

A Novel Evolution-Based Recommendation System

Yi-Cheng Chen

Department of Information anagement
National Central University, Taiwan
ycchen@mgt.ncu.edu.tw

Yen-Lung Chu

Department of Computer Science and
Information Engineering
Tamkang University, Taiwan
yen199565@gmail.com

Lin Hui

Department of Innovative Information
and Technology
Tamkang University, Taiwan
121678@mail.tku.edu.tw

Sheng-Chih Chen

Master's Program of Digital Content
and Technologies, College of
Communication
National ChengChi University, Taiwan
cchen@nccu.edu.tw

Tipajin Thaisutikul

Department of Computer Science and
Information Engineering
National Central University, Taiwan
106582601@cc.ncu.edu.tw

Kai-Ze Weng

Department of Computer Science and
Information Engineering
Tamkang University, Taiwan
403850281@gms.tku.edu.tw

Abstract—Matrix factorization (MF) technique has been widely utilized in recommendation systems due to the precise prediction of users' interests. Prior MF-based methods adapt the overall rating to make the recommendation by extracting latent factors from users and items. However, in real applications, people's preferences usually vary with time; the traditional MF-based methods could not properly capture the change of users' interests. In this paper, by incorporating the recurrent neural network (RNN) into MF, we develop a novel recommendation system, M-RNN-F, to effectively describe the preference evolution of users over time. A learning model is proposed to capture the evolution pattern and predict the user preference in the future. The experimental results show that M-RNN-F performs better than other state-of-the-art recommendation algorithms. In addition, we conduct the experiments on real world dataset to demonstrate the practicability.

Keywords—deep learning, recommendation system, matrix factorization, Long Short-Term Memory

I. INTRODUCTION

Undoubtedly, the exponential growth of information have emerged in this era. How to efficiently and effectively retrieve relevant information for users from a massive data is a challenging issue. The goal of recommendation system is to capture the user's potential interests and suggest relevant items by users' personalize preferences. It has been proven to improve the user satisfaction and content provider revenue in many applications.

A considerable amount of research effort has been put forth on the recommendation system. Matrix factorization (MF) is one technique widely applied on recommendation. It could reduce dimensions by finding features from each user and item to improve recommendation accuracy. However, prior MF-based studies mainly focus on extracting latent factors from users' explicit feedbacks which implicitly reflected the user preferences and item characteristics. These methods may suffer from several critical issues. Clearly, in several applications, user preference usually vary with time. The latent factors extracted from MF may not properly capture the evolution of users' interests. The traditional MF-based techniques do not include the concept of time progress.

In this paper, based on matrix factorization (MF) and recurrent neural network (RNN), we develop a novel recommender system, Matrix-Based Recurrent Neural Network Factorization (abbreviated as M-RNN-F), to capture the evolving pattern from the preference evolution. We also propose a dependent learning model to describe the behavior pattern and predict the user preference in the future.

Dependent learning considers the context awareness of user and item concurrently; the evolutions of users' latent interests and items' latent characteristics are learning together. We fuse the latent factors of users and items together and learn the evolutions. The main contributions of proposed method are as follows:

- A transformation method is proposed to separate the huge user rating matrix into smaller feedback matrices based on the rating time. The feedback matrix reveals the accumulated users' rating behaviors without any information loss. Obviously, the smaller matrix we tackle, the less computation time and memory resource are required for processing.
- In advance, the relationship among users and items is a critical issue for recommend "right" things to "right" users. By incorporating the extracted latent factors of users and items in M-RNN-F architecture, the cross-relation among users and items also could be effectively learned and estimated by dependent learning model.
- We perform extensive experiments on several real-life datasets. Our results demonstrate that proposed method outperforms the other state-of-the-art recommendation models. The proposed M-RNN-F exhibits outstanding generalization ability and is robust across real datasets with different natures.

The remainder of this paper is organized as follows. Sections 2 presents related works and preliminary. Section 3 provides the proposed M-RNN-F recommendation system. Section 4 details the experiments and the performance study, while Section 5 presents the conclusion.

II. RELATED WORK

2.1 Matrix Factorization

Matrix is one of the most important basic concepts in mathematics. Matrix factorization (MF) refers to the decomposition of a matrix into the product of several matrices or the sum of some matrices by some algorithm according to a certain principle. The special forms of these factorizations can obviously reflect some characteristics of the original matrix; moreover, the methods and processes of factorizations provide some effective numerical calculation methods and theoretical analysis.

