Journal of Engineering and Applied Sciences 12 (Special Issue 10): 8858-8863, 2017
ISSN: 1816-949X

Enhancing Content-Based Recommender System by Using Enriched User Profile

Einas Al-Turkey, Ghaidaa Al-Sultany and Huda Almamory
Department of Software, Faculty of Information Technology, University of Babylon, Hillah, Iraq

Abstract: Recommend personalized contents to users be strongly related to their profiles which reflected on the accuracy of recommendation system. In this study, the user profile has been built based on content and context of his preferences. The improved user profile is constructed by finding the correlation among the content and context of items. Given the user and item profiles, Jaccard's coefficient is applied to compute the similarity between pair of profiles, hence, k-nearest neighbors of items have been suggested to be top N in recommendation list. The system has been evaluated using LDOS-CoMoDa dataset. In conclusion, it can offer promising approach in recommendation system field if taking the correlations among content and context of items in consideration.

Key words: Item content, context, correlation, prediction model, recommender, evaluation

INTRODUCTION

Recommender Systems (RS) have been gainfully applied to different applications and domains such as e-Commerce (e.g., Amazon and e-Bay), social networks (e.g., Facebook and Twitter) and streaming media (e.g., Pandora and Netflix), etc. (Zheng, 2015). Recommender Systems (RS) are strong at mitigate information overload by submitting recommendations to user's personal preferences (Zheng *et al*., 2016).

Different methods (Raj and Suja, 2015) are used in content recommendations for item recommender. Kazienko and Kolodziejski (2006) recommendation systems typically output a roster of recommendations in one of two ways-through collaborative or content-based filtering, Content Based (CB) in this method recommendation is from a set of items with similar characteristics. Content based predictor method is based on description of item and a profile of user's preference (Ricci *et al*., 2011).

Affording users with an adapted information that satisfy their preferences and interests is the target of most the profitable businesses. The significant substance in this respect is how to discover the personalized information from various sources in particular from the users past interest's history and the product's content that the users like or dislike (Resnick and Varian, 1997). Content-based recommendation system predicts recommended items that were rated or admired by a given user positively in the past (Schafer *et al*., 2007; Rajaraman and Ullman, 2011). It analyzes a set of item's descriptions (that were previously purchased by the users) for enhancing either a prediction model to recommend items that harmonize the users tendencies (Mladenic, 1999). A user profile is a structured representation of user interests that has been adopted in many research to recommend the users with new interesting items. The recommendation process basically considers a matching up process among the attributes of the user profile against the attributes of a content object. The result is a relevance judgment that represents the user's level of interest in that object. Consequently, the more accurate a profile reflects user preferences, the better recommendation prediction process will be Thorat *et al*. (2015). In this research, enriching a user's profile with social information (context based) aligned with the features of the items that he/she likes or dislike (content based) has been adopted for enhancing the prediction of recommendation process.

Literature review: The most existing works use information as collaborative recommendation system based on prediction rating or behavior of other users. By (Uluyagmur *et al*., 2012) recommended movies list was predicted based on content as a features set, actor sets, director sets, genre sets and keyword sets and prediction rating for user and movie by given feature set and joined user specific weight for movie features. Systems depended on content analysis, must filtering for each information of items which is a record or collection of records representing important characteristics of that item. In simple cases, the profile consists of some characteristics of the item that are easily discovered, not only need to build vectors describing items, need to

Corresponding Author: Einas Al-Turkey, Department of Software, Faculty of Information Technology, University of Babylon, Hillah, Iraq

build vectors with the same components that describe the user's preferences (Rajaraman and Ullman, 2011).

A recommendation system proposed by Bae *et al*. (2012) analyzed the relationships between user ratings of movies and a user profile such as age, gender and genre of movies. While recommender system represented by Han *et al*. (2014) used retrieval information based on content and context for recommender tasks for user mobile and compared with a baseline algorithm for enhanced system. By Inzunza *et al*. (2016), the researchers used contextual information such as time, location and social information to generate an accurate recommender system that applied on music and movies dataset. Recommender system services provider is proposed by Ikawa used actor and keyword information of the users movies. The time of the day that the movies were watched by the users was taken into account also. The researchers used the ratio of the number of times in which a certain user watched a movie with a certain feature (such as actor, key word) to the number of times the feature is observed in all the movies.

MATERIALS AND METHODS

Proposed system: The proposed system includes four main phases as shown in Fig. 1.

