International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 17, Number 1 (2024), pp. 21-27

A Recommendation System for E-Commerce Based On Similarity Measure and Clustering Technique

Mohammed Abdullah Al-Hagery {Email: hajry@qu.edu.sa}
Department of Computer Science, College of Computer, Qassim University, KSA
ORCID: 0000-0001-6939-013X

Afnan Almuqbil {**Email:**4459@qu.edu.sa},
Department of Computer Science, College of Computer, Qassim University, KSA
ORCID: 0000-0001-5565-3103

Bashayer Alharbi {**Email:**bas.alharbi@uoh.edu.sa}
Department of computer Science, Applied college,University of Hail,KSA
ORCID: 0000-0002-9092-3364

Amjad Alowageel {**Email:**Amjad.alsultan@ju.edu.sa},
Computer and Information Sciences Department, Jouf University, Sakaka 72388, Saudi Arabia
ORCID: 0009-0004-8964-188X

Abstract
There are huge tons of transactions being done online every day and this implies that e- commerce is facing the problem of information overloads. This problem can easily be alleviated by having recommendation systems in this regard. There are two main techniques of recommendation; content-based filtering and collaborative filtering. Implementation of one of these two techniques at a time on an e-commerce system for product recommendations can have huge limitations. The proposed method aims at a hybrid approach that would exploit the capabilities of both the techniques to achieve optimal output for product searching. In this paper, we will implement three techniques clustering, user-based collaborative filtering, and similarity measure, to get the final recommended products. This approach was applied to a random data. It presents an approach in which past user experienced in content-based filtering and user-based collaborative filtering with a clustering algorithm to improve search accuracy. The accuracy of search results matters to the user as it is what makes they pick on one e- commerce website over the other. The high-level accuracy of product recommendations is achievable through the proposed approach.
There are
Keywords: Recommendation Systems, E-Commerce, Clustering, Content-Based Filtering, Collaborative Filtering, Similarity Measure, User Preferences.

1. Introduction

what songs to listen [4]. Making decisions on the websites is time-consuming so by an effectiveRS that will resolve the problem and save the time and effort when the user is looking for a particular product.

Recommendation systems, in general, make recommendations list in various approaches

- through different strategies like collaborative filtering and content-based filtering[1], graph- based algorithm[5], or

The objective of analysis of transactions data is to investigate and extract hidden patterns from data of transactions that kept as large databases by using diverse data mining strategies and methods. Data are available in a large number in the information industry; these data are useless until it converted to helpful information. Therefore, it is fundamental to analyze this data and extract valuable information from it. That permits clients to use data from wide distinctangels [1]. The recommendation system is an application example of data mining and it is important nowadays because it provides the easiness for the users to find the products that theysought to buy. Recommendation systems are used in many sites to recommend products to thecustomers [2]. Nowadays E-commerce websites are developing so quickly, for that it is a difficult action for online buyers to select a proper category. To deal with such a broad-ranging commercial problem, most electronic retailing sites merge the Internet services with buyer datato evolve a recommendation system, to predict their desire they use buyers background and actions, then it helps E-commerce sites to make appropriate recommendations [3]. Often the search for a product on an online site is by writing keywords such as “red skirt,yellow t-shirt”. This search can be exhausting and annoying if the site does not have a useful Recommender System (RS).The RS assists users to discover items that meet their desires by recommending services or products with the goal of supporting users in making decisions in different domains, like what products to buy, what film to look at, what painting to observe, or concept level method[6]. In [7], a collaborative filtering strategy was suggested in personalized recommendation method for E-commerce platform. The collaborative filtering has been widely used in data mining and other fields such as the improved cooperative filtering algorithm that merges the user review text and user rating [8] and for recommending items to users [9], [10], [11]and in film recommendation [12], [13]. Oursuggestion in this paper is based on using a cluster

21

approach on both of the filtering methods, which are content-based filtering and collaborative filtering as an extension for developing the recommendation system presented by [14]. The importance of the proposed approach is facilitating the search for a recommended product, and it depends on two merged techniques works with a similarity measure.

