GENERALIZED INTERVAL VALUED NONNEGATIVE MATRIX FACTORIZATION

Masahiro Kohjima *Tatsushi Matsubayashi* *Hiroyuki Toda*

NTT Service Evolution Laboratories, NTT Corporation, Japan

ABSTRACT

In this paper, we propose a probabilistic model for analyzing the generalized interval valued matrix, a matrix that has scalar valued elements and bounded/unbounded interval valued elements. We derive a majorization minimization algorithm for parameter estimation and prove that the objective function is monotonically decreasing by the parameter update. An experiment shows that the proposed model well handles interval-valued elements and offers improved performance.

***Index Terms*—** interval valued matrix, nonnegative matrix factorization

1. INTRODUCTION

Nonnegative matrix factorization (NMF) [1, 2] have been applied to various research fields such as signal processing and text mining [3, 4, 5, 6, 7]. NMF and its variants are also regarded as versatile tools for data analysis and are applied to e.g., recommendation, social data analysis and purchase log analysis [8, 9, 10, 11, 12, 13]. However, some extension of NMF is still required since researchers at the field of data analysis are faced with various types of data that are collected in many different ways e.g., membership cards and survey questionnaires. Especially, the need to analyze data containing both precise information and imprecise information is increasing.

Let us consider the example of jointly analyzing membership card data and survey questionnaire results (Fig. 1). Membership card records are precise data since they provide exact visit counts such as "user A visited a shop 4 times". In contrast, survey questionnaires often yield imprecise data since the range of the count is imprecise, e.g., "user B visited more than 3 times". When combined, these data are represented by a matrix that has scalar-valued elements and bounded/unbounded interval-valued elements. We call this the *generalized interval valued matrix*. Note that this matrix is needed to represent various types of data containing both precise and imprecise information such as data containing privacy protected and non-protected details, e.g., the exact ages of some users are hidden by conversion into a range such as between 30 and 39 years old or more than 50 years old. Thus, we tackle the problem of analyzing the generalized interval valued matrix.

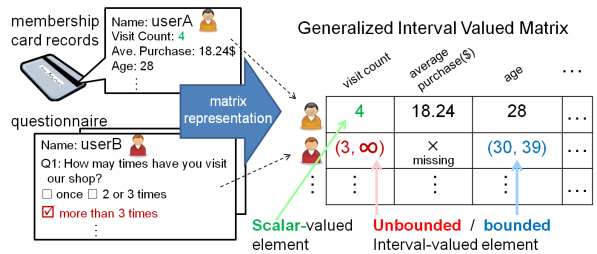


Fig. 1: Example of generalized interval valued matrix

This paper proposes a new probabilistic model called generalized interval valued nonnegative matrix factorization (GIV-NMF). GIV-NMF is derived by extending nonnegative matrix factorization (NMF) [1], which can handle only scalar-valued elements. The key to model formulation is the use of the cumulative density function (CDF) and latent variables indicating the scalar-values underlying interval-valued elements. We develop the majorization minimization (MM) algorithm [14, 15] for parameter estimation and prove that the objective function is monotonically decreasing by the parameter update.

Our approach using CDF can be found in the model for survival analysis (e.g., [16]) since survival data contain *censored* samples whose values are not observed when the values exceed certain threshold. Similar to the existing method for survival data [17, 18, 19], *truncated* Gaussian distribution [20] has a important role for the derivation of the proposed algorithm. In the context of interval-valued data analysis, the work most related to ours comes from Shen et al. [21]. They adopt a heuristics approach that constructs two scalar-valued matrices by extracting lower/upper limit of the interval-valued elements. Therefore, it can deal only with bounded interval-valued elements and not with unbounded interval elements. To the authors' knowledge, our proposal is the first factorization method that can deal with the matrix with *unbounded* interval elements and so should be seen as a fundamental method for interval-valued matrix analysis.

We conduct experiments on both synthetic and real data to confirm the effectiveness of the proposal. Using test mean squared error as a performance measure, we show that the proposed method can well handle the interval-valued elements with improved performance.

