

CDMF: A Deep Learning Model based on Convolutional and Dense-layer Matrix Factorization for Context-Aware Recommendation

Mingxin Gan*, Yingxue Ma and Kejun Xiao

Department of Management Science and Engineering
Donlinks School of Economics and Management
University of Science and Technology Beijing, Beijing 100083, China
ganmx@ustb.edu.cn

Abstract

We propose a novel deep neural network based recommendation model named Convolutional and Dense-layer Matrix Factorization (CDMF) for Context-aware recommendation, which is to combine multi-source information from item description and tag information. CDMF adopts a convolution neural network to extract hidden feature from item description as document and then fuses it with tag information via a full connection layer, thus generates a comprehensive feature vector. Based on the matrix factorization method, CDMF makes rating prediction based on the fused information of both users and items. Experiments on a real dataset show that the proposed deep learning model obviously outperforms the state-of-art recommendation methods.

1. Introduction

The last few years have witnessed an explosion of information caused by the exponential growth of the Internet and World Wide Web, which confronted us with information overload and brought about an era of big data, appealing for efficient personalized recommender systems to assist the screening of useful information from various sources. To further improve user experiences and reflect personalization of individual users, various recommender systems have also been proposed recently in an active way towards personalized nomination of resources and to assist individuals for efficiently identifying their potential interest, leading to a variety of successful stories in the online recommendations of books [1], movies [2], news [3], TV programs [4], microblogs [5] and many others [6, 7, 8].

Classic recommendation methods do have some shortcomings and deficiencies. The traditional collaborative filtering methods rely on user's rating

matrix and it is inferior in recommendation accuracy. Secondly, such traditional methods as regression-based models and network based methods focus on similarity calculation, therefore lack associations extracting of multi-source information from both users and items. Thirdly, such matrix factorization methods currently developed resort to feature extraction, and lack the use of multi-source context information from items, as tags, description, and so on. Existing studies try to use machine learning and deep learning models in recommendation [9, 10]. Such methods based on machine learning use Bayesian methods [11], clustering, ANNs, linear regression to minimize the error between predicted ratings and historical ratings.

Deep learning is an emerging technique in the big data era [12], with lots of powerful models developed by utilizing deep neural networks. Although deep learning models have achieved outstanding performance in extracting hidden features, the problem of multi-source information fusion has not been effectively solved in the application fields, especially in recommendation. Therefore, to overcome these limitations, we propose a recommendation method based on a deep convolutional neural network, a dense-layer and a matrix factorization.

The remaining of the paper is structured as follows. In Section 2 we present the overview of the related work. In Section 3 we present details of the proposed model. Then we present the experiments and results in Section 4. Finally, we conclude the paper in Section 5.

2. Related Work

2.1 Classical Methods in Recommendation

Collaborative filtering is the most typically designed recommendation method. A user-based design uses historical data to calculate similarities

between users and relies on such information to calculate discriminant scores for recommending candidate objects [3, 13]. An item-based design is formally equivalent to its user-based counterpart by exchanging the roles of users and objects [14]. To make use of information of objects, a content-based method further uses properties of objects to characterize their similarities [15]. To promote respective advantages of these two categories of approaches, hybrid approaches have also been proposed [16].

One of the most important aspects in a collaborative filtering method is the quantification of associations that measure relatedness of users or objects. Such a measure is typically derived from a transformation of matched relationships between characteristics of objects and preferences of users [17, 18]. For example, in widely used methods such as the cosine vector similarity, users are represented as vectors of objects according to historical data, and similarities are calculated as the cosine of the angle between the vectors.

Recent studies have also shown that recommendation methods based on the simulation of a diffusion process exhibit higher performance over classical collaborative filtering methods [8, 19]. For instance, network based methods have been widely used in recent years, which has been demonstrated that the simulation of interactive between users and items effectively improve recommendation accuracy. Besides, the heat-spreading method enhances recommendation diversity [20]. Social network among users extends the user-item relationship network used in recommendation [21]. However, these methods are short in using multi-source information, such as item description and tag information.

