i\cdot}|0,\sigma_1^2 \boldsymbol{I}), \ p(\boldsymbol{V}|\sigma_1^2) = \prod_j G(\boldsymbol{V}_{\cdot j}|0,\sigma_1^2 \boldsymbol{I})$$
(12)

The U and V are computed via over the observed ratings

$$\mathcal{L}_{\text{PMF}} = \|X - UV\|_{\text{F}}^2 + \lambda \left[\|U\|_{\text{F}}^2 + \|V\|_{\text{F}}^2 \right]$$
 (13)

where $\lambda = \sigma^2/\sigma_1^2$. A local minimum of (13) can be found via gradient decent in U_i and V_{ij}

$$\frac{\partial \mathcal{L}_{\text{PMF}}}{\partial U_{i.}} = \sum_{j \in \mathbf{j}_{i}} (U_{i.} V_{.j} - X_{ij}) V_{.j}^{T} + \lambda U_{i.}
\frac{\partial \mathcal{L}_{\text{PMF}}}{\partial V_{.j}} = \sum_{i \in \mathbf{i}_{j}} (U_{i.} V_{.j}^{T} - X_{ij}) U_{i.}^{T} + \lambda V_{.j}$$
(14)

Based on the learning of U and V, we can estimate the unknown ratings in X via

$$\hat{X}_{ij} = U_{i} \cdot V_{\cdot j} \tag{15}$$

IV. INTERVAL-VALUED MATRIX FACTORIZATION

In this section, we introduce the IMF framework. The proposed framework is based on the Min-Max representation of the interval-valued matrix: $I(\boldsymbol{X}) = [\boldsymbol{X}^{\text{low}}, \boldsymbol{X}^{\text{up}}]$. We can extend the original MF over X to the joint MF over $\boldsymbol{X}^{\text{low}}$ and $\boldsymbol{X}^{\text{up}}$. Firstly, we assume each X_{ij} is drawn from a uniform distribution with parameters X_{ij}^{low} and X_{ij}^{up} .

$$X_{ij} \sim \operatorname{uniform}(X_{ij}^{\mathrm{low}}, X_{ij}^{\mathrm{up}})$$
 (16)

Base on this assumption, we have

$$E(X_{ij}) = \frac{1}{2}(X_{ij}^{\text{low}} + X_{ij}^{\text{up}})$$
 (17)

Therefore, we propose to estimate the bounds of I(X) first via the following joint MF

$$X^{\text{low}} \to UV^{\text{low}}, \quad X^{\text{up}} \to UV^{\text{up}}$$
 (18)

We fix the weight matrix U to make a unique profile for each data instance and use $V^{\mathrm{low}}, V^{\mathrm{up}}$ to maintain the data approximation. The reconstructions of X^{low} and X^{up} could be calculated as follows

$$\hat{X}^{\text{low}} \leftarrow UV^{\text{low}}, \quad \hat{X}^{\text{up}} \leftarrow UV^{\text{up}}$$
 (19)

According to (17) and (19), we can reconstruct X via

$$\hat{\boldsymbol{X}} \leftarrow \frac{1}{2}(\hat{\boldsymbol{X}}^{\text{low}} + \hat{\boldsymbol{X}}^{\text{up}}) \tag{20}$$

A. Interval-valued NMF

According to (9) and (18), the L_2 loss function of intervalvalued NMF (I-NMF for short) is

$$\mathcal{L}_{\text{I-NMF}} = \|\boldsymbol{X}^{\text{low}} - \boldsymbol{U}\boldsymbol{V}^{\text{low}}\|_{\text{F}}^{2} + \|\boldsymbol{X}^{\text{up}} - \boldsymbol{U}\boldsymbol{V}^{\text{up}}\|_{\text{F}}^{2}$$
s.t.
$$\boldsymbol{U} \geq 0, \boldsymbol{V}^{\text{low}} \geq 0, \boldsymbol{V}^{\text{up}} \geq 0$$
(21)

