Proceedings of the International Conference on Intelligent Computing and Sustainable System (ICICSS 2018)
DVD Part Number: CFP18O03-DVD; ISBN; 978-1-5386-4344-0

A Content Based Filtering and Negative Rating Recommender System for E-learning Management System*

Dr.Mahesh Kandakatla
Department of Computer Science & Engineering
Vaagdevi College of Engineering
Warangal, India
maheshkandakatla@gmail.com

Krishna Bandi
Department of Computer Science & Engineering
Vaagdevi College of Engineering
Warangal, India
bandikrishna007@gmail.com

***Abstract*—In online e-learning, there is a lot of shared information for learners, but discovering the learner's interest is more difficult because of information overload. Traditional approaches are used to generate recommendations for learners by using content features and ratings of Collaborative Filtering (CF) and content based approaches. The personalized and accurate learning resource recommendations are presented by incorporating the sequential access patterns and learner's context into the Recommender System (RS). In this research, semantic Content-Based filtering and learner's Negative Ratings (CBNR) is presented to recommend the interesting messages to the learners. The experiments are evaluated to validate the learners' performance and recommendation accuracy of the proposed system with existing e-learning recommendation system. Furthermore, the obtained outcome proves that the learning performance has been increased in terms of accuracy and quality of recommendations when compared with the similar recommendation techniques.**

Keywords—E-learning, Collaborative Filtering, Recommender Systems, Content-Based filtering, Negative Ratings, Accuracy.

I. INTRODUCTION

As there are very fast development and wide range of application of the internet, World Wide Web has become an interesting medium for pool, exchange, sharing of information and efficient channel for collaborative work [1]. RS has been used to serve more people by understanding their usage behaviours and recommends products or items to them based on their activity and also improves the computational process of a website [2]. RS aims at assisting users in their selection or purchase of items, by suggesting those items that fit their needs. In the last ten years, many online services consider RS as a classical tool in various domains such as such as e-commerce, tourism, e-learning, social networks, and internet advertising, etc. [3]. At the present time, RS has a massive amount of data on the social websites, and finding the appropriate resource based on the user's preferences is getting more attention [4]. The potential performance of customers is estimated by the RS that are adopted by big companies such as eBay, Amazon and Netflix, etc. used to recommend relevant products or items to the user. Performance of recommendation systems have a huge impact on the commercial success of these companies in terms of revenue and user satisfaction [5]. Some methods have been widely used for general recommendation, such as Content-Based (CB), CF and hybrid solution [6]. The quality of recommendations and usability of six online recommender systems are examined and the results show that RS provided better recommendations [7]. The one of the challenges of many recommendation tasks is the presence of implicit feedback where users' explicit preferences (e.g., ratings) on items are unavailable. In the real world, often only implicit feedbacks are available to learn a recommendation model [8]. The features and attributes are created by utilizing the contents of items to match the user profiles that are widely used in CB filtering. The previous items liked by users are compared with the items to match the best items are then recommended.

CF utilizes numerous data collected from websites about user behaviour in the past and predicts which items users will like. The CF does not need to analyse the content of the items [9]. In CF, two strategies are mainly applied, namely model-based and memory-based approach. The model is built from the database which is used for making recommendations in model-based approach. The second approach will work on the ratings from the whole dataset by the system. [10]. However, the method widely affected from cold start (CS) and sparsity and problems and still CS problem remain as a big challenge which hasn't been effectively solved by most of the existing methods. When a network newly imports some items which are associated with no user feedback, the CS problem arises from the data sparsity problem and noticed in RS. There are three types of CS problems such as the new system-when the system has just been used/created. The new user-when a user has just joined the system and the new item-when a new item has been recently introduced into the system [11]. The information which is related to the previous understanding of learner can be avoided by the idea of learning theory, moreover, the smooth and contextualized learning process will promote by keeping the recommendations progressively [12]. The main goal of this study is to use the negative ratings by implementing a semantic CBNR in e-learning discussion forum. The CBNR system is used to provide active learners for increasing their performance and saves their time. The recommendations didn't change within the current learner's context which is proven by the CBNR method.

The paper is organized as discuss below: Section II detailed about the research work of existing technologies in e-learning recommender system. The proposed method uses the negative rating to predict the recommendation are discussed in Section III. The section IV mainly described about the dataset used for experiments, the results and the parameters used for predicting the performance of the CBNR system. At last, the conclusion of the work and the future work of this research are made in Section V.