The paper organized as follows, section 2 discusses the historical background, section 3 illustrates the used tools, section 4 explains methodology in details, and section 5 discusses the results, section 6 demonstrates the conclusions, finally section 7 for future work.

2. Historical Background

An example of web usage mining is Recommendation system [15]. Content recommendation system suggests products to the customers based on the content of their buying history because it provides the content of the overview of the products, in which customer is generally attracted from many products. Finding out the quality of the item cannot be done via content-based filtering. To solve this issue cooperative filtering system are included(sometimes called "social filtering" or recommender system) because they are based on the opinion of the other customers[16]. According to [3], they introduced a new strategy merging sentiment assessment with cooperative filtering to improves the accuracy of personalized recommendation and attempts to overcome the problem of cold start and data sparsity. Cooperative filtering systems applications provide customers with a good experience, but they still face some of the main issues such as data sparsity. Data sparsity is a very big problem because the accuracy of the recommendation from the cooperative filtering algorithm declines. Building a recommendation system have a variety of methods such as content-based filtering, and hybrid collaborative filtering Recommender System. Content-Based Filtering has been successfully implemented using algorithms TF-IDF in [17], as a comparison of data existing cosmetics in the database with the user input to increase their sales. A new clothing Recommendation System has been presented by [4] System based on the combination of visual features, visual attention and textual attributes. They conducted the tests according to the recommendation and classification mechanisms. The outputs showed that their strategy reached the best outputs when compared to the standard item KNN methods.

Recommendation system can be used in different stores or sites like in [1] they use the RS in books to help people like learners to find the best available books from the database that meet their preferences. Collaborative filtering and content-based filtering are used to find out the desired books based on ranking and content. In the recommendation system, data sparsity is one of the common issues. According to [18], there is a solution to overcome the data sparsity in RS and to enhance accuracy and scalability of recommender systems they introduce a clustering-based matrix factorization method. Their idea is to identify rating-patterns by mapping all customers/products to their corresponding customer/product clusters. This method helped in solving the sparse data by applying three points including; Factorization; Clustering and Approximation. According to [19], the collaborative filtering is implemented to obtain more accurate prediction based on user's preferences by combining of user's comments and scoring of the items.

Recommendation system application in different areas in our life .one of the application like[20] Introduced a content-based mobile recipe that is able to recommend a customer preferred recipe using content-based filtering algorithm(CBFA). The solution helped those who have no idea what to cook with only one click of a button. With this service, the customer does not need to spend a lot of time to get his/her desired recipe. To recommend a recipe the weight of each feature for the viewed recipes by the customer is computed. The recipe that has the same features besides the highest weight and not has been seen before will be displayed to the customer as a recommendation

Information system us everywhere and the recommendation system helped to lighten the problem of overloaded information and allow the customers to access the relevant information and services according to [21], their objective was to personalize the mediation system by providing recommendations to customers. They followed a clustering algorithm of the customer profiles basing on a variety of customer's dimension that they grouped into categories depending on the kinds of values contained on each dimension in a view of the comparison between the different user profiles and generate clusters of customers. From the concerns in implementing recommendation, system is providing good quality to the registered users.

A set of features that may enhance the quality and impact of collaborative quality- improvement identified in the suggested approach [22]. The proposed recommendation system by [23] minimizes the false positive error that occurs frequently in the traditional system. The results proved that accuracy achieved using improved k-means that are 82% to 85% is better than the old k-means algorithm. The recommendation system has the potential to attract the new customer and maintain the existing one. Their suggested work represents the age-based clustering method that improved K-means clustering algorithm performance and accuracy in the area of recommending products such as books to users. They conclude that increasing efficiency of K-mean algorithm and Users find better results corresponding to their views and purchases of books. The proposed system is also used for other recommendation like movie, music, electronic items etc.

According to [24], the content recommender system limitation in the early stage is insufficient data and to suggest relevant content that meets the consumer need have to obtain more information related to the consumer such as the profile. The profile helped to obtain personalized content and recommended similar content products. Mobile content recommendation system for revisiting consumer is presented. It integrated the system with the content-based filtering method. That addressed the problem of the insufficient information and it provided more efficiency.

