S-ISFMF: SOCIAL NETWORK AND IMAGE SHAPE FEATURE MATRIX FACTORIZATION

Takuya Tamada[1], Mingtian Gao[2] and Ryosuke Saga[1]
[1]*Osaka Metropolitan University, Graduate School of Informatics, Sakai, Japan*
[2]*Osaka Prefecture University, Graduate School of Humanities and Sustainable System Sciences, Sakai, Japan*

ABSTRACT

Recommender system is a useful tool to help users to select the best item in today's information-overloaded society. Probabilistic Matrix Factorization (PMF) is one of successful recommender system methods that introduces a probabilistic algorithm to matrix factorization. Various auxiliary information can be used in PMF. However, there is no example of using images and social networks at the same time. In this paper, we describe a model that extracts information from images and social networks and incorporates it into PMF. In our method, we extracted contour information from item images and optimized the user and item matrix in matrix factorization. We further optimized the user matrix using explicit and implicit social networks among users. As a result, we achieved an accuracy improvement of 0.75%~7.25% on three real world datasets compared to existing methods.

KEYWORDS

Recommender System, Machine Learning, Matrix Factorization, Deep Learning

1. INTRODUCTION

Recommender System is a useful tool for presenting appropriate information to users from among a large amount of information in today's information-overloaded society(Saga et al., 2008; Saga & Duan, 2018). Especially in recent years, with the development of online review sites such as Amazon and Yelp, referring to the ratings of items that one is considering purchasing has become a natural part of everyday consumption. In addition, people often rely on the opinions of others, such as close friends and family members, to help them decide on the purchase of an item. In recommender systems, collaborative filtering, which makes recommendations based on the preference information of similar users, is a widely used technique (Resnick et al., 1994; Su & Khoshgoftaar, 2009).

Matrix factorization has been used successfully in rating prediction tasks. Among the collaborative filtering methods, Probabilistic Matrix Factorization (PMF), which employs a stochastic model to matrix factorization, works well on sparse and unbalanced datasets and contributes to solving the cold start problem (Salakhutdinov & Mnih, 2007). PMF decomposes the rating matrix obtained from the user's rating information into two matrices, a user matrix and an item matrix, which represent the characteristics of each user and item, respectively. There are other examples of attempts to improve accuracy by using auxiliary information in matrix factorization to solve the cold start problem(Kim et al., 2016; Lu et al., 2021; Ma et al., 2008; Saga & Duan, 2018; Tamada & Saga, 2022; Wang et al., 2019). For example, ConvMF is a typical example of research that uses text information in matrix factorization to improve recommendation accuracy (Kim et al., 2016). In this method, features are extracted from the text representing the features of the item using CNN and incorporated into PMF. Among the various auxiliary information, images and social networks are considered particularly useful in the recommender system (Li et al., 2015; Lu et al., 2021; Ma et al., 2008, 2008, 2008; McAuley et al., 2015; Saga & Duan, 2018; Wang et al., 2019).

One of recommender systems using images is the method by McAuley (McAuley et al., 2015). It used image information to compute models of higher-level relationships between objects and to recommend complementary clothing. There are also models that extract features from images and apply collaborative filtering. Verma tried to understand the preferences of individual users and to take into account user evaluation information using image information for collaborative filtering(Verma et al., 2020). It was method to construct a visual preference model by inputting several images that represent user preferences. One method that incorporates image information into PMF is ISFMF(Saga & Duan, 2018). In this method, item contour information is captured by CNN and incorporated into the item vector *V* in the PMF to improve accuracy. In Bi-ISFMF, the user vector *U*, which is not considered in ISFMF, is also incorporated into PMF by capturing features based on the items purchased by users in the past (Lu et al., 2021). our proposed method uses Bi-ISFMF model in image feature extraction.