Similar to traditional NMF, we have the following multiplicative update rule for U, V^{low} and V^{up} :

$$U_{ij}^{t+1} \leftarrow U_{ij}^{t} \frac{[\boldsymbol{X}^{\text{low}}(\boldsymbol{V}^{\text{low}})^{T} + \boldsymbol{X}^{\text{up}}(\boldsymbol{V}^{\text{up}})^{T}]_{ij}}{[\boldsymbol{U}\boldsymbol{V}^{\text{low}}(\boldsymbol{V}^{\text{low}})^{T} + \boldsymbol{U}\boldsymbol{V}^{\text{up}}(\boldsymbol{V}^{\text{up}})^{T}]_{ij}}$$

$$V_{ij}^{\text{low},t+1} \leftarrow V_{ij}^{\text{low},t} \frac{(\boldsymbol{U}^{T}\boldsymbol{X}^{\text{low}})_{ij}}{(\boldsymbol{U}^{T}\boldsymbol{U}\boldsymbol{V}^{\text{low}})_{ij}}$$

$$V_{ij}^{\text{up},t+1} \leftarrow V_{ij}^{\text{up},t} \frac{(\boldsymbol{U}^{T}\boldsymbol{X}^{\text{up}})_{ij}}{(\boldsymbol{U}^{T}\boldsymbol{U}\boldsymbol{V}^{\text{up}})_{ij}}$$

$$(22)$$

Similar to traditional NMF, we also have that the L_2 loss function $\mathcal{L}_{\rm I-NMF}$ as shown in (21) is nonincreasing under the multiplicative update rules as shown in (22).

Traditional NMF decomposes the original data matrix into two low-rank factor matrices: one profiles the data instances while the other profiles the features. In I-NMF, the proposed the joint matrix factorization framework makes the feature profile factor matrices $\boldsymbol{V}_{.j}^{\text{low}}$ and $\boldsymbol{V}_{.j}^{\text{up}}$ contain the data approximation while preserving a unique profile \boldsymbol{U}_i . for each data instance. We can directly apply the face analysis techniques over \boldsymbol{U} .

B. Interval-valued PMF

In this section we introduce the interval-valued PMF (I-PMF for short). Analogously to (13) and according to (18), we have the following regularized L_2 loss

$$\mathcal{L}_{\text{I-PMF}} = \| \boldsymbol{X}^{\text{low}} - \boldsymbol{U} \boldsymbol{V}^{\text{low}} \|_{\text{F}} + \| \boldsymbol{X}^{\text{up}} - \boldsymbol{U} \boldsymbol{V}^{\text{up}} \|_{\text{F}}^{2} + \lambda \left(\| \boldsymbol{U} \|_{\text{F}}^{2} + \| \boldsymbol{V}^{\text{low}} \|_{\text{F}}^{2} + \| \boldsymbol{V}^{\text{up}} \|_{\text{F}}^{2} \right)$$
(23)

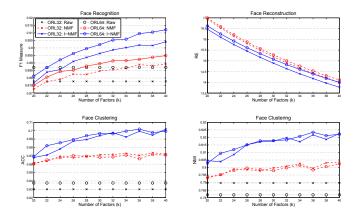


Figure 3. Performance comparison in face analysis.

It is easy to derive a gradient decent in U_i , $V_{\cdot j}^{\text{low}}$ and $V_{\cdot j}^{\text{up}}$ to find a local minimum of (23).

$$\frac{\partial \mathcal{L}_{\text{I-PMF}}}{\partial U_{i}} = \sum_{j \in \mathbf{j}_{i}} [(U_{i}.V_{.j}^{\text{low}} - X_{ij}^{\text{low}})V_{.j}^{\text{low}T} + (U_{i}.V_{.j}^{\text{up}} - X_{ij}^{\text{up}})V_{.j}^{\text{up}T}] + \lambda U_{i}.$$

$$\frac{\partial \mathcal{L}_{\text{I-PMF}}}{\partial V_{.j}^{\text{low}}} = \sum_{i \in \mathbf{i}_{j}} (U_{i}.V_{.j}^{\text{low}} - M_{um}^{\text{low}})U_{i}^{T} + \lambda V_{.j}^{\text{low}}$$

$$\frac{\partial \mathcal{L}_{\text{I-PMF}}}{\partial V_{.j}^{\text{up}}} = \sum_{i \in \mathbf{i}_{j}} (U_{i}.V_{.j}^{\text{up}} - M_{um}^{\text{up}})U_{i}^{T} + \lambda V_{.j}^{\text{up}}$$

$$\frac{\partial \mathcal{L}_{\text{I-PMF}}}{\partial V_{.j}^{\text{up}}} = \sum_{i \in \mathbf{i}_{j}} (U_{i}.V_{.j}^{\text{up}} - M_{um}^{\text{up}})U_{i}^{T} + \lambda V_{.j}^{\text{up}}$$
(24)

For CF application, we can used the learned U, V^{low} and V^{up} to compute the unknown ratings via (19) and (20).

