An Item-based Multi-Criteria Collaborative Filtering Algorithm for Personalized Recommender Systems

Qusai Shambour, Mou'ath Hourani, Salam Fraihat

Department of Software Engineering, Faculty of Information Technology
Al-Ahliyya Amman University
Amman, Jordan

Abstract—**Recommender Systems are used to mitigate the information overload problem in different domains by providing personalized recommendations for particular users based on their implicit and explicit preferences. However, Item-based Collaborative Filtering (CF) techniques, as the most popular techniques of recommender systems, suffer from sparsity and new item limitations which result in producing inaccurate recommendations. The use of items' semantic information besides the inclusion of multi-criteria ratings can successfully alleviate such problems and generate more accurate recommendations. This paper proposes an Item-based Multi-Criteria Collaborative Filtering algorithm that integrates the items' semantic information and multi-criteria ratings of items to lessen known limitations of the item-based CF techniques. According to the experimental results, the proposed algorithm prove to be very effective in terms of dealing with both of the sparsity and new item problems and therefore produce more accurate recommendations when compared to standard item-based CF techniques.**

Keywords—*Collaborative Filtering; Recommender Systems; Multi-Criteria; Sparsity; New Item*

I. INTRODUCTION

The information overload problem occurs due to the increasing growth of web information, which makes it difficult for web users to locate relevant information, products or services according to their needs and preferences. Recommender systems have been broadly utilized to address the information overload problem by helping web users in finding the most related information, products or services in diverse application domains such as e-commerce, e-learning, e-government and e-tourism [1-6]. Recommender systems are personalized decision support tools employed to exploit the users' explicit and implicit preferences to recommend to them the most relevant information, products or services. Collaborative filtering (CF) is one of the most known techniques in recommender systems to generate personalized recommendations. The CF technique can be further classified into user-based and item-based CF techniques. In user-based CF, recommendations for users are generated based on items that are liked by other similar users. In item-based CF, recommendations for users are generated based on items that are similar to those they have liked in the past [1, 7].

However, the item-based CF technique is proved to be more successful in terms of the prediction accuracy than the user-based CF technique [8, 9]. Regardless of its efficiency, the item-based CF does not perform well and may produce inaccurate recommendations when there is a lack of users' ratings due to two key obstacles: the sparsity and the new item problems [8, 9]. To solve such problems, recent recommender systems have focused on the integration of additional information, thus, allowing recommender systems to exploit the added information as a supplementary to the insufficient users' ratings to generate more accurate recommendations. Examples of such additional information are: the semantic relationships that are exist among users or items [10-14]; and the multi-criteria ratings which can imply more complex users' preferences [7, 15-17].

Semantic information associated with users or items can be represented by taxonomies or ontologies and has an important task, by including concepts and their relationships, in accurately representing the item information and the user model [10-14]. In addition, recent studies acknowledge that multi-criteria ratings of users can be utilized to find the actual correlations between users as it based on more than one criterion [18-25]. To sum up, the additional information of users and items would assist in precisely model users' preferences and items' relations, and accordingly can result in more accurate recommendations.

This paper proposes an Item-Based Multi-Criteria CF (IMCCF) algorithm for personalized recommender systems. The proposed algorithm is a hybrid of MC item-based CF and an item-based semantic filtering techniques. The proposed algorithm exploits the additional information provided by both the semantic relationships among items and the multi-criteria ratings of users to address the sparsity and new item problems. The proposed algorithm proved to be more effective in dealing with the above limitations and therefore produce more accurate recommendations when compared to standard item-based CF techniques. The rest of this paper is ordered as follows. Section 2 describes the related work of the research. Section 3 demonstrates the proposed Item-Based Multi-Criteria CF algorithm. Section 4 shows the experimental setup and results. The conclusion and future work are revealed in Section 5.