2. PROPOSED METHOD

2.1. Formulation

Our proposal factorizes the $I \times J$ generalized interval valued nonnegative matrix $\boldsymbol{X}$. $\boldsymbol{X}$ contains nonnegative scalar-valued elements, x_{ij}, and interval-valued elements, (x_{ij}^L, x_{ij}^R) $(0 \leq x_{ij}^L \leq x_{ij}^R)$, i.e., $\boldsymbol{X} = \{x_{ij}\}_{(i,j)\in\Omega_{sv}} \cup \{(x_{ij}^L, x_{ij}^R)\}_{(i,j)\in\Omega_{iv}}$, where Ω_{sv} and Ω_{iv} is the index set whose elements are scalar values and the index set whose elements are interval values, respectively. We also denote all observed elements as $\Omega = \Omega_{sv} \cup \Omega_{iv}$. The interval element (x_{ij}^L, x_{ij}^R) indicates that the exact value, x_{ij}, is unknown but it is within the interval, $x_{ij}^L \leq x_{ij} \leq x_{ij}^R$. It is allowed that x_{ij}^R be an infinitely large value.

Let Θ be the model parameter. Θ consists of factor matrices $\boldsymbol{A} = \{a_{ir}\}_{i,r=1}^{I,R}$, $\boldsymbol{B} = \{b_{jr}\}_{j,r=1}^{J,R}$ and precision parameter τ. R is the number of factors. Following standard NMF formulation, we assume that the (scalar-valued) element of $\boldsymbol{X}$ follows the normal distribution

$$P(x_{ij}|\Theta = \{\boldsymbol{A}, \boldsymbol{B}, \tau\}) = f(x_{ij}\big|\hat{x}_{ij}, \tau), \tag{1}$$

where $\hat{x}_{ij} = \sum_{r=1}^{R} a_{ir} b_{jr}$ and f is the probability density function (PDF) of the normal distribution:

$$f(x|\mu,\tau) = \frac{1}{\sqrt{2\pi\tau^{-1}}} \exp\Big\{-\frac{\tau}{2}(x-\mu)^2\Big\}. \tag{2}$$

Note that we can adopt other distributions such as the Poisson distribution in an analogous manner.

The key of our model lies in its ability to handle interval-valued elements. This is done by using the following cumulative density function (CDF):

$$F(C|\mu,\tau) = \int_{-\infty}^{C} f(x|\mu,\tau)dx \tag{3}$$

$F(C|\mu,\tau)$ indicates the probability that a random variable following f takes value less than C. Therefore, the probability that x_{ij} following $f(x_{ij}\big|\hat{x}_{ij},\tau)$ is within the interval (x_{ij}^L, x_{ij}^R) is given by

$$P(x_{ij}\in(x_{ij}^L, x_{ij}^R)|\Theta) = F(x_{ij}^R\big|\hat{x}_{ij},\tau) - F(x_{ij}^L\big|\hat{x}_{ij},\tau). \tag{4}$$

Use of PDF and CDF enable us to handle unbounded interval-valued elements since the limits are defined as $\lim_{x\to\infty} f(x|\mu,\tau) = 0$, $\lim_{C\to\infty} F(C|\mu,\tau) = 1$. To summarize, the probability of generating interval valued matrix $\boldsymbol{X}$ is written as

$$P(\boldsymbol{X}|\Theta) = \prod_{(i,j)\in\Omega^{sv}} P(x_{ij}|\Theta) \prod_{(i,j)\in\Omega^{iv}} P(x_{ij}\in(x_{ij}^L, x_{ij}^R)|\Theta).$$

Model parameter Θ is estimated by optimizing the following log-likelihood function.

$$\arg\min_{\Theta} \mathcal{L}(\Theta) = -\log P(\boldsymbol{X}|\Theta), \text{ s.t. } \boldsymbol{A} \geq 0, \boldsymbol{B} \geq 0, \tau \geq 0 \tag{5}$$

where $\boldsymbol{A} \geq 0$ means that all elements of $\boldsymbol{A}$ are nonnegative.

