Online Product Recommendation System Using Multi Scenario Demographic Hybrid (MDH) Approach

R. V. Karthik, Sannasi Ganapathy

Online Product Recommendation System Using Multi scenario Demographic Hybrid (MDH) Approach

Karthik.R.V[0000-0003-1579-5750] and Sannasi Ganapathy[0000-0001-9177-5378]

School of Computer Science and Engineering, VIT University-Chennai Campus, Chennai-127, India

karthikr.v2016@vitstudent.ac.in, sganapathy@vit.ac.in

Abstract . The recommendation system plays very important role in the ecommerce domain to recommend the relevant products / services based on end user preference or interest. This helps end users to easily make the buying decision from a vast range of product and brands are available in the market. A lot of research is done in recommendation system, aim to provide the relevant product to the end user by referring end user past purchase history, transaction details etc. In our Multi scenario demographic hybrid (MDH) approach, important demographic influence factors like the user age group and located area are considered. The products are also ranked with associated age group category. The experimental results of the proposed recommendation system have proven that it is better than the existing systems in terms of prediction accuracy of relevant products.

Keywords: ecommerce; user reviews; recommendation;

1 INTRODUCTION

Recommendation systems are applied in almost all the applications like Movie recommendation, book recommendation, product recommendation in online shopping sites, travel package recommendation etc. The recommendation system helps to propose a subset of relevant item/ service that user are interested from the large amount of datasets. In online shopping sites, Recommendation system understands the user behavior and purchase interest and gives the customized recommendation based on each user.

A recommendation system is commonly classified in to a content based recommendation system and collaborative filtering approach. Content based recommendation system considers user past purchased data to understand the user interest and propose new product list, which are relevant or similar to the user past purchased product. Another recommendation system collaborative Filtering method focus on the similarities of other users. Based on similarity between the users, product purchased by one user is recommended to another user if both users are having same interest. Both these approaches has advantages and drawbacks. Content based recommenda-

tion system requires past history of the user. On other hand, collaborative filtering face challenge to recommend the product for some new customers.

To overcome this problem, the hybrid recommendation system is developed which combines both content based and collaborative filtering method. The results of the hybrid recommendation system also give better result when compared to individual recommendation systems. Now a day deep learning techniques are applied for recommendation system to improve the prediction accuracy and also the performance of the recommendation system.

In this paper, a new recommendation system is proposed to recommend suitable products by considering demographic information. The major contribution of this paper are 1) identify the end user age group category from customer review 2) computes the rating score for each user category 3) Apply fuzzy logic to get the relevant product. The remainder of this research article is formulated as below: The related worksin this area are elaborated in section 2. The proposed recommendation working model is described in section 3. Section 4explains the performance metrics and experimental results of the new MDH approach. Conclusions on the proposed work with future focus are described in Section 5.

2 LITERATURE SURVEY

There are various recommendation systems are developed to predict the user interested item/ service from a large amount of items. Soe-Tsyr Yuan et al [1] proposed customized product recommendation by considering clustering algorithm. Jae Kyeong Kim et al [2] used associate rule mining and decision tree to improve the recommendation system by selecting the relevant taxonomy and the preference of the user.

Duen-Ren Liu et al [3] proposed Hybrid based approach by combining preference based collaborative filtering and Weighted RFM methods. Clustering is done based on the similar user preference and interest. Finally, recommendation is performed by grouping users with similar weighted RFM value.

Sang Hyun Choi et al [4] proposed utility based recommendation system. In this approach multi attribute decision making procedure is used to predict the best product using utility value among other similar products. Utility value is computed by considering all the multi influence attributes. This approach shows better results and predicting accuracy when compared to distance based similarity product identification.

Kun Chang Lee et al [5] proposed new casual mapping approach which consider qualitative characteristics that influence the quantitative factors. By considering this relationship helps to predict the interest level of product and which in turn consider for later product recommendation. Fuzzy based product recommendation is proposed by Yukun Cao et al [6] where the user needed features are noted. Based on the importance, raking is done. Fuzzy rules are framed by considering both product feature and customer expectation. Final recommendation is done based on the optimal combination.