II. RELATED WORK

In general, most of current recommender systems use single-criterion CF recommendation approaches, which have been deployed highly successfully for many years. Recently, a number of research studies [18-25] have employed multi-criteria ratings in their recommender systems taking into account that multi-criteria ratings would facilitate the accurate modeling of users' preferences, and thus provide more precise

recommendations. Examples of recent studies on Multi-Criteria Recommender Systems are:

Ebadi & Krzyzak [18] develop an intelligent hybrid multi-criteria hotel recommender system that suggests a number of hotels that are tailored for the preferences of a given user. To enhance the recommendation accuracy, the proposed system utilizes a multi-criteria rating technique to better capture and learn the preferences of users. TripAdvisor data is used to train the proposed system. Experimental results based on different settings and scenarios confirm the outstanding performance of recommendation accuracy of the system.

Jhalani et al. [19] propose the employment of multi linear regression approach for determining the weights for each criterion and calculating the overall ratings predictions of each item. Experimental results on Yahoo movie dataset show the effectiveness of the proposed method in generating quality recommendation compared with single criterion and multi-criteria CF benchmark algorithms.

Nilashi et al. [20] propose a novel recommendation algorithm using expectation maximization (EM) and classification and regression tree (CART) in order to improve the recommendation accuracy of multi-criteria recommender systems. The authors also employ the principal component analysis as a dimensionality reduction technique to alleviate the multi-collinearity limitation due to the interdependencies between different criteria in multi-criteria CF datasets. TripAdvisor and Yahoo! Movies datasets are used to validate the performance of the proposed algorithm. Experimental results show that the proposed algorithm extensively enhances the accuracy of recommendations in multi-criteria CF.

Farokhi et al. [21] propose a tourism recommender system that employs a recommendation method that integrates both multi criteria user-based and multi criteria item-based CF approaches. Fuzzy C-means algorithms beside k-means algorithms have been used to improve the recommendation accuracy of user-based and item-based CF approaches. The authors acknowledge that the use of multi-criteria rating in producing recommendations can improve the recommendation accuracy by providing more realistic recommendation that are very close to users' interests. Experimental results on the TripAdvisor dataset confirm the high performance in accuracy of the proposed method.

Nilashi et al.[22] incorporate the multi-criteria ratings in a new hybrid method for hotel recommendation using prediction and dimensionality reduction techniques to improve the predictive accuracy. The proposed method is a hybrid of the expectation maximization (EM) clustering, adaptive neuro-fuzzy inference system (ANFIS) and the principal component analysis (PCA) techniques. These techniques are combined to boost the predictive accuracy of the multi-criteria CF in tourism domain by exploiting the extra knowledge hidden in the multi-criteria ratings and reducing the dimensionality of a dataset to deal with the multi-collinearity problem presents in the multi-criteria ratings. Experimental results on TripAdvisor dataset show that the proposed method achieved high recommendation accuracy in the tourism sector.

Bokde et al.[23] propose a university recommendation system that provides students, of the Engineering College, with recommendations derived from their past preferences. The proposed system employ a hybrid method of multi-criteria item-based CF and dimensionality reduction approaches to produce high quality recommendations. The hybrid method decreases the computational cost and increases the prediction accuracy, thus overcoming the scalability and sparsity limitations.

Bilge & Kaleli [24] propose a multi-criteria item-based CF framework that extends the conventional item-based CF algorithm to make use of the benefits of multi-criteria rating systems. The authors determine the most suitable neighborhood selection approach and examine the performance of accuracy of statistical regression-based predictions. Experimental results on Yahoo Movies dataset affirm the assumption that multi-criteria item-based CF algorithms can accurately generate more reliable recommendations than single criterion rating item-based CF algorithms.

Shambour & Lu [25] propose a hybrid multi-criteria trust-enhanced CF (MC-TeCF) method that addresses the limitations of single criterion user-based CF techniques by integrating the MC user-based CF and the MC user-based Trust filtering techniques. Empirical results of the proposed MC-TeCF method prove its significance over single criterion user-based CF techniques, in improving the accuracy and coverage of recommendations, when faced with extreme sparse data sets or new users.