II. Literature Review

This section presents a brief review of some past researches conducted by researchers on RS in e-learning domain.

Z. Zhao, *et al.,* [13] proposed a framework Online Graph Regularized user Preference Learning (OGRPL) for social recommendation in online. The approach OGRPL integrated both the item features and collaborative user-item relationship for learning process into a unified preference. The problem of online optimization problem was solved by further developing the method into an iterative procedure effectively, OGRPL-Frank-Wolfe algorithm (OGRPL-FW). In user-item rating matrix, the prediction of online rating for missing values were further improved by the approach. The method was unable to explore the non-linear user PL function to solve the online recommendation problem.

R. S. Varma, *et al.,* [14] proposed a Collaborative Trust (CT) based content RS that generates the recommendations to the target user based on his trusted friends. The method developed the E-Learning solution which provided users with recommendations based on their preferences and content consumed by similar students. The method provided all the facilities like course sharing between two universities, online tests, analytics etc. in one software. CF found people with similar interests, analysed their behaviour derived from their ratings, and recommend target user the same items. This paper proposed a video RS for generating the outcomes from the association rule mining and CT based target user for capturing trends in the network.

W. Chen, *et al.,* [15] implemented a hybrid RS to recommend learning items in users' learning processes. The method contains two steps such as discovering content-related item sets using item-based CF. According to common learning sequences, the other step included applying the item sets to Sequential Pattern Mining (SPM) algorithm for filtering the items. The two approaches were combined to recommend potentially useful learning items to guide users in their current learning processes. The proposed method was evaluated by several experiments in peer-to-peer and centralized online system and the outcomes proved the good performance. If the values of TTL were large, the precision values were unstable and affected by the number of peers in the network, lead to poor performance.

M. Salehi, *et al.,* [16] a new material RS framework based on SPM and multidimensional attribute-based CF was proposed Learner Preference Tree (LPT) was implemented to consider the dynamic model, multi-preference of learners and the ratings of learners and attribute materials in multidimensional based CF framework. In terms of classification accuracy, the proposed method outperformed than the existing algorithms and the learner's real learning preference was satisfied accurately according to the real-time up dated contextual information. However, the framework provided higher precision and higher execution time.

J. K. Tarus, *et al.,* [17] implemented a hybrid RS combining Context Awareness (CA), SPM and CF algorithms to recommend learners about resources. The contextual information such as learning goals and knowledge level was incorporated about learners by CA. The sequential access patterns were discovered and the web logs were mined by using the SPM algorithm, whereas CF computed predictions and generated the outcomes for the target learner. The evaluation of the hybrid RS indicated that the framework outperformed other RS in terms of accuracy and quality of recommendations. The method mainly concentrated on the learning recommendation, the problem of the cold-start problem was not noticed by the method.

To overcome the problem of existing method, the paper proposes the CBNR system for predicting the better recommendations. The RS system uses the negative rating given by the learners in e-learning system.

III. Proposed Methodology

In this section, the introduction of our e-learning RS, consists of four phases, as represented by Figure 1. The first step of the process is retrieving the learners' ratings from the database. The pre-processing step contains two steps namely removing and stemming the negative rating-based profiles from the retrieved data. In the next phases, the processing time is reduced by these two processes for preparing the input data. The vector space dimensionality are reduced by the processed data in semantic indexing phase by performing main processes. In the next phase, learners profile was built in the system by modelling the negative ratings of learners. The integration of negative rating-based and semantic-based profiles built the learner's profile in the builder process. The prediction phase uses these profiles for computing the recommendation.

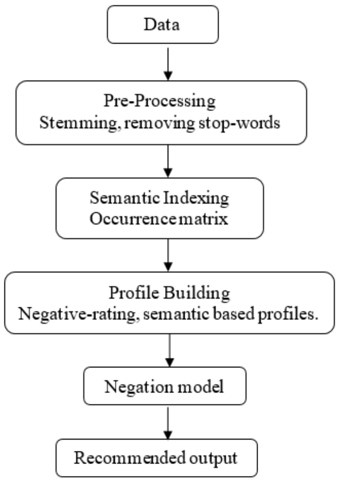

Fig. 1. Block Diagram of the proposed methodology

A. Pre-processing Phases

The contextual data and learners' rating are taken as an input which is retrieved from the dataset for preparing these input to the next phase. In this phase, parsing each messages, every special characters, numbers and stop-words are removed from each message, moreover, each word is converted into its root. The use of this phase is to prepare the input data to indexing phase.

