Abstract: The matrix factorisation approach computes a low-rank approximation of the incomplete user-item rating matrix. Existing approaches suffer from under-fitting due to the use of global information for all users and items. In this study, the authors propose a three-way recommendation model that integrates global and local information. This new model has a number of main aspects. The first is rating prediction with global and local information. A clustering and two matrix factorisation algorithms are employed for this purpose. The second is the computation of recommendation thresholds based on the decision-theoretic rough set model. Misclassification and promotion costs are considered simultaneously to build the cost matrix. The last is the determination of the recommender actions based on the prediction and thresholds. Experimental results on the well-known datasets show that authors’ proposed model improves recommendation quality in terms of average cost.

Three-way recommendation integrating global and local information

Yuan-Yuan Ma\(^1\), Heng-Ru Zhang\(^1\) E-mail: zhanghrswpu@163.com
Yuan-Yuan Xu\(^1\), Fan Min\(^1\), Lei Gao\(^1\)
\(^1\)School of Computer Science, Southwest Petroleum University, Chengdu 610500, People’s Republic of China

Nomenclature

| Symbol | Description |
|--------|-------------|
| \(U\) | the set of all users |
| \(V\) | the set of all items |
| \(R\) | the rating matrix |
| \(p\) | predicting rating |
| \(m\) | the number of users |
| \(n\) | the number of items |
| \(r_{ij}\) | the rating of \(u_i\) to \(v_j\) |
| \(c\) | rating levels |
| \(d(\cdot)\) | feature vector |
| \(D\) | feature vector set |
| \(r_{\text{max}}\) | the maximum rating level |
| \(r_{\text{min}}\) | the minimum rating level |
| \(g\) | the user group number |
| \(q\) | the item group number |
| \(G_g\) | the set of the \(g\)th user group |
| \(Q_q\) | the set of the \(q\)th item group |
| \(NU\) | the total number of user groups |
| \(NI\) | the total number of item groups |
| \(a\) | MF latent factor |
| \(a_p\) | the positive domain |
| \(a_N\) | the negative domain |
| \(a_B\) | the boundary domain |
| \(\alpha\) | three-way recommended threshold |
| \(\beta\) | three-way non-recommended threshold |
| \(PP\) | the domain for the recommended item that the user also likes |
| \(BN\) | the domain for the promoted item that the user also likes |
| \(NP\) | the domain for the promoted item that the user does not like |
| \(NN\) | the domain for the non-recommended item that the user likes |
| \(\lambda_{PP}\) | the cost for PP |
| \(\lambda_{PN}\) | the cost for PN |
| \(\lambda_{BP}\) | the cost for BP |
| \(\lambda_{BN}\) | the cost for BN |
| \(\lambda_{NP}\) | the cost for NP |
| \(x\) | random sample point |

1 Introduction

The matrix factorisation (MF) approach commonly produces a low-rank approximation based on the rating matrix of recommender systems [1, 2]. Its advantages including fast speed and ability to effectively alleviate the data sparsity issue. The probabilistic MF (PMF) [3] algorithm establishes a model from the perspective of probability by observing a large amount of linear data. Singular value decomposition (SVD) [4, 5] uses the characteristics of users and items to give an optimal low-rank approximation to the real-valued matrix based on the squared error. Non-negative MF (NMF) uses non-negative constraints to prevent mutual cancellation between basis functions and generates part-based representations [6, 7]. However, existing MF-based collaborative filtering (CF) methods use global information about all users and items, which can suffer from an under-fitting problem.

In this paper, we propose to build a three-way recommendation model that integrates global and local information (3WGL). Three-way recommender systems partition a set of items into positive, negative and boundary domains [8–10]. Items in these domains are divided into recommended, not recommended and promoted [9, 11]. Misclassification costs [11, 12] result from wrong recommendations, including recommending items to users who dislike them, and not recommending items to users who like them. From the view of cost sensitivity [13], the cost of false-positive errors is different from the cost of false-negative errors. In order to recommend items with fewer errors, the systems introduce promotion cost, e.g. distributing coupons to users. We construct an improved MF approach that integrates global and local information to predict ratings. According to the three-way decision and prediction ratings, the recommendation behaviours of items can be obtained.