However, compared to the huge amount of research carried on in the last years on single-criteria recommender systems, the adoption and employment of multi-criteria ratings in recommender systems has received limited attention [15, 22]. Thus, the need of more research in the area of multi-criteria recommender systems has provoked our interest toward the development of an Item-Based Multi-Criteria CF algorithm in this study.

III. THE ITEM-BASED MULTI-CRITERIA CF (IMCCF) ALGORITHM

The proposed IMCCF algorithm takes a raw matrix of user-item MC ratings, as input, which consists of multi-criteria ratings of M users on N items, and a hierarchical tree structured item taxonomy. The item taxonomy, given by the domain experts, has a set of main items' categories where items should belong to as leaf nodes. It should be noted that each item can be a member of one or more items' categories. The process of recommendation of the proposed IMCCF algorithm is demonstrated by the subsequent three main tasks:

A. The Computation of MC Item-based CF Similarity

The MC item-based CF similarity between a given target item i and an item neighbor j is computed in this step through: 1) the calculation of the partial similarities between each of the rating criteria c, then 2) the use of an aggregation function to get the overall similarity value. According to [24], the use of Euclidean distance as similarity measure proved to be an excellent choice for item-item similarity computation in comparison with the traditional item-based CF similarity techniques. Thus, the *Euclidean Distance* similarity measure

[16, 24] is used here to calculate the MC item-based CF similarity values between the target item i and the item neighbor j based on each individual criterion as shown below:

$$Dis_{i,j}^{c} = \sqrt{\sum_{u=1}^{n} abs\,(r_{u,i}^{c} - r_{u,j}^{c})^2} \quad , \tag{1}$$

where $r_{u,i}^{c}$ and $r_{u,j}^{c}$ denote the user u ratings on items i and j with regard to criteria c correspondingly. n is the number of users who commonly rated items i and j. The smaller is the distance between two items are, the larger the similarity value between them is. Therefore, the following metric is needed to convert the resultant distance into the similarity value based on each individual criterion:

$$Sim_{i,j}^{c} = \frac{1}{1 + Dis_{i,j}^{c}}, \tag{2}$$

Then, we use the worst-case (i.e., smallest) similarity [16, 24] as an aggregation approach on the partial similarities to find out the overall similarity value between a given target item i and an item neighbor j as follows:

$$Sim_{i,j} = \min_{c=1,\dots,x} sim_{i,j}^{c} \tag{3}$$

where $Sim_{i,j}^{c}$ is the value of partial similarity based on criteria c, x is the number of individual criterion.

Nevertheless, the Euclidean Distance similarity measure that is used to calculate the similarity values between items based on each individual criterion considers only the absolute value of ratings between users who have commonly rated items i and j. This could produce unreasonable similarity values between items since two items can have a high similarity value even though they have obtained an extremely limited amount of ratings. This issue can be improved by taking into account the amount of the users who have rated both items while computing the similarity between them. To solve this issue, we employ the *Dice* coefficient [26], as shown in (4), as a weighting factor to consider the percentage of users who have commonly co-rated both items i and j to the total number of users who have rated items i and j separately. Thus, the final MC item-based CF similarity is given by (5).

$$Dice_{i,j} = \frac{2 \times |U_i \cap U_j|}{|U_i| + |U_j|}, \tag{4}$$

$$CFSim_{i,j} = Sim_{i,j} \times Dice_{i,j} \tag{5}$$

B. The Computation of Item-based Semantic Similarity

The item taxonomy is used to exploit the semantic relationships among items. To form such taxonomy in a particular domain: 1) The total number of main items' categories should be identified ; 2) The main items' categories should be created; 3) each item should be assigned to one or more appropriate main category. Formally, every item is modeled as a vector of binary values [0,1], as depict by (6).

$$\vec{V}_{(i)} = (y_{i,1}, y_{i,2}, \dots, y_{i,t}),\ where \tag{6}$$

$$y_{i,t} = \begin{cases} 1 & Indicates\ that\ Item\ i\ is\ attached\ to\ a\ Category\ t \\ 0 & Indicates\ that\ Item\ i\ doesn't\ attached\ to\ a\ Category\ t \end{cases}$$