Yuanchun Jiang et al [7] provides the importance of capturing the end user feedback after purchase of the product for recommendation system. Simply buying the

product would not mean that it can be recommended for other end users. Instead recommendation system supposes the capture the real feedback or review comments after they used the product. This approach improves the customer satisfaction level by considering the user profile and review comments. Sung-Shun Weng et al [8] propose new multidimensional recommendation model which improves the accuracy of recommendation system. The approach is applied to the movie recommendation system, it identifies the users key influence preference factors and compute rating by considering all this multidimensional input factor. User profile and product rating are considered in this approach.

Ruben Gonzelez Crespo et al [9] done detail analysis on challenge in the education sector for both students and teachers in getting the needed information and shown the importance of recommendation system to extract the relevant information and topics to the knowledge seekers. Using recommendation system also improves the satisfaction of the student's expectations. Another approach rule based recommendation system is proposed by Vicente Arturo Romero Zaldivar et al [10] for education applications. In this approach Meta based rule recommendation is used to understand the profile, the interest of the user and helps to recommend the details for similar kind of users.

Mohanraj et al [11] proposed Ontology driven bees foraging technique to understand the user navigation pattern and predict the user interest. The score is calculated based on user navigation or selection in web pages. By learning this user profile and interest, proposed approach capture the navigation pattern and helps to recommend the relevant information in the webpages.Sunita et al [12] proposed a course recommendation system to recommend some interested course to a new student. Clustering technique is used to group the related courses and then apriori algorithm is then applied with associate rule to find the best combination and recommend courses that is interested by the new student.

Jian-Wu Bi et al [13] used type 2 Fuzzy numbers for sentimental analysis, which helps to improve the accuracy rate and also predict the relevant sub set of products from the large value of the product set for the recommendation.Chao Yang et al [14] proposed neural network based recommendation method which considers both user item interaction and also user list interaction. This improves the accuracy of the recommendation list and also found more relevant products to recommend.

Aminu Dau et al [15] used deep learning method for product recommendation. In this method sentiment score is calculated based on each aspect and then finally overall product score is computed and used for product recommendation.Xiaofeng Zhang et al [16] proposed deep recommendation system by considering user rating and also user textual information or feedback.

Karthik et al [17] proposed feature and product ranking approach to predict most suitable product for recommendation based on user preference and interested feature set. The proposed approach is an extension of this feature and product ranking approach. By considering the demographic information like user age group and delivery location, the recommendation can be further improved. For that purpose, this paper introduces a new recommendation system for recommending the right products to the customers in e-commerce.

3 PROPOSED WORK

Proposed multi demographic approach contains following main modules. 1) Preprocessing the customer review 2) Identify the End user age group 3) Compute product rating 4) Apply Fuzzy logic and computes the recommended level. All the modules are explained in the following chapters. Section 3.1 describes the preprocessing technique applied for the customer review. User reviews play important role in identifying the real end user of the product. By processing each customer review, rating of the product can be computed which gives better value when compared to the overall rating.

Section 3.2 explains the method to identify the end user age group. In order to recommend the relevant product to the right customer. It is important to focus the end user to whom the product is likely to purchase. Section 3.3 explains the technique to identify the product rating. In order to provide better recommendation, it is important to rank the product based on the user reviews and comments. Section 3.4 provides the fuzzy logic details to compute the recommended level. Fuzzy rule is framed by considering the product overall rating, end user age group and customer location or delivery location.