First, we predict the ratings with MF considering both global and local information. For this purpose, we divide users and items into a number of groups using the \(k\)-means clustering [14] method coupled with the Kullback–Leibler divergence [15]. In this way, we obtain different sub-matrices of the original rating matrix. Sub-matrices of users and items indicate user local information and item local information. Global information [16] is used to obtain users’ preferences based on all users and items involved. By integrating global and local information, we build two kinds of MF algorithms of local-user–global-item (LUGI) and global-user–local-item (GULI). The results of these two algorithms are integrated as the predicting rating \(p\).
Second, we calculate the threshold pair $\alpha$ and $\beta$ based on the three-way decision. We construct a cost matrix according to the misclassification cost and promotion cost mentioned above. The cost matrix is a $3 \times 2$ matrix that stores the cost of recommendation actions in different states [17]. We calculate the threshold pair $\alpha$ and $\beta$ based on the cost matrix and the rating level. $\alpha$ and $\beta$ represent certain tolerance levels for making a decision of recommendation, non-recommendation or promotion [8, 17].

Third, recommendation actions are determined based on $p$, $\alpha$ and $\beta$. If $p \geq \alpha$, the item is recommended. The recommended items that the user does not like cause the wrong acceptance cost ($\lambda_{PN}$). If $p \leq \beta$, the item is not recommended. Not recommending the user’s favourite items causes the wrong rejection cost ($\lambda_{NP}$). If $\beta < p < \alpha$, the item is promoted. In boundary-positive and boundary-negative regions, items need to pay the promotion cost ($\lambda_{PN}$ and $\lambda_{NP}$), i.e. the businesses will distribute the coupons to the consumers. For the correct recommendation, recommending the user’s favourite items causes the correct acceptance cost ($\lambda_{PP}$), and items that users do not like that are not recommended cause the correct rejecting cost ($\lambda_{NN}$). Finally, the average cost is used as a measure of recommendation quality.

Experiments on the well-known datasets show that (a) the integration of global and local information can effectively improve the quality of recommendation; and (b) our three-way method outperforms the existing three-way approaches in terms of average cost.

The rest of the paper is arranged as follows: Section 2 briefly reviews some of the studies related to our paper, as well as the definition of the rating system. Section 3 discusses how to build a three-way recommendation system model based on global and local information. Section 4 shows experimental results in real-world datasets. Section 5 concludes the paper.

## 2 Related work

### 2.1 Matrix factorisation

MF is one of the most commonly used CF methods. It was initially designed to resolve matrix sparseness problems in machine learning [16]. In 1998, it was first introduced to the CF domain [18] and widely adopted. Srebro et al. [19] provided a simple and efficient algorithm for solving weighted low-rank approximation problems. Lee et al. [20] proposed a new MF model where the observation matrix is represented by a weighted sum of low rank matrices.

Salakhutdinov and Mnih [1, 2] presented a fully Bayesian treatment of the PMF model in which model capacity is controlled automatically by integrating over all model parameters and hyper-parameters. SVD [21] is used to establish an implicit high-order structure associated with a document to detect related documents based on the terms found in queries. Billsus et al. [18] proposed an algorithm based on the SVD of a user rating matrix, which uses the latent structure to substantially eliminate the limitations of previous approaches. Ma et al. [22] proposed a semi-NMF method with global statistical consistency, emphasising the consistency between the predicted values and the statistics given by the data.

### 2.2 Three-way recommender systems

A three-way decision focuses on ternary problems, that is using three regions instead of two to represent a concept [11, 23]. Yang and Yao [24] proposed a multi-agent decision-theoretic rough set model and expressed it in the form of three-way decisions. Li et al. [25] proposed a sequential three-way decision method based on granular computing for cost-sensitive face recognition. Li et al. [26] presented a cost-sensitive sequential three-way decision model that simulates a gradual decision process from rough granules to precise granules.

In order to reduce the recommendation errors, three-way decisions have been introduced into recommender systems. A three-way recommendation classifies items into three parts, namely, recommended items, promoted items and rejected items. Scharff et al. [27] combined the implicit profile and the feedback profile with suitably greater weight being given to the feedback information to generate three-way media recommendations. Liu et al. [28] adopted a three-way decision approach based on the decision-theoretic rough set model adopted to risk government decision-making. However, they did not consider cost-sensitive issues. Zhang and Min [9] proposed a framework that integrates three-way decision and random forests to build cost-sensitive recommender systems, but they do not consider pair-wise preference of items. Huang et al. [8] introduced pair-wise preference into cost-sensitive three-way recommendations that forms the basis of a ranking strategy, using a user’s preference to estimate the user-item relevance probability.

