



Discrete Deep Learning Based Collaborative Filtering Approach for Cold Start Problem

Archana Kalidindi^{1*}Prasanthi Yavanamandha¹Anusha Nagina Kunuku¹

¹*Department of Information Technology,
Gokaraju Rangaraju Institute of Engineering and Technology, India*

* Corresponding author's Email: archana.buddaraju@gmail.com

Abstract: The recommendation system plays a major role in e-commerce services and online applications such as social-networking, product recommendation, etc. The recommendation system uses the most popular techniques namely Collaborative Filtering (CF) and Deep Learning Neural Network (DLNN) approach. The Cold-Start (CS) problem and the recommendation efficiency are considered as crucial challenges and affected the efficiency of the most popular techniques. In this paper, a Discrete Deep Learning (DDL) based on hashing framework is implemented to achieve Hamming space for mapping Users and Items (UIs). Hamming distance used here to calculate the user's preference for an item, and the efficiency of online recommendation significantly improved by this computation technique. The performance of DDL framework evaluated by conducting various experiments on the large Netflix rating dataset. The experimental outcomes showed that online recommendation efficiency and CS recommendation accuracy of DDL outperformed the existing methods.

Keywords: Cold-start problem, Collaborative filtering, Discrete deep learning, Deep learning neural network, Recommendation system.

1. Introduction

In recent years, the context of Machine Learning (ML) made a significant progress in the field of Artificial Intelligence. While making decisions, the features obtaining in the practical applications support the users, e.g. making the decisions fully autonomous in recommender system [1]. Recommender Systems (RS) have been developed for more than twenty years, which aims to overcome the information overload problems and to provide users with personalized services. The RS mostly use in knowledge management systems and e-commerce applications. RS is responsible for delivering accurate and reliable information to the specific users [2]. The RS model collects the relevant information and the preferences are given to the individual or several users. There are several techniques available for generating recommendations. The CF and content-based methods mainly useful for recommendation techniques [3]. Collaborative

filtering is a traditional technique that mainly use in areas such as books, music, and movies, where users express their preferences by rating those items. It mainly focused on the prediction of user interest by exploiting information about preferences of other users automatically. The users in past with Similar Preferences (SP) will tend to have the SP in the future and this is the basic assumption of the CF approach [4].

The memory-based algorithms and the model-based algorithms are the two classified algorithms that mainly use in CF methods [5]. In memory-based techniques, an aggregate function of the ratings of some other user calculated as the value of unknown ratings (usually, the top-N most similar ratings of a user). The other technique known as model-based CF estimates the parameters of statistical models for ratings of user's items to provide the recommendations [6]. Although, the CF-based approach suffers from the most challenging issues like CS problem or sparsity. If a system fails to

provide practical support, new users will quickly lose interest and stop using it [7]. The CS problem can be categorized as new-user and new-item. The traditional methods aim to find an additional information for the CS problem, unfortunately, additional information is not available in common [8]. Recently, DL techniques attract a huge interest in the RS field. The problem of overfitting and time-consuming is the two important issues in the learning method, even though DL considered more an efficient than other typical neural networks [9]. The existing methods develop a hierarchical Bayesian model called Collaborative Deep Learning (CDL) as a novel tightly coupled method for RS to address the above mentioned challenges. The CDL model allows the two-way interaction by tightly coupled deep representation learning for collaborative filtering and content information in the ratings (feedback) matrix [10]. The existing methodology needs to improve the method's performance by using other DL techniques and also requires additional computational resources of the RS.

To overcome the above issues, the paper provides a Discrete Deep Learning (DDL), which combines a DL method, and CF framework based on the hashing recommendation framework. The DDL method helps learn the binary representations of UIs collected from content data and rating. The representation of items from the content information integrated into the proposed DDL scheme by using Deep Belief Network (DBN).

The proposed paper organized as follows. Section 2 describes the related works that represent the DL methods with limitations by using several techniques. Section 3 presents the proposed DDL model with DBN for rating the frequently used items in detail. Section 4 explains the database used in this framework and the extensive experiments with their outcomes. Section 5 concludes the paper.