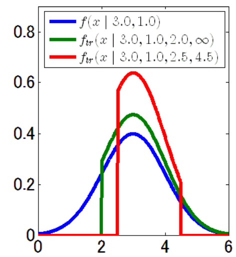


Fig. 2: distributions

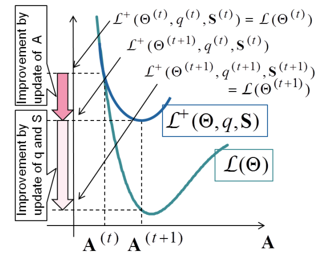


Fig. 3: MM scheme

2.2. Algorithm

As shown in the next subsection, the following algorithm can be used to solve the optimization problem given by Eq. (5).

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j\in\Omega_i^{sv}} x_{ij} b_{jr} + \sum_{j\in\Omega_i^{iv}} \bar{y}_{ij} b_{jr}}{\sum_{j\in\Omega_i} \hat{x}_{ij} b_{jr}}, \tag{6}$$

$$b_{jr} \leftarrow b_{jr} \frac{\sum_{i\in\Omega_j^{sv}} x_{ij} a_{ir} + \sum_{i\in\Omega_j^{iv}} \bar{y}_{ij} a_{ir}}{\sum_{i\in\Omega_j} \hat{x}_{ij} a_{ir}}, \tag{7}$$

$$\tau^{-1} \leftarrow \frac{1}{N}\Big\{\sum_{(i,j)\in\Omega_{sv}} (x_{ij} - \hat{x}_{ij})^2 + \sum_{(i,j)\in\Omega_{iv}} (\bar{y}_{ij}^{(2)} - 2\bar{y}_{ij}\hat{x}_{ij} + \hat{x}_{ij}^2)\Big\}, \tag{8}$$

where Ω_i^{sv}/ Ω_i^{iv} and Ω_j^{sv}/ Ω_j^{iv} is the set of scalar/interval-valued elements in that i-th row and j-th column, respectively. $\bar{y}_{ij}$ and $\bar{y}_{ij}^{(2)}$ are defined as

$$\bar{y}_{ij} = \hat{x}_{ij} + \frac{1}{\tau} \frac{f(x_{ij}^L|\hat{x}_{ij},\tau) - f(x_{ij}^R|\hat{x}_{ij},\tau)}{F(x_{ij}^R|\hat{x}_{ij},\tau) - F(x_{ij}^L|\hat{x}_{ij},\tau)}, \tag{9}$$

$$\bar{y}_{ij}^{(2)} = \hat{x}_{ij}^2 + \frac{1}{\tau} + \frac{1}{\tau} \frac{(x_{ij}^L + \hat{x}_{ij}) f(x_{ij}^L|\hat{x}_{ij},\tau) - (x_{ij}^R + \hat{x}_{ij}) f(x_{ij}^R|\hat{x}_{ij},\tau)}{F(x_{ij}^R|\hat{x}_{ij},\tau) - F(x_{ij}^L|\hat{x}_{ij},\tau)}. \tag{10}$$

Note that if (i,j)-th interval element is unbounded, i.e., $x_{ij}^R \to \infty$, $\bar{y}_{ij} = \hat{x}_{ij} + \frac{1}{\tau}\frac{f(x_{ij}^L|\hat{x}_{ij},\tau)}{1-F(x_{ij}^L|\hat{x}_{ij},\tau)}$, $\bar{y}_{ij}^{(2)} = \hat{x}_{ij}^2 + \frac{1}{\tau} + \frac{1}{\tau}\frac{(x_{ij}^L+\hat{x}_{ij})f(x_{ij}^L|\hat{x}_{ij},\tau)}{1-F(x_{ij}^L|\hat{x}_{ij},\tau)}$. Update rules for $\boldsymbol{A}$, $\boldsymbol{B}$ are given in "multiplicative form". The right hand side of the update for $\boldsymbol{A}$ is (I) always nonnegative and (II) equals a_{ir} when $x_{ij} = \hat{x}_{ij} = \bar{y}_{ij}$[1]. By updating the parameters following Eq. (6)-(8), the objective function is monotonically decreasing (non-increasing); proof is provided in § 2.4.