Preprocessing: The preprocessing phase has been applied on the dataset through two stages: firstly, the missing values in the dataset are estimated by average the values. Secondly, extracting history for each user as a sequence of items.

Features selection: Basically, the purchased items are represented in terms of their content and context however, the items have many contents (features) in addition to the context. Accordingly, the selection procedure for the most effective and significant features is essentially required. One of the methods that are widely used for features selection is the information gain as illustrated in Eq. 1:

$$\text{InfoGain} = \text{H(class)-H(class/atrribute)} \quad (1)$$

Where:

H = Entropy
Class = User Id
Attribute = Item content and context

User profile formation: Commonly, a user profile is formed from a set of focuses such as the demographical information, the preferred items and occasionally the preferred item's features. However, to the best of our knowledge considering the preferred content based items and the contextual features accompanied with in addition

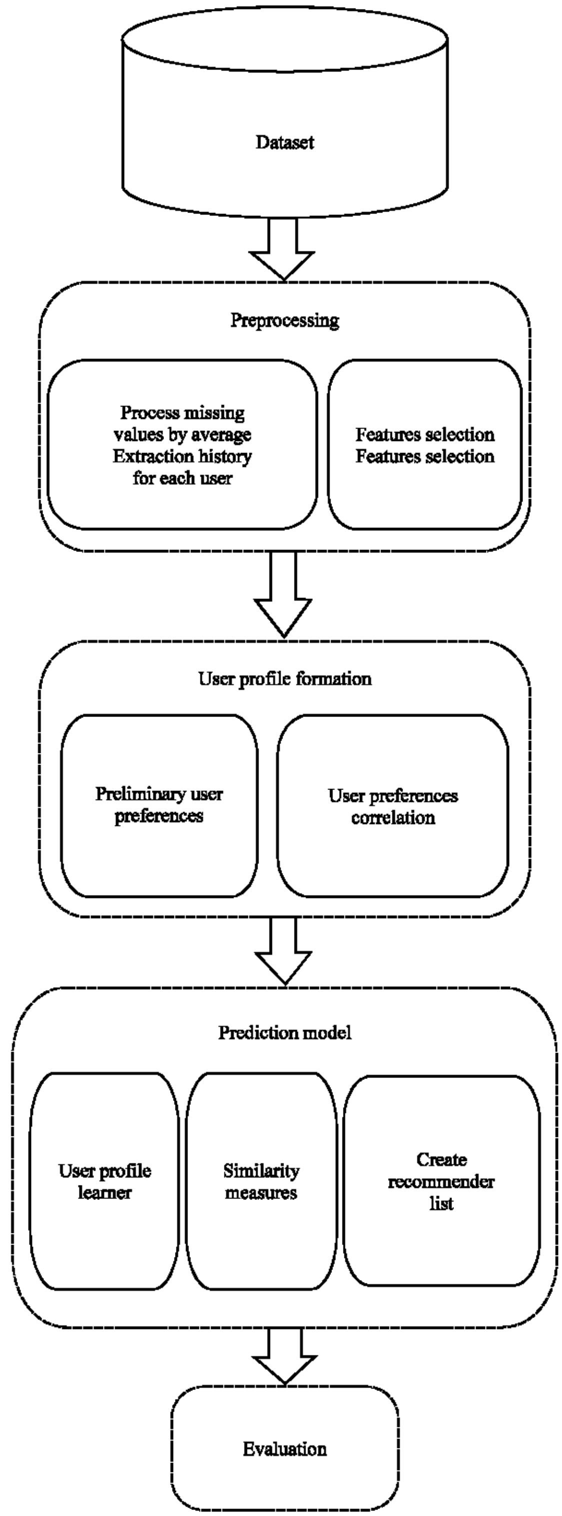

Fig. 1: Proposed system

to, finding the interrelation among those content and context have not taken into consideration in the recent literature.

Preliminary user's preferences: For each user u in the dataset, the frequency weights (count of weight feature) of the items features have been calculated. In precise, the number of occurrence of every feature in the all items that were preferred by the user himself has been counted as clarified in Eq. 2:

$$\forall\{f, u, ui: f \in IF, u \in U, ui \in I\} \text{Fwght} = \sum_{\substack{j=1 \\ if(f_j^{ut}=1)}}^{ui} fj^u \tag{2}$$

Where:

f = An item's feature
ui = The number of a user's items
$Fwght_{u,f}$ = The Frequency weight (count weight feature) of the item's feature to the current user u

On the same direction, the contextual features of each user that related with his past preferred item's list have been counted also as in Eq. 3:

$$\forall\{cx, u, ui: cx \in Cx, u \in U, ui \in I\} \text{Fwght} = \sum_{\substack{j=1 \\ if(f_j^{ut}=1)}}^{ui} fj^u \tag{3}$$

Where:

Cx = Auser context's feature
$Fwght_{u,Cx}$ = The frequency weight (count weight feature) of the context's feature
Cx = The current user u

Equation 2 and 3 have been applied on all the selected item's features and the contextual features, respectively for all the users U. The resulted weights of the features for each user have been sorted in descending order from the high weight value to the low one. The high weight values for the all features have been chosen for forming the preliminary user's preferences.