V. EXPERIMENTAL RESULTS

We divide the experiments into two parts: In Section V-A we conduct the comparison of I-NMF against the basic NMF for FA applications, and in Section V-B we compare the performance of I-PMF and PMF over CF applications.

A. Comparison of I-NMF against NMF

We compare the performance of NMF and I-NMF on various FA applications including *face recognition*, *face reconstruction* and *face clustering*.

1) Data Description and Evaluation Setting: We use the Olivertti Research Laboratory (ORL) face data sets to evaluate the NMF and I-NMF models, which contain ten different images of each of 40 distinct persons, $(n=10\times40=400$ in total). Two versions of processed data sets¹: one with resolution 32×32 (ORL32) and the other with 64×64 (ORL64), are used for our experimental evaluation. In ORL32, each face image is represented by a vector with dimensionality $d=32\times32=1024$ while in ORL64, $d=64\times64=4096$.

We implement I-NMF based on multiplicative update rules introduced in Section IV-A. The experiments for NMF

and are based on the DTU NMF toolbox². Various classifiers has been adopted for face recognition and in this paper, we apply the the nearest neighbor method for its simplicity. For face clustering, we choose the popular K-means algorithm. All the classification and clustering algorithms are applied on the output weight matrices U from NMF and I-NMF and we also give the performance of these algorithms over the raw data matrix X as the baseline. In the construction of interval-valued matrices (4), we set r = 5 and $\alpha = 2.5$.

We evaluate the proposed models in terms of the face recognition and clustering effectiveness. Note that face recognition is actually a classification problem. To evaluate the effectiveness of classification (FR), we use the standard F1 measure. We adopt two popular metrics Normalized Mutual Information (NMI) [11] and Clustering Accuracy (ACC) for cluster evaluation. Based on NMF, the faces are reconstructed with the weighted summation of basis vectors. We use the following Reconstruction Error (RE):

$$\mathrm{RE}(\hat{X}, X) = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{d} (\hat{X}_{ij} - X_{ij})^2}{n \times d}}$$
 to evaluate the goodness of reconstruction matrix \hat{X} according to the original data matrix X .

Note that larger values of F1, NMI and ACC indicate better face recognition or clustering results while small values of RE indicate better performance of face reconstruction.

2) Evaluation Results: We compare the models with varying rank of factor matrix k and interval sizes.

Evaluation with varying k: The face clustering and face reconstruction tasks are evaluated over entire data sets. For the face recognition task, we make ten rounds of random sampling of 50% data for training. In general, the performance of NMF and I-NMF for all the face analysis tasks varies with the number of latent factors (k). For each value of k, we run 100 rounds of NMF and I-NMF. The average values of the performance metrics plotted for each model as shown in Figure 3 where each sub-figure corresponds to a face analysis task with the specific evaluation metric and each line corresponds to a model on a specific data set. From Figure 3, we see that I-NMF outperforms NMF with statistical significance over all evaluation metrics on both two data sets.

B. Comparison of I-PMF against PMF

1) Data Description and Evaluation Setting: In this part of experiments, we also use two data sets for evaluation. Movielens data set³ is downloaded from the web-site of GroupLens research group and we use the subset which contains 100,000 ratings for d=1682 movies by n=943 users of the online movie recommender service. We name this data set as Movielens-100K. Netflix data set⁴ is the official data set used in the Netflix Prize competition. Again,

¹http://www.cs.uiuc.edu/homes/dengcai2/Data/FaceData.html

²http://isp.imm.dtu.dk/toolbox/nmf/nmf toolbox ver1.4.zip

³http://www.grouplens.org/system/files/ml-data 0.zip

⁴http://archive.ics.uci.edu/ml/datasets/Netflix+Prize

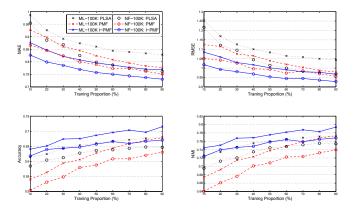


Figure 4. Performance comparison in collaborative filtering.

we sample a subset of this data set which also contains about $100,\!000$ ratings for about d=1999 movies from about n=1319 users. We name this data set as Netflix-100K.