2.3. Algorithm Derivation

In this subsection, we derive the update rules given by Eq. (6)(7)(8). We minimize $\mathcal{L}(\Theta)$ following the optimization scheme of majorization-minimization (MM) [14][15]. In the scheme of MM, minimization of function $\mathcal{L}$ is indirectly conducted by minimizing the majorizing functional $\mathcal{L}^+$.

[1] nonnegativity of $\bar{y}_{ij}$ is confirmed later

Let us define latent variable $\boldsymbol{Y} = \{y_{ij}\}_{(i,j)\in\Omega^{iv}}$ whose elements $y_{ij} \in (x^L_{ij}, x^R_{ij})$ indicate elements with "scalar" values where only interval observations are given. Using $\boldsymbol{Y}$, we define the majorizing functional $\mathcal{L}^+$ as

$$\mathcal{L}^+(\Theta, q, \boldsymbol{S}) = \mathbb{E}_q[-\log h(\boldsymbol{X}, \boldsymbol{Y}, \Theta, S)] + \int q(\boldsymbol{Y}) \log q(\boldsymbol{Y}) d\boldsymbol{Y},$$

$$\log h(\boldsymbol{X}, \boldsymbol{Y}, \Theta, \boldsymbol{S}) = -\frac{N}{2}\log(2\pi\tau^{-1})$$
$$-\sum\nolimits_{(i,j)\in\Omega^{sv}} \frac{\tau}{2}\Big\{x^2_{ij} - 2x_{ij}\hat{x}_{ij} + \sum_{r=1}^{R}\frac{(a_{ir}b_{jr})^2}{s_{ijr}}\Big\}$$
$$-\sum\nolimits_{(i,j)\in\Omega^{iv}} \frac{\tau}{2}\Big\{y^2_{ij} - 2y_{ij}\hat{x}_{ij} + \sum_{r=1}^{R}\frac{(a_{ir}b_{jr})^2}{s_{ijr}}\Big\},$$

where $q(\boldsymbol{Y})$ is the auxiliary probability distribution of latent variable $\boldsymbol{Y}$ and $\boldsymbol{S} = \{s_{ijr}\}$ is an auxiliary variable satisfying $\sum_r s_{ijr} = 1$ $(\forall(i,j))$. It can be verified that the majorizing functional $\mathcal{L}^+$ has the following two properties:

1. $\mathcal{L}(\Theta) \le \mathcal{L}^+(q, \boldsymbol{S}, \Theta)$. 2. $\mathcal{L}(\Theta) = \min_{q,\boldsymbol{S}} \mathcal{L}^+(q, \boldsymbol{S}, \Theta)$. Note that the equality holds if and only if

$$q(y_{ij}) = f_{tr}(y_{ij}|\hat{x}_{ij}, \tau, x^L_{ij}, x^R_{ij}), \quad s_{ijr} = \frac{a_{ir}b_{jr}}{\sum_{r'} a_{ir'}b_{jr'}}. \quad (11)$$

where $f_{tr}(x|\mu, \tau, a, b)$ is the *truncated* normal distribution (Fig. 2) defined as the distribution whose PDF is given by:

$$f_{tr}(x|\mu,\tau,a,b) = \begin{cases} \frac{f(x|\mu,\tau)}{F(b|\mu,\tau)-F(a|\mu,\tau)} & (\text{if } x \in (a,b]) \\ 0 & (\text{otherwise}) \end{cases}$$

The r.h.s of equations (9)(10) come from the moment of the truncated normal, $\bar{y}_{ij} = \mathbb{E}_{q(y_{ij})}[y_{ij}]$ and $\bar{y}^{(2)}_{ij} = \mathbb{E}_{q(y_{ij})}[y^2_{ij}]$, respectively [20].