User's preferences interrelation: It has been noticed that the item features could be involving with each other to enrich the user profile with the most interrelated features. In this research a correlation process has been performed on the item features in terms of the feature with the high count weight feature as shown in Eq. 4:

$$CC_{ct,f} = \frac{\text{Frequency of item content weight features}}{\text{Total number of the correlation feature}} 100\% \tag{4}$$

On the same direction, correlation between user context that prefer item contents, calculated from count weight feature of the context features to total number of correlation feature as clarify (Eq. 5):

$$CC_{cx,f} = \frac{\text{Frequency of context weight features}}{\text{Total number of the correlation feature}} 100\% \tag{5}$$

Prediction model: To sum up the prediction model consist of three stages.

Step 1; user profile learner: User profile can be created in response to feedback which represents his preferred items. Generally, the user profile can be represented by the same features has been used in item profile. However, to improve the user profile, it has to find correlations among the sub features of his favorite items content.

Step 2; similarity measures: Given item and user profiles, Jaccard coefficient measure similarity can be used to compute the closeness between the profiles of a new item with that of a user as clarify in Eq. 6:

$$J(A, B) = \frac{|A \cap B|}{|A \cap B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{6}$$

Where:

A = An improved user profile
B = Items profile

Step 3; create recommender list: Subsequently, the similarity matrix that obtained from the above step, k-nearest neighbors of items can be elected for each user of which can create list of recommendations with different lengths.

RESULTS AND DISCUSSION

Implementation: Going in details with the implementation of the proposed system we applied it for performing a movies recommendation system.

Data description: In this research, the projected enhanced recommender system has been performed on the LDOS-CoMoDa dataset (Kosir *et al.*, 2011; Odic *et al.*, 2013) which it is produced for the purposes of prediction applications. It comprises of a rated 1232 Movies (M) by 121 Users (U). Table 1 illustrate the description of the LDOS-CoMoDa.

The following Table 2 show movie contents and context aspect or in other concept context condition that user follow him for movies watch.

Applying equation1 as explained on some from 11 movie features and 12 features for context aspects that

Table 1: The description LDOS-CoMoDa dataset

Variables	Values
Number of users-U	121
Number of movies (items)-M	1232
Number of actors (actor 1-3)	2542
Number of directors	815
Number of genres (genre 1-3)	24
Context aspects	45

Table 2: The description movie contents and context conditions

Movie contents	Context condition
Director	Time
Movie country	Day type
Movie language	Season
Movie year	Location
Gener 1	Weather
Gener 2	Social
Gener 3	EndEmo
Actor 1	DominanEmo
Actor 2	Mood
Actor 3	Physical
Movie budget	Decision
	Interaction

user preferred for movies watch, information gain as high ranking filtering for movie contents are actor 1 and 3, director, actor 2, respectively on same direction on context aspects that user preferred for movies watch such as day type, location and social, furthermore has been selected also support by survey.

System fulfillment: The proposed system implementation can be summarizes as follows: creating profile for each movie in terms of content and context according to the features have been elected previously:

$$\text{Item\_profile} = [G_1, G_2, G_3, D, A_1, A_2, A_3, C_1, C_2, C_3]$$

Creating profile for each user relate to content and context of his preferences inferred from past history. Similarly, the above features have been represented for each user as well. Worth to mention, there are sub features for genre (darma, action, etc.) and context (location, day type and social) indeed each genre consists of 24 sub-genres, hence, one sub-genre has been chosen from each genre. Among of 2542 actors, only three actors have been elected (Abdullah *et al*., 2016). However, that does not applicable to actor and director features. Accordingly, the correlations among fine-grained features have been found to obtain what user prefers perfectly. Given the genre which is considered important feature in profile of user it can find the correlation between it and the other features such as actor, director and context in feedback of user. Finally, it can obtain the improved profile of users.