Several studies focus on effectively factorizing the target matrix. He et al. [5] design a new learning algorithm based on the eALS technique, for efficiently optimizing a MF model with variably-weighted missing data, and then seamlessly

update the MF model as new entries are given. ML-JMF [13] introduces a multi-label answer aggregation approach, which not only selectively and jointly factorizes the sample-label association matrices and shared low-rank matrices, but also takes advantage of the correlation among labels and of the similarity between annotators to guide the factorization.

To solve the extreme sparsity problem of the target rating-matrix, several MF-based methods constrain the decomposition results to be non-negative. Luo et al. [8] propose a regularized single element based NMF (RSNMF) model which focuses on the non-negative up-dated process depending on each involved feature rather than on the whole feature matrices. RSNMF utilizes the single-element-based update rule to integrate the Tikhonov regularizing terms. Semi-non-negative MF [15] is a low-dimensional representation technique for learning data sets. The representation may contain quite complex hierarchical information and implicit low-level hidden attributes, which cannot be explained by classical first-level clustering methods. Thus, Deep Semi-NMF [12] is proposed to learn hidden representations and interpret clusters based on different unknown properties of a given dataset.

Chua et al. [2] utilize the dynamic matrix factorization (DMF) techniques to de-riive different time factorization models and predict the missing time steps in different historical steps. DMF technique extends the concept from non-negative matrix factorization (NMF) and Linear Dynamics System (LDS). Sorkunlu et al. [11] design a dynamic matrix factorization algorithm, called DynamicMF, with a three-dimensional delay. DynamicMF automatically extracts low-dimensional signals in the data to track system efficiency and identify performance anomalies.

2.2 Recommendation on MF

With the impressive achievement of latent factor models (LFM) in many domains, various MF-based approaches [4, 14, 16, 17] have been proposed to leverage on user-item rating recommendation systems. Actually, the rating patterns could be inferred through the factor vectors extracted from matrix decomposition. A high degree of correspondence between user and item factors leads to facilitating the recommendations. Koren et al. [6] decompose the rating matrix into two low-dimensional matrices, which are used to describe the characteristics of the users and items, and then the inner product of the two low-dimensional matrices to reconstruct the user's score for the unrated items.

Abdi et al. [1] utilize MF approaches to improve the performance of processing large scale datasets. Authors also incorporate contextual information on the quality and accuracy for recommendation. Park et al. [10] devise an RWR method combining global bias term to MF. Authors also discuss various aspects of qualified recommendation to tackle the cold-start problem. Ma et al. [9] construct a factor analysis approach based on the probabilistic matrix factorization (PMF) to solve the data sparsity and poor prediction accuracy problems. PMF also employs both users' social network information and rating records.

Moreover, Lin et al. [7] proposed a hybrid real-time incremental stochastic gradient descent (RI-SGD) to derive the implicit Matrix Factorization, which consists of alternating least squares (ALS) with weight regularization in the training phase and stochastic gradient descent (SGD) in the updating phase. Gemulla et al. [3] develop a stratified SGD variant

(SSGD) to apply on a new MF algorithm which can handle a wide variety of Factorizations. Authors also establish sufficient conditions for convergence of SSGD using results from stochastic approximation theory and regenerative process theory.

III. M-RNN-F RECOMMENDATION SYSTEM

The architecture of M-RNN-F consists of four major components: (1) Feedback sequence transformation & matrix factorization, (2) Evolution learning, and (3) Prediction and recommendation, as shown in Fig. 1.

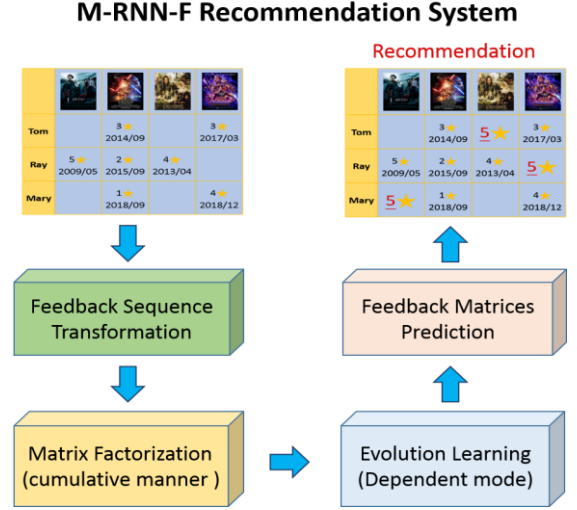


Fig. 1. The architecture of M-RNN-F system.