### 2.3 Rating system

In this subsection, the rating system [10] is defined as follows. Let $U = \{u_1, u_2, ..., u_n\}$ be the set of users and $V = \{v_1, v_2, ..., v_n\}$ be the set of all possible items that can be recommended to users. The rating function is given as follows:

$$R: U \times V \rightarrow R_c$$

where $R_c = \{r_1, r_2, ..., r_s\}$ is the rating domain employed by users to evaluate items. For convenience, we denote the rating system as an $m \times n$ rating matrix $R = (r_{ij})_{m \times n}$, where $r_{ij} = R(u_i, v_j)$, $1 \leq i \leq m$ and $1 \leq j \leq n$.

Table 1 is a simple rating system of size $4 \times 4$, where $R_c = \{1, 2, 3, 4, 5\}$. The numbers 1–5 represent the five rating levels. Some of these elements are zero, indicating that the items are not rated by the users.

## 3 Proposed approach

In the section, we illustrate how to build a three-way recommendation model integrating global and local information. For the sake of clarity, Nomenclature introduces the symbols in this paper.

### 3.1 Three-way framework

In order to obtain the recommendation behaviours of items, we calculate the prediction rating and three-way threshold pair $(\alpha, \beta)$. Fig. 1a depicts our approach framework. It is divided into three parts:

- Fig. 1a constructs LUGI and GULI, two MF algorithm models using the rating matrix. Two types of prediction ratings can be obtained based on these models. By integrating these two ratings, we can get the prediction rating $p$.
- Fig. 1b constructs a cost matrix to calculate the threshold pair $(\alpha, \beta)$. $\alpha$ is the threshold of recommended items and $\beta$ is the threshold of not recommended items.
Fig. 2 Example of item-based feature-vector extraction

In the section, we describe how to extract local information. Local information corresponds to user subsets and item subsets clustered from the rating matrix. We employ the k-means clustering algorithm to extract local information. The clustering algorithm includes two steps: feature vector extraction and user/item clustering.

First, we introduce feature vector extraction. Fig. 2 depicts an example of item-based feature vector extraction corresponding to Table 1. This is the same way to extract user feature vectors. We extract the item-level connections to obtain a feature vector of length $e$. Each element of the feature vector is the number of users’ connections to the corresponding level. In this example, $u_1$ rates $v_2$ as 2, $u_2$ rates $v_2$ as 2, $u_3$ rates $v_2$ as 1, and $u_4$ rates $v_2$ as 1 and nobody rates $v_2$ as 3, 4 or 5. Hence, we have $d(v_1) = (0, 0, 1, 1, 0)$, $d(v_2) = (2, 2, 0, 0, 0)$, $d(v_3) = (0, 1, 1, 1, 1)$, and $d(v_4) = (0, 1, 0, 2, 0)$.

Second, we use the feature vectors to represent the users/items in the k-means clustering. k-means clustering is a prototype-based clustering algorithm, where $k$ is a constant that indicates the number of categories of clustering. Here, we employ the Kullback-Leibler divergence method to calculate the distance between two items. There is a similar way for user-based clustering.

3.3 Rating prediction and ensemble through MF

In the section, we introduce a sub-matrix construction, reconstruct the MF optimisation objective function based on these sub-matrices and design two MF algorithms integrating global and local information.

Firstly, we construct two kinds of sub-matrices using the user/item cluster and the $R$. Assume we have $NU \times NI$ clusters. Let $g \in [1, NU]$ be the user cluster number and $q \in [1, NI]$ be the item cluster number. $R^{g*}$ is a sub-matrix with historical ratings for users in $G_q$ and all items. $R^{*q}$ is a sub-matrix with historical ratings for all users and items in $Q_g$.