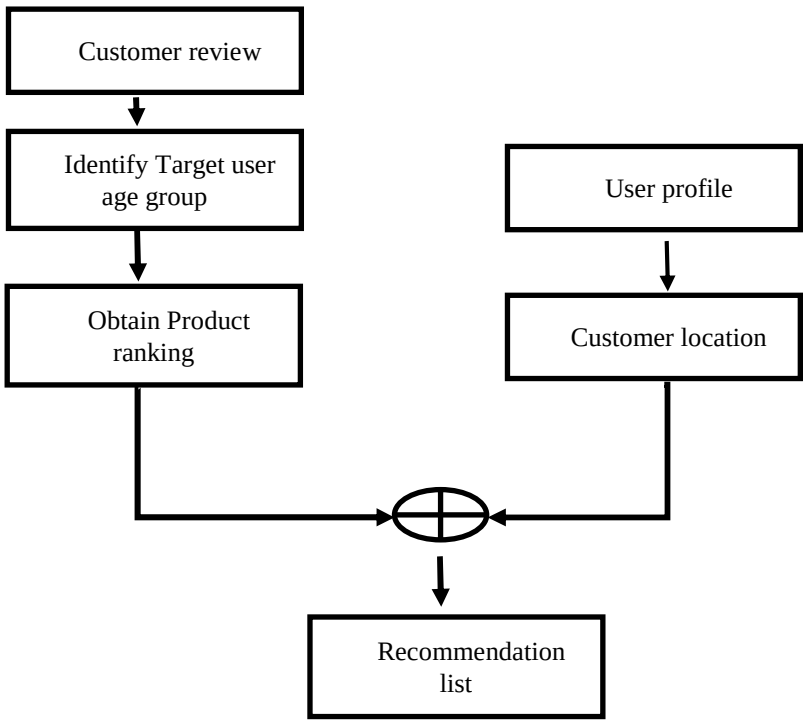

Fig. 1. Proposed model

Figure 1 explains the overall proposed model. It takes customer reviews as input and provides the recommendation list based on customer present interest.

3.1 Pre processing

In most of the ecommerce sites, customer review section is a free text area. Customers text their user experience and feedbacks. These text reviews are required to preprocess it, remove all unnecessary information's, stop words etc.

In preprocessing phase, stop words are removed from the review text and POS tagging is applied. Applying the preprocessing reduce the computation time and also improve the accuracy of the recommendation list.

3.2 Identify end user age group

In most of the customer reviews, real end users are mentioned by the purchased customers. By capturing this user age group helps recommendation system to propose the liked product to the similar kind of age group customers.

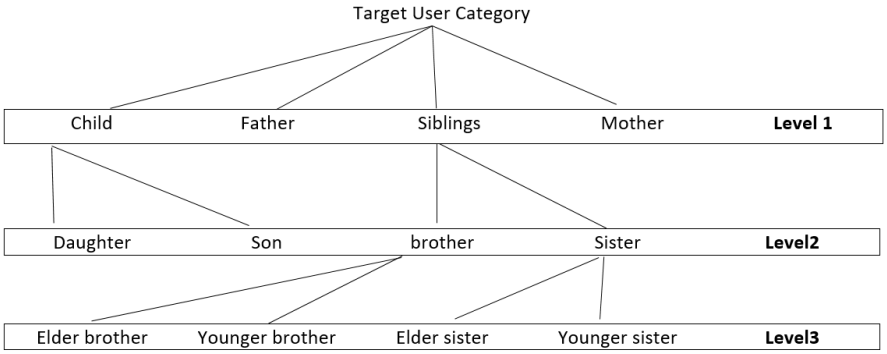

Fig. 2. User category

Figure 2 demonstrates the levels of age group captured in processing the review comments. Sibling category is divided into brother and sister. Further it is sub divided into elder brother and younger brother. If the end user information is not present in the review, then the MDH approach considers the shopping customer as an end user. This assumption avoids the reviews that don't contain any end user information.

3.3 Calculate product ranking

Overall ranking provided by the user is noted with associated end user age group.

Table 1 represents the user age group and overall rating given by user on each reviews.

Table 1. Product ranking

Item Id	User age group	Review Overall rating (Out of 5)
5834CS01	son	4
5834CS01	daughter	2
5834CS01	son	4
5834CS01	son	4

The average ranking value for each age group is computed and noted as product ranking for respective age group.

$$ranking_{ageGroup} = \frac{\sum_{k=1}^{n} review_{ranking}}{Totalnumberofreviewsofsameagegorup}$$

Overall product ranking for each group helps to recommend the product based on the current customer interest.

3.4 Fuzzy logic

Fuzzy logic helps to obtain the recommendation level by taking computed product ranking and customer delivery location. The MDH recommendation system gets the age group to whom like to purchase as input from the shopping user. Fuzzy rule is written to decide the recommendation level based on the rating of the product and the location of the customer.