2. Literature review

Researchers suggested several techniques on the web-page recommendation system. In this scenario, a brief evaluation of some important contributions to the existing techniques presented below.

H. Wu, Z. Zhang, K. Yue, B. Zhang, J. He, and L. Sun, [11] implemented a model for reducing the data sparsity problem by using deep neural networks with Dual-Regularized Matrix Factorization (DRMF). The distributed representations of the contents in UIs were independently generated by adopting the DRMF model with neural networks like convolutional network and gated recurrent. The outcome statement showed that the DRMF model

was very effective in the task of predicting the rates and also achieved gains in the task of top-n recommendation. The method considered only the temporal dependency by screening the UIs with reviews. The temporal dynamics on ratings were not considered that was a major drawback of the DRMF method.

M. Fu, H. Qu, Z. Yi, L. Lu, and Y. Liu [12] proposed a DL method which imitated an effective intelligent recommendation system to avoid the disadvantage of CF by understanding the UIs beforehand. The framework contained two stages such as generation of predicted ratings by using a multi-view Feed-Forward NN (FFNN) and learning low-dimensional embedding for both UIs. The co-occurrences of UIs grasped by using the representation of low-dimensional embedding. Multi-View FFNN were responsible to proceed each factor into consideration for capturing the interaction between UI. The advantage of the method was pre-trained representations contained co-occurrence information with respect to UI and this information were flexible, easily used in the multi-views networks since it is weakly coupled. The overall performance of the method was affected by the training time which was a major limitation of this approach.

J.S. Guan, M. Xu, and X.S. Kong [13] implemented Neural Language Models (NLM) in the social recommendation task for appropriate UI representation learning. The UI interaction matrix mainly focused on modelling and the framework consider only the item-level context. The user social relationship information was modelled in the paper and moreover, a layer is added to incorporate UI context. The social regularization established recommendation based on matrix factorization using derived representation, which was more flexible context. The advantage of the framework was when users with an observed social relationship had similar representation with their potentially interesting items. The disadvantage was the method didn't address the fundamental challenges when applying NLM in the recommendation.

T.K. Paradarami, N.D. Bastian, and L.J. Wightman [14] presented a DLNNs for utilizing reviews to generate model-based predictions combined with content-based features for the business user combinations. The method showed that the combination of content and CF allowed for the improvement of an NN model. The goal of the methodology was rating misclassification error and minimize log loss using optimization algorithm such as stochastic gradient descent. The set of experiments was conducted on the method that proved the hybrid approach was a very promising solution compared to

stand-alone memory-based collaborative filtering method. The method identified the latent factors within the review text provides the poor quality of recommendations by applying natural language processing method.

J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang [15] proposed two models for solving the CS problem which was categorized into Complete CS (CCS) and Incomplete CS (ICS) problems for new items known as Integrated Recommendation CF and Deep Learning (IRCD). The framework that combined CF approach and Stacked Denoising Auto Encoder (SDAE) of DLNNs were used for extracting the Content Features Items (CFI). Some experiments were tested in the framework and the outcome of the experiments showed that the combined approach was very effective and feasible to solve the CS problems compared to the other existing approaches. The CF-SDAE was more general and might be applied in other recommender systems such as social networking applications and online shopping. The model required more resources like memory storage and computational resource which is the major limitation of the method. Moreover, the RSME rating predictions evaluated by the CF-SDAE model that leads to poor performance of the recommendation systems.

To overcome the above issued presented by several techniques, the paper presented the DDL methods to solve the CS problems in recommendation systems.

3. Proposed methodology

The paper proposed a common framework by integrating the CF approach and machine learning algorithms for improving the performance of recommendation for CS items. Based on the general framework, specific system design and models are presented, that DDL model extracted the CFI using DLNN from the content that is considered as an important factor for the recommendation model. Moreover, the model effectively reduces the sparse and CS issues by implementing hashing-based hybrid recommendation framework with the combination of hashing and DBN techniques. The investigation of the DDL model carried out on Netflix movie recommendation with nearly 100 million ratings. The proposed methodology explained in the following sections.