Secondly, we reconstruct the new MF optimisation objective functions based on $R^{g*}$ and $R^{*q}$. According to the idea of MF, we decompose each sub-matrix into the product of two low rank matrices. We adopt the principle of square error to design optimisation functions as

$$\min_{U,V} L(R^{g*}, G_q V) = \sum_{i,j} (R^{g*}(i,j) - G_q(i,a)V(a,j))^2,$$

and

$$\min_{G_q} L(R^{*q}, UQ_q) = \sum_{i,j} (R^{*q}(i,j) - U(i,a)Q_q(a,j))^2.$$

The learning goal can be viewed as minimising the integrating of these two functions,

$$\min_{U,V} \min_{G_q, Q_q} \left[ L(R^{g*}, G_q V) \leq \epsilon_1, L(R^{*q}, UQ_q) \leq \epsilon_2 \right],$$

where $L(R^{g*}, G_q V)$ and $L(R^{*q}, UQ_q)$ are constraints, and $\epsilon_1$ and $\epsilon_2$ are almost close to 0. Using Lagrange multipliers [29], the objective function with the $L_2$ regulariser [30, 31] of our algorithm can be further presented as

$$\min_{U,V} \min_{G_q, Q_q} \left[ L(R^{g*}, G_q V) + \lambda_1 \| U \|^2 + \alpha_V \| V \|^2 + \pi_1 L(R^{*q}, UQ_q) + \lambda_2 \| U \|^2 \right],$$

where $\lambda_1$, $\lambda_2$, $\alpha_V$ and $\pi_2$ are the regularisation parameters. $\pi_1$ and $\pi_2$ are hyperparameters that are used to control the contribution of user and item features in an embedded model.

Finally, we design two MF algorithms integrating global and local information. Here, we use the root mean square error (RMSE) [32] as the loss function in (5). By computing the partial derivatives of parameters, we can get the updated formulas:

$$V(a, j) \leftarrow V(a, j) + \gamma (\pi_1 \Delta G_q(i,a) - l_V V(a,j)),$$

$$U(i,a) \leftarrow U(i,a) + \gamma (\pi_2 \Delta Q_q(a,j) - l_U U(i,a)),$$

where $\gamma$ is the learning rate. We set $\pi_1$ and $\pi_2$ as the RMSE in $R^{g*}$ and $R^{*q}$, respectively, and $\lambda_1$, $\lambda_2$ as the prediction errors $R(i, j) - G_q(i,a)V(a,j)$ and $R(i, j) - U(i,a)Q_q(a,j)$. Based on the updated formulas, we can obtain two prediction ratings, $R_{PLC}$ and $R_{PL}". The average rating of these two ratings as $p$ is involved in determining the recommendation behaviours of the items.

3.4 Three-way recommendation

The traditional decision model is the two-way decision model, that is, accept and reject the two options. However, this model can easily lead to erroneous decisions, and in some cases, making wrong decisions can be costly. On this basis, Yao et al. put forward a three-way decision theory. Moreover, introduced the Bayesian risk decision [33] into the three-way decision method, which makes it have the risk-cost-sensitive feature [34].

We define the entire domain as $A$. The three-way decision set is

$$A = \{a_0, a_0, a_1, a_2, a_3\},$$

which indicates that the judgement objects belong to the positive domain, the negative domain, or the boundary domain, respectively.

Different domains correspond to different decisions, and different decisions result in different costs. Different decision costs are divided into six situations: correct acceptance cost $\Upsilon_{PP}$, false rejection cost $\Upsilon_{NN}$, boundary domain acceptance cost $\Upsilon_{PN}$, correct rejection cost $\Upsilon_{NN}$, false acceptance cost $\Upsilon_{PN}$ and boundary domain rejection cost $\Upsilon_{NN}$. Table 2 is a cost matrix constructed on these six kinds of decision costs.
To simplify the rules, some constraints should be added so that each item can be put into only one domain. The first condition is
\[
\lambda_P \leq \lambda_B < \lambda_N, \\
\lambda_N \leq \lambda_B < \lambda_P.
\] (9)

The second condition is
\[
(\lambda_P - \lambda_B)(\lambda_P - \lambda_B) > (\lambda_N - \lambda_B)(\lambda_N - \lambda_B).
\] (10)

The final condition is
\[0 \leq \beta^* < \alpha^* \leq 1.\] (11)

\(\alpha^*\) is the positive domain threshold and \(\beta^*\) is the negative domain threshold. Here, \(\alpha^*\) and \(\beta^*\) are computed as follows:
\[
\alpha^* = \frac{\lambda_P - \lambda_N}{(\lambda_P - \lambda_B) + (\lambda_B - \lambda_P)},
\] (12)

\[
\beta^* = \frac{\lambda_B - \lambda_N}{(\lambda_B - \lambda_N) + (\lambda_N - \lambda_B)}.
\]