Table 2. Fuzzy rule

Rule	Fuzzy rule description	Outcome
Rule 1	IF Age group Exactly matches AND Rating is High AND Location Exactly matches	Highly recommended
Rule 2	IF Age group Exactly matches AND Rating is High AND Location Matches	Highly recommended
Rule 3	IF Age group Exactly matches AND Rating is Medium AND Location Matches	Highly recommended
Rule 4	IF Age group Matches AND Rating is Medium AND Location Exactly Matches	Recommended

Rule 5	IF Age group Matches AND Rating is Medium AND Location Somewhat Matches	Recommended
Rule 6	IF Age group Not matches AND Rating is Medium AND Location Not matches	Not Recommended
Rule 7	IF Age group Not matches AND Rating is Low AND Location Not matches	Not Recommended

Algorithm (MDH approach)

Input: Online review comments of each product and age group interested to purchase

Output: Recommended product for the requested target user

Step1: Process all the reviews for each product

Step2: Apply preprocess technique and remove unwanted data from the reviews

Step 3: Identify the target end user category for each customer review

Step 4: Compute overall product rating score for each associated target user category

Step 5: Apply Fuzzy rule and determine the recommend level

Step 6: Share the highly recommended product based on target user

Overall product ranking computed for each age group looks like in Table 3.

Table 3. Product Recommendation

Item	Overall rating	End user age group recommended	Overall rating (Age group 1)	Overall rating (Age group 2)
5834CS01	3	brother	3.2	4.2
5834CS01	4	sister	3.5	3.1

4 RESULTS AND DISCUSSION

Dataset: Amazon review dataset is considered for this work. The data set contains more than 10,000 of products and 25,000 customer reviews. Each entry in the dataset contains the details present in the Table 4. Customer id is part of the user profile. Review comments and user rating are available for each customer reviews.

Table 4. Dataset details

S.No	Dataset items	Example
1	Customer id	239845
2	Overall rating	4
3	Product id	5834CS01

4	Review comments	"Awesome product, worth for money"
5	User rating	3.5
6	Purchased date	April/2016

4.1 Performance Metrics

In the evaluation, most widely used content based and collaborative filtering recommendation system approach is considered. In addition to these conventional recommendation systems, Hybrid recommendation system also taken in to account for comparison. Hybrid recommendation system is combination of both content and collaborative filtering technique. Hybrid approach outstands when compare to the individual recommendation system. Precision and Recall metrics are evaluated with all these recommendation techniques. Recommendation system specific metrics Item, User and Catalog coverage are also considered for evaluation.

$$Precision = \frac{Numberofrelvantrecommendationproduct}{Totalnumberofrecommendedproducts}$$

$$Recall = \frac{Number\ of\ relavant\ recommended\ prodcut}{Total\ number\ of\ product\ that\ to\ be\ recall}$$

Item coverage: This metric helps to know the percentage of products available in recommendation list over the number of potential available products.

$$Coverage_{item} = \frac{n_r}{N} * 100$$

Where, n_r represents the number of recommended products and N represents the number of potential available products

User coverage: This metric helps to obtain the percentage of users or customer that the system able to provide the recommendation list over the number of potential users.

$$Coverage_{user} = \frac{n_u}{N} * 100$$

Where, n_u represents the number of users that system can capable to generate recommended products and N represents the number of potential users

Catalog coverage: This metric helps to obtain the percentage of user – product pair over the total number of possible user product pair.

$$Coverage_{catalog} = \frac{n_{up}}{N * U} * 100$$

Where, n_{up} represents number of user pair product considered in product recommended products and N represents the number of possible user product pair.

4.2 Experimental Results

The dataset contains more than 80,000 reviews and 1000 products. Customers purchased more than 4 times are grouped separately and used for predicting the accuracy of the productrecommendation. For example, if the customer done shopping ten times in the history, then First 4 transactions are considered as test data to understand the behavior of the user. Based on these initial transactions, recommendation list is computed by the MDH approach and checked how it's relevant to the reaming transaction that were really purchased by the customer X. This method helps to identify the accuracy of the proposed MDH recommendation system.

Totally 5 experiment results are recorded by considering a different number of products and user list. Table 5 represents the number of product and users included in each experiment.