3.1 Feedback matrix transformation & factorization

Let $U = \{u_1, u_2, \dots, u_m\}$ be a set of users and $V = \{v_1, v_2, \dots, v_n\}$ be a set of items. A rating record is a pair (r_{ij}, t_{ij}) where r_{ij} and t_{ij} denote the rating value and the rating time generated by user u_i for item v_j , respectively. A rating matrix $R (m \times n)$ consists of all rating records (r_{ij}, t_{ij}) have been generated by $u_i \in U$. Given a rating matrix $R (m \times n)$, the task of recommendation system is to predict the rating value $r_{ij} \notin R$, i.e., the items that have not yet been rated by users.

For preprocessing, the rating matrix R is transformed into several smaller feedback matrix based on the rating time t_{ij} of a rating record as the Definition 1.

(Definition 1 Feedback Matrix and Sequence)

Given a rating matrix $R (m \times n)$ consists of all records (r_{ij}, t_{ij}) , with the user-specified started time t_0 and time interval length Δt , a feedback matrix M^k is a matrix consisting of all the rating value r_{ij} , where $t_0 \leq t_{ij} < (k \times \Delta t)$ and $k \geq 1$. The sequence $\langle M^1, \dots, M^k, \dots \rangle$ is called feedback sequence.

After feedback sequence generation, we factorize the feedback matrix to discover the latent preferences of users and characteristics of items. Due to the characteristic of the accumulated manner of feedback matrices in feedback sequence, we could easily reuse the result of previous matrix to facilitate the computation of next matrix.

(Definition 2 Preference and Characteristic Matrix)

Suppose d dimension vectors $p_i \in \mathbb{R}^d$ and $q_j \in \mathbb{R}^d$ denote the vectors of latent preference of each user and characteristic of each item, respectively. $P (m \times d)$ represents the preference

matrix of all user latent vectors $p_i, 1 \leq i \leq m$; $Q (d \times n)$ represents the characteristic matrix of all item latent vectors $q_j, 1 \leq j \leq n$.

For M_k , we compute the new user preference p_i^k and q_j^k by minimizing the loss function (1).

$$p_i^k, q_j^k = \underset{p_i, q_j}{\operatorname{argmin}} \left(\frac{1}{|\Delta M^k|} \sum_{r_{ij} \in \Delta M^k} (r_{ij} - p_i^T q_j)^2 + \lambda_p \sum_{i=1}^m \|p_i\|^2 + \lambda_q \sum_{j=1}^n \|q_j\|^2 \right) \quad (1)$$

where λ_p and λ_q are two positive constants. In this paper, we adopt gradient decent method to find the optimization in (1). We iteratively adapt the p_i and q_j by the functions (2) and (3),

$$p_i^\ell = p_i^{\ell-1} - \eta \cdot \left(-\frac{2}{|\Delta M^k|} \sum_{r_{ij} \in \Delta M^k} q_j (r_{ij} - p_i^T q_j) + 2\lambda_p p_i^{\ell-1} \right) \quad (2)$$

$$q_j^\ell = q_j^{\ell-1} - \eta \cdot \left(-\frac{2}{|\Delta M^k|} \sum_{r_{ij} \in \Delta M^k} p_i (r_{ij} - p_i^T q_j) + 2\lambda_q q_j^{\ell-1} \right) \quad (3)$$

where ℓ is the learning iteration index and η is the learning rate.

Definitely, from (2) and (3), we could find that η controls how fast that p_i and q_j change to reduce the mean squared error in (1).

3.2 Evolution learning

After factorizing the feedback sequence, we could derive the latent preference of each user and characteristic of each item. In next step, the latent features are processed and trained by the learning module to capture the evolution patterns. In this paper, by extending the idea of recurrent neural network (RNN), a learning model is developed to capture and describe the complicated sequential information and long-range evolutions contained in user preferences and item characteristics. The recurrent layer takes the sequence of latent vectors (including preference and characteristics) factorized by MF as inputs. Then, it outputs the hidden state step by step. These hidden states are called as the current status of the latent vectors. During this processing, memory cell will capture the long-term evolution by memorizing not only the order of users' and items' variations, but also the relationships among them.

We utilize the series of factorized latent matrices including the information of users' latent preferences and items' latent characteristics to model the short-term and long-term evolutions simultaneously. The idea of learning is extended and modified from LSTM model; intuitively, the pattern of variation of both user preference and item characteristic could be captured by recurrent component in standard LSTM architecture. The concept of learning is given as Fig. 2. The main learning component is D-LSTM (Dependent LSTM). When using dependent learning for the item recommendation, the combination of user preference and item characteristic can be exploited to learn user's and item's short-term interests.