3.1 Deep learning of content features

The proposed model DDL needs additional content descriptions for the items, but, the traditional CF models were unable to estimate the CS item

ratings. The CS rating estimations predicted by item features collected from the content and used with a CF model. The generation of vectors from the raw content information of all items processed based on the approach called bag of words. SDAE helps to obtain the CFI by learning the associated vectors of items that used in the CF models. SDAE is a deep network that used in the CF models. SDAE is a deep network that stacked by multiple Denoising Auto-Encoders (DAEs). In SDAE, each layer trained as a DAE and reconstructing its input that is previous layer's output for minimizing the error. The method contains main parts namely encoding part and decoding part in which the encoding part consider as a first half layer of the network whereas the decoding part takes the last half layer. The encoding part learned the feature representations of the noise-corrupted input and the process of cleaning the input itself can be reconstructed by the decoding part in the output. An example for the construction of SDAE presented in the following Fig. 1.

Initially, an L-Layer SDAE and the combination of C vectors considered as a raw content information of all items that solves the problem of optimization as described in Eq. (1).

$$\min_{W_l, b_l} \|C - C_L\|^2 + \lambda \sum_l \|W_l\|^2 \quad (1)$$

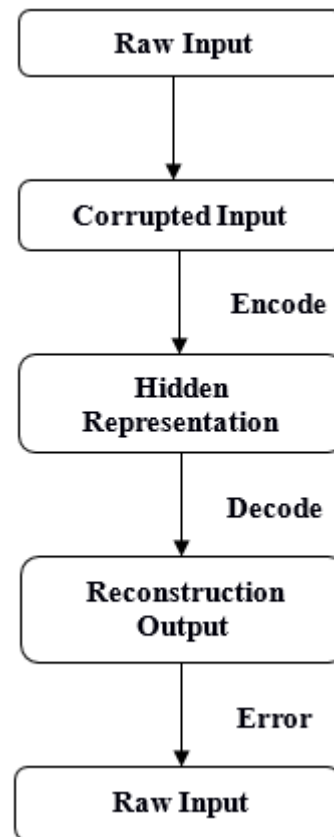


Figure. 1 A graphic structure of SDAE

Where C_L indicates the outcome of layer L of the network

W_l and b_l demonstrate the matrix weight and bias vector of layer l of the network.

Once the model is trained, the CFI could be obtained from the hidden layer $CL/2$ of the network. The feature representation which is represented by θ_i is a vector with low dimensions for a given item i .

The goal of learning the features achieved by the proposed method from the rating records, and the dimensionality reduction of the item based on content vectors with latent factor using SDAE is achieved and then the vectors fused into CF process.

3.2 Discrete deep learning (DDL)

Discrete Deep Learning (DDL) is a framework of hybrid recommendation system based on hashing techniques which adds hashing technique into the hybrid recommender framework that consists of DBN and CF by exploiting rating and item content data. Hence, DDL can obtain effective hash codes by joining the two objectives such as DBN based objective and CF-based objective. The following part describes the methodology of deep belief networks.

3.2.1. Deep belief network

The paper uses the DBN technique to obtain an effective item representation by extracting deep hierarchical item representations from item content data. DBN is a generative probabilistic method contains one input layer and multiple hidden layers, whereas training of all layers in a joint process is computationally intractable. The greedy algorithm is adopted for training the DBN in a greedy layer-wise manner and training the hidden layers once at a time in a bottom-up way moreover, the above mentioned-unsupervised procedure known as pre-train. After adding an extra learning objective, the unsupervised representation converts into a supervised learning process by pre-train stage, then tune the all specification of the DBN parameters in favour of concerning the additional learning objective. The fine-tuning process usually implemented by Back-Propagation (BP) algorithm.