2.2 Deep Learning Models in Recommendation

Deep learning methods have been currently adopted in various areas in recommendation, including rating prediction [22], text recommendation [23], image recommendation [24], and location recommendation [25, 26].

Deep neural network (DNN) creates a denser high-level semantic abstraction by combining low-level features to automatically discover distributed representations of data, which has been mainly used to learn the hidden features from users and items in recommendation [27]. It usually reconstructs user or item-related information (including rating data, text and images, *etc.*) to obtain an implicit representation for a user or for an item, supporting further

predictions of user preferences. DNNs fit complex nonlinear relationships by identifying and learning deep features, often used as tools to extract hidden-layer features in complex associations between users and items [28, 29]. Convolutional neural network (CNN) is a focus in current research for image understanding. It works well when processing image and dealing with feature learning via local perception, pooling dimension reduction.

The idea of local perception has been widely applied in the feature-extracted full-channel and local attention channels, used as feature extraction tools to obtain text features [30]. Multi-mode information also has been adopted in the recommendation of microblogging [31], using CNN and RNN to extract features from images and texts, and then combining them to make tag recommendations. In addition, CNN learns implicit representations of users and items from user reviews [32, 33]. Moreover, CNN has also been used for feature learning and extraction in image-related and music-related recommendations [34], generating better performance to traditional methods.

The application of neural networks in recommendation is mainly in modeling the sequence impact from data, thus helping to obtain effective representation of users and items. It mainly relies on two aspects: one is the sequence pattern in users' behavior, such as *ratings*, *purchasing*, should be considered in recommendation, and the other is that the use of *words* in modeling text information of both user and item should be extracted with implicit representations [35].

3. Convolutional and Dense-layers Matrix Factorization

3.1. Overview of CDMF

The overview of our proposed method CDMF is illustrated in Figure 1. We firstly initialize a latent vector from Gaussian distribution randomly for the representation of user information. Secondly, as for the item description, we resort to a CNN model in feature extraction and a vector that represents hidden features is generated in Figure 1(2). Thirdly, combined with the tag vectors as the input, we fuse multi-source information and generate feature vectors that represent items. Finally, the loss function is defined as the error between predicted rating and actual rating to adjust model parameters based on a back propagation process, and then obtain the optimal parameters for recommendation. Then,

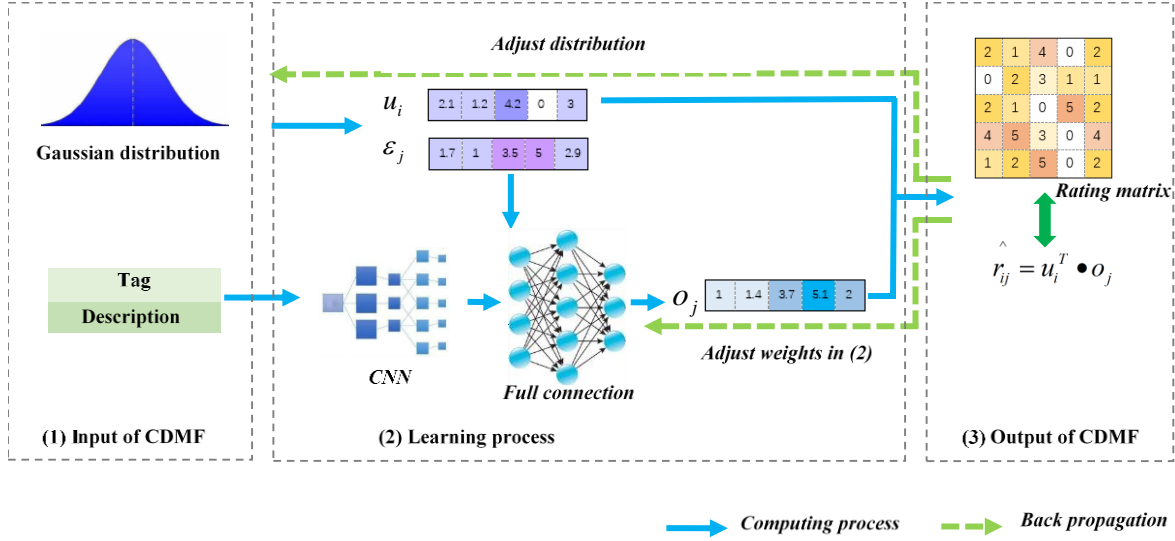


Figure 1: Convolutional and dense-layers matrix factorization (The black arrow indicate tensor forward flows and green is update parameters backpropagation)

combined with the user vector, we obtain predicted ratings via matrix factorization.