Similar to original LSTM, D-LSTM consists of one cell state and three controlled gates to keep and update the memory. While memory cell C_{t-1} contains the historical combination of user's preference and item's characteristic, which reflect the long-term interest evolution, the h_{t-1} stores the previous output result. The objective functions of D-LSTM are as follows,

$$f_t = \sigma(W_f(P_t + Q_t) + U_f h_{t-1} + b_f) \quad (4)$$

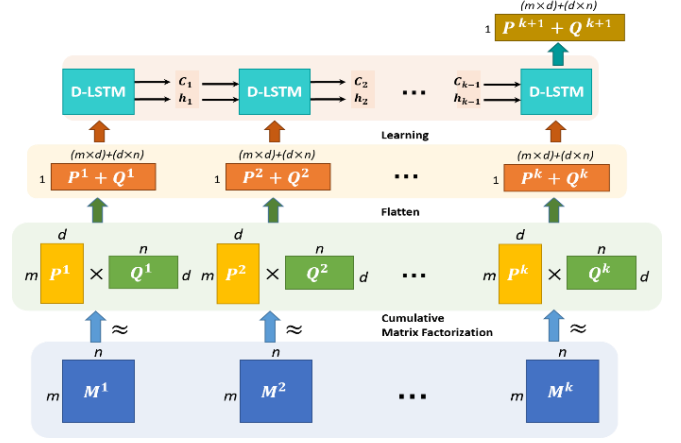


Fig. 2. The concept of dependent learning model.

$$i_t = \sigma(W_i(P_t + Q_t) + U_i h_{t-1} + b_i) \quad (5)$$

$$o_t = \sigma(W_o(P_t + Q_t) + U_o h_{t-1} + b_o) \quad (6)$$

$$Z_t = \sigma(W_z(P_t + Q_t) + U_z h_{t-1} + b_z) \quad (7)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ Z_t \quad (8)$$

$$h_t = o_t \circ \tanh(C_t) \quad (9)$$

The inputs are $P^t + Q^t$: the concatenation of the flatten preference matrix and flatten characteristic matrix, C_{t-1} : memory cell of the combination of preference and characteristic history, h_{t-1} : previous output result. We first derive the forget gate f_t , input gate i_t , and output gate o_t , deciding how much information will be forgotten, input, and output by (4) to (6), respectively. σ represents a sigmoid function to map the values between 0 to 1, where 0 represents completely ignoring the content and 1 represents completely keeping this content. W and U are the learning weights, and b is the bias of each gate. Z_t is learning from the input $P^t + Q^t$ and previous result h_{t-1} in Eq. (7). Finally, in Eqs. (8) and (9), we derive the new memory cell C_t and output h_t . Note that with the data-driven concept, the parameters of the network learning (i.e., W , U and b) are all adjusted concurrently with the rest of network parameters by back-propagation.

3.3 Prediction and recommendation

After feeding a feedback sequence $\langle M^1, \dots, M^k \rangle$ into D-LSTM, we could have the final output result $P^{t+1} + Q^{t+1}$. Then M-RNN-F decompose $P^{t+1} + Q^{t+1}$ into two matrices, preference matrix $P^{t+1} (m \times d)$ and characteristic matrix $Q^{t+1} (d \times n)$. By multiply P^{t+1} and Q^{t+1} together, we could derive the prediction matrix $R' = P^{t+1} \times Q^{t+1}$. The rating value $r'_{ij} \in R'$ but $\notin R$ is called the prediction value of user u_i for item v_j . Given a user-specified number N , for the query about user u_i , M-RNN-F will list the top- N largest prediction values in i -th-row of R' .

IV. PERFORMANCE EVALUATION

In this section, we evaluate the effectiveness and robustness of proposed M-RNN-F recommendation system

comparing with the static MF model. The dataset and data preprocessing procedure are discussed as in Table 1.

TABLE I. THE MOVIELENS DATASETS

Dataset	#Users	#Movies	#Ratings	Avg #Rating	Collected Interval	#Months	Rating Zone
MovieLens 1M	6,040	3,706	100,209	166	2000/04 ~ 2003/03	36	1-5
MovieLens 100K	943	1,682	100,000	106	1997/09 ~ 1998/04	8	1-5

Since both two datasets are sparse, to analyze the impact of each user on the movie ratings of each month, we first group the data according to the timestamp of month, and transform into the feedback matrices and sequence. For MovieLens 100K, the transformed feedback sequence has 8 feedback matrices; for MovieLens 1M, the transformed feedback sequence has 36 feedback matrices. Then we segment the 80% of dataset as training set, and the remaining 20% one for the test set.