Trust is an important issue in building a proper recommendation system some of the researchers study this issue in "TruCom" that uses the social network for the user in a specific domain to recommend for them the items the is approbated for them[25]. Trust and secure recommendation system can solve problems such as the cold star as in[26] via using the location of the user to recommend items for the cold user.

3. The Proposed Techniques

In order to solve the information overload problem, it is imperative to establish an understanding of how content-based and collaborative filtering techniques work. The content-based filtering the suggested items to the buyer based on their experience with the system, which involves product ratings and previously purchased items. Therefore, in order to develop the hybrid system that incorporates the technique Term Frequency (TF) and Inverse Document Frequency (IDF) must be accounted for as they are essential in information retrieval systems [27]. While TF is the definition of how frequent a word appears, IDF defines how frequent the
document holding the word appears in the whole collection of documents. TF/IDF is a Vector Space Model. It is just one of the preferred models for developing the recommendations system, other models include probabilistic models such as Decision Trees and Naïve Bayes Classifier [28]

The collaborative algorithms appropriate include neighborhood-based algorithm and correspondence mean algorithm[29]. In the neighborhood-based algorithm, find of different users who share common interests, therefore forming a neighborhood, this implies that products liked by one user can be recommended to a different user in the neighborhood. With correspondence mean algorithm, the frequency of a product in a neighborhood is utilized in the recommendation. In addition, we used the following software's: SPSS and Excel.

4. The Methodology

In the proposed approach takes into consideration the similarity between customers to improve the accuracy of the recommendation system and make more choices. This approach proposes three techniques; clustering, user-based collaborative filtering, and measuring the similarity between items to obtain the final recommended products as shown in Figure 1.

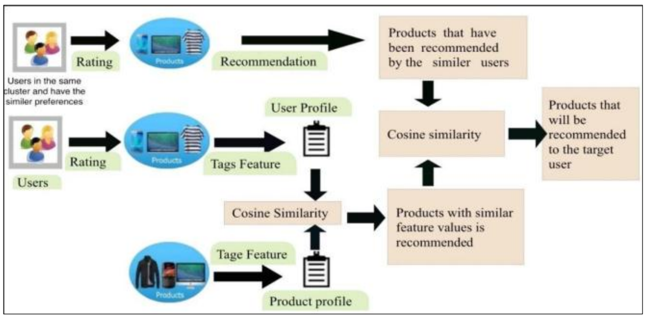

Figure 1: Proposal Recommendation System Approach

Figure 1 shows the processes implemented in that approach. First, Similar users rating some products that will be recommended to the target user. Second, measure the similarity between the products that will be recommended by a similar user. Third, select the top recommended products that have similar features to the user preferences.

These additions to the proposed approach implemented as follows:

1. Clustering: by dividing customers based on their personal Information to include similar users in the same cluster.
2. User-based collaborative filtering: by measuring the similarity between the users based on their preferences.
3. Measure the similarity between the recommended items where the similarity distance between the top recommended products resulted from the original approach and recommended items of a proposed approach is calculated using Cosine similarity. The next section showing in detail how these techniques implemented.

4.1. Clustering

Clustering is a grouping set of customers where the users in the same cluster more similar to each other than to those in other groups. By this technique, the users divided based on two attributes (age and income) into four clusters using K-mean algorithm. Table (1) shows an example of the personal information data of the customers to categorize similar customers together.

Table 1: Personal Information of the Users

Customer_Id	Age	Income	Weight	Occupation	Sex	Race
1	22	2,000	70	writer	F	w
2	44	10,000	55	banker	M	b
3	34	9,000	44	nurse	F	b
4	67	1,500	67	singer	F	a
5	65	1,000	63	teacher	M	a
6	32	5,000	89	biologist	F	i
7	44	8,000	50	chemical	M	w
8	20	2,500	67	teacher	F	b
9	30	4,500	76	nurse	F	a
10	35	5,000	87	doctor	M	i
11	22	3,000	65	waiter	M	w
12	19	1,500	67	singer	M	b
13	32	4,200	97	nurse	F	a
14	27	6,000	78	teacher	F	i
15	50	3,000	87	banker	M	w
16	45	9,000	76	writer	M	b
17	55	10,000	98	teacher	F	a
18	39	8,500	78	doctor	M	w
19	23	2,000	76	writer	F	b
20	52	9,500	87	chemical	F	a
21	15	1,000	62	singer	F	i
22	28	6,000	83	waiter	M	w
23	26	5,500	92	banker	F	b
24	33	7,000	100	writer	F	a
25	38	7,500	69	nurse	F	i
26	47	9,000	56	doctor	M	w
27	16	800	76	biologist	F	b
28	46	9,000	87	manger	M	a
29	36	4,000	84	singer	M	i
30	41	2,000	57	doctor	F	w