3.2. Vector Representation in CDMF

We partitioned the item information based on one of the following scenarios. When two items are similar in their descriptions, the combined consideration of tag information and item description can distinguish them obviously. For example, storylines of two stories, which are very close to each other, may not be distinguished obviously by their descriptions. While tag information, such as different year and actors, can help. Therefore, we adopt two kinds of contextual information for items, as listed in Table 1. We think tag information is a structured indicate as item feature and usually easy to compute while not easy to understand, even some tags need abundant semantic explanation. Item description is an unstructured summary for overall feature of items, which is usually a piece of neural language. The transformation from unstructured data to structured data is key and difficult in feature extraction process.

Item Tags. An item generally has multiple attributes with different values. For example, a certain attribute set. Thus we denote an item matrix $A_{(n, k)}$ as a binary matrix, where $A = \{T_{11}, T_{12}, \dots, T_{1n}; T_{21}, T_{22}, \dots, T_{2m}; \dots, T_{k1}, T_{k2}, \dots, T_{kn}\}$ and the element $A_{i,j}=1$ presents that item i has attribute j .

Test Description. Such contexts for items, as the summary of a movie, reflect an overall profile with

meaningful descriptions, which are comprehensive for human but need to be automatically extracted and converted into formal and uniform vectors for machine learning. To achieve this, we resort to a widely used natural language processing method Word2Vec [36], to get feature vectors for items and further calculate their similarities.

Table 1. Tag information and item description.

Feature	Value	Examples
Tag	Year, genre, actor	1998, Adventure Animation Children Jones
Description	Summary, storyline	Four 12-year-old girls grow up together during an eventful small-town summer in 1970.

3.3. Information Processing in CDMF

We illustrate the data processing in our proposed method Convolutional and Dense-layers Matrix Factorization (CDMF), by firstly illuminating the feature extraction process from item description, secondly explaining the dimension reduction process via a full connection layer, and finally developing the rating prediction process.

3.3.1 Hidden feature extraction via an improved convolutional neural network (CNN). A *general convolutional neural network* includes an input layer, a convolution layer, a pooling layer, a full connection layer and an output layer. The convolutional layer performs convolution calculation on input data

through convolution kernels, adds bias to obtain the feature map after feature extraction, and adds the weighted sum of multiple feature mappings as the input of pooling layer. After many repeats of similar operation, the features from bottom layer to the high layer are finally delivered to full connected layer to get significant feature extraction. The pooling layer is also called subsampling layer, which reduces the dimension of data, and retains representative features to reduce the amount of computation in the model.

The improved convolutional neural network. Different from traditional methods of text similarity calculation, we take item description as the input to realize the automatically extraction process for hidden features that are first represented as vectors and then learned via different convolution kernels.

3.3.2 Fusing tags and text descriptions via dimension reduction. To effectively use multi-source information in recommendation, we fuse information of item tags, text description and user-item ratings.

First, we merge two obtained vectors directly, one of which represents full information extracted from item tags, and the other is the low-dimensional vector as text descriptions, which is extracted by the improved convolutional neural network, mentioned in the above sub-section. Thus, two vectors are merged into an o -dimension one.

Then, to avoid the sparsity and high-dimension of the vectors and to emphasize the most important features for the item, we use a full connection layer to make dimensional deduction for the merged vector and obtain an r -dimension vector.