Where $\vec{V}_{(i)}$ is the binary vector representation of item i, t is the overall number of the major items' categories. The value of item-based semantic similarity among two items i and j is computed using the standard vector-based cosine similarity [8], as shown in (7).

$$SemSim_{i,j} = \frac{\sum_{u=1}^{n} v_i \times v_j}{\sqrt{\sum_{u=1}^{n}(v_i)^2} \times \sqrt{\sum_{u=1}^{n}(v_j)^2}} \quad , \tag{7}$$

C. The Computation of Rating Predictions

The prediction process of unrated item x by an active user a consists of two major steps. First, we use the weighted sum of deviations from the mean approach [27] to compute the rating predictions for each unrated item twice: 1) using the MC item-based CF similarity as specified by (8); and 2) using the item-based semantic similarity as specified by (9).

$$P_{a,x}^{CFSim} = \overline{r_x} + \frac{\sum_{n=1}^{NN^{CFSim}} (CFSim_{x,n} \times (\overline{r_{a,n}} - \overline{r_n}))}{\sum_{n=1}^{NN^{CFSim}} CFSim_{x,n}}, \tag{8}$$

$$P_{a,x}^{SemSim} = \overline{r_x} + \frac{\sum_{n=1}^{NN^{SemSim}} (SemSim_{x,n} \times (\overline{r_{a,n}} - \overline{r_n}))}{\sum_{n=1}^{NN^{SemSim}} SemSim_{x,n}}, \tag{9}$$

where, $\overline{r_x}$ and $\overline{r_n}$ denote the mean values of ratings of items x and n respectively. $CFSim_{x,n}$ and $SemSim_{x,n}$ represent the CF-based and semantic similarities between the items x and n respectively. The most Nearest Neighbors of items to the target item x identified according to the MC item-based CF and item-based semantic similarity weights denoted by NN^{CFSim} and NN^{SemSim} respectively. $\overline{r_{a,n}}$ denotes the mean rating value based on all rating criteria of item n by the active user a.

Finally, the above rating predictions is merged using the weighted harmonic mean aggregation method as revealed by (10) to guarantee that a high rating value of the $P_{a,x}$ will be attained only if both $P_{a,x}^{CFSim}$ and $P_{a,x}^{SemSim}$ have high prediction values.

$$P_{a,x} = \frac{2 \times P_{a,x}^{CFSim} \times P_{a,x}^{SemSim}}{P_{a,x}^{CFSim} \times P_{a,x}^{SemSim}} \tag{10}$$

IV.   EXPERIMENTAL SETUP AND RESULTS

A. Dataset and Evaluation metrics

To validate the performance of the proposed IMCCF recommendation algorithm, we use the Yahoo! Movies MC dataset [28] which was collected from the Yahoo! movies website (http://movies.yahoo.com). Each record of the rating data includes ratings for four criteria: story, acting, direction and visuals, in addition to an overall rating, user ID, and movie ID. The Yahoo! Movies MC dataset consists of 34,800 ratings from 1,716 users on 965 movies. The ratings are on the scale from 1 to 5. We built a movie taxonomy hierarchical tree structure with two levels. The main categories of items, referred to as movie genres, in which every item should be attached to are included in the first level. Whereas, the second level includes the items, referred to as movies, as leaf nodes. The movie genres has 32 attributes such as Action, Drama, Fantasy, … etc.

To evaluate the quality of the proposed algorithm, the recommendations produced were evaluated using: 1) the Mean Absolute Error (MAE) metric to measure the prediction accuracy (Note that the lower MAE is, the higher is the prediction accuracy), and 2) the Coverage metric to evaluate the capability of a given recommendation algorithm to produce recommendations (refer to [29] for more details on the metrics).

B. Benchmark algorithms

For benchmark purposes, we compare the results of the proposed IMCCF algorithm with the results of two widely used item-based CF algorithms: 1) The item-based CF based on cosine similarity proposed by [8] (denoted as VC-ICF); and 2) The item-based CF based on adjusted cosine similarity (denoted as AVC-ICF) proposed by [30].