Clustering was applied in Table (1), using the SPSS to find clustering results based on two steps; First, choose the number of clustering K=4. Second, run the clustering task according to two attributes (age and income). The clustering result will be shown in Tables (2 - 7). In Table (2), initial cluster centres are evaluated by applying the first estimate of the variable means to find the k centres for each of the clusters. Initial cluster centres are used for the first round of clustering then will be updated.

23

Table 2: Initial Cluster Centers

	Clusters			
	1	2	3	4
age	38	55	36	16
income	7500	10000	4000	800

In Table (3), final cluster centres are generated as the mean for each variable within each final cluster. The final cluster centres represent the characteristics of the typical case for each cluster.

Table 3: Final Cluster Centers

	Clusters			
	1	2	3	4
age	35	46	31	34
income	7167	9357	4078	1475

Table (4) represents the distance between the final cluster centres using Euclidean distances. Larger distances between clusters centre result of greater dissimilarities. Where the most different clusters are cluster 2 and cluster 4. In addition when the clusters are compared to itself, the result is empty like between the cluster 1 and 1.

Table 4: Distance between final cluster centers

Cluster	1	2	3	4
1		2190.505	3088.891	5691.667
2	2190.505		5279.386	7882.153
3	3088.891	5279.386		2602.779
4	5691.667	7882.153	2602.779	---

Table (5) illustrates the number of cases that are assigned to each cluster. Where the larger number of customers included in the third group, which they have an average income.

Table 5: Number of Cases in each Cluster

Cluster	1	6.000
	2	7.000
	3	9.000
	4	8.000
Valid		30.000
Missing		.000

Table (6) categorizes the customers to four groups based on their personal information. It includes a clustered column that is referred to the cluster category for each customer. Also, it has distances column that shows the distance between the initial cluster centroids and the object in the same cluster, i.e., customer 17 is very close to the initial cluster centre of cluster 2, where his age is 55 and has an income of 10000$.

Table 6: Categorization of the Customers Based on Personal Information

Case Number	Customer_id	Cluster	Distance
1	1	4	1200.015
2	2	2	11.0000
3	3	2	1000.220
4	4	4	701.855
5	5	4	205.915
6	6	3	1000.008
7	7	1	500.036
8	8	3	1500.085
9	9	3	500.036
10	10	3	1000.000
11	11	3	1000.098
12	12	4	700.006
13	13	3	200.040
14	14	1	1500.040
15	15	3	1000.098
16	16	2	1000.050
17	17	2	.000
18	18	1	1000.000
19	19	4	1200.020
20	20	2	500.009
21	21	4	200.002
22	22	1	1500.033
23	23	3	1500.033
24	24	1	500.025
25	25	1	.000
26	26	2	1000.032
27	27	4	.000
28	28	2	1000.040
29	29	3	.000
30	30	4	1200.260

4.2. Collaborative Filtering

Collaborative filtering is automatic prediction procedure that is carried out based on the history of user's product interaction integrated with the interaction history of all other customers on a site.

User-based collaborative filtering will be used by applying similarity computation to find set N of other users whose ratings are similar to target user using centred cosine (Pearson correlation) that expressed in formula1.

$$sim(A, B) = corr(A, B) = \frac{\sum u \in U (R_{u,A} - R_A)(R_{u,B} - R_B)}{\sqrt{\sum u \in U (R_{u,A} - R_A)^2} \sqrt{\sum u \in U (R_{u,B} - R_B)^2}} \quad \textbf{(1)}$$

Where:

- $R_{u,A}$ Denotes the rating of user u for item A.
- RA is the average rating of the item A.
- Ru,B Denotes the rating of user u for item B.
- RB is the average rating of the item B.