Finding the similarity between user and item profiles pairs using Jaccard coefficient measure. Recommend list of items recommend list of items according to k-nearest neighbor technique, lists arrange of item with nearest neighbor for user profile with different length such as tow users select list with k = 5, 10, respectively:

$$\text{user } 55 = [0.7, 0.7, 0.6, 0.6, 0.6]$$

$$\text{user } 15 = [0.8, 0.7, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6]$$

Table 3: Represent precision measures

Variables	P@5	P@10	P@15	P@20	P@25	P@50
10% of orginal data						
Baseline 1	0.13	0.09	0.10	0.08	0.05	0.040
Baseline 2	0.14	0.11	0.20	0.17	0.14	0.130
Proposed system	0.21	0.26	0.25	0.25	0.22	0.170
20% of orginal data						
Baseline 1	0.14	0.09	0.08	0.06	0.03	0.020
Baseline 2	0.23	0.17	0.16	0.14	0.11	0.100
Proposed system	0.21	0.21	0.20	0.22	0.19	0.140
30% of orginal data						
Baseline 1	0.10	0.07	0.05	0.04	0.02	0.010
Baseline 2	0.17	0.15	0.12	0.11	0.09	0.070
Proposed system	0.20	0.18	0.15	0.16	0.14	0.110
100% of data						
Baseline 1	0.06	0.04	0.02	0.02	0.01	0.004
Baseline 2	0.08	0.07	0.05	0.04	0.03	0.030
Proposed system	0.08	0.07	0.05	0.06	0.05	0.040

Evaluation process: The proposed system has been evaluated using accuracy measures such as precision, recall and F-measure. In comparison with two baselines, the results of the proposed system have been evaluated. Firstly, the results are reported for the proposed system over the recommender based on only content in terms of building the user profile (baseline 1). Secondly, the same results are reported over the recommender based on content and context when building the user profile (baseline 2). It must be noted again that the proposed recommender is based on content, context and correlation among the sub features when building the profile user.

The chosen dataset have been divided into 70% for training set and 30% for testing set, Table 3 and 4, respectively show the precision and recall for the proposed system over baselines. However, the tables illustrate the results with different percentages of dataset where the 10% includes most wealthy users in terms of viewing whereas the 100% represents all users regardless if they have rich or poor history interms of viewing. In addition, the precision and recall have been calculated for lists of recommendations with different length for instance at 5, 10, 15, 20, 25 and 50.

Accuracy of system have been measured with F-measure which is the harmonic mean of precision and recall clarify (Eq. 7):