For each item j , the real-valued representation is f_j that can be automatically learned by the pre-train procedure of DBN from the bag-of-words c_j :

$$f_j = DBN(c_j, \theta) \quad (2)$$

The framework DDL addresses the issues of sparsity of data and CS items by exploiting the content information of items using DL with the help

of DBN. One objective is to minimize the difference between the learning representation f_j and the binary representation (hash code) d_j for each item j , and it is given by Eq. (3).

$$\underset{D, \theta}{\operatorname{argmin}} \sum_{j=1}^m \|d_j - DBN(c_j, \theta)\|_F^2 \quad (3)$$

where $\theta = \{W, b\}$ is the parameter of an l -layer DBN that contains weight matrices $W = \{W(1), W(2), \dots, W(l)\}$ and bias vectors $b = \{b(1), b(2), \dots, b(l)\}$, D is stacked by d_j , where $j \in J$. Therefore, by reducing the objective function in Eq. (3), the model contains effective items' hash codes by the deep hierarchical framework from content data.

3.2.2. Collaborative filtering approach

The CF-based method's objective is to minimize the difference between the detected rating s_{ij} and the predicted preference. These predicted preferences of user i for item j is defined in Eq. (4),

$$p_{ij} = \frac{1}{2} + \frac{1}{2r} b_i^T d_j \quad (4)$$

We assume that $s_{ij} \in S$ is in the range of 0 to 1 denoting the rating of user i to item j , and get the following objective from Eq. (4),

$$\underset{B, D}{\operatorname{argmin}} \sum_{(i,j) \in V} \left(s_{ij} - \frac{1}{2} - \frac{1}{2r} b_j^T d_j \right)^2 \quad (5)$$

where V is the index set of S, B, D respectively which are the hash codes of UIs, and r is the hash code length. By reducing the objective in Eq. (5), the method can obtain an effective hash code of UIs from the observed rating data.

By considering the above two objectives of Eq. (3) and Eq. (5), the main function of the DDL is explained in Eq. (6).

$$\underset{B, D, \theta}{\operatorname{argmin}} \sum_{(i,j) \in V} (2rs_{ij} - r - b_j^T d_j)^2 + \frac{\lambda}{2} \sum_{j=1}^m \|d_j - DBN(c_j, \theta)\|_F^2 \dots, s.t. B \in \{\pm 1\}^{r \times n}, D \in \{\pm 1\}^{r \times m} \quad (6)$$

where $\lambda > 0$ is a tuning parameter that weights the precedence of two objectives. The entropy of each binary bit need to increase by adding a balance constraint, so that each bit carries as much information as possible. In addition, learning the compact binary codes, irrelevant constraints also

need to impose, that guarantees each bit is independent of others. In other words, there is no redundant information in the obtained hash codes.

3.3 Rating prediction model

Next, the DDL based hybrid techniques presented for the prediction rating of CS items. First, two baseline rating prediction approaches based on content similarity presented. At last, the framework for CS item is predicted by integrating the DLL model and content similarities based approach. The rating evaluation for CS items first uses similarity measure for relating the CS items with non-CS items and then evaluate ratings from their most related non-CS items. The method uses Pearson's correlation coefficient formula for computing the similarity between non-CS items and CS items, which depends upon the item features obtained from the SDAE DL process. The similarity is predicted for any two feature vectors for user's items in the Eq. (7).

$$S_{ij} = \frac{\sum_{k=1}^d (\theta_{ik} - \bar{\theta}_i)(\theta_{jk} - \bar{\theta}_j)}{\sqrt{\sum_{k=1}^d (\theta_{ik} - \bar{\theta}_i)^2 \cdot \sum_{k=1}^d (\theta_{jk} - \bar{\theta}_j)^2}} \quad (7)$$

Where θ_i, θ_j are the mean values of vectors θ_i and θ_j .

The framework predicts the ratings for CS items by considering two baseline approaches. The first approach called Top-of-All (ToA) is used to predict the CS items rating of missing values by using their M most similar items within the whole non-CS item set are examined. The prediction of ratings by user u on CS item j can be computed by the Eq. (8).

$$r_{uj} = \frac{\sum_{i \in S^M(j)} r_{ui} \cdot S_{ij}}{\sum_{i \in S^M(j)} S_{ij}} \quad (8)$$

The ratings r_{uj} can be both real and predicted ones.