Using the majorizing functional $\mathcal{L}^+$, minimization of function $\mathcal{L}$ is conducted in the following manner: (Step 1) Minimize $\mathcal{L}^+(\Theta, q, \boldsymbol{S})$ w.r.t. $\boldsymbol{A}$ or $\boldsymbol{B}$. (Step 2) Minimize $\mathcal{L}^+(\Theta, q, \boldsymbol{S})$ w.r.t. $\boldsymbol{S}$ and q. (Step 3) Minimize $\mathcal{L}^+(\Theta, q, \boldsymbol{S})$ w.r.t. τ. See Fig. 3 for visual understanding.

For the first step, we compute the partial derivative of $\mathcal{L}^+$ w.r.t. $\boldsymbol{A}$. The necessary condition of the local minima, which is the partial derivative $\frac{\partial\mathcal{L}^+}{\partial a_{ir}} = 0$, can be simplified to

$$a_{ir} = \Big(\sum_{j\in\Omega^{sv}_i} x_{ij}b_{jr} + \sum_{j\in\Omega^{iv}_i} x_{ij}b_{jr})\Big)\Big/\Big(\sum_{j\in\Omega_i}\frac{b^2_{jr}}{s_{ijr}}\Big). \quad (12)$$

By substituting Eq. (11) into Eq. (12), we obtain the multiplicative update rules for $\boldsymbol{A}$ given by Eq. (6). The update rules for $\boldsymbol{B}$ are derived in analogous manner.

For the second step, we consider the update for $\boldsymbol{S}$ and q. The update for $\boldsymbol{S}$ (Eq. (11)) is derived by the method of Lagrange multipliers which uses the Lagrange function $\mathcal{F}$:

$$\mathcal{F}(\boldsymbol{S}, \Lambda) = \mathcal{L}^+(\Theta, q, \boldsymbol{S}) + \sum\nolimits_{i,j}\lambda_{ij}\big(\sum\nolimits_r s_{ijr} - 1\big), \quad (13)$$

where $\Lambda^S = \{\lambda_{ij}\}$ are Lagrange multipliers. The update for q is derived by the variational method and the optimal q is given by $q(Y) \propto \exp\big(\log h(X, Y, \Theta, S)\big)$. The optimal q is the one shown by Eq. (11). Update for τ in the third step is also derived by solving $\frac{\partial\mathcal{L}^+}{\partial\tau} = 0$.

2.4. Theoretical Analysis

Here we show the property of the proposed algorithm [2].

Theorem 1. *Objective function $\mathcal{L}(\Theta)$ is monotonically decreasing under the update by Eq.* (6)(7) *. $\mathcal{L}(\Theta)$ is invariant if and only if Θ is at a stationary point.*

This theorem indicates that the parameters are "improved" by the update. The theorem is proven by showing that $\mathcal{L}^+$ decreases with each optimization step. We need to prove following two lemmas to prove the theorem.

Lemma 1. *$\mathcal{L}^+(q, \boldsymbol{S}, \Theta)$ is a convex function w.r.t. $\boldsymbol{A}$ and thus $\boldsymbol{A}$ satisfing Eq.* (12) *is the global minimum while the other parameters are fixed.*

Proof. The second derivative of $\mathcal{L}^+$ is given by $\frac{\partial\mathcal{L}^+}{\partial a_{ir}a_{i'r'}} = \delta_{ii'}\delta_{rr'}\{\sum_j b^2_{jr}/s_{ijr}\}$, where $\delta_{ii'} = 1$ if $i = i'$, otherwise $\delta_{ii'} = 0$. Since $\sum_{i,r,i',r'} v_{ir}\frac{\partial\mathcal{L}^+}{\partial a_{ir}a_{i'r'}}v_{i'r'} \ge 0$ is satisfied for arbitrary $v_{ir} \in \mathbb{R}^{I\times R}$, $\mathcal{L}^+$ is a convex function w.r.t. $\boldsymbol{A}$. □