C. Experimental results

Two main experiments have been performed to prove the improvement of the proposed IMCCF recommendation algorithm with respect to the prediction accuracy and recommendation coverage when faced with the challenges of sparsity and new item.

1) Evaluating the Prediction Accuracy and Recommendation Coverage of the IMCCF on the Sparsity problem. On this experiment, we verify the efficiency of the proposed IMCCF algorithm compared with the benchmark algorithms in reducing the impact of the sparsity problem. As shown in Fig. 1 and Fig. 2, the proposed IMCCF algorithm has proven its superiority over other benchmark algorithms by obtaining the highest prediction accuracy (i.e., lowest MAE) and the maximum recommendation coverage at all sparsity levels.
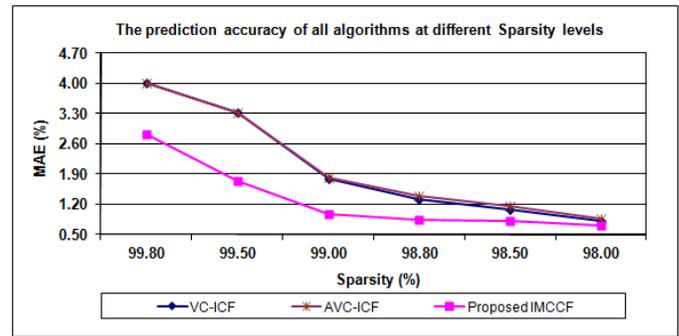

Fig. 1.   Comparing the predictuion accuracy of each algorithm on different levels of sparsity
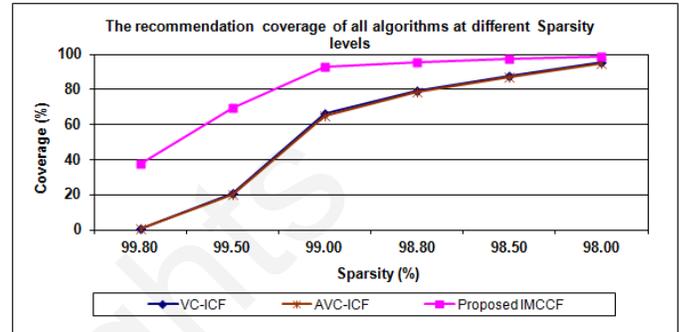

Fig. 2.   Comparing the coverage of recommendation of each algorithm on specific levels of sparsity

2) Evaluating the Prediction Accuracy and Recommendation Coverage of the IMCCF on the new item problem. The aim of this experiment is to verify the efficiency of the proposed IMCCF algorithm compared with the benchmark algorithms in reducing the impact of the new item problem. As shown in Fig. 3 and Fig. 4, the proposed IMCCF algorithm has proven its superiority over other benchmark algorithms by obtaining the highest prediction accuracy (i.e., lowest MAE) and the maximum recommendation coverage at every specified amount of ratings of new items.
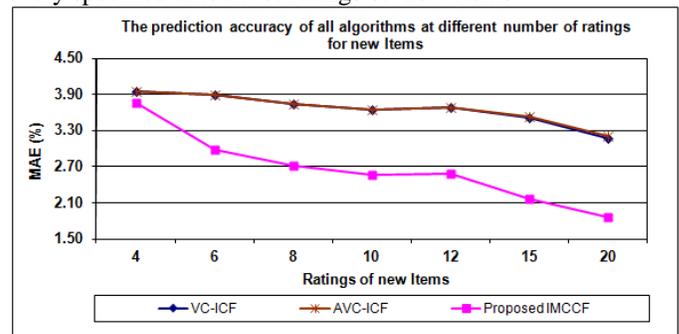