Social networks have long been used in recommender systems. Generally, when we purchase an item, we often ask friends and acquaintances for their recommendations. Because we usually share interests and tastes with our friends and acquaintances. Social networks include explicit social ties such as friendships, siblings, superiors, and subordinates, and implicit social ties are those with users who share common interests; similarity calculations show they are related. Therefore, research has been conducted to systematically leverage explicit social networks among users to make recommendations (Ma et al., 2008, 2009). Ma et al. proposed an MF method that takes into account social trust among users (Ma et al., 2008) and a method that takes into account not only trust but also distrust between users (Ma et al., 2009). Li et al. proposed a method that calculates similarity from scoring data and builds an implicit social network to be incorporated into the recommendation system, claiming that it is only slightly inferior to a recommender system that uses an explicit social network in terms of recommendation accuracy, but has the advantage of being less prone to accuracy degradation when scoring data increase (Li et al., 2015). Recommender systems that consider social network and textual information have also emerged. Wang et al. proposed a method that incorporates explicit social network and review information into the PMF (Wang et al., 2019).

As we have seen, there have been researches that have incorporated images and social networks into PMF independently to improve their accuracy, however no research has clarified the effectiveness of using images and social networks at the same time. In this paper, we propose *Social network and Image Shape Feature Matrix Factorization* (S-ISFMF) which incorporates image shape information and social relationships into PMF for fields where visual information and social relationships are important, such as apparel and restaurants.

The contributions of this paper are as follows:

- S-ISFMF, which incorporates image shape information and social relationships into PMF, achieves improvements on three datasets over other models that add auxiliary information to matrix factorization.
- We used two types of social relationships—explicit social networks derived from real-world friendships and acquaintances, and implicit social networks derived from the degree of similarity between users and examined the difference in effectiveness between these two types of relationships.

2. METHOD

2.1 Probabilistic Matrix Factorization

Salakhutdinov et al. proposed a recommendation method called PMF, which is a matrix factorization method (Salakhutdinov & Mnih, 2007). PMF assumes users *N*, items *M*, an arbitrary integer *D*, and a scoring matrix $R \in \mathbf{R}^{N\times M}$ obtained from the user's rating information. The PMF is a matrix factorization *R* into a user matrix $U \in \mathbf{R}^{D\times N} = \{u_1, u_2, ..., u_N\}$ and an item matrix $V \in \mathbf{R}^{D\times M} = \{v_1, v_2, ..., v_M\}$. The measured score r_{ij} is made by user *i* for item *j*. In this case, *R* is expressed by the following equation.

$$p(R|U,V,\sigma^2) = \prod_i^N \prod_j^M \left[N(r_{ij}|u_i^T v_j, \sigma^2)\right]^{I_{ij}} \tag{1}$$

where σ is the Gaussian noise of *R*. I_{ij} is the indicator function that is equal 1 if user *i* rated item *j* and equal to 0 otherwise. The optimal matrix *U*, *V* minimizes the loss function ε, as shown in the following:

$$\min \mathcal{E}(U,V) = \sum_i^N \sum_j^M \frac{I_{ij}}{2}(r_{ij} - u_i^T v_j)^2 + \frac{\lambda_U}{2}\sum_i^N \|u_i\| + \frac{\lambda_V}{2}\sum_j^M \|v_j\| \tag{2}$$

where λ_u and λ_v are the L_2 regularization terms derived from the Gaussian noise of *R*, *U*, and *V*, and I_{ij} is an indicator function of 1 if user *i* evaluated item *j* and 0 otherwise.

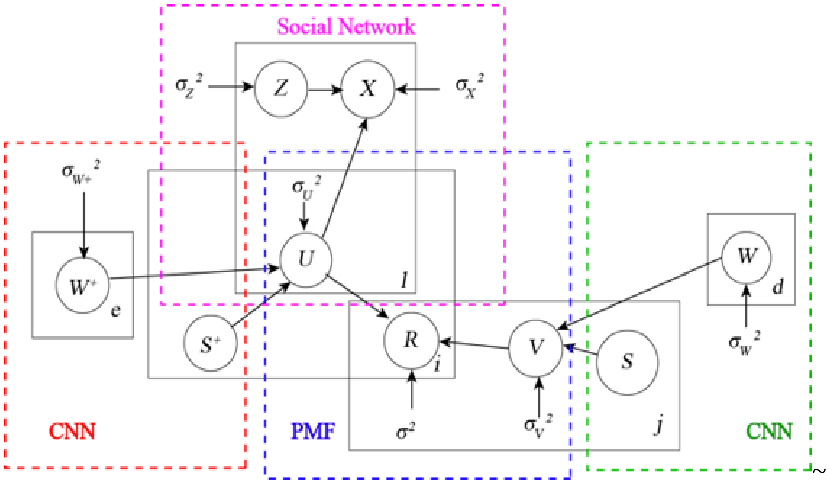

Figure 1. Graphical model of S-ISFMF

2.2 Overview of *S-ISFMF*

S-ISFMF is a model that incorporates images and social networks into PMF at the same time, referring to Bi-ISFMF (Lu et al., 2021), which extracts and integrates shape information from images representing item and user features into PMF, and SoRec (Ma et al., 2008), which incorporates social networks into PMF.