We map \(\alpha^*\) and \(\beta^*\) to the rating interval to obtain the recommendation behaviour demarcation threshold pair \((\alpha, \beta)\). The rules for mapping are:
\[
\alpha = \alpha^*(r_{\min} - r_{\max}) + r_{\max},
\]
\[
\beta = \beta^*(r_{\max} - r_{\min}) + r_{\max}.
\] (13)

Table 2 Cost matrix

| Action | \(\alpha_P\) | \(\alpha_B\) | \(\alpha_N\) |
|--------|--------------|--------------|--------------|
| L      | \(\lambda_P\) | \(\lambda_B\) | \(\lambda_N\) |
| D      | \(\lambda_P\) | \(\lambda_B\) | \(\lambda_N\) |

Table 3 Datasets

|        | Users | Items | Ratings    |
|--------|-------|-------|-----------|
| MI-100k | 943   | 1682  | 100,000   |
| MI-1M   | 6040  | 3900  | 1,000,209 |
| DouBan  | 2964  | 39,694| 912,479   |

Table 4 Average cost comparisons for different strategies

| Method | MI-100k | MI-1M | DouBan |
|--------|---------|-------|--------|
| LULI   | 13.8580 | 13.2630| 13.6656|
| LUGI   | 13.8580 | 13.2629| 13.6657|
| GULI   | 13.8580 | 13.2628| 13.6658|
| GUGI   | 13.8580 | 13.2628| 13.6656|
| GULI + LUGI | 13.8565 | 13.2627| 13.6655|

Table 5 Average cost comparisons for 3WGL and random forests

| Method               | MI-100k | MI-1M | DouBan |
|---------------------|---------|-------|--------|
| 3WGL                | 13.8565 | 13.2627| 13.6655|
| random forests      | 17.2500 | 15.9100| —      |
| Improvement, %      | 19.65   | 16.63 | —      |

Table 6 Average cost comparisons for 3WGL and kNN

| Method   | MI-100k | MI-1M | DouBan |
|----------|---------|-------|--------|
| 3WGL     | 13.8565 | 13.2627| 13.6655|
| kNN      | 16.1997 | 13.8511| 17.1621|
| Improvement, % | 14.46  | 4.24  | 20.37  |

where \(r_{\min}\) is the user's minimum rating of the item and \(r_{\max}\) is the maximum rating. By comparing the relationship between \((\alpha, \beta)\) and the prediction rating \(p\), we can get the recommendation behaviours of items:

(a) if \(p > \alpha\), the item can be recommended;
(b) if \(p \leq \alpha\) and \(p \geq \beta\), the item can be promoted;
(c) if \(p < \beta\), the item can not be recommended.

4 Experiments

In the section, we conduct several experiments on real-word datasets and evaluate the proposed performance of our algorithm based on the average cost [9]. For each dataset, we randomly divided it into a training set and a testing set in a ratio of 8:2. First, we use training data to build our three-way recommendation model. Then, we use the three-way model to recommend items to each user in the testing data and evaluate the recommendation performance. The average cost is used as the evaluation measure. The recommendation effect is good and the average cost is low.

Our experiments are performed on two MovieLens [35] datasets and one DouBan dataset shown in Table 3. MI-100k and MI-1M are two MovieLens datasets. MI-100k consists of 100,000 ratings (1–5) from 943 users on 1682 movies. MI-1M contains 1,000,209 anonymous ratings of about 3900 movies made by 6040 MovieLens users. DouBan contains 912,479 ratings of about 39,695 movies from 2965 users.

For parameters, we set learning rate \(\gamma = 0.0008\) as the gradient descend iteration step; \(\lambda_U = \lambda_V = \lambda_1 = \lambda_2 = 0.05\) are the penalty factors; and hyperparameters \(\pi_1 = 0.5\) and \(\pi_2 = 0.3\). The cost matrix is set as \(\lambda_P = \lambda_N = 0, \lambda_B = \lambda_{BP} = 20, \lambda_{PN} = 40\) and \(\lambda_{BP} = 50\). Then, we can obtain \(\alpha^* = 0.6\) and \(\beta^* = 0.5\). Based on (13), the recommendation behaviour demarcation thresholds are \(\alpha = 3.8\) and \(\beta = 3.4\).

We design two sets of comparative experiments to verify the recommendation performance of our algorithm. We observe the average cost on three datasets.