Top-of-User (ToU) approach is a second approach for predicting the ratings of CS items within the set of non-CS items rated by a user from their similar items. The Eq. (9) describes the prediction of values given by user u on CS item j for the ToU approach,

$$r_{uj} = \frac{\sum_{i \in S^M(u,j)} r_{ui} \cdot S_{ij}}{\sum_{i \in S^M(u,j)} S_{ij}} \quad (9)$$

In this case, r_{ui} are real ratings in the training set.

4. Experimental outcome

In this section, the performance of the proposed recommendation model called DDL based hybrid techniques evaluated. This section discussed about the experiments, data preparation and results.

4.1 Experimental dataset

The evaluation of the DDL model presented using a real dataset which created by the Netflix Prize (<https://www.kaggle.com/netflix-inc/netflix-prize-data>). The Netflix dataset contains more than 100 million explicit ratings on a scale of 1 to 5 stars for 17770 movies defined by 480189 anonymous users. The DDL method chose some values for experimental settings to predict the overall performance of the model. The models can work with the general configurations of parameters. As there is a number of values for the parameters, the experiments consume more time to test all the combinations of settings for the parameters. The method considered only a small value of parameters for finding the performance to overcome the time-consuming process (e.g., setting feature vector dimensionality d to 50). Table 1 represents the rating statistics for training and testing sets of Netflix datasets.

Table 1. Statistics of both data for experiment datasets

	L= 100		L=300	
	Data for training	Data for testing	Training Dataset	Testing dataset
Number of users	476691	11764	476691	51171
Number of movies	14557	100	14357	300
Number of ratings	95959733	16112	95874146	101699
Mean rates	3.6042	3.4975	3.6043	3.5652

4.2 Result and analysis

In this subsection, the model considers the performance metrics such as Root Mean Square Error (RMSE) for evaluating the model and compared with the other existing recommendation methods. The prediction outcome of the DDL model can be evaluated by using RMSE, which is defined as:

$$RMSE = \sqrt{\frac{1}{N_p} \sum_{u,i} (r_{ui} - \tilde{r}_{ui})^2} \quad (10)$$

Where N_p denotes the total number of predictions, \tilde{r}_{ui} are the real predicted rating in the user's item.

4.2.1. Performance evaluation of proposed model on Netflix datasets

In this section, first evaluated the predicting performance of DDL based hybrid recommendation model for CS items. Table 2 provides the prediction result of RMSE which calculated a number of related items M , the size of test dataset and L for CS movies. A comparison done with ToA model, ToU model, IRCD for CCS (IRCD-CCS) model [15] and the proposed model. The number of significant items designed between 20 to 100, and the L size test

dataset for CS movies is configured as 100 to 300, respectively. The Figs. 2 and 3 represent the outcome measure of RMSE for both $M = 20$ and $M = 100$.

The results represented in Table 2 shows that the DDL model performs well by comparing with the existing methods such as ToA, ToU, IRCD-CCS method [15]. The outcome of the framework provides better results than the baseline models. The existing method ToA achieved nearly 1.15 RMSE values in both large size dataset, when $M=20$, whereas ToU achieved nearly 1.14 RMSE values. When $M=100$, the existing methods such as ToA, ToU achieved nearly 1.25 RMSE values in both size of the datasets like $L=100$ and 300. Because of less computation resources and memory storage, the existing methods leads poor performance. The result shows an improvement of about 0.126 on RMSE compared to the existing model of IRCD-CCS. The outcome of both IRCD-CCS model and proposed approach improves largely as M increases. The performance of the existing models affects due to the non-continuous of the item recommendation, therefore there is no need to retrain the ICS model frequently because it incurs a large computational cost. The issues can be overcome by improving the RMSE values in the proposed framework.

Table 2. Performance comparison of prediction models with Netflix dataset.