Figure 1 shows an overview of the probabilistic model of S-ISFMF, which integrates CNN and social networks that capture image contour information into PMF. As shown in this figure, S-ISFMF requires as data, in addition to the evaluation matrix *R* required by the PMF, the image of the item (S, S^{+}) and also *X* representing the social network. Then, features are extracted from these matrices and data and integrated to build a latent model of *U* and *V*. Specifically, to capture image contour information, we detect contours from item images and incorporate those features into the item matrix *V* in the PMF. In addition, we consider that the images of items purchased by the user in the past represent the user's characteristics, and we assume that those characteristics have an impact on the user matrix *U* and the item matrix *V*. We also consider that the social relationship matrix *X* is decomposed by the latent social network matrix $Z \in \mathbf{R}^{D \times N} = \{z_{k1}\ z_{k2}, \ldots, z_{kN}\}$ and *U*, just as the evaluation matrix *R* is decomposed by *U* and *V*. These elements are combined and adjusted *U* and *V* to finally decompose the evaluation matrix *R* by its *U* and *V*. However, there are cases where the social relationship matrix may or may not exist as explicit data. If an explicit network exists, then an adjacency matrix *X* can be constructed based on the relationships. For example, *X* is constructed by an adjacency matrix that represents directed graphs representing follow/follower relationships such as Yelp and Twitter, and undirected graphs representing friendship relationships such as Facebook. If the social network does not exist as data, then we assume that the social network is hidden in the user's purchase history, etc. (i.e., an implicit social network exists), and we represent the network by calculation and use it as *X*.

2.3 Mathematical Model of *S-ISFMF*

In this section, the mathematical model of S-ISFMF, is explained. In Figure 1, $S = \{s_1, s_2, \ldots, s_M\}$ is the set of images of items, and W is the weight vector of the CNN architecture of items. $S^{+} = \{s^{+}_1, s^{+}_2, \ldots, s^{+}_N\}$ is the set of images of items the user has purchased in the past. *X* represents the social network matrix, where x_{ik} takes the value 1 if a social relationship exists between user *i* and user *k*, and 0 otherwise. *Z* is also the latent social network matrix, and *U* and *Z* are obtained by decomposing *X*.

Then, from a probabilistic point of view, *X* can be expressed by the following equation and by *R* shown in Equation (1):

$$p(X|U,Z,\sigma_X^2) = \prod_i^N \prod_k^N [N(x_{ik}|u_i^T z_k, \sigma_X^2)]^{X_{ik}} \quad (3)$$

Next, the matrix decomposes *R*, *X* into latent models *U*, *V*, *Z*. In PMF, *R* is generated from *U*, *V*, and σ. In S-ISFMF, *V* is generated from *S*, W, and σ_V representing the Gaussian noise, and *U* is generated from S^+, W^+ and σ_U representing the Gaussian noise. *Z* can also be expressed in terms of σ_Z, which represents the Gaussian noise. The shape features of the image are then incorporated into *U* and *V*. The shape features of the image are denoted as $cnn(w, s_j)$ and $cnn(w^+, s_i^+)$. Here, $cnn(w, s_j)$ is the item feature vector obtained by applying the product image s_j of item *j* to a CNN that extracts image shape features (Shen et al., 2015), and $cnn(w^+, s_i^+)$ is the user feature vector obtained by applying the image s_i of the item purchased by user *i* in the past to the previously mentioned CNN.