Table 5. Experiment sets

Experiment	Number of products	Number of users
Exp-1	2000	1200
Exp-2	4500	2000
Exp-3	6000	2800
Exp-4	8000	3500
Exp-5	10000	5000

Inference: By referring all the experiment outcomes, Hybrid approach result are better when compare to the individual content based or collaborative filtering techniques. Whereas Feature and product based ranking performs betters on comparing the Hybrid approach. Enhancement done on top of Feature and product ranking approach – MDH approach is further shown better result when compared to feature and product ranking feature. MDH approach improves 8% of accuracy of predicting relevant product for recommendation when compare to the existing recommendation system.MDH approach also shows better results when considering the coverage evaluation outcome. All three User, item and catalog coverage shows better results when compare to the other existing recommendation system.

Figure 3 and Fig 4gives the comparative results of precision and recall evaluation. Both precision and recall evaluation outcome shows MDH approach shows better results when compare to other existing recommendation approach.

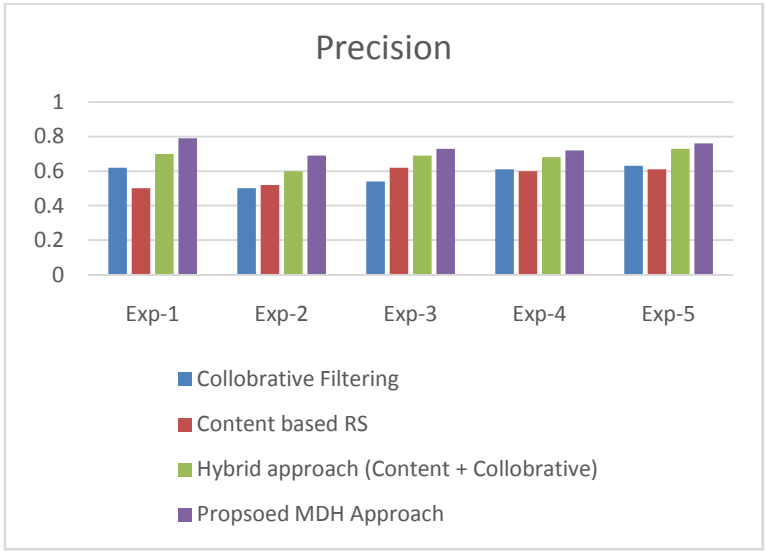

Fig. 3. Precision value comparative analysis

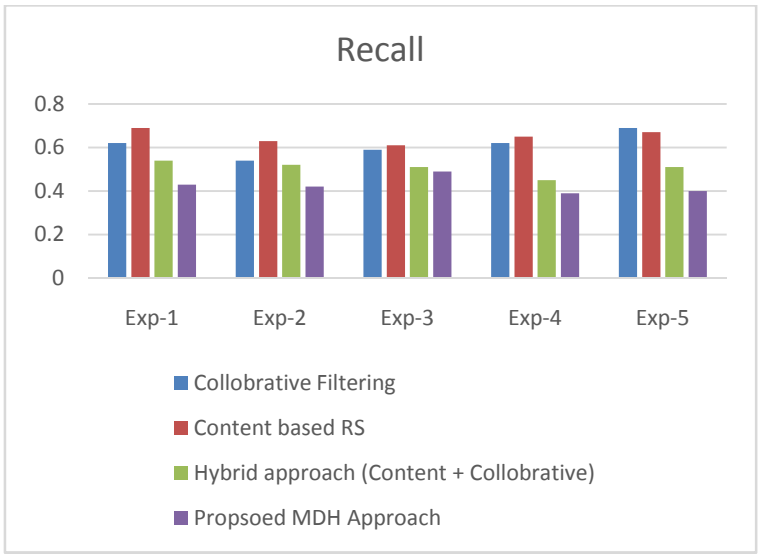

Fig. 4. Recall value comparative analysis

Figure 5 and Figure 6 gives the comparative results of user and item coverage evaluation. Coverage is recommendation system specific metrics. Both user and item coverage results imply MDH approach shows better results when compare to other existing recommendation approach.