Table (7) gives an example of a user rating of a product-rating matrix). Each row in this table represents a user who has rating different products, where rating start from zero (for low rating) to five (for high rating). The empty fields refer to the user does not have been making a rating for this product. The implementation carried out in two steps. Normalizing the rating data then calculating similarity distance between users based on their rating.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 17, Number 1 (2024), pp. 21-27
© International Research Publication House. http://www.irphouse.com

Table 7: Customers Rating of the Products

Item 5	Item 4	Item 3	Item 2	Item 1	Customer _id
1	4	3	2		1
4		3			2
		2		3	3
3	1		2	1	4
2			2		5
1	4	3		5	6
		3	1		7
	5		2	5	8
5	2	2		1	9
5			4		10
1	2	4		2	11
1		4			12
5	1	2		1	13
	1	1		4	14
1	2	3		1	15

4.2.1. Normalization

The first step is illustrated in Table (8) where the user-based collaborative filtering is applied, where rating data is normalized by subtracting row mean to get the modified rating matrix. The main purpose of this normalization is to avoid calculating the products that have not been rating as a low rating.

Table 8: Normalized Customer Rating of the products

Item 5	Item 4	Item 3	Item 2	Item 1	Customer_id
1	4	3	2		1
4		3			2
		2		3	3
3	1		2	1	4
2			2		5
1	4	3		5	6
		3	1		7
	5		2	5	8
5	2	2		1	9
5			4		10
1	2	4		2	11
1		4			12
5	1	2		1	13
	1	1		4	14
1	2	3		1	15

4.2.2. Similarity Measures

The similarity distance is obtained using the similarity measure, which is a measurement tool to quantify the similarity between two data objects. For the data mining perspective, the similarity measure is a distance with dimensions representing features of the objects. When the similarity degree is increased then the distance will be reduced. As the second step, we will find the similarity between users using the pearson correlation coefficient expressed in formula 2.

$$sim(a,b) = \frac{\sum_{p\in P}(r_{a,p}-\bar{r}_a)(r_{b,p}-\bar{r}_b)}{\sqrt{\sum_{p\in P}(r_{a,b}-\bar{r}_a)^2}\sqrt{\sum_{p\in P}(r_{b,p}-\bar{r}_b)^2}} \quad \textbf{(2)}$$

Where:

- a, b: Refers to the users.
- p=$\{p1, \ldots\ldots, pm\}$Denotes the sets of products.
- $r_{a,p}$: Denotes the rating of the user a for product p.
- $\bar{r}_a$ Refers to the average rating of the user a.

Table (9) and Table (10) represent the similarity distance between fifteen Users based on their product rating in Table (7), where great distances between user testes correspond to higher dissimilarities.