$$\text{F-measure} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

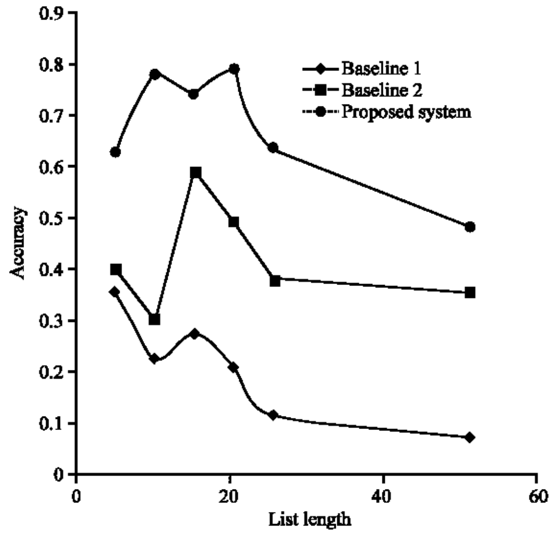

Fig. 2: Represent accuracy for 10% of data size

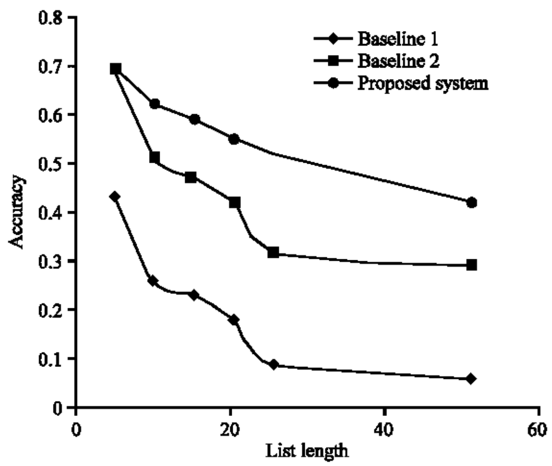

Fig. 3: Discern accuracy for 20% of data size

Table 4: Represent recall measures

Variables	P@5	P@10	P@15	P@20	P@25	P@50
10% of orginal data						
Baseline 1	0.01	0.02	0.16	0.16	0.04	0.06
Baseline 2	0.02	0.03	0.20	0.37	0.09	0.17
Proposed system	0.04	0.19	0.08	0.12	0.13	0.20
20% of orginal data						
Baseline 1	0.02	0.03	0.11	0.11	0.02	0.03
Baseline 2	0.04	0.06	0.14	0.28	0.11	0.18
Proposedsystem	0.04	0.13	0.08	0.13	0.15	0.23
30% of orginal data						
Baseline 1	0.02	0.07	0.07	0.11	0.02	0.02
Baseline 2	0.04	0.10	0.13	0.26	0.12	0.18
Proposedsystem	0.05	0.12	0.12	0.16	0.16	0.22
100% of data						
Baseline 1	0.04	0.05	0.04	0.06	0.02	0.02
Baseline 2	0.09	0.13	0.24	0.14	0.12	0.15
Proposedsystem	0.07	0.11	0.11	0.11	0.15	0.21

Figure 2-5 show F-measure for 10, 20, 30 and 100% of the data size. Given the profile user that has been built

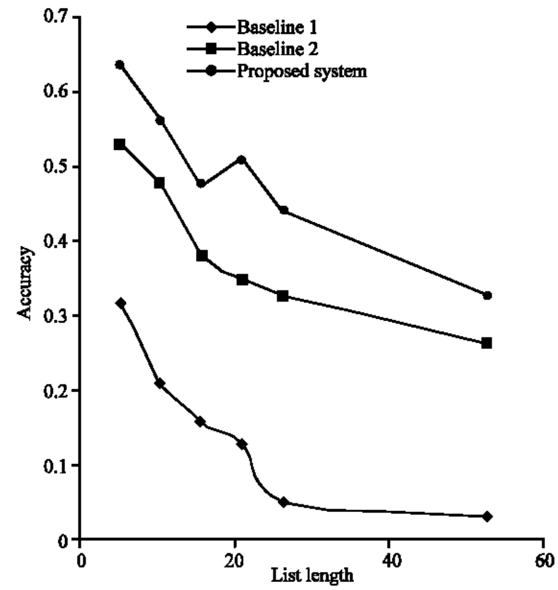

Fig. 4: Represent accuracy for 30% of data size

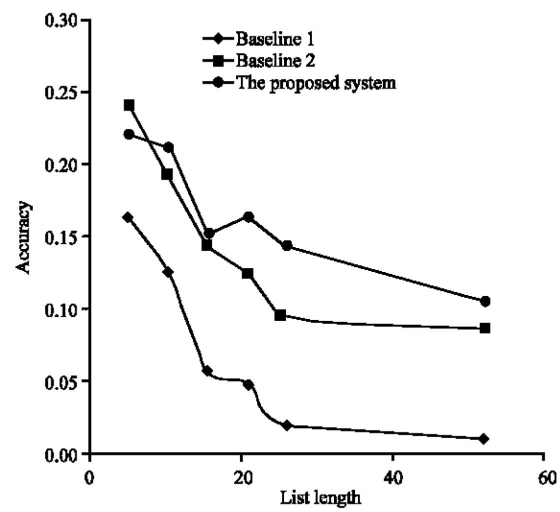

Fig. 5: Discern accuracy for 100% data size

based on content and context, the recommendation system (baseline 2) is superior over that (baseline 1) which the profile user is structured depend on only content. Iother side, the improved profile user in proposed system which is constructed based on content, context and correlations among sub features enhances the accuracy of recommender over baseline 2.

Figure 6 shows the comparison among the performance of the system at 10, 20, 30 and 100% of original data size. As it is illustrated, the performance of the system is best when 10% of the original dataset has been used then it gradually decreasing till 100% in which the system records less performance.