When $cnn(w, s_j)$, $\mathrm{cnn}(\mathrm{w}^+, s_i^+)$, and σ_Z are used, *V*, *U*, and *Z* in the PMF probability model can be expressed by the following prior distribution equations:

$$p(U|W^+, S^+, \sigma_U^2) = \prod_i^N N\,(u_i|cnn(W^+, s_i^+), \sigma_U^2 I), p(V|W, S, \sigma_V^2) = \prod_j^M N\left(v_j\middle|cnn\left(W, s_j\right), \sigma_V^2 I\right) \quad (4,5)$$

$$p(Z|\sigma_Z^2) = \prod\nolimits_k^N N(z_k|0, \sigma_z^2) \quad (6)$$

where *I* is the diagonal matrix.

The weights w_d and w_e for each of *W* and W^+ can be expressed by the following two prior distribution formulas:

$$p(W|\sigma_w^2) = \prod_d N\,(w_d|0, \sigma_w^2), p\left(W^+\middle|\sigma_{w^+}^2\right) = \prod_e N\left(w_e^+\middle|0, \sigma_{w^+}^2\right) \quad (7,8)$$

Equations (4) to (8) can be rewritten as follows:

$$p\big(U, V, Z, W, W^+ \mid R, X, S, S^+, \sigma^2, \sigma_X^2, \sigma_U^2, \sigma_V^2, \sigma_Z^2, \sigma_W^2, \sigma_{W^+}^2\big)$$
$$\propto p(R \mid U, V, \sigma^2) p(X \mid U, Z, \sigma_X^2) p(U \mid W^+, S^+, \sigma_U^2) p(V \mid W, S, \sigma_V^2) p(Z \mid \sigma_Z^2) p(W \mid \sigma_W^2) p(W^+ \mid \sigma_{W+}^2) \quad (9)$$

To optimize (9), we use the following maximum a posteriori estimation:

$$\max_{U,V,Z,W,W^+} p\big(U, V, Z, W, W^+ \mid R, X, S, S^+, \sigma^2, \sigma_X^2, \sigma_U^2, \sigma_V^2, \sigma_Z^2, \sigma_W^2, \sigma_{W^+}^2\big)$$
$$= \max_{U,V,Z,W,W^+} [p(R \mid U, V, \sigma^2) p(X \mid U, Z, \sigma_X^2) p(U \mid W^+, S^+, \sigma_U^2) p(V \mid W, S, \sigma_V^2) p(Z \mid \sigma_Z^2) p(W \mid \sigma_W^2) p(W^+ \mid \sigma_{W+}^2)] \quad (10)$$

By taking the negative algorithm in (10), we can reformulate it as follows:

$$min\,\varepsilon(U, V, Z, W, W^+) = \sum_i^N \sum_j^M \frac{I_{ij}}{2}\left(r_{ij} - u_i^T v_j\right)^2 + \frac{\lambda_X}{2} \sum_i^N \sum_k^N I_{ik}(x_{ik} - u_i^T z_k)^2 + \frac{\lambda_Z}{2} \sum_k^N \|z_k\|^2 + \frac{\lambda_V}{2} \sum_j^M \left\|v_j - cnn\left(W, s_j\right)\right\|^2$$
$$+ \frac{\lambda_W}{2} \sum_d^{|W_d|} \|w_d\|^2 + \frac{\lambda_U}{2} \sum_i^N \|u_i - cnn(W^+, s_i^+)\|^2 + \frac{\lambda_{W^+}}{2} \sum_e^{|w_e^+|} \|w_e^+\|^2 \quad (11)$$

$\lambda_X, \lambda_Z, \lambda_V, \lambda_W, \lambda_U,$ and λ_{W^+} are the regularization terms derived from the Gaussian noise in *X*, *Z*, *V*, *U*, *W*, and W^+, respectively. Partial differentiation of (11) by *U* and *V* respectively yields the following equation:

$$u_i = (VI_i V^T + \lambda_X Z I_i^X Z^T + \lambda_U)^{-1} (VR_i + \lambda_U cnn(W^+, s_i^+) + \lambda_X Z x_i) \quad (12)$$

$$v_j = \left(UI_j U^T + \lambda_V I\right)^{-1} \left(UR_j + \lambda_V cnn\left(W, s_j\right)\right), z_k = (\lambda_X U I_k^X U^T + \lambda_Z I)^{-1} (\lambda_X U x_k) \quad (13,14)$$