Exp 1: We use LUGI and GULI, two MF algorithms, to predict ratings and use different integration strategies between them to predict ratings. The average cost is used as an evaluation metric to observe which strategy can obtain better predictions on different datasets.

Exp 2: Our algorithm is compared with the classical three-way recommendation based on random forests and kNN to judge how well the algorithm is improved.

Table 4 shows the recommendation performance using different strategies on different datasets. For the MI-100k and DouBan datasets, we clustered users and items into three clusters to extract local information. For the MI-1M dataset, we have five user clusters and five item clusters. Through the different combinations of two MF algorithms, we can get five strategies. From the comparison of the results in the table, we can find that the recommendation strategy of integrating LUGI and GULI is the best. This is because our recommendation strategy captures global and local information for users and items.

Table 5 compares the performance of our algorithm 3WGL with the traditional three-way recommendation based on random forests. For the MI-100k and MI-1M datasets, our algorithm reduces the average cost and improves performance by 19.65 and 16.63%, respectively. The random forests approach uses users, items, and rating information to build multiple subtrrees. The DouBan dataset does not contain users and items information; so, it is not possible to build random forests for recommendation. In comparison, our approach uses only rating information to build the MF algorithms. It reduces the recommendation error caused by unreliable sources of users and items information.

Table 6 contrasts the performance between 3WGL and kNN approaches. By comparing the results on three datasets, we can see that our algorithm performance is greatly improved compared to kNN.
5 Conclusions

We construct a new three-way recommendation approach integrating global and local information. The recommendation strategy of integrating LUGI and GULI is better than other strategies. Our approach outperforms the existing three-way recommender systems.

6 Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (grants nos. 61379089, 41604114), the Innovation and Entrepreneurship Foundation of Southwest Petroleum University (Grant SWPUSC16-003), and the Natural Science Foundation of the Department of Education of Sichuan Province (grant no. 16ZA0060).