Fig. 3.   Comparing the predictuion accuracy of each algorithm on specific number of ratings of new items
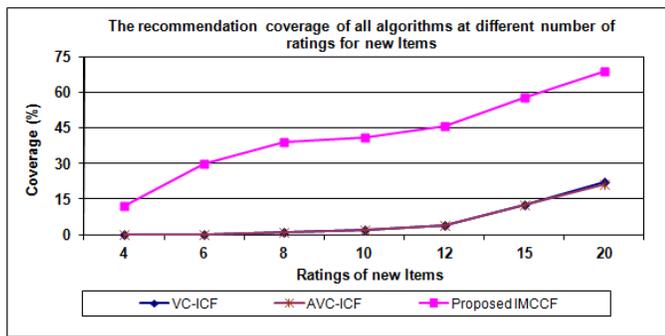
Fig. 4. Comparing the recommendation coverage of each algorithm on different number of ratings for new items

To conclude, it can be proven that the proposed IMCCF algorithm has a considerable improvement in lessen the effect of the sparsity and new item problems in comparison to the benchmark algorithms.

V. CONCLUSION AND FUTURE WORK

This paper proposes an Item-based Multi-Criteria Collaborative Filtering algorithm that integrates the items' semantic information and multi-criteria ratings of items to lessen known obstacles of the item-based CF techniques. The experimental results of the proposed algorithm, in comparison to the benchmark item-based CF algorithms, prove that the proposed IMCCF algorithm is very effective in dealing with both of the sparsity and new item problems with respect to the prediction accuracy and recommendation coverage. The proposed IMCCF algorithm enhances the quality of produced recommendations by exploiting the added information obtained from both the multi-criteria ratings of users and the semantic relationships among items to address the sparsity and new item limitations. In future, we will focus on further validating the performance of the proposed algorithm against more benchmark CF-based algorithms on larger data sets.

REFERENCES

[1] J. Lu, D. Wu, M. Mao et al., "Recommender system application developments: A survey," Decision Support Systems, vol. 74, pp. 12-32, 2015.

[2] Q. Shambour, and J. Lu, "A trust-semantic fusion-based recommendation approach for e-business applications," Decision Support Systems, vol. 54, no. 1, pp. 768-780, 2012.

[3] Q. Shambour, and J. Lu, "An effective recommender system by unifying user and item trust information for B2B applications," Journal of Computer and System Sciences, vol. 81, no. 7, pp. 1110-1126, 2015.

[4] S. Fraihat, and Q. Shambour, "A Framework of Semantic Recommender System for e-Learning," Journal of Software, vol. 10, no. 3, pp. 317-330, 2015.

[5] Q. Shambour, and J. Lu, "A framework of hybrid recommendation system for government-to-business personalized e-services," in Proceedings of the 7th International Conference on Information Technology: New Generations (ITNG 2010), Las Vegas, Nevada, USA, 2010, pp. 592-597.

[6] J. Lu, Q. Shambour, and G. Zhang, "Recommendation technique-based government-to-business personalized e-services," in Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS 2009), Cincinnati, OH, 2009, pp. 1-6.

[7] C. C. Aggarwal, "An Introduction to Recommender Systems," Recommender Systems: The Textbook, pp. 1-28, Cham: Springer International Publishing, 2016.

[8] B. Sarwar, G. Karypis, J. Konstan et al., "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th International Conference on World Wide Web, Hong Kong, China, 2001, pp. 285-295.

[9] C. C. Aggarwal, "Neighborhood-Based Collaborative Filtering," Recommender Systems: The Textbook, pp. 29-70, Cham: Springer International Publishing, 2016.

[10] E. Peis, J. M. del Castillo, and J. A. Delgado-López, "Semantic recommender systems. analysis of the state of the topic," Hipertext. net, vol. 6, no. 2008, pp. 1-5, 2008.

[11] M. Pozo, R. Chiky, and E. Métais, "Enhancing Collaborative Filtering Using Implicit Relations in Data," Transactions on Computational Collective Intelligence XXII, pp. 125-146: Springer, 2016.

[12] M. I. Martín-Vicente, A. Gil-Solla, M. Ramos-Cabrer et al., "A semantic approach to improve neighborhood formation in collaborative recommender systems," Expert Systems with Applications, vol. 41, no. 17, pp. 7776-7788, 2014.