We implement I-PMF based on the gradient decent method introduced in Section IV-B. The experiments for PMF are based on the code downloaded from the homepage of the authors⁵. Since the PMF and I-PMF are both model based techniques for CF, we also implement the Probabilistic Latent Sematic Analysis (PLSA)[12], a representative model based CF approach, as the baseline. The original rating matrix \boldsymbol{X} are separated into two parts. One, denoted by $\boldsymbol{X}^{\text{train}}$, is for training. The other, denoted by $\boldsymbol{X}^{\text{test}}$, is for testing. The interval-valued matrix are constructed from $\boldsymbol{X}^{\text{train}}$ with $\alpha=1$.

We use two series of metrics: predictive error and ranking accuracy to evaluate the model performance. Mean Absolute Error (MAE) and Rooted Mean Squared Error (RMSE) are widely used in CF literatures to evaluate the accuracy of the rating prediction, while Normalized Discounted Cumulative Gain (NDCG) [13] measures the ranking performance. Note that smaller values of MAE and RMSE indicate better prediction result while larger values of NDCG $_N$ indicate better ranking of predictions.

2) Evaluation Results: We compare the models with varying training proportion and the interval sizes.

Evaluation with varying training proportion: For each value of training proportion, we make six rounds of random partitions and train the models on the training set. The first round was used to tune the number of latent factors in PLSA, and the regularization parameters λ in PMF and I-PMF. The average values of the performance metrics are plotted as shown in Figure 4 where each sub-figure corresponds to a evaluation metric and each line corresponds to a model on a specific data set. From Figure 4, we see that I-PMF outperforms PMF in all cases.

VI. CONCLUSIONS

In this paper we propose the IMF framework which injects data approximation into traditional MF via taking intervalvalued matrices as input. Based on the two representative MF techniques: NMF and PMF, we proposed their intervalvalued counterparts according to a joint matrix factorization framework. We analyze the data approximation phenomenons in two data mining applications: FA and CF, and introduce the empirical interval-matrix construction method specific to these applications. The evaluations over multiple real-life data sets with different experimental settings show that I-NMF and I-PMF, which take these interval-valued matrices as input, significantly outperform their corresponding single-valued counterparts.

REFERENCES

- C. C. Aggarwal and P. S. Yu, "A survey of uncertain data algorithms and applications," *IEEE Trans. on Knowl. and Data Engin.*, 2008.
- [2] ——, Privacy-Preserving Data Mining: Models and Algorithms. Springer Publishing Company, Incorporated, 2008.
- [3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. of NIPS*, 2000. [Online]. Available: citeseer.ist.psu.edu/lee01algorithms.html
- [4] —, "Learning the parts of objects by non-negative matrix factorization," *Nature*, 1999.
- [5] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. of NIPS*, 2008.
- [6] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Advan. in Artif. Intel., 2009.
- [7] H. H. Bock, Analysis of Symbolic Data: Exploratory Methods for Extracting Statistical Information from Complex Data. Springer-Verlag New York, Inc., 2000.
- [8] E. d. A. Lima Neto and F. d. A. T. de Carvalho, "Centre and range method for fitting a linear regression model to symbolic interval data," *Comput. Stat. Data Anal.*, 2008.
- [9] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.
- [10] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold," in *Proc. of ICDM*, 2008.
- [11] T. M. Cover and J. A. Thomas, "Elements of information theory," Wiley-Interscience, 1991.
- [12] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, 2001.
- [13] J. Kalervo and K. Jaana, "Cumulated gain-based evaluation of ir techniques," ACM Trans. Inf. Syst., 2002.

⁵http://www.mit.edu/~rsalakhu/BPMF.html