B. Semantic Indexing Phase

The messages are represented by vector space model in this method, the m-dimensional vectors of each message are described in [19]. The message representations are more complicated and also lead to shortcomings in identification of similar messages. In this regard, the method explained that the latent semantic structures are captured into messages by using Latent Semantic Analysis (LSA).

In similar contexts, the terms having the similar messages are assumed by LSA. The creation of occurrence matrix is used for this phase, in which the frequency vector of a message is represented by each column. After the occurrence matrix, the weighting algorithm called Term Frequency-Inverse Document Frequency (TF-IDF) is used for normalizing its elements are represented in Eq. (1)

$$w_{i,j} = \frac{f_{i.j}}{\max_z f_{z,j}} * \log(\frac{D}{d_i}) \quad (1)$$

Where $w_{i,j}$ represents the weighted frequency value of i and j, $f_{i,j}$ describes the frequency value of i and j, $max_z f_{z,j}$ is the maximum frequency, D is the total number of messages and d_i is the message number.

The low-rank approximation is determined by Singular Value Decomposition (SVD) to the occurrence matrix A. Consider, a $m \times n$ rectangular matrix, the product of three other matrices are decomposed from a matrix A by using SVD are given in Eq. (2)

$$A = U \sum V^T \quad (2)$$

Where U is an $m \times n$ column-orthonormal matrix, Σ is an $n \times n$ diagonal matrix whose diagonal elements $\sigma_1, \sigma_2, ..., \sigma_n$ are non-negative singular values sorted in descending order, and V is an $n \times n$ orthonormal matrix.

The similarity structure are preserved among columns by truncated SVD, and the approximate matrix $\hat{A}$ is constructed by the reduction of rows. The quality of filtering is enhanced by mitigating the problems of polysemy and synonymy, and the semantic concepts are treated as one component by merging the similar contexts terms in a low-rank semantic vector space. In other hand, the process of factorization became economical and quicker in processing memory and time allocation. The truncated SVD is explained in Eq. (3),

$$\hat{A} = U_k \sum_k V_k^T \quad (3)$$

In the proposed system, the value of $qS_r = \sum_{i=1}^{[p^{'}]} P_i^{'}$ is set to 100 dimensions for dimensionality reduction. The messages of semantic vectors in 100-dimensional semantic spaces are used for learner profile and then evaluating the similarity values in the next phases.

C. Modelling phase

According to the negative ratings of learners, the negations are modelled in this phase and also the learner profiles are created. The profile contains negative rating-based profile (i.e. conjunction of negated messages) and the semantic-based profile (i.e. positively rated images).

Consequently, qS_r is the semantic rating profile and qN_r is the negative rating profiles are integrated with the help of projecting profile vector of semantic-based approach. The negative rating-based and semantic-based profile vectors are explained in the following Eq. (4) and (5) as below.

$$qS_r = \sum_{i=1}^{[p^{'}]} P_i^{'} \quad (4)$$

$$qN_r = \sum_{i=1}^{[p^{-}]} P_i^{-} \quad (5)$$

The negative rating profile features are irrelevant in semantic-based profile for creating the single profile which is explained in Eq. (6),

$$q_r = qS_r \neg qN_r \quad (6)$$

Where q_r is the profile of learner r that integrates the features of qS_r in which the features of qN_r are irrelevant. This profile is used for calculating the similarity values between the messages in the next phase.

D. Recommendation prediction phase

In this phase, the similarity values are calculated between the learner profile and messages and finally, top-N messages are recommended to the learners. In a vector space, the angle of cosine between the vectors of both message and learner profiles are calculated by cosine similarity. The values range from 0 to 1, where the value 1 represents the highest similarity between them and the 0 represents there is no similarity between the learner profile features and messages. Eq. (7) defines the cosine similarity measure.

$$S = \cos(q_r, P_i) = \frac{q_r . P_i}{\|q_r\| \|P_i\|} \quad (7)$$

Where q_r represents the weighted vector of r, P_i represents the weighted vector of i, $\|q_r\|$ and $\|P_i\|$ are the magnitudes vectors respectively. Eq. (5) is used to obtain the similarity values between profile and message, and these top-N similar messages are recommended to the learners.

IV. Experimental Results

The main objective is used to recommend the resources to learners in e-learners RS system. In subsection, the outcomes of the CBNR system is presented by validating the proposed RS with the help of various experiments and the results are tabulated. The performance of the system is tested by selected metrics are also explained.