3.3.3 Rating prediction via matrix factorization and back propagation. We adopt matrix factorization for rating prediction, which is based on the mathematical matrix eigenvalue factorization to decompose a rating score matrix $R_{(|U|, |V|)}$ into a user matrix $U_{(|U|, k)}$ and an item matrix $O_{(k, |V|)}$, and uses the product of $U_{(|U|, k)}$ and $O_{(k, |V|)}$ to represent $R_{(|U|, |V|)}$, where $|U|$ is the number of user and $|V|$ is the number of item.

Thus, a high-dimensional sparse rating matrix is decomposed into two matrices with lower dimensions and less sparseness, so that the rating of each user on the item $\hat{r}_{ij} = u_i \times o_j^T$ can be obtained by the inner product of the two low-dimensional vectors. According to the descending order of the predicted rating, we obtain recommendation lists for all user [37].

4. Experiments

4.1 Dataset

The dataset was obtained from MovieLens-10m (<https://grouplens.org/>) which consist of users' explicit ratings on items on range of 1 to 5 and tags information of items (movies), such as year, genres, actors (Table 1). Description documents such as summaries and storylines of movie and clerks table of item, we obtained them from movies information pages on IMDB provided by MovieLens researchers (<https://www.imdb.com/>). For optimizing results, we drop movies that are lack of above information and rated less than twice. Thus, the dataset contains 9,390,713 ratings by 69,878 users on 9,492 movies, the mean length of description of each movie is 72.5 and the overall density is 1.42%.

We preprocessed item description according to [37, 38], by firstly sentences to set maximum length of raw documents to 195 (the maximum length of all cleaned sentences in the dataset), then removing stop words, stemming word by Porter stemmer and thirdly selecting most frequency occurring 8,000 words to build vocabulary, and finally setting all out-of-vocabulary words to special label. To fuse tags information of movies into our model, we gathered three important tags for each movie, they are year, genres and actors. We only select actors who occurs twice or more in the dataset and top 10 actors for each movie. Figure 2 shows the distribution of the tag information.

4.2 Evaluation criteria

To evaluate recommendation performance of each model on the MovieLens dataset, we randomly divided the data set into a training set (80%), a validation set (10%) and a test set (10%). We use RMSE as the evaluation criteria, and the difference between the predicted and actual ratings represents the performance of the model. We use results in validation set to tune our model and finally report the results evaluating in test set.

$$RMSE = \sqrt{\frac{\sum_i \sum_j \left(\hat{r}_{ij} - r_{ij} \right)^2}{n_ratings}} \quad (4)$$

4.3 Methods comparison

To evaluate our proposed method, we select three comparison methods and illustrate them in brief as below.

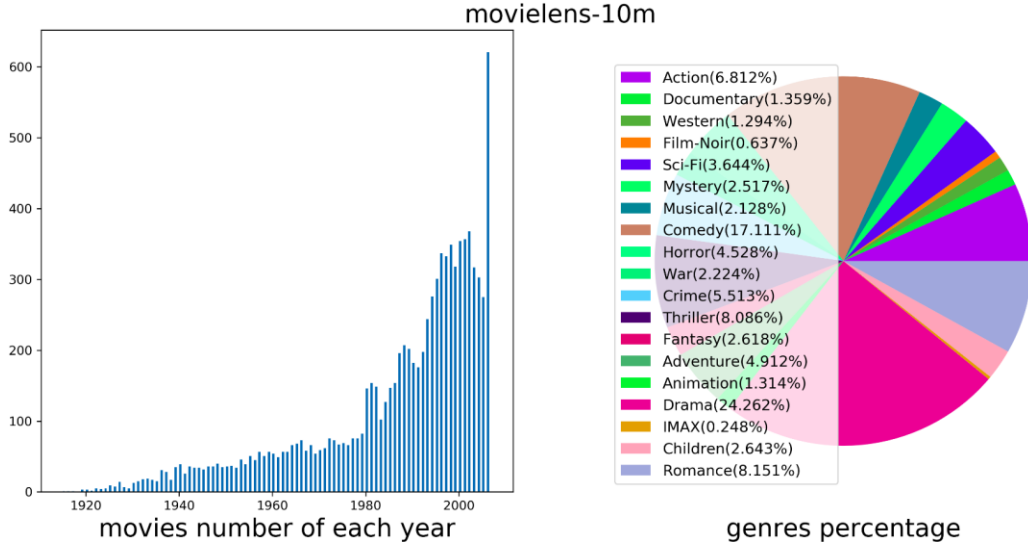


Figure 2: Distribution on year and genre.