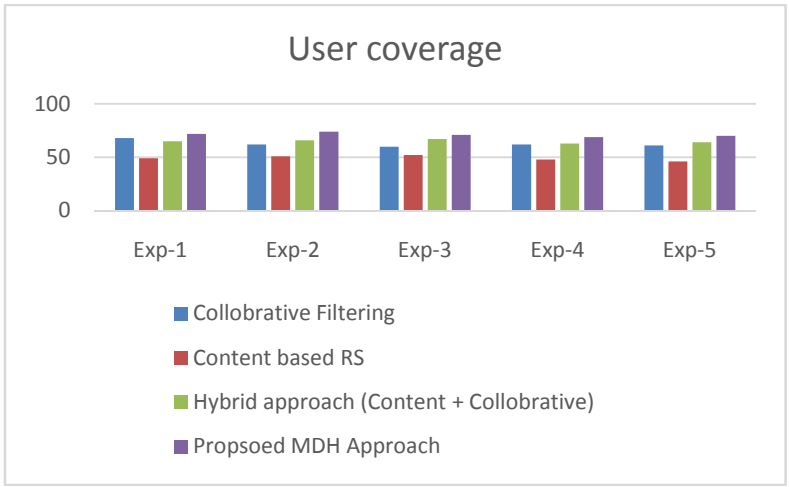

Fig. 5. User coverage comparative analysis

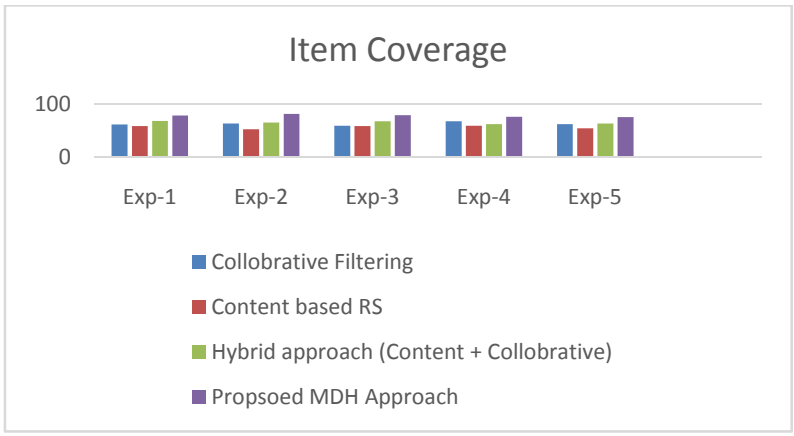

Fig. 6. Item coverage comparative analysis

Figure 7 gives the comparative results of catalog coverage evaluation. The outcome of the of all 5 experiment set, shows better results with MDH approach when compared to other existing recommendation approach.

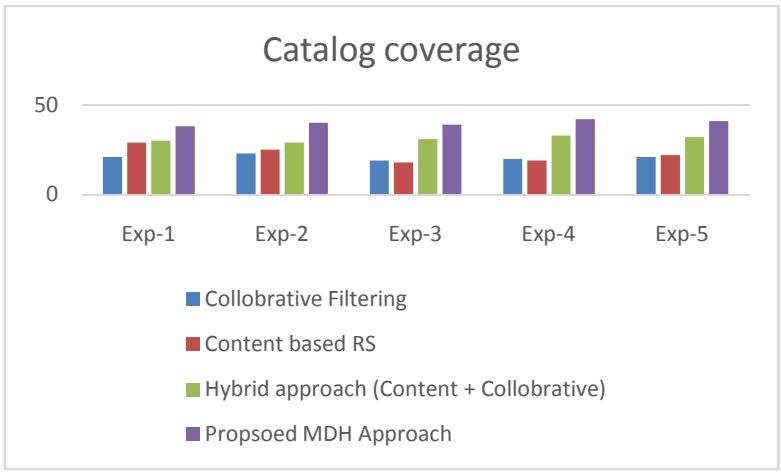

Fig. 7. Catalog coverage comparative analysis

5 CONCLUSION AND FUTURE WORK

In order to provide interesting and relevant products to the customer, new Multi demo graphic hybrid based product recommendation system is developed. Enhancing the recommendation system fuzzy approach with Demographic information, gives better results.Fuzzy rule is written to consider the product overall ranking and also the delivery location of the customer. Considering demo graphic information end user age and location is further improving the accuracy of the recommendation system. In the MDH approach, the first end user group is determined and product rank is computed. Later product list is further filtered based on the delivery location. Experimental results show MDH approach better results prediction accuracy is improved by 8 percentage when considering the user location and age group. In addition to accuracy, coverage also evaluated and found product and user coverage of MDH gives better results.