Table 9: Similarity Measurement between Users Based on Products Rating

Table 9: Similarity Measurement between Users Based on Products Rating

User 1	User 2	User 3	User 4	User 5	User 6	User 7
Sim(1,2)=-0.632	Sim(2,3)=0.5	Sim(3,4)=-0.319	Sim(4,5)=0.825	Sim(5,6)=-0.694	Sim(6,7)=-0.059	Sim(7,8)=0.577
Sim(1,3)= -0.158	Sim(2,4)=0.533	Sim(3,5)=0	Sim(4,6)=-0.955	Sim(5,7)=-0.645	Sim(6,8)=0.345	Sim(7,9)=-0.117
Sim(1,4)= - 0.842	Sim(2,5)=0.645	Sim(3,6)=0.478	Sim(4,7)=-0.1066	Sim(5,8)=0.745	Sim(6,9)=-0.957	Sim(7,10)=0.5
Sim(1,5)=-0.816	Sim(2,6)=-0.48	Sim(3,7)=-0.5	Sim(4,8)=-492	Sim(5,9)=0.760	Sim(6,10)=-0.537	Sim(7,11)=0.31
Sim(1,6)=0.661	Sim(2,7)=-0.5	Sim(3,8)=0.288	Sim(4,9)=0.929	Sim(5,10)=0	Sim(6,11)=0.259	Sim(7,12)=0.5
Sim(1,7)=0.316	Sim(2.8)=0	Sim(3,9)=-0.235	Sim(4,10)=0.426	Sim(5,11)=-0.32	Sim(6,12)=0.478	Sim(7,13)= - 0.05
Sim(1,8)=0.457	Sim(2,9)=0.707	Sim(3,10)=0	Sim(4,11)=-0.264	Sim(5,12)=-0.645	Sim(6,13)=0.462	Sim(7,14)=0.29
Sim(1,9)=-0.708	Sim(2,10)=0.5	Sim(3,11)=-0.62	Sim(4,12)= -0.533	Sim(5,13)=0.765	Sim(6,14)=-0.432	Sim(7,15)=0.53
Sim(1,10)=- 0.316	Sim(2,11)= - 0.93	Sim(3,12)=-0.5	Sim(4,13)=0.977	Sim(5,14)=0	Sim(6,15)=0.168	
Sim(1,11)=0.392	Sim(2,12)=-1	Sim(3,13)=-0.22	Sim(4,14)=-0.184	Sim(5,15)= -0.41		
Sim(1,12)=0.632	Sim(2,13)=0.65	Sim(3,14)=0.866	Sim(4,15)=-0.204			
Sim(1,13)=- 0.835	Sim(2,14)=0.29	Sim(3,15)=-0.852				
Sim(1,14)=- 0.365	Sim(2,15)= - .85					
Sim(1,15)=0.573						

25

Table 10: Similarity Measurement between Users Based on Product Rating

User 8	User 9	User 10	User 11	User 12	User 13	User 14
Sim(8,9)=-0.27	Sim(9,10)=-0.3	Sim(10,11)=-0.62	Sim(11,12)=0.93	Sim(12,13)= -0.65	Sim(13,14)=-0.13	Sim(14,15)=-0.74
Sim(8,10)=0.58	Sim(9,11)=-0.51	Sim(10,12)= -0.5	Sim(11,13)=-0.43	Sim(12,14)= -0.29	Sim(13,15)=-0.32	
Sim(8,11)=-0.36	Sim(9,12)=-0.58	Sim(10,13)=0.593	Sim(11,14)=-0.36	Sim(12,15)=0.85		
Sim(8,12)=0	Sim(9,13)=0.97	Sim(10,14)=0	Sim(11,15)=0.86			
Sim(8,13)=-0.31	Sim(9,14)=-0.27	Sim(10,15)=-0.32				
Sim(8,14)=0.17	Sim(9,15)=-0.3					
Sim(8,15)=-0.12						

Let consider user 9 is a target user, we notice that the user 9 and user 13 have more similar taste and in the same cluster, so the products that preferred by user 13 will be recommended to user 9.

4.2.3. Similarity Of Given Items

Measure similarity between items that are given by two approaches based on their description. By taking top recommended items that is a result of the first approach and find similar items (based on the product description) that will be also recommended to the target user. Table (11) shows an example of a product profile matrix. it knows that there are six products, three given by the first approach, namely item A, item B, and item C. Three given by the proposed approach, namely item 1, item 2, and item 3 and the combined features of these products are 'short', 'occasion', 'skinny’, and bright.

Table 11: Description of The Items That are Resulted from Both Approaches

Item Name	Bright	Skinny	Occasion	Short
Item A	1		1	
Item B	1	1	1	
Item C	1			1
Item 1	1	1	1	
Item 2	1			1
Item 3	1		1	1

4.2.4. Normalization

First, we will perform the normalization by dividing the term occurrence (1/0) by sorting the number of attributes in the item description and the result will be shown in Table (12).

Table 12: Normalized items Description

Item Name	Bright	Skinny	Occasion	Short
Item A	0.7071	-	0.7071	-
Item B	0.5773	0.5773	0.5773	-
Item C	0.7071		-	0.7071
Item 1	0.5773	0.5773	0.5773	-
Item 2	0.7071	-	-	0.7071
Item 3	0.5773	-	0.5773	0.5773

4.2.5. Similarity Measures Using Cosine

Second, the similarity between the items will be measured using centred cosine (Equation1). In the original approach, the result was the product B, A, and C respectively, So, we will take product B first and find the similar products that given by the proposed approach, then we will take item A and find their similarity items and so on. For item B, we find that item1 is the most similar item by calculating their similarities: Sim(B,1) = 1. Therefore, item1 will be the next recommended item after item B.