NMF: Non-negative Matrix Factorization is a state-of-the-art recommendation model [39].

PMF: Probabilistic Matrix Factorization is a standard rating prediction model that only uses ratings [40].

CTR: Collaborative topic regression (CTR) is a recently proposed tightly coupled method, which was proposed for recommendation on implicit feedback data. CTR is a probabilistic graphical model that seamlessly integrates a topic model, Latent Dirichlet Allocation (LDA), and a model-based CF method, probabilistic matrix factorization (PMF) [41]. In our experiment, CTR is for processing text topics in movie description but without using tag information. Because we use explicit feedback datasets, we set the rating to 1 if u_i rate on o_j and 0 otherwise.

5. Results

Experiment results on MovieLens-10m dataset are shown in Table 2. In general, our proposed model CDMF obviously outperforms others. For instance, the improvements of CDMF compared to NMF and PMF are around 11.00%, and 4.26%, respectively.

Table 2. RMSE of different model

Model	RMSE	Improvement
NMF	0.885	11.00%
PMF	0.822	4.26%
CTR	0.819	3.93%
CDMF	0.787	-

NMF is very similar to PMF, but the latent factors of users and items are positive. So NMF only reflects users' preferences, and cannot reflect users' negative opinions on items. Consequently, the performance of NMF is worst. Although PMF is good at avoiding overfitting, the ability of learning latent factors is a little bit poor, and history ratings are adopted only in PMF, so less information is utilized than CTR and CDMF.

The comparison between CDMF and CTR indicates that the representation of sentences learnt by topic model is worse than convolutional neural network. The limitation of LDA topics mixture of CTR is the lack of consideration of the relative positions among words, perhaps caused by its underlying unigram text pattern [41].

The results of both experiments demonstrate that CDMF integrating tags and documents simultaneously outperforms other methods. It is proved that the fused comprehensive feature vector effectively reflects external characteristics of items, *e.g.* tags, and help to generate item latent vector.

It indicates that using tag information and textual modeling together addresses the problems lacking of clearly source of text. In a word, CDMF inherits the idea of classical recommendation method on using rating matrix, and at the same time, the integration of the comprehensive item information by deep learning has obtained a more accurate prediction result, and the RMSE was decreased over 3.93% by the best competitor CTR.

6. Conclusions

We proposed a novel deep learning recommendation model, namely CDMF, to combine item tags and text descriptions for personalized recommendation. Firstly, text descriptions are embedded to a matrix via the embedding method word2vec, then hidden features are extracted by a convolution neural network, and the item tags are integrated. After a full connection layer for fusing different features, a comprehensive feature vector for items is generated. We set deviation adjustment for vectors of both users and items, and finally get the predicted user ratings with higher accuracy by the inner product of the item and user vector according to the idea of matrix factorization.

Though the experimental validations, the result of our model CDMF shows that methods considering both text descriptions and tags of an item are supposed to be more effective than those using partial characteristics of an item, which indicates the necessity of fusion multiple source information is a promising research focus in future studies.

As for future work, firstly, one can consider other prediction methods as the alternatives of the basic matrix factorization method, such as PMF, SVD and deep neural network models. Secondly, we choose a simple structure when we use CNN to extract hidden features from text description. Therefore, the development of CNN structure can be further considered, especially, the idea of three-channel in image recognition for information fusion.

7. Acknowledgements

This work was partly supported by the National Natural Science Foundation of China (Nos. 71871019, 71471016, 71531013, 71729001).

8. References

[1] B. Jeong, J. Lee, and H. Cho, "Improving memory-based collaborative filtering via similarity updating and prediction modulation", *Information Sciences*, 2010, pp. 602-612.

[2] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering", *IEEE Internet Computing*, 2003, pp. 76-80.