To evaluate the effectiveness of our proposed M-RNN-F, we use accuracy metric with different size of latent vector d . d setting controls the sizes of two factorized results at month t , preference matrix P^t and characteristic matrix Q^t . We choose $d = 100, 200, 300, 400, 500$, and 800 to illustrate the different accuracy results. To measure the statistical accuracy of prediction, we use the mean absolute error (MAE) and root mean square error (RMSE) as the metrics to evaluate the quality of prediction results. The MAE and RMSE are defined as,

$$MAE = \frac{\sum_{i=1}^n |\hat{r}_i - r_i|}{n} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{r}_i - r_i)^2}{n}} \quad (11)$$

where n is the number of total predicted rating, \hat{r}_i the predicted rating for the i th movie, r_i is the user's true rating value for the i th movie. MAE is defined as the average absolute difference between predicted ratings and actual ratings; likewise, RMSE is the average root square difference between predicted values and actual values. Both measures are frequently used to assess the goodness of predicted values by a model or an estimator.

We demonstrate the performance of proposed M-RNN-F comparing to traditional static MF method. All models are implemented on keras. First, we apply M-RNN-F on MovieLens 100K and 1M datasets to discuss the MAE experimental results. The performance of MAE metric evaluation is shown in Tables 2 and 3. We could observe that M-RNN-F has overall best performance comparing to static MF method. In advance, without any doubt, the d setting is also an important issue for movie recommendation.

TABLE II. MAE@D ON THE MOVIELENS 100K DATASET

	MovieLens 100K dataset					
	MAE@ d=100	MAE@ d=200	MAE@ d=300	MAE@ d=400	MAE@ d=500	MAE@ d=800
Static MF	0.855	0.878	0.892	0.982	0.963	0.837
M-RNN-F	0.900	0.860	0.845	0.899	0.785	0.795

As shown in Tables 2 and 3, when $d=200$ to 800 , M-RNN-F has better MAE on MovieLens 100K dataset; however, when $d=100$ to 500 , M-RNN-F has better MAE on MovieLens 1M dataset.

TABLE III. MAE@D ON THE MOVIELENS 1M DATASET

	MovieLens 1M dataset					
	MAE@ d=100	MAE@ d=200	MAE@ d=300	MAE@ d=400	MAE@ d=500	MAE@ d=800
Static MF	0.576	0.584	0.611	0.614	0.622	0.541
M-RNN-F	0.564	0.571	0.606	0.611	0.564	0.609

Next, we have the experiment on the RMSE metric on MovieLens 100K and 1M datasets. The experimental results are as Tables 4 and 5. We could find that M-RNN-F has overall better performance comparing to static MF method. Similar to the above case, as shown in Tables 4 and 5, when $d=200$ to 800 , M-RNN-F has better RMSE on MovieLens 100K dataset; especially when $d=500$, the proposed method has the best result. However, M-RNN-F has better RMSE on MovieLens 1M dataset when $d=100$ to 500 .

TABLE IV. RMSE@D ON MOVIELENS 100K DATASET

	MovieLens 100K dataset					
	RMSE@ d=100	RMSE@ d=200	RMSE@ d=300	RMSE@ d=400	RMSE@ d=500	RMSE@ d=800
Static MF	0.924	0.937	0.945	0.991	0.991	0.915
M-RNN-F	0.948	0.927	0.917	0.948	0.886	0.892

TABLE V. RMSE@D ON MOVIELENS 1M DATASET

	MovieLens 1M dataset					
	RMSE@ d=100	RMSE@ d=200	RMSE@ d=300	RMSE@ d=400	RMSE@ d=500	RMSE@ d=800
Static MF	0.759	0.764	0.781	0.784	0.788	0.736
M-RNN-F	0.753	0.756	0.762	0.778	0.751	0.768

In summary, we set the pre-trained matrix factorization size of $d=100$ to 800 , and the training batch size and epoch to $200,000$ and 100 , respectively. From the experimental results, the proposed D-LSTM model in M-RNN-F is significantly superior to the baseline method in MAE and RMSE.