Lemma 2. *The objective $\mathcal{L}^+(q, \boldsymbol{S}, \Theta)$ is minimized w.r.t. $\boldsymbol{S}$ and q when $\boldsymbol{S}$ and q equals Eq.* (11) *and $\mathcal{L}(\Theta) = \min_{q,\boldsymbol{S}} \mathcal{L}^+(q, \boldsymbol{S}, \Theta)$ holds.*

Proof. When auxiliary variable $\boldsymbol{S}$ satisfies Eq. (11), $\mathcal{L}^+$ equals $\mathcal{L}'$ and $\mathcal{L}(\Theta) = \mathcal{L}'(q, \Theta) - KL(q||p)$, where $\mathcal{L}'(q, \Theta) = -\mathbb{E}_q[\log P(X, Y|\Theta)] + \int q(Y)\log q(Y)dy$, $P(\boldsymbol{X}, \boldsymbol{Y}|\Theta) = \prod_{(i,j)\in\Omega_{sca}}\mathcal{N}(x_{ij}|\hat{x}_{ij})\prod_{(i,j)\in\Omega_{int}}\mathcal{N}(y_{ij}|\hat{x}_{ij})$, $KL(q||p) = -\int q(Y)\log\{\frac{P(\boldsymbol{Y}|\boldsymbol{X},\Theta)}{q(\boldsymbol{Y})}\}$. Since $KL(q||p) > 0$, $\mathcal{L}'(q, \Theta)$ is the upper bound of $\mathcal{L}(\Theta)$. $\mathcal{L}(\Theta)$ does not depend on q, and thus $\mathcal{L}'(q, \Theta)$ is minimized when $KL(q||p) = 0$, i.e., $q(\boldsymbol{Y}) = P(\boldsymbol{Y}|\boldsymbol{X}, \Theta) = \prod_{(i,j)\in\Omega^{iv}} f_{tr}(y_{ij}|\hat{x}_{ij}, \tau, x^L_{ij}, x^R_{ij})$. Since $\mathcal{L}'(q, \Theta) \le \mathcal{L}^+(q, \boldsymbol{S}, \Theta)$ is shown by Jensen's inequality, this concludes the proof. □

The theorem follows from the application of the lemmas.

Proof. Let us denote the parameter and the auxiliary distribution and variables which satisfy $\mathcal{L}(\Theta) = \mathcal{L}^+(\Theta, q, \boldsymbol{S})$ as Θ^{old}, q^{old}, $\boldsymbol{S}^{old}$. We also denote $\boldsymbol{A}$ after the first step of the MM given by Eq. (12) as $\boldsymbol{A}^{new}$ and $q, \boldsymbol{S}$ after the second step given by Eq. (11) as $q^{new}, \boldsymbol{S}^{new}$. From lemma 1 and lemma 2, $\mathcal{L}^+(\boldsymbol{A}^{new}, q^{old}, \boldsymbol{S}^{old}) \le \mathcal{L}^+(\boldsymbol{A}, q^{old}, \boldsymbol{S}^{old})$ $(\forall\boldsymbol{A})$ and $\mathcal{L}^+(\boldsymbol{A}^{new}, q^{new}, \boldsymbol{S}^{new}) \le \mathcal{L}^+(\boldsymbol{A}^{new}, q, \boldsymbol{S})$ $(\forall q, \boldsymbol{S})$. Note that we omit the notation of $\boldsymbol{B}$ and τ. Since $\mathcal{L}(\boldsymbol{A}^{old}) =$

[2] We omit proof of τ since it is analogous to the variance estimation of Normal distribution.