5. Results Discussion

An increased accuracy of search results that leads to user satisfaction is based on the assumption that users are a neighbourhood have similar tests and that user preferences never change. With a hybrid recommender system in place, expected results for product recommendations will be optimal due to increased accuracy, for instance, let take in consideration user13 is a target user. First, the recommendation system will search for the products that are similar to the products that user13 has been rating. The output items will be listed as following: product A, product B, and product C. The second, is the system, which will search for the similar users to user13 by similarity computation, it will be found that the user 9 is the most similar user where the similarity distance between user 9 and user 13 is 0.965. In the same time, it has been found that user 9 and user 11 in the same cluster (cluster 3), so the products that have been rating by user 9 are recommended to user 13. The output items will be listed as follows: product E, Product D and product F. Third, the recommendation system will calculate the similarities between all the output results. Where the most similar product to the product A (let take product D for example) is listed after product A.

6. Conclusion

People in general, are attracted to the sites that have a powerful recommendation system because it makes it easier for them in searching. Our approach involves the clustering algorithm and user-based collaborative filtering to recommend the products that are favourable by one of the users to another user who has a similar preference to each other. Then measure the similarity between the products to recommend the most similar one for them. The paper also addresses some of the critical concerns in product recommendation: insufficient data for recommendation and novelty. Factorization, clustering, and approximation as perceived as the constructs of a better recommender system capable of addressing the issue of insufficient data for the recommendation. The issue of novelty, which mainly arises from collaborative filtering techniques, can be overcome by measuring the similarity between recommended items and clustering hence amplifying the significance of the proposed hybrid system. The recommendation works under the assumption that the preferences of a user never change when this may not be true in some cases since some people are just explorative.

7. Future Works

The future works focus on improving product recommendation through the process improvement. This recommender system can be developed by making changes to the algorithms that drive the recommendation process. Process improvement is highly dependent on the availability of information and the ability to analyze and manipulate the data in a productive manner. For instance, the proposal points out that the product recommendation works under the assumption that consumer preference never changes. In this case, calculating the similarity between all output results can help to recommend something new for consumers who are explorative with a very high likelihood that they would be impressed by the recommendation. Thus, the basis of future work can be attributed to improving working on the underlying assumption of product recommendation. Understanding consumers through feedback at a personal level can also be a priority in future studies regarding product recommendation.

References

[1] Mathew P, Kuriakose B, Hegde V. Book Recommendation System through content based and collaborative filtering method. Data Min. Adv. Comput. (SAPIENCE), Int. Conf., IEEE; 2016, p. 47–52.

[2] Tewari AS, Kumar A, Barman AG. Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining. Adv. Comput. Conf. (IACC), 2014 IEEE Int., IEEE; 2014, p. 500–3.

[3] Jing N, Jiang T, Du J, Sugumaran V. Personalized recommendation based on customer preference mining and sentiment assessment from a Chinese e-commerce website. Electron Commer Res 2018;18:159–79.

[4] de Melo EV, Nogueira EA, Guliato D. Content-based filtering enhanced by human visual attention applied to clothing recommendation. Tools with Artif. Intell. (ICTAI), 2015 IEEE 27th Int. Conf., IEEE; 2015, p. 644–51.

[5] Shaikh S, Rathi S, Janrao P. Recommendation system in E-commerce websites: A Graph Based Approached. Adv. E-commerce based on product description and user profile. 7th Int Conf Innov Comput Technol INTECH 2017 2017:95–100. doi:10.1109/INTECH.2017.8102435.

[15] Baraglia R, Silvestri F. Dynamic personalization of web sites without user intervention. Commun ACM 2007;50:63–7.

[16] Shardanand U, Maes P. Social information filtering: algorithms for automating "word of mouth." Proc. SIGCHI Conf. Hum. factors Comput. Syst., ACM Press/Addison- Wesley Publishing Co.; 1995, p. 210–7.