7 References

[1] Salakhutdinov, R., Mnih, A.: ‘Probabilistic matrix factorization’. Proc. of the 20th Int. Conf. on Neural Information Processing Systems, Vancouver, Canada, 2008, pp. 1257–1264
[2] Salakhutdinov, R., Mnih, A.: ‘Bayesian probabilistic matrix factorization using Markov chain Monte Carlo’. Proc. of the 25th Int. Conf. on Machine Learning, 2008, pp. 880–887
[3] Yang, W.F., Wang, M., Chen, Z.: ‘Fast probabilistic matrix factorization for recommender system’. Int. Conf. on Mechatronics and Automation (ICMA), 2014, pp. 1889–1894
[4] Ma, C.C.: ‘A guide to singular value decomposition for collaborative filtering’. Computer, Long Beach, CA, 2008, pp. 1–14
[5] Sarwar, B., Karypis, G., Konstan, J., et al.: ‘Incremental singular value decomposition algorithms for highly scalable recommender systems’. Int. Conf. on Computer and Information Science, 2002, pp. 27–28
[6] Lee, D.D., Seung, H.S.: ‘Learning the parts of objects by non-negative matrix factorization’, Nature, 1999, 401, (6755), pp. 788–791
[7] Li, T., Wang, J., Chen, H., et al.: ‘A NMF-based collaborative filtering recommendation algorithm’. The Sixth World Congress on Intelligent Control and Automation, 2006, vol. 2, pp. 6082–6086
[8] Huang, J., Wang, J., Yao, Y., et al.: ‘Cost-sensitive three-way recommendations by learning pair-wise preferences’, Int. J. Approx. Reason., 2017, 86, pp. 28–40
[9] Zhang, H.R., Min, F.: ‘Three-way recommender systems based on random forests’, Knowl.-Based Syst., 2016, 91, pp. 275–286
[10] Zhang, H.R., Min, F., Shi, B.: ‘Regression-based three-way recommendation’, Inf. Sci., 2017, 378, (1), pp. 444–461
[11] Yao, Y.: ‘Three-way decisions with probabilistic rough sets’, Inf. Sci., 2010, 180, (3), pp. 341–353
[12] Yao, Y.: ‘The superiority of three-way decisions in probabilistic rough set models’, Inf. Sci., 2011, 181, (6), pp. 1080–1096
[13] Min, F., He, H., Qian, Y., et al.: ‘Cost-sensitive attribute reduction’, Inf. Sci., 2011, 181, (22), pp. 4928–4942
[14] Hartigan, J.A., Wong, M.A.: ‘Algorithm AS 136: A k-means clustering algorithm’, J. Roy. Stat. Soc., 1979, 28, (1), pp. 100–108
[15] Joyce, J.M.: ‘Kullback-Leibler divergence’, in Mood, L. (Ed.): ‘International encyclopedia of statistical science’ (Springer, Berlin Heidelberg, 2011), pp. 720–722
[16] Smola, A.J., Schölkopf, B.: ‘Sparse greedy matrix approximation for machine learning’, Int. Conf. on Machine Learning, 2000, pp. 911–918
[17] Yao, Y.: ‘Three-way decision: an interpretation of rules in rough set theory’. Int. Conf. on Rough Sets and Knowledge Technology, 2009, pp. 642–649
[18] Bilous, D., Pazzani, M.J.: ‘Learning collaborative information filters’. AAAI Tech. Report, 1998, vol. 98, pp. 46–54
[19] Srebro, N., Jaakkola, T.: ‘Weighted low-rank approximations’. Proc. of the 20th Int. Conf. on Machine Learning (ICML-03), 2003, pp. 720–727
[20] Lee, J., Kim, S., Lebanon, G., et al.: ‘Local low-rank matrix approximation’. Int. Conf. on Machine Learning, 2013, pp. 82–90
[21] Furnas, G.W., Deerwester, S., Dumais, S.T., et al.: ‘Information retrieval using a singular value decomposition model of latent semantic structure’. Proc. of the 11th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 1988, vol. 51, (2), pp. 465–480
[22] Ma, H., Yang, H., King, J., et al.: ‘Semi-nonnegative matrix factorization with global statistical consistency for collaborative filtering’. Proc. of the 18th ACM Conf. on Information and Knowledge Management, 2009, pp. 767–776
[23] Liu, X.Y., Zhou, Z.H.: ‘The influence of class imbalance on cost-sensitive learning: An empirical study’. Sixth Int. Conf. on IEEE, 2006, pp. 970–974
[24] Yang, X., Yao, J.: ‘Modelling multi-agent three-way decisions with decision-theoretic rough sets’, Fundam. Inform., 2012, 115, (2–3), pp. 157–171
[25] Li, H., Zhang, L., Huang, B., et al.: ‘Sequential three-way decision and granulation for cost-sensitive face recognition’, Knowl.-Based Syst., 2016, 91, pp. 241–251
[26] Li, H., Zhou, X., Huang, B., et al.: ‘Cost-sensitive three-way decision: a sequential strategy’. Int. Conf. on Rough Sets and Knowledge Technology, 2013, pp. 325–337
[27] Schaffer, J.D., Lee, K.P., Gutta, S.: ‘Three-way media recommendation method and system’. U.S. Patent 7,979,792, May 2011
[28] Liu, D., Li, T., Liang, D.: ‘Three-way government decision analysis with decision-theoretic rough sets’, Int. J. Uncertain. Fuzziness Knowl.-Based Syst., 2012, 20, (5), pp. 715–737
[29] Bertekas, D.P.: ‘Constrained optimization and Lagrange multiplier methods’ (Academic Press, New York, 2014, 1st edn.)
[30] Chu, D., Mehrmann, V., Nichols, N.K.: ‘Minimum norm regularization of descriptor systems mixed output feedback’, Linear Algebra Appl., 1999, 296, (1–3), pp. 39–77
[31] Peng, T.K., Tseng, P., Ji, S., et al.: ‘Trace norm regularization: reformulations, algorithms, and multi-task learning’, SIAM J. Optim., 2010, 20, (6), pp. 3465–3489
[32] Willmott, C.J., Matsuura, K.: ‘Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance’, Clim. Res., 2005, 30, (1), pp. 79–82
[33] Fenton, N., Neil, M.: ‘Risk assessment and decision analysis with Bayesian networks’ (CRC Press, Boca Raton, 2012, 1st edn.)
[34] Ji, S., Carin, L.: ‘Cost-sensitive feature acquisition and classification’, Pattern Recognit., 2007, 40, (5), pp. 1474–1485
[35] Harper, F.M., Konstan, J.A.: ‘The MovieLens datasets: history and context’, ACM Trans. Inter. Intell. Syst. (TIIIS), 2016, 5, (4), pp. 1–20