In future, planned to improve the preprocessing technique and apply neuro fuzzy to improve the accuracy further. In the neural implementation, negative product list and outcome of the earlier recommendation list can be considered, which will improve the result of the product recommendation.

References

1. Soe-Tsyr Yuan, Chiahsin Cheng, “Ontology-based personalized couple clustering for heterogeneous product recommendation in mobile marketing”, Expert Systems with Applications 26, pp. 461–476,2004
2. Jae Kyeong Kim, Yoon Ho Cho, Woo Ju Kim, Je Ran Kim and Ji Hae Suh, “A personalized recommendation procedure for Internet shopping Support”, Electronic Commerce Research and Applications, pp. 301–313,2002
3. Duen-Ren Liu, Ya-Yueh Shih, “Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences” , The Journal of Systems and Software, Volume 77, pp. 181–191, 2005
4. Sang Hyun Choi, Sungmin Kang and Young Jun Jeon, “Personalized recommendation system based on product specification values”, Expert Systems with Applications, Volume 31,pp. 607–616,2006
5. Kun Chang Lee and Soonjae Kwon , “Online shopping recommendation mechanism and its influence on consumer decisions and behaviors: A causal map approach”, Expert Systems with Applications, volume 35, pp. 1567–1574, 2008
6. Yukun Cao and Yunfeng Li , “An intelligent fuzzy-based recommendation system for consumer electronic products”, Expert Systems with Applications, volume 33, pp. 230–240, 2007
7. Yuanchun Jiang, Jennifer Shang and Yezheng Liu, “Maximizing customer satisfaction through an online recommendation system: A novel associative classification model”, Decision Support Systems, Volume 48, pp. 470–479, 2010

8. Sung-Shun Weng, Binshan Lin and Wen-Tien Chen , “Using contextual information and multidimensional approach for recommendation”, Expert Systems with Applications, volume 36, pp. 1268–1279, 2009
9. Ruben Gonzelez Crespo, Oscar Sanjuán Martínez, Juan Manuel Cueva Lovelle, B. Cristina Pelayo García-Bustelo, José Emilio Labra Gayo and Patricia Ordoñez de Pablos, “Recommendation System based on user interaction data applied to intelligent electronic books”, Computers in Human Behavior, volume 27, pp. 1445–1449, 2011
10. Vicente Arturo Romero Zaldivar and Daniel Burgos,” Meta-Mender: A meta-rule based recommendation system for educational applications, Procedia Computer Science, pp. 2877–2882,2010
11. V. Mohanraj, M. Chandrasekaran, J. Senthilkumar, S. Arumugam and Y. Suresh,” Ontology driven bee’s foraging approach based self-adaptive online recommendation system”, The Journal of Systems and Software, volume 85, pp. 2439– 2450, 2012
12. Sunita B. Aher and L.M.R.J. Lobo, “Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data”, Knowledge-Based Systems, Volume 51, pp. 1–14,2013
13. Jian-Wu Bi, Yang Liu and Zhi-Ping Fan,"Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking",Information Sciences, volume 504, pp. 293–307,2019
14. Aminu Dau, Naomie Salim, Idris Rabiu and Akram Osman ,"Recommendation system exploiting aspect-based opinion mining with deep learning method", Information Sciences, volume 512, pp. 1279–1292,2020
15. Xiaofeng Zhang,Huijie Liu, Xiaoyun Chen, Jingbin Zhong and Di Wang, "A novel hybrid deep recommendation system to differentiate user’s preference and item’s attractiveness",Information Sciences, volume 519, pp. 306–316, 2020
16. Karthik R V, Ganapathy Sannasi and Kannan Arputharaj, “A Recommendation System for Online Purchase Using Feature and Product Ranking”, Eleventh International Conference on Contemporary Computing (IC3), 2018
17. Chao Yang, Lianhai Miao, Bin Jiang, Dongsheng Li and Da Cao,"Gated and attentive neural collaborative filtering for user generated list recommendation",Knowledge-Based Systems ,volume 187, pp. 104839,2020