Approaches	ToA [15]		ToU [15]		IRCD-CCS [15]		Proposed Method	
	M=20	M=100	M=20	M=100	M=20	M=100	M=20	M=100
RMSE (L=100)	1.155	1.224	1.133	1.113	1.075	1.053	0.948	0.927
RMSE (L=300)	1.134	1.218	1.140	1.127	1.096	1.082	0.992	0.965

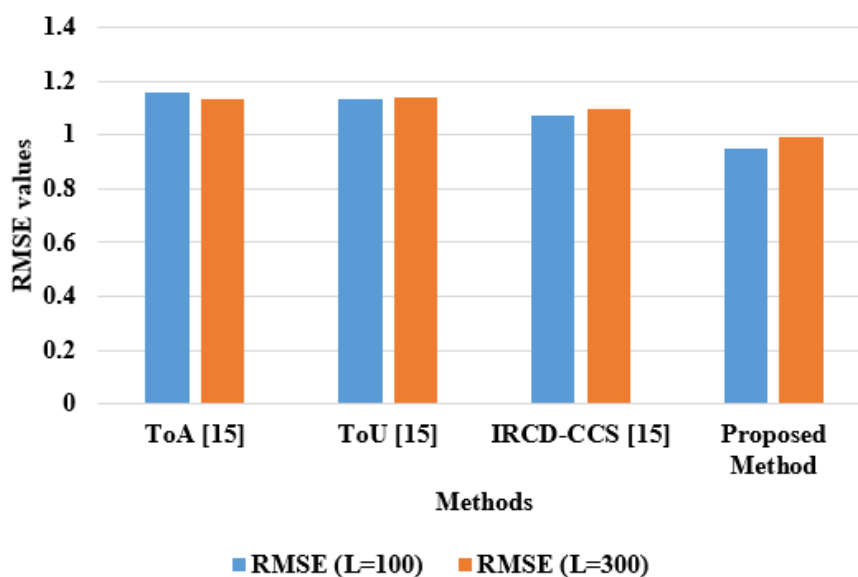


Figure. 2 Performance measure of RMSE at M=20

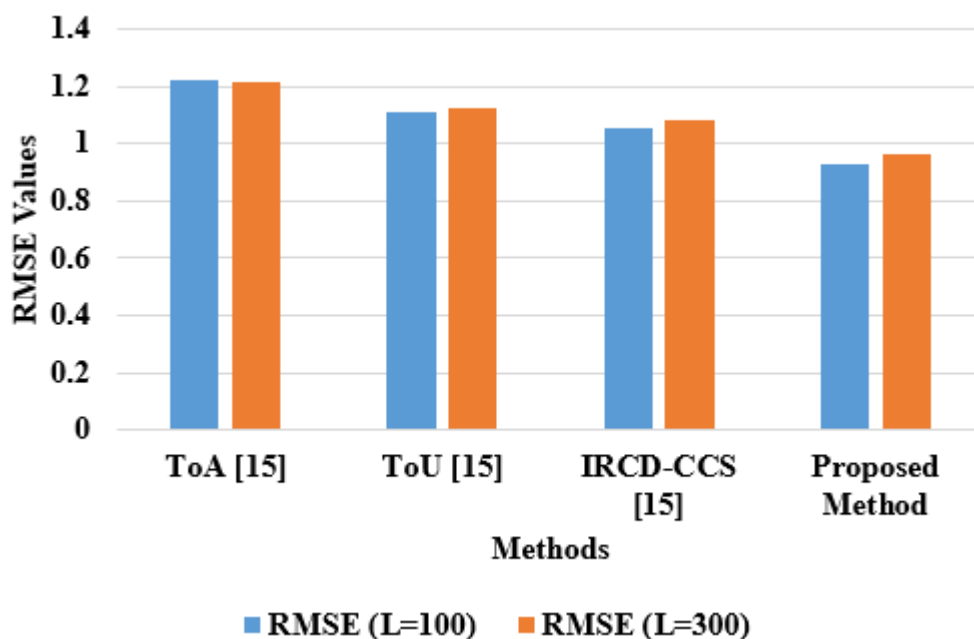


Figure. 3 Performance measure of RMSE at M=100

5. Conclusion

The recommendation system still facing challenges and issues for CS items. In this DDL framework, the two models used for recommending the items are implemented to resolve the problems of CS items. The models combine DLN with DDL which is made up of DBL and CF approach. The method is responsible for extracting the CFI by using DLNN SDAE and the process of prediction of unknown ratings can be evaluated by DDL model. First, the method formulates a preference model based on hash codes. Second, depending upon the preference model, the framework presents the DDL recommendation model by adding a binary constraint on the combined objective of the supervised DBN and CF-based objectives. The temporal dynamics of item features and user preferences are considered by the method and a number of experiments are conducted on Netflix dataset for predicting the recommendation error RMSE. The DDL models performed well compared to existing baseline approaches that are proven by the results of CS item recommendation. The proposed method achieved 0.948 RMSE value for L=100, and 0.992 RMSE value for L=300, when M=20. The proposed method achieved nearly 0.95 RMSE value for both L=100 and 300, when M=100. The impact of including the time and item content information are very large from the analysis and experimental results. The existing method reflects the poor performance because of extra storage and computation resources. The proposed DDL presented a better trade-off between

efficiency and accuracy of the recommendation system. In future, the recommendation of Netflix Price data will be improved with other deep learning techniques to reduce the problem of cold start in effective way.

References

- [1] N. Ingrid and D. Jannach, "A systematic review and taxonomy of explanations in decision support and recommender systems", *User Modeling and User-Adapted Interaction*, Vol.27, No.3-5, pp.393-444, 2017.
- [2] K. Rahul and O.P. Verma, "Recommender system with grey wolf optimizer and FCM", *Neural Computing and Applications*, pp.1-9, 2016.