[3] G. Nie, H. Xia, and X. Li, "An Ontology-Based Approach on Intelligent Recommendation in Movie Field", *Proceedings of the 6th International Conference on Innovation and Management*, 2009, pp. 1489-1494.

[4] T. Bogers, and A. V. D. Bosch, "Fusing Recommendations for Social Bookmarking Web Sites", *International Journal of Electronic Commerce*, 2011, pp. 31-72.

[5] S. Prawesh, and B. Padmanabhan, "Probabilistic news recommender systems with feedback", *Proceedings of the sixth ACM conference on Recommender systems*, 2012, pp. 257-260.

[6] A. B. Barragáns-Martínez, "A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition", *Information Sciences*, 2010, pp. 4290-4311.

[7] J.S. Cheng, A. Sun, D. N. Hu, and D. Zheng, "An Information Diffusion Based Recommendation Framework for Micro-Blogging", *Social Science Electronic Publishing*, 2011, pp. 463-486.

[8] B. Sarwar, G. Karypis, G. Badrul, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms", *International Conference on World Wide Web ACM*, New York, 2001, pp. 285-295.

[9] X. X. Wang, and Y. Wang, "Improving Content-based and Hybrid Music Recommendation using Deep Learning." *ACM International Conference on Multimedia ACM*, Orlando Florida, USA, 2014, pp. 627-636.

[10] A. Singhal, P. Sinha, and R. Pant, "Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works", *International Journal of Computer Applications*, 2017, pp. 17-22.

[11] A. M. Elkahky, Y. Song, and X. He, "A Multi-View Deep Learning Approach for Cross Domain User Modeling in Recommendation Systems", *International Conference on World Wide Web*, Florence, Italy, 2015, pp. 278-288.

[12] C. Ono, Y. Motomura, and H. Asoh, "Recommendation Algorithm based on Bayesian Network", *Ieice Technical Report Neurocomputing*, 2004, pp. 55-60.

[13] O. Georgiou, and N. Tsapatsoulis, "The Importance of Similarity Metrics for Representative Users Identification in Recommender Systems", *Artificial Intelligence Applications and Innovations Proceedings DBLP*, Ayia Napa, Cyprus, 2010, pp. 12-21.

[14] S.Y. Wei, N. Ye, S. Zhang, X. Huang, and J. Zhu, "Item-Based Collaborative Filtering Recommendation Algorithm Combining Item Category with Interestingness Measure", *International Conference on Computer Science & Service System IEEE*, 2012, pp. 2038-2041.

[15] I. J. Pérez, F. J. Cabrerizo, and E. Herrera-Viedma, "Group decision making problems in a linguistic and dynamic context", *Expert Systems with Applications*, 2011, pp. 1675-1688.