[13] F. S. Gohari, and M. J. Tarokh, "New Recommender Framework: Combining Semantic Similarity Fusion and Bicluster Collaborative Filtering," Computational Intelligence, pp. 1-26, 2015.

[14] Q. Shambour, and J. Lu, "Government-to-business personalized e-services using semantic-enhanced recommender system," EGOVIS 2011, LNCS, K. Andersen, E. Francesconi, Å. Grönlund et al., eds., pp. 197–211, Heidelberg: Springer, 2011.

[15] G. Adomavicius, N. Manouselis, and Y. Kwon, "Multi-Criteria Recommender Systems," Recommender Systems Handbook, pp. 769-803, , US: Springer, 2010.

[16] G. Adomavicius, and Y. O. Kwon, "New recommendation techniques for multicriteria rating systems," IEEE Intelligent Systems, vol. 22, no. 3, pp. 48-55, 2007.

[17] Q. Shambour, and J. Lu, "A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations," in Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence (WI'11), Lyon, France, 2011, pp. 71-78.

[18] Ebadi, and A. Krzyzak, "A Hybrid Multi-Criteria Hotel Recommender System Using Explicit and Implicit Feedbacks," International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 10, no. 8, pp. 1237-1245, 2016.

[19] T. Jhalani, V. Kant, and P. Dwivedi, "A Linear Regression Approach to Multi-criteria Recommender System," Data Mining and Big Data: First International Conference, DMBD 2016, Bali, Indonesia, June 25-30, 2016. Proceedings, Y. Tan and Y. Shi, eds., pp. 235-243, Cham: Springer International Publishing, 2016.

[20] M. Nilashi, M. D. Esfahani, M. Z. Roudbaraki et al., "A Multi-Criteria Collaborative Filtering Recommender System Using Clustering and Regression Techniques," Journal of Soft Computing and Decision Support Systems, vol. 3, no. 5, pp. 24-30, 2016.

[21] N. Farokhi, M. Vahid, M. Nilashi et al., "A Multi-Criteria Recommender System for Tourism Using Fuzzy Approach," Journal of Soft Computing and Decision Support Systems, vol. 3, no. 4, pp. 19-29, 2016.

[22] M. Nilashi, O. bin Ibrahim, N. Ithnin et al., "A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS," Electronic Commerce Research and Applications, vol. 14, no. 6, pp. 542-562, 2015.

[23] D. K. Bokde, S. Girase, and D. Mukhopadhyay, "An Approach to a University Recommendation by Multi-criteria Collaborative Filtering and Dimensionality Reduction Techniques," in Proceedings of 2015 IEEE International Symposium on Nanoelectronic and Information Systems, 2015, pp. 231-236.

[24] Bilge, and C. Kaleli, "A multi-criteria item-based collaborative filtering framework," in Proceedings of the 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chon Buri, 2014, pp. 18-22.

[25] Q. Shambour, and J. Lu, "Integrating multi-criteria collaborative filtering and trust filtering for personalized recommender systems," in Proceedings of the 2011 IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making (MCDM 2011), Paris, France, 2011, pp. 44-51.

[26] W. B. Frakes, and R. Baeza-Yates, Information retrieval: data structures and algorithms: Prentice Hall, 1992.

[27] J. Herlocker, J. A. Konstan, and J. Riedl, "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms," Information Retrieval, vol. 5, no. 4, pp. 287-310, 2002.

[28] K. Alodhaibi, "Decision-guided recommenders with composite alternatives," Information Technology, George Mason University, Virginia, 2011.

[29] C. Aggarwal, "Evaluating Recommender Systems," Recommender Systems: The Textbook, pp. 225-254, Cham: Springer International Publishing, 2016.

[30] M. Deshpande, and G. Karypis, "Item-based top-N recommendation algorithms," ACM Transactions on Information Systems, vol. 22, no. 1, pp. 143-177, 2004.