where I_i is a diagonal matrix whose diagonal components are the evaluation vectors $\{I_{i1}, I_{i2}, \ldots, I_{i1}\}$ that indicate whether user *i* evaluated each item. Similarly, I_j is a diagonal matrix whose diagonal components are the evaluation vectors $\{I_{1j}, I_{2j}, \ldots, I_{Nj}\}$ that indicate whether item *j* was evaluated each user. Furthermore, I_i^X is a diagonal matrix with the vector $x_i = \{x_{i1}, x_{i2}, \ldots, x_{iN}\}$ representing the social relationship with each user of user *i* in its diagonal components. I_k^X is also a similar diagonal matrix for user *k*.

On the basis of (12), (13), and (14), *U*, *V*, and *Z* are updated by stochastic gradient descent to obtain the optimal user matrix *U*, item matrix *V*, and social network matrix *Z*. However, *W* and W^+ cannot be optimized in the same ways as *U*, *V*, and *Z* because they are closely related to the features of CNN architecture, such as the max pooling layer and nonlinear activation function. Therefore, we temporarily fix *U*, *V*, and *Z*, and use the error backpropagation method to estimate *W* and W^+.

$$\varepsilon(W) = \frac{\lambda_V}{2}\sum_j^M \left\|v_j - cnn(W, s_j)\right\|^2 + \frac{\lambda_W}{2}\sum_d^{|W_d|} \|w_d\|^2 + constant\,, \tag{15}$$

$$\varepsilon(W^+) = \frac{\lambda_U}{2}\sum_i^N \|u_i - cnn(W^+, s_i^+)\|^2 + \frac{\lambda_{W^+}}{2}\sum_e^{|W_e^+|^2} \|w_e^+\|^2 + constant \tag{16}$$

3. EXPERIMENT

In this section, we compare the performance of the proposed and existing methods using three real-world datasets. We also discuss how the performance of the proposed method changes when the parameters are varied. First, we describe experiment settings, including the details of the dataset, evaluation index, and parameter settings. Then, we discuss the experiment results.

3.1 Experiment Settings

For this experiment, we used two Amazon datasets, namely, Clothes and Accessories, for the apparel sector (McAuley et al. 2015). For the restaurant dataset, we used data from the Yelp dataset for the British Columbia region. These datasets include rating information, images of each item, and user reviews. The Yelp dataset has an explicit social network, but the Amazon dataset does not. For this reason, an implicit social network was created in the Amazon dataset by calculating the cosine similarity derived from the user's rating history and assuming that a network exists between users whose relationships are above a threshold value. The evaluation value for each dataset was taken from 1 to 5. This experiment used image data; thus, items without an image were excluded from these datasets. Users with fewer than two evaluations were deleted because they could not be divided into training data and test data, and thus could not be evaluated properly. The statistics for each dataset as a result of these processes are shown in Table 1. In actual experiments, 80% of these datasets were divided randomly into training data, 10% into validation data, and the remaining 10% into test data. The best values, λ_U, λ_V and λ_Z were obtained in advance using the validation data, and the test data were evaluated using these parameters.

In this experiment, we adopted root mean square error (RMSE) as an evaluation index and took an average of five trials to ensure reliability. We compared our proposed method with following models. PMF (Salakhutdinov & Mnih, 2007), ConvMF (Kim et al., 2016), SRCMF (Wang et al., 2019), ISFMF (Saga & Duan, 2018), Bi-ISFMF (Lu et al., 2021). In addition, we performed grid search for λ_U, λ_V and λ_Z in the range [0.1,1,10,20,30,40,50,60,70,80,90,100] to find best parameters. We set other parameters as follows: Dimensionality *D* of the user matrix *U* and the item matrix *V* is set to 50. The size of the item image is set to 45 × 45. The images representing the user's characteristics were images of two previously purchased items. The cosine similarity threshold was 0.5 when using implicit social networks. These parameters were chosen because they were the optimal values when comparing accuracy in prior experiments.