A. Database Description

Learning Management System (LMS) provides the data for validating the CBNR framework. The LMS is used for encouraging the teaching and learning for students by e-learning. The time period for collection of data is nearly six months for predicting the CBNR performance. Total 1200 data were used for the experiments using LSM and allows to rate the learning resources on a scale of 1-5 values for learners (1–very irrelevant, 2–fairly irrelevant, 3–irrelevant, 4–relevant, 5–very relevant).

The preferences and contextual information are matched by the RS to suggest the learning resources to the learners. The knowledge of the learners can be changed according to beginner, intermediate or advanced. The learners' rating and their contextual information can be extracted from the RS database during collection periods. The dataset contains more than 1000 learners, 57153 ratings, and 756 of total LO. The training data can be split at 80% and the remaining data is used for purposes of experimental evaluation as testing data.

B. Evaluation of Performance

The learning resources are recommended for the learners by RS which is a task in e-learning. For measuring the outcome of the CBNR RS, the experiment use recall, F1 measure, accuracy and precision metrics. The method calculates and compare the outcome of the CBNR RS against other existing RS, namely GSP–CA–CF and CF–CA, in terms of the performance metrics.

1) Accuracy measure

The optimum size of neighbourhood are established by varying the sizes in experiments for predicting the better results. The accuracy are calculated for three recommendation algorithm by numerous test, in addition to Mean Absolute Error (MAE) are used for computing the accuracy which is described in Eq. (8). The high accuracy are predicted by lower the value of MAE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \quad (8)$$

Where n describes the test case number, p_i details the item predicted rating and r_i is the true ratings. The table 2 describes the prediction value of accuracy which is measured using MAE. The figure 2 represents the prediction accuracy and sensitivity to neighbourhood size.

TABLE I. THE PREDICTED ACCURACY VALUE USING MAE

Number of Neighbors	MAE values		
	CF-CA [17]	*GSP-CA-CF [17]*	*Proposed Method*
5	0.88	0.83	0.79
10	0.83	0.76	0.72
15	0.79	0.69	0.62
20	0.75	0.58	0.53
25	0.69	0.55	0.50
30	0.66	0.54	0.48
35	0.68	0.58	0.45
40	0.65	0.60	0.47
45	0.64	0.62	0.52
50	0.68	0.64	0.54
55	0.70	0.63	0.57
60	0.72	0.67	0.60

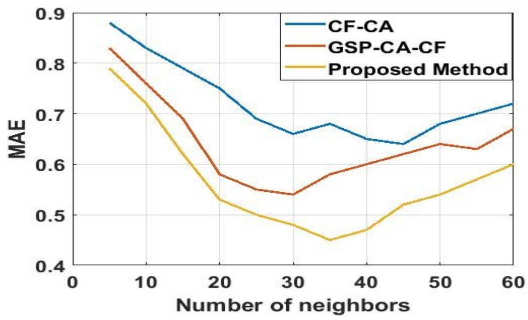

Fig. 2. Prediction of accuracy using MAE

The number of neighbours from 5 to 25 increases, the prediction accuracy of CBNR with existing RS increases steadily as shown from Fig. 2. The optimum solution is obtained, when the neighbour reaches 25. At small intervals, the curves start to rises slowly for three models. Beyond 25, the accuracy starts to decreases as clearly shown in figure. Hence, from the experiments, the optimal size of the neighbourhood is selected at 25. The results proved that the CBNR algorithm attains better accuracy, when compared with the other RS system for any number of nearest neighbours.

2) Precision and Recall

The confusion matrix is used for computing the recall and precision values. The confusion matrix is described as True Positive (TP) is used for recommendation and False Negative (FN) for non-recommendation in retrieved data. Moreover, False Positive (FP) and True Negative (TN) is used for both recommended and non-recommended data in non-retrieval data.