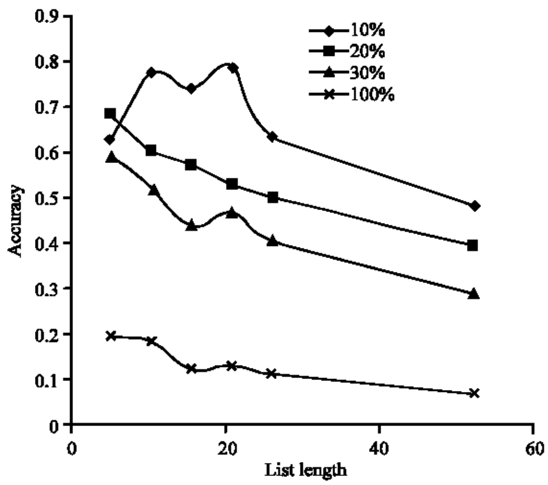

Fig. 6: Represent accuracy for proposed system

CONCLUSION

In conclude, by making the user profile more personality the accuracy of system is enhanced. Moreover, the improvement of system increases whenever the history of user is richer. In other words, the system can successfully infer the behavior of user when he has a wealthy history. Minutely, the users who have more than 10 items as feedback forms 10% of the dataset where they consider have rich history by which can detect a behavioral pattern for a user.

REFERENCES

Abdullah, E., H. Nawaf and G. Bilal, 2016. User profile enrichment with correlative item contents and user context. Intl. J. Soc. Sci., 11: 4320-4322.

Bae, S.M., K.H. Han and J.H. Park, 2012. Movie recommendation based on user profile and reviews. Proceedings of the WSEAS 1st International Conference on Latest Trends in Information Technology and Computer Engineering, November 10-12, 2012, WSEAS, Vienna, Austria, ISBN: 978-1-61804-134-0, pp: 168-191.

Han, J., H.R. Schmidtke, X. Xie and W. Woo, 2014. Adaptive content recommendation for mobile users: Ordering recommendations using a hierarchical context model with granularity. Pervasive Mobile Comput., 13: 85-98.

Inzunza, S., R. Juarez-Ramirez and A. Ramirez-Noriega, 2016. User and Context Information in Context-Aware Recommender Systems: A Systematic Literature Review. In: New Advances in Information Systems and Technologies, Rocha, A., M.C. Ana, A. Hojjat, P.R. Luis and M.T. Marcelo (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-31231-6, pp: 649-658.

Kazienko, P. and P. Kolodziejski, 2006. Personalized integration of recommendation methods for E-commerce. Intl. J. Comput. Sci. Appl. IJCSA., 3: 12-26.

Kosir, A., A. Odic, M. Kunaver, M. Tkalcic and J.F. Tasic, 2011. Database for contextual personalization. Elektrotehniski Vestnik, 78: 270-274.

Mladenic, D., 1999. Text-learning and related intelligent agents: A survey. IEEE. Intell. Syst. Appl., 14: 44-54.

Odic, A., M. Tkalcic, J.F. Tasic and A. Kosir, 2013. Predicting and detecting the relevant contextual information in a movie-recommender system. Interacting Comput., 25: 74-90.

Raj, N. and R.M.S. Suja, 2015. An overview of content recommendation methods. Intl. J. Innovative Res. Comput. Commun. Eng., 3: 334-339.

Rajaraman, A. and J.D. Ullman, 2011. Mining of Massive Datasets. Cambridge University Press, UK., ISBN-13: 978-1107015357, Pages: 326.

Resnick, P. and H.R. Varian, 1997. Recommender systems. Commun. ACM., 40: 56-58.

Ricci, F., L. Rokach and B. Shapira, 2011. Introduction to Recommender Systems Handbook. In: Recommender Systems Handbook, Ricci, F., L. Rokach, B. Shapira and P.B. Kantor (Eds.). Springer, New York, USA., ISBN:978-0-387-85819-7, pp: 1-35.

Schafer, J.B., D. Frankowski, J. Herlocker and S. Sen, 2007. Collaborative Filtering Recommender Systems. In: The Adaptive Web, Brusilovsky, P., K. Alfred and N. Wolfgang (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-72079-9, pp: 291-324.

Thorat, P.B., R.M. Goudar and S. Barve, 2015. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. Intl. J. Comput. Appl., 110: 31-36.

Uluyagmur, M., Z. Cataltepe and E. Tayfur, 2012. Content-based movie recommendation using different feature sets. Proceedings of the World Congress on Engineering and Computer Science Vol. 1, October 24-26, 2012, WCECS, San Francisco, USA., ISBN:978-988-19251-6-9, pp: 17-24.

Zheng, Y., 2015. Context suggestion: Solutions and challenges. Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), November 14-17, 2015, IEEE, Atlantic City, New Jersey, USA., ISBN:978-1-4673-8493-3, pp: 1602-1603.

Zheng, Y., B. Mobasher and R. Burke, 2016. User-oriented context suggestion. Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, July 13-17, 2016, ACM, Halifax, Nova Scotia, Canada, ISBN:978-1-4503-4368-8, pp: 249-258.