Table 1. Detailed statistics of the datasets

Dataset	**Users**	**Items**	**Ratings**	**Density**
Clothes	12062	14182	29547	0.0173%
Accessories	4049	4503	8824	0.0484%
Food	53391	2634	224461	0.160%

3.2 Experiment Results

3.2.1 Comparing the Proposed Method with Existing Methods

Table 2 shows the RMSE of each model. Here, S-ISFMF with I is a model that uses implicit social network, S-ISFMF with E is a model that uses explicit social network, and Improve is the percentage improvement between our proposed method and the best value of the compared methods. This table indicates that the best accuracy is obtained by the proposed method S-ISFMF in any dataset. This result suggests that incorporating image contour information and social network into PMF at the same time is useful. For each dataset, we achieved an improvement of 7.25% for the Amazon Clothes dataset, 6.97% for the Amazon Accessories dataset, and 0.75% for the Yelp dataset. Improvement was lower in the Yelp dataset than in the other two datasets possibly because textual information such as item descriptions and user reviews is as important as image information in restaurant recommendations. In fact, the next most accurate model after the proposed method was the SRCMF, which uses text and social networks. However, both the proposed method and SRCMF use social networks. This finding indicates that using social networks together with other auxiliary information is effective. As can be seen from Table 2, the RMSE for the Yelp dataset with explicit social networks was 1.451, while the accuracy with implicit social networks was 1.454. The explicit social network works slightly better than the implicit social network, but the difference is not very large. Therefore, we believe that implicit social networks can be used in the same way as explicit social networks.

Table 2. Overall RMSE on each dataset

Dataset	**Clothes**	**Accessories**	**Food**
PMF	1.781	1.761	1.834
ConvMF	1.514	1.422	1.654
SRCMF	1.336	1.239	1.462
ISFMF	1.394	1.372	1.526
Bi-ISFMF	1.255	1.233	1.521
S-ISFMF with I	**1.164**	**1.147**	1.454
S-ISFMF with E	-	-	**1.451**
Improve	7.25%	6.97%	0.75%

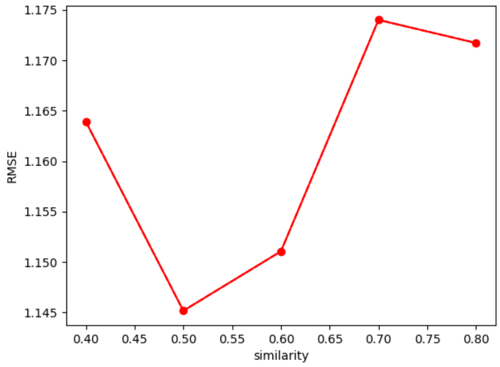

Figure 2. Impact of the similarity on implicit social network

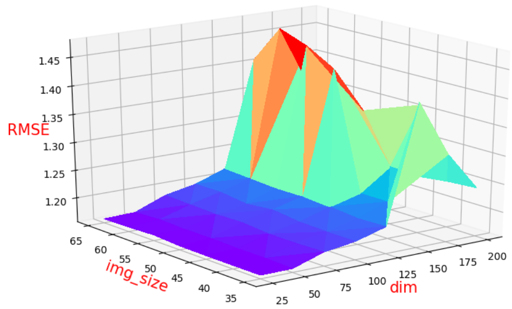

Figure 3. Impact of the dimension and image size

3.2.2 Parameter Analysis

The previous experiment showed the usefulness of using an implicit network. In this section, we first discuss the threshold in the creation of that implicit network. Figure 2 shows the relationship between similarity on threshold of implicit social network and RMSE. if the threshold is too high, then the accuracy of similarity determination will increase, but the accuracy will also decrease because the density of the social network will be smaller. Next, experiments were conducted to examine the robustness of the latent matrix U, V by changing the number of dimensions D and the size of the images used as input. Figure 3 shows the relationship between RMSE when changing D and image size using the best parameters (λ_U:100, λ_V:10, λ_Z:0.1) obtained using validation data in Amazon Clothes. This figure shows that the overall trend is that the accuracy worsens as the number of dimensions D increases. When D is sufficiently large, the accuracy worsens as the image size increases. However, when D is sufficiently small, the image size has no effect on accuracy.