V. CONCLUSION

In this information explosion era, how to recommend right thing to right user is a challenging issue. In this paper, we point out the importance of preference evolution. By combining the recurrent neural network and Matrix Factorization, we develop a novel recommendation system, M-RNN-F, to effectively describe the preference evolution of users over time. A learning model is also proposed to capture the evolution pattern and predict the user preference in the future. The experimental results show that M-RNN-F performs better than static MF-based recommendation algorithm. In addition, we conduct the experiments on real world dataset to demonstrate the practicability.

VI. ACKNOWLEDGMENT

Yi-Cheng Cheng was supported in part by the Ministry of Science and Technology of Taiwan, Project, No. MOST-106-2221-E-008-104-MY2.

REFERENCES

- [1] M. Abdi, G. Okeyo and R. Mwangi, "Matrix Factorization Techniques for Context-Aware Collaborative Filtering Recommender Systems: A Survey," *Computer and Information Science*, vol. 11, no. 2, 2018.

- [2] F. CHUA, R. Oentaryo and E. LIM, "Modeling Temporal Adoptions Using Dynamic Matrix Factorization," *IEEE 13th International Conference on Data Mining (ICDM)*, pp. 91-100, 2013.
- [3] R. Gemulla, E. Nijkamp, P. Haas and Y. Sismanis, "Large-scale matrix factorization with distributed stochastic gradient descent," *17th ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD)*, pp. 69-77, 2011.
- [4] J. He, X. Li, L. Liao, D. Song, and K. Cheung, "Inferring a personalized next point-of-interest recommendation model with latent behavior patterns," *Proceedings of the 13th AAAI Conference on Artificial Intelligence (AAAI 2016)*, pp. 137-143, 2016.
- [5] X. He, H. Zhang, M. Kan and T. Chua, "Fast matrix factorization for online recommendation with implicit feedback," *39th International ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 549-558, 2016.
- [6] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *IEEE Computer*, 42: 30-37, 2009.
- [7] C. Lin, L. Wang, K. Tsai, "Hybrid Real-Time Matrix Factorization for Implicit Feedback Recommendation Systems," *IEEE Access*, 6: 21369-21380, 2018.
- [8] X. Luo, M. Zhou, Y. Xia and Q. Zhu, "An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems," *IEEE Transactions on Industrial Informatics*, vol. 10, issue: 2, pp. 1273-1284, 2014.
- [9] H. Ma, H. Yang, M. Lyu and I. King, "SoRec: social recommendation using probabilistic matrix factorization," *17th ACM conference on Information and knowledge management (CIKM)*, pp. 931-940, 2008.
- [10] H. Park, J. Jung and U. Kang, "A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems," *IEEE International Conference on Big Data (IEEE BigData)*, pp. 756-765, 2017.
- [11] N. Sorkunlu, D. Luong and V. Chandola, "dynamicMF: A Matrix Factorization Approach to Monitor Resource Usage in High Performance Computing Systems," *IEEE International Conference on Big Data (IEEE BigData)*, 2018.
- [12] G. Trigeorgis, K. Bousmalis, S. Zafeiriou and B. Schuller, "A Deep Matrix Factorization Method for Learning Attribute Representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, issue: 3, pp. 417-429, 2017.
- [13] J. Tu, G. Yu, C. Domeniconi, J. Wang, G. Xiao and M. Guo, "Multi-Label Answer Aggregation based on Joint Matrix Factorization," *IEEE 18th International Conference on Data Mining (ICDM)*, pp. 517-526, 2018.
- [14] Z. Wu, H. Tian, X. Zhu, and S. Wang, "Optimization Matrix Factorization Recommendation Algorithm Based on Rating Centrality," *Third International Conference on Data Mining and Big Data (DMBD)*, LNCS 10943: 114-125, 2018.
- [15] H. Yu, H. Huang, I. Dihillon and C. Lin, "A Unified Algorithm for One-Class Structured Matrix Factorization with Side Information," *31st AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2845-2851, 2017.
- [16] J. Zhang and C. Chow, "CRATS: An LDA-Based Model for Jointly Mining Latent Communities, Regions, Activities, Topics, and Sentiments from Geosocial Network Data," *IEEE Transactions on Knowledge and Data Engineering (TKDE 2016)*, vol. 28, no. 11, pp. 2895-2909, 2016.
- [17] S. Zhao, M. Lyu, and I. King, "STELLAR: Spatial-Temporal Latent Ranking Model for Successive POI Recommendation," *SpringerBriefs in Computer Science Point-of-Interest Recommendation in Location-Based Social Networks*, pp. 79-94, 2018.