$\mathcal{L}^+(\boldsymbol{A}^{old}, q^{old}, \boldsymbol{S}^{old})$ and $\mathcal{L}(\boldsymbol{A}^{new}){=}\mathcal{L}^+(\boldsymbol{A}^{new}, q^{new}, \boldsymbol{S}^{new})$, $\mathcal{L}(\boldsymbol{A}^{new}) \leq \mathcal{L}(\boldsymbol{A}^{old})$ holds. The proof for the update of $\boldsymbol{B}$ is analogous. □

3. EXPERIMENT

3.1. Setting

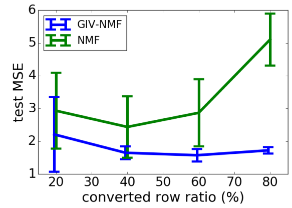
We conducted experiments to confirm the effectiveness of the proposal. Since our method is, to the best of the author's knowledge, the first that can deal with unbounded interval-valued elements, we investigate whether the proposal can well handle unbounded interval-valued elements and improves the performance. The evaluations use synthetic and real data.

For the synthetic data (Synth), we generated factor matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ whose sizes are $I = J = 30, R = 6$ using a Gaussian with mean of 1.0 and precision of 0.3^{-2}. Computing their product and adding Gaussian noise with mean 0.0 and precision 0.1^{-2}, yielded a matrix with $\boldsymbol{X}_{sc}$ whose elements are all scalar. We prepared five data sets by dividing the elements of $\boldsymbol{X}_{sc}$ into five, using 20% of the data as a training set and the remaining 80% as a test set so that the matrix contains many missing elements as would occur in real scenarios. Note that test data are treated as missing elements in training. Assuming the situation shown in Fig. 1, we converted the elements in the training set whose value is larger than 4.0 to the interval valued-element $(4.0, \infty)$ in chosen $20, 40, 60, 80\%$ rows (excepting the first 3 logs). Intuitively, the ratio corresponds to the ratio of imprecise information.

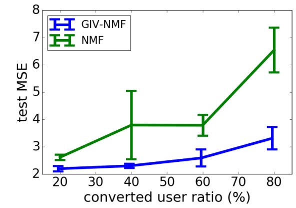
For the real data, we use Yelp Academic dataset (YA)[3] including users' review scores of businesses (shops); they range from 1.0 (min) to 5.0 (max). Using the review log of Montreal and extracting users and businesses that appear more than 10 times, we construct a matrix with all scalar-valued elements; its size is $I = 166, J = 382$. Similar to the synthetic data, we prepared five data sets by dividing the data and using 80% of the data as a training set and 20% as a test set. We convert the elements whose value is equal to or larger than 3.0 to the interval valued-element $(3.0, \infty)$ for chosen $20, 40, 60, 80\%$ users (excepting the first 3 logs).

We use test mean squared error (test MSE) as a performance metric. Test MSE is defined as $\frac{1}{|\mathcal{T}|}\sum_{(i,j)\in\mathcal{T}}(x_{ij} - \hat{x}_{ij})^2$, where $\mathcal{T}$ is the set of element indexes in the test and $|\cdot|$ indicates the number of elements in the set. By comparing standard NMF [1], which handles only scalar-valued elements, we investigate the effectiveness of the proposal. Moreover, we also report the latent MSE for evaluating the prediction of latent scalar values, which were hidden by conversion into interval-valued elements. The proposal offers two approaches to prediction: using $\hat{x}_{ij}$ or $\bar{y}_{ij}$ (Eq.(9)) since the corresponding interval-valued elements (x_{ij}^L, x_{ij}^R) is available. When using $\bar{y}_{ij}$, latent MSE is computed as $\frac{1}{|\mathcal{T}'|}\sum_{(i,j)\in\mathcal{T}'}(x_{ij} - \bar{y}_{ij})^2$, where $\mathcal{T}'$ is the set of element indexes whose values are converted into interval-valued elements. We compare the results of the two approaches and NMF. The number of factors is set to 6 for Synth and to 2 for YA in common based on preliminarily experiments.