In [20], the rated scales of 1-5 are used for learning resources in recall and precision parameters. The "not relevant" is given to learning resources which are having the scales of 1-3, while the "related" are considered for having more than 4 scales of resources. The ratio of recommended resources of learning to the total number of resources are selected is called Precision, which is given in Eq. (9),

$$Precision = \frac{Recommended\ learning\ resources}{Total\ learning\ resources} = \frac{tp}{tp + fp} \quad (9)$$

The ratio of correctly recommended learning resources to the relevant learning resources is described as Recall, which is represents in Eq. (10)

$$Recall = \frac{Correctly\ recommended\ learning\ resources}{Relevant\ learning\ resources} = \frac{tp}{tp + fn} \quad (10)$$

3) F1 Measure

The combination of recall and precision into single values for getting the better view of performance, which is represented by F1 measure. These measure gave equal weight to the both metric described in Eq. (11),

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{11}$$

Table 2 shows the outcome of the proposed approach CBNR, when compared with another recommendation algorithm, namely CF-CA and GSP-CA-CF, in terms of recall and precision of different numbers of recommendations. The figure 3 and 4 represent the performance measures in terms of precision and recall for CBNR method with the GSP-CA-CF methods.

TABLE II. THE EVALUATION OF PRECISION AND RECALL FOR CBNR WITH EXISTING METHOD

No. of Recs	CF-CA [17]		GSP-CA-CF [17]		CBNR	
	Precision	*Recall*	*Precision*	*Recall*	*Precision*	*Recall*
4	0.449	0.231	0.481	0.244	0.448	0.248
8	0.443	0.242	0.476	0.256	0.473	0.259
12	0.438	0.248	0.462	0.272	0.458	0.278
16	0.419	0.267	0.445	0.290	0.440	0.303
20	0.388	0.301	0.420	0.336	0.416	0.341
24	0.359	0.321	0.396	0.373	0.393	0.376
28	0.339	0.352	0.367	0.405	0.362	0.413
32	0.314	0.379	0.348	0.451	0.343	0.458

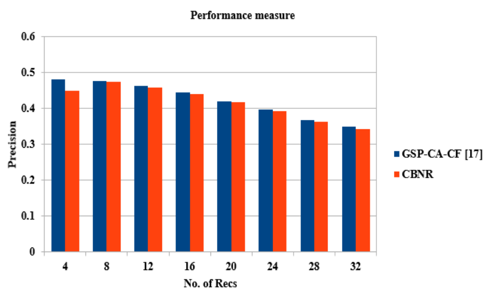

Fig. 3. Precision measure for proposed method with GSP-CA-CF

From figure 3, it is proved that the proposed recommendation algorithm CBNR achieved the better precision for any number of recommendation, while comparing with other RS approaches. The diagram shows that decreasing the precision values, increasing the recommendation results in CBNR approach.

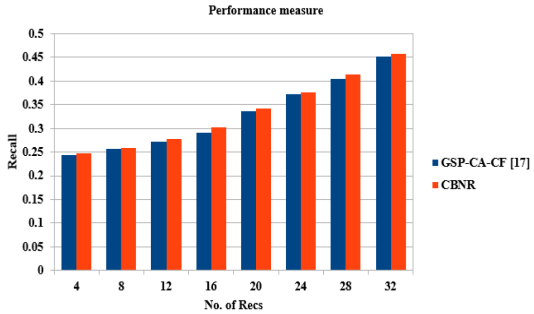

Fig. 4. Performance measure of CBNR method for Recall values with existing method

The figure 4 represents the prediction of recall values for the proposed CBNR method with the existing method GSP-CA-CF. In contrast, the values for recall increases, the recommendation also increases for numerous algorithm rather than CBNR approach. From the proposed approach, the evaluation of learner's satisfaction was obtained with the help of various experiments for recommendation. The "user satisfaction" is considered as a major evaluation measures for e-learning RS which are described in [18]. The performance of the CBNR method is evaluated using several experiments and the results stated that the method improved the e-learning process when compared with the existing methods in terms of precision, recall and accuracy.

V. CONCLUSION

In e-learning environments, this work achieved learning personalization by developing new RS based on CBNR approach. The crucial step is fulfilled by exploiting the proposed RS in e-learning system. The system accuracy is increased and the performance of learners are improved by ensuring the recommended items in current learning context. An experiments are conducted for evaluating the performance of the proposed CBNR method with other similar e-learning RS. Furthermore, the obtained results also showed that the learning performance has been increased in terms of accuracy and quality of recommendations compared to the existing techniques. The implement of learner's negative ratings into e-learning RS has a major impact for improving the performance of learners and also the accuracy of the RS in e-learning system. In future work, the results of recommendations are optimized and improved by using emerging tools that are mainly depends on the hybridization of RS with negative ratings.

REFERENCES

[1] S. S. Ahila, and K. L. Shunmuganathan, “Role of agent Technology in Web Usage Mining: homomorphic encryption based recommendation for E-commerce applications,” Wireless Personal Communications, vol. 87, no. 2, pp. 499-512, 2016.