- [16] R. Burke, "Hybrid Recommender Systems: Survey and experiments", *User Modeling and User-Adapted Interaction*, 2002, pp. 331-370.
- [17] F. Francois, A. Pirotte, J.M. Renders, and M. Saerens, "Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation", *IEEE Transactions on Knowledge & Data Engineering*, 2007, pp. 355-369.
- [18] Q. M. Zhang, A. Zeng, and M. S. Shang, "Extracting the information backbone in online system", *Plos One*, 2013, pp. e62624.
- [19] F. François Fouss, K. Francoise, L. Yen, A. Pirotte, and M. Saerens, "An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification", *Neural Networks*, 2012, pp. 53-72.
- [20] T. Zhou, Z. Kuscsik, J. G. Liu, M. Medo, J. R. Wakeling, and Y. C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems", *Proceedings of the National Academy of Sciences of the United States of America*, 2008, pp. 4511-4515.
- [21] J. C. Ying, W. N. Kuo, V. S. Tseng, and H. C. Lu, "Mining user check-in behavior with a random walk for urban point-of-interest recommendations", *Acm Transactions on Intelligent Systems & Technology*, 2014, pp. 1-26.
- [22] S. Seo, J. Huang, H. Yang, and Y. Liu, "Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction", *Eleventh ACM Conference on Recommender Systems ACM*, Como, Italy, 2017, pp. 297-305.
- [23] Y. Li, J. Nie, Y. Zhang, B. Wang, B. Yan, and F. Weng, "Contextual recommendation based on text mining", *International Conference on Computational Linguistics: Posters Association for Computational Linguistics*, 2010, pp. 692-700.
- [24] C. Lei, D. Liu, W. Li, Z. J.Zha, and H. Li, "Comparative Deep Learning of Hybrid Representations for Image Recommendations", *Computer Vision and Pattern Recognition IEEE*, Las Vegas, Nevada, 2016, pp. 2545-2553.
- [25] R. F. Feitosa, S. Labidi, and A. L. S. D. Santos, "Social Recommendation in Location-Based Social Network Using Text Mining", 2013, pp. 67-72.
- [26] Zeng, Xuelin, and W. U. Bin, "Parallelized recommendation algorithm in location-based social network", *Journal of Computer Applications*, 2016.
- [27] S. David, H. Aja, J. M. Chris, G. L. S. Arthur, and D. George, "Mastering the game of go with deep neural networks and tree search", *Nature*, 2016, pp. 484-489.
- [28] J. Peng, and D. D. Zeng, "Exploring Information Hidden in Tags: A Subject-Based Item Recommendation Approach", *Social Science Electronic Publishing*, 2010.
- [29] Q. T. Truong, and H. W. Lauw, "Visual Sentiment Analysis for Review Images with Item-Oriented and User-Oriented CNN", *ACM on Multimedia Conference ACM*, California, USA, 2017, pp. 1274-1282.
- [30] S. Novoselov, O. Kudashev, V. Schemelinin, I. Kremnev, and G. Lavrentyeva, "Deep CNN based feature extractor for text-prompted speaker recognition", *International Conference on Acoustics, Speech and Signal Processing*, 2018.
- [31] Y. Y. Gong, and Q. Zhang, "Hashtag recommendation using attention-based convolutional neural network", *International Joint Conference on Artificial Intelligence AAAI Press*, New York, 2016, pp. 2782-2788.
- [32] S. Seo, J. Huang, and H. Yang, "Representation Learning of Users and Items for Review Rating Prediction Using Attention-based Convolutional Neural Network", *International Workshop on Machine Learning Methods for Recommender Systems*, 2017.
- [33] X. Wang, L. Yu, and K. Ren K, "Dynamic Attention Deep Model for Article Recommendation by Learning Human Editors' Demonstration", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM*, 2017, pp. 2051-2059.
- [34] S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation", *International Conference on Neural Information Processing Systems Curran Associates Inc*, 2013, pp. 2643-2651.
- [35] L. H. Li, W. Chu, Langford, and John, "A contextual-bandit approach to personalized news article recommendation", 2010, pp. 661-670.
- [36] Y. A. Chung, C. C. Wu, C. H. Shen, H. Y. Lee, and L. S. Lee, "Audio Word2Vec: Unsupervised Learning of Audio Segment Representations Using Sequence-to-Sequence Autoencoder", *INTERSPEECH*, 2016, pp. 765-769.
- [37] Y. Koren, R. Bell, and C. Volinsky. "Matrix Factorization Techniques for Recommender Systems", *Computer*, 2009, pp. 30-37.
- [38] C. Wang, and D. M. Blei, "Collaborative topic modeling for recommending scientific articles", *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM*, San Diego, U.S.A, 2011, pp. 448-456.
- [39] H. J. Duan, "Matrix Factorization Techniques with Temporal Dynamics for Recommender Systems," *Microcomputer Applications*, 2013.

[40] Z. J. Zhang, and H. Liu, "Application and Research of Improved Probability Matrix Factorization Techniques in Collaborative Filtering. Int J Control Autom", International Journal of Control & Automation, 2014, pp. 79-92.

[41] S. Purushotham, Y. Liu, and C. C. J. Kuo, "Context-aware collaborative topic regression with social matrix factorization for recommender systems", Palo Alto California AAAI Press, 2014, pp. 239-242.