[3] https://www.yelp.com/academic_dataset

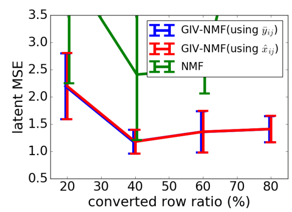


(a) Synth

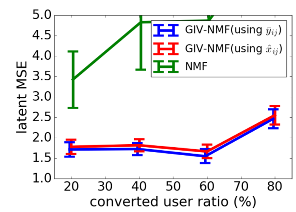


(b) YA

Fig. 4: Missing value prediction. Lower is better.



(a) Synth



(b) YA

Fig. 5: Latent value prediction. Lower number is better.

3.2. Result

Figure 4 shows the results of missing value prediction. GIV-NMF and NMF have competitive performance when the ratio of converted users is small while GIV-NMF outperforms NMF as the ratio increases. This indicates GIV-NMF well handles the (unbounded) interval-valued elements and offers improved performance. Figure 5 shows the results of latent scalar value prediction. Similar to the above, both GIV-NMF variants outperform NMF. GIV-NMF (using $\bar{y}_{ij}$) shows slightly better performance than its sibling. This implies the use of the interval can contribute to improved performance.

4. CONCLUSION

In this paper, we proposed a probabilistic model for analyzing the generalized interval valued matrix, a matrix that has scalar-valued elements and bounded and unbounded interval-valued elements. We provide the theoretical proof of the proposed algorithm and confirmed its effectiveness by experiments on synthetic and real world data. Future work is to evaluate and to clarify how the length of the interval affects performance since it reflects the certainty of information.

5. REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[3] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE workshop on applications of signal processing to audio and acoustics*, 2003, pp. 177–180.

[4] C. Févotte, N. Bertin, and J. L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[5] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Non-negative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[6] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 267–273.

[7] F. Shahnaz, M. W. Berry, V. P. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Information Processing & Management*, vol. 42, no. 2, pp. 373–386, 2006.

[8] S. Zhang, W. Wang, J. Ford, and F. Makedon, "Learning from incomplete ratings using non-negative matrix factorization," in *Proceedings of the 2006 SIAM international conference on data mining*. SIAM, 2006, pp. 549–553.

[9] Y. Chi, S. Zhu, X. Song, J. Tatemura, and B. L. Tseng, "Structural and temporal analysis of the blogosphere through community factorization," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 163–172.

[10] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, 2011.

[11] K. Takeuchi, K. Ishiguro, A. Kimura, and H. Sawada, "Non-negative multiple matrix factorization," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2013, vol. 13, pp. 1713–1720.

[12] P. Lahoti, K. Garimella, and A. Gionis, "Joint non-negative matrix factorization for learning ideological leaning on twitter," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 2018, pp. 351–359.

[13] M. Kohjima, T. Matsubayashi, and H. Sawada, "Probabilistic non-negative inconsistent-resolution matrices factorization," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 1855–1858.

[14] D. R. Hunter and K. Lange, "A tutorial on mm algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[15] J. De Leeuw, "Block-relaxation algorithms in statistics," in *Information systems and data analysis*, pp. 308–324. Springer, 1994.

[16] D. G. Kleinbaum and M. Klein, *Survival analysis*, vol. 3, Springer, 2010.

[17] D. Chauveau, "A stochastic em algorithm for mixtures with censored data," *Journal of statistical planning and inference*, vol. 46, no. 1, pp. 1–25, 1995.

[18] G. Lee and C. Scott, "Em algorithms for multivariate gaussian mixture models with truncated and censored data," *Computational Statistics & Data Analysis*, vol. 56, no. 9, pp. 2816–2829, 2012.

[19] M. Kohjima, T. Matsubayashi, and H. Toda, "Variational bayes for mixture models with censored data," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 605–620.

[20] NL. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Probability Distributions,(Vol. 1)*, John Wiley & Sons Inc., NY, 1994.

[21] Z. Shen, L. Du, X. Shen, and Y. Shen, "Interval-valued matrix factorization with applications," in *IEEE International Conference on Data Mining*. IEEE, 2010, pp. 1037–1042.