[2] R. Mahajan, J. S. Sodhi, and V. Mahajan, “Usage patterns discovery from a web log in an Indian e-learning site: A case study,” Education and Information Technologies, vol. 21, no. 1, pp. 123-148, 2016.

[3] M. Aleksandrova, A. Brun, A. Boyer, and O. Chertov, “Identifying representative users in matrix factorization-based recommender systems: application to solving the content-less new item cold-start problem,” Journal of Intelligent Information Systems, vol. 48, no. 2, pp. 365-397, 2017.

[4] M. Rawashdeh, M. Shorfuzzaman, A. M. Artoli, M. S. Hossain, and A. Ghoneim, “Mining tag-clouds to improve social media recommendation,” Multimedia Tools and Applications, vol. 76, no. 20, pp. 21157-21170, 2017.

[5] J. Yan, K. Wang, Y. Liu, K. Xu, L. Kang, X. Chen, and H. Zhu, H. “Mining social lending motivations for loan project recommendations,” Expert Systems with Applications, 2017.

[6] G. Zhu, J. Cao, C. Li, and Z. Wu, “A recommendation engine for travel products based on topic sequential patterns,” Multimedia Tools and Applications, vol. 76, no. 16, pp. 17595-17612, 2017.

[7] X. Qian, H. Feng, G. Zhao, and T. Mei, “Personalized recommendation combining user interest and social circle,” IEEE transactions on knowledge and data engineering, vol. 26, no. 7, pp. 1763-1777, 2014.

[8] T. Ebesu, and Y. Fang, “Neural Semantic Personalized Ranking for item cold-start recommendation,” Information Retrieval Journal, vol. 20, no. 2, pp. 109-131, 2017.

[9] 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.

[10] U. Ocepek, J. Rugelj, and Z. Bosnić, “Improving matrix factorization recommendations for examples in cold start,” Expert Systems with Applications, vol. 42, no. 19, pp. 6784-6794, 2015.

[11] J. Zhu, J. Zhang, C. Zhang, Q. Wu, Y. Jia, B. Zhou, and S. Y. Philip, “CHRS: Cold Start Recommendation Across Multiple Heterogeneous Information Networks,” IEEE Access, vol. 5, pp. 15283-15299, 2017.

[12] F. Peng, X. Lu, C. Ma, Y. Qian, J. Lu, and J. Yang, “Multi-level preference regression for cold-start recommendations,” International Journal of Machine Learning and Cybernetics, pp. 1-14, 2017.

[13] Z. Zhao, H. Lu, D. Cai, X. He, and Y. Zhuang, “User preference learning for online social recommendation,” IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 9, pp. 2522-2534, 2016.

[14] R. S. Varma, B. M. Gosavi, A. M. Chawrai, and O. M. Deshpande, “Providing accurate course/video recommendations in E-Learning environment using association rule mining and collaborative filtering,” Journal of Data Mining and Management, vol. 3, no. 1, 2018.

[15] W. Chen, Z. Niu, X. Zhao, and Y. Li, “A hybrid recommendation algorithm adapted in e-learning environments,” World Wide Web, vol. 17, no. 2, pp. 271-284, 2014.

[16] M. Salehi, I. N. Kamalabadi, and Mohammad Bagher Ghaznavi Ghoushchi, “Personalized recommendation of learning material using sequential pattern mining and attribute based collaborative filtering,” Education and Information Technologies vol. 19, no. 4, pp. 713-735, 2014.

[17] J. K. Tarus, Zhendong Niu, and Dorothy Kalui, “A hybrid recommender system for e-learning based on context awareness and sequential pattern mining,” Soft Computing, vol. 22, no. 8, pp. 2449-2461, 2018.

[18] M. Erdt, Alejandro Fernandez, and Christoph Rensing, “Evaluating recommender systems for technology enhanced learning: a quantitative survey,” IEEE Transactions on Learning Technologies vol. 8, no. 4, pp. 326-344, 2015.

[19] G. Salton, Anita Wong, and Chung-Shu Yang, “A vector space model for automatic indexing,” Communications of the ACM vol. 18, no. 11, pp. 613-620, 1975.

[20] F. Ricci, Lior Rokach, and Bracha Shapira, “Introduction to recommender systems handbook,” Recommender systems handbook, springer US, pp. 1-35, 2011.