- [3] Y. Zuo, J. Zeng, M. Gong, and L. Jiao, "Tag-aware recommender systems based on deep neural networks", *Neurocomputing*, Vol.204, pp.51-60, 2016.
- [4] F. Zhang, N.J. Yuan, D. Lian, X. Xie, and W. Y. Ma, "Collaborative knowledge base embedding for recommender systems", In: *Proc. of the 22nd ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, pp.353-362, 2016.
- [5] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, "On deep learning for trust-aware recommendations in social networks", *IEEE Transactions on Neural Networks and Learning Systems*, Vol.28, No.5, pp.1164-1177, 2017.
- [6] K. Rasoul, A. Nanopoulos, and L. Schmidt-Thieme, "A supervised active learning

- framework for recommender systems based on decision trees”, *User Modeling and User-Adapted Interaction*, Vol.25, No.1, pp.39-64, 2015.
- [7] E.A. Mamdouh, Y. Song, and X. He, “A multi-view deep learning approach for cross domain user modeling in recommendation systems”, In: *Proc. of the 24th International Conf. on World Wide Web*, pp.278-288 2015.
- [8] Q. Zhang, D. Wu, J. Lu, F. Liu, and G. Zhang, “A cross-domain recommender system with consistent information transfer”, *Decision Support Systems*, Vol.104, pp.49-63, 2017.
- [9] M.M. Hassan, M.Z. Uddin, A. Mohamed, and A. Almogren, “A robust human activity recognition system using smartphone sensors and deep learning”, *Future Generation Computer Systems*, Vol.81, pp.307-313, 2018.
- [10] W. Hao, N. Wang, and D.Y. Yeung, “Collaborative deep learning for recommender systems”, In: *Proc. of the 21st ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, pp. 1235-1244, 2015.
- [11] H. Wu, Z. Zhang, K. Yue, B. Zhang, J. He, and L. Sun, “Dual-Regularized Matrix Factorization with Deep Neural Networks for Recommender Systems”, *Knowledge-Based Systems*, Vol.145, pp.46-58, 2018.
- [12] M. Fu, H. Qu, Z. Yi, L. Lu, and Y. Liu, “A Novel Deep Learning-Based Collaborative Filtering Model for Recommendation System”, *IEEE Transactions on Cybernetics*, Vol.PP, No.99, pp.1-13, 2018.
- [13] J.S. Guan, M. Xu, and X.S. Kong, “Learning social regularized user representation in recommender system”, *Signal Processing*, Vol.144, pp.306-310, 2018.
- [14] T.K. Paradarami, N.D. Bastian, and L.J. Wightman, “A hybrid recommender system using artificial neural networks”, *Expert Systems with Applications*, Vol.83, pp.300-313, 2017.
- [15] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, “Collaborative filtering and deep learning based recommendation system for cold start items”, *Expert Systems with Applications*, Vol.69, pp.29-39, 2017.