3.3 Experiment Summary

Experimental results on the Amazon product dataset and the Yelp dataset yield the following findings:

- The proposed S-ISFMF obtains better results than those of existing methods.
- The predicted results are best when the similarity threshold for creating an implicit social network is 0.5.
- Implicit social networks can be used in the same way as explicit social networks.
- As *U* and *V* dimensions increase, the experimental results gradually deteriorate, with abrupt changes and fluctuations occurring in higher dimensions.
- Image size is stable and highly accurate when *U* and *V* are low dimensional. When *U* and *V* are high dimensional, the accuracy is affected by the image size.

These are some of the useful details and characteristics of the S-ISFMF that we have identified.

4. CONCLUSION

In this paper, we proposed S-ISFMF, which extracts contour information features from images that represent item and user features and incorporates them into a matrix factorization method with social networks. The proposed method achieves an improvement of 0.75% to 7.25% on three real-world datasets over other existing models that add auxiliary information to matrix factorization. The effectiveness and robustness of the implicit social network compared with the explicit social network are also discussed.

One of the current challenges is that we are not able to consider features such as item descriptions and user reviews, which are highly likely to represent item and user characteristics. Future work will include clarifying the effectiveness of extracting features from other auxiliary information, such as textual information, and incorporating them into matrix factorization.

REFERENCES

Kim, D., Park, C., Oh, J., Lee, S., & Yu, H. (2016). Convolutional Matrix Factorization for Document Context-Aware Recommendation. *Proceedings of the 10th ACM Conference on Recommender Systems*, 233–240. https://doi.org/10.1145/2959100.2959165

Li, Y., Song, M., & E, H. (2015). Recommender System Using Implicit Social Information. *IEICE TRANSACTIONS on Information and Systems*, *E98-D*(2), 346–354.

Lu, Y., Gao, M., & Saga, R. (2021). Apparel Recommender System based on Bilateral image shape features. *ArXiv:2105.01541 [Cs]*. http://arxiv.org/abs/2105.01541

Ma, H., Lyu, M. R., & King, I. (2009). Learning to recommend with trust and distrust relationships. *Proceedings of the Third ACM Conference on Recommender Systems*, 189–196. https://doi.org/10.1145/1639714.1639746

Ma, H., Yang, H., Lyu, M. R., & King, I. (2008). SoRec: Social recommendation using probabilistic matrix factorization. *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, 931–940. https://doi.org/10.1145/1458082.1458205

McAuley, J., Targett, C., Shi, Q., & van den Hengel, A. (2015). Image-Based Recommendations on Styles and Substitutes. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 43–52. https://doi.org/10.1145/2766462.2767755

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 175–186. https://doi.org/10.1145/192844.192905

Saga, R., & Duan, Y. (2018). Apparel Goods Recommender System Based on Image Shape Features Extracted by a CNN. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2365–2369. https://doi.org/10.1109/SMC.2018.00406

Saga, R., Hayashi, Y., & Tsuji, H. (2008). Hotel recommender system based on user's preference transition. *2008 IEEE International Conference on Systems, Man and Cybernetics*, 2437–2442. https://doi.org/10.1109/ICSMC.2008.4811660

Salakhutdinov, R., & Mnih, A. (2007). Probabilistic Matrix Factorization. *Proceedings of the 20th International Conference on Neural Information Processing Systems*, 1257–1264.

Shen, W., Wang, X., Wang, Y., Bai, X., & Zhang, Z. (2015). DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3982–3991. https://doi.org/10.1109/CVPR.2015.7299024

Su, X., & Khoshgoftaar, T. M. (2009). A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, *2009*, e421425. https://doi.org/10.1155/2009/421425

Tamada, T., & Saga, R. (2022). Book Recommender System Using CNN Capturing Feature of Synopses and Reviews. In J. Y. C. Chen, G. Fragomeni, H. Degen, & S. Ntoa (Eds.), *HCI International 2022 – Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence* (pp. 518–530). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-21707-4_37

Verma, D., Gulati, K., & Shah, R. (2020). Addressing the Cold-Start Problem in Outfit Recommendation Using Visual Preference Modelling. *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*. https://doi.org/10.1109/BigMM50055.2020.00043

Wang, X., Yang, X., Guo, L., Han, Y., Liu, F., & Gao, B. (2019). Exploiting Social Review-Enhanced Convolutional Matrix Factorization for Social Recommendation. *IEEE Access*, *7*, 82826–82837. https://doi.org/10.1109/ACCESS.2019.2924443