[17] Patty JC, Kirana ET, Giri MSDK. Recommendations System for Purchase of Cosmetics Using Content-Based Filtering n.d.

[18] Shajarian Z, Seyedi SA, Moradi P. A clustering-based matrix factorization method to improve the accuracy of recommendation systems. Electr. Eng. (ICEE), 2017 Iran. Conf., IEEE; 2017, p. 2241–6.

[19] Zhu Y, Song W, Liu L, Zhao X, Du C. Collaborative Filtering Recommender Algorithm Based on Comments and Score. Comput. Intell. Des. (ISCID), 2017 10th Int. Symp., vol. 1, IEEE; 2017, p. 304–7.

Comput. Conf. (IACC), 2017 IEEE 7th Int., IEEE; 2017, p. 931–4.

[6] Lah NSBC, Dahlan HM. A concept-level approach in analyzing review readership for E-Commerce persuasive recommendation. Res. Innov. Inf. Syst. (ICRIIS), 2017 Int. Conf., IEEE; 2017, p. 1–5.

[7] Yan L. Personalized Recommendation Method for E-Commerce Platform Based on Data Mining Technology. Smart Grid Electr. Autom. (ICSGEA), 2017 Int. Conf., IEEE; 2017, p. 514–7.

[8] Wang X, Wang C. Recommendation System of E-commerce Based on Improved Collaborative Filtering Algorithm n.d.

[9] Xia J. E-Commerce Product Recommendation Method Based on Collaborative Filtering Technology. Smart Grid Electr. Autom. (ICSGEA), Int. Conf., IEEE; 2016, p. 90–3.

[10] Wang P. A Collaborative Filtering Recommendation Algorithm Based on Product Clustering. Appl. Mech. Mater., vol. 267, Trans Tech Publ; 2013, p. 87–90.

[11] Sarwar B, Karypis G, Konstan J, Riedl J. Item-based collaborative filtering recommendation algorithms. Proc. 10th Int. Conf. World Wide Web, ACM; 2001, p. 285–95.

[12] Zhou T, Chen L, Shen J. Movie Recommendation System Employing the User-Based CF in Cloud Computing. Comput. Sci. Eng. Embed. Ubiquitous Comput. (EUC), 2017 IEEE Int. Conf., vol. 2, IEEE; 2017, p. 46–50.

[13] Pal A, Parhi P, Aggarwal M. An improved content based collaborative filtering algorithm for movie recommendations. Contemp. Comput. (IC3), 2017 Tenth Int. Conf., IEEE; 2017, p. 1–3.

[14] Badriyah T, Wijayanto ET, Syarif I, Kristalina P. A hybrid recommendation system for

[20] Cheng TL, Yusof UK, Khalid MNA. Content-based filtering algorithm for mobile recipe application. Softw. Eng. Conf. (MySEC), 2014 8th Malaysian, IEEE; 2014, p. 183–8.

[21] Ouaftouh S, Zellou A, Idri A. CARMS: Clustering based approach for recommendation in mediation systems. Innov. Creat. Inf. Technol. (ICITech), 2017 Int. Conf., IEEE; 2017, p. 1–6.

[22] Schouten LMT, Grol RPTM, Hulscher MEJL. Factors influencing success in quality- improvement collaboratives: development and psychometric testing of an instrument. Implement Sci 2010;5:84.

[23] Saklecha A, Raikwal J. Dynamic Recommendation System Using Enhanced K-means Clustering Algorithm for E-commerce. Int J 2017;8.

[24] Paireekreng W. Mobile content recommendation system for re-visiting user using content-based filtering and client-side user profile. Mach. Learn. Cybern. (ICMLC), 2013 Int. Conf., vol. 4, IEEE; 2013, p. 1655–60.

[25] Liu H, Xia F, Chen Z, Asabere NY, Ma J, Huang R. TruCom: Exploiting domain- specific trust networks for multicategory item recommendation. IEEE Syst J

27

2017;11:295–304.

[26] Ramesh B, Reeba R. Secure recommendation system for E-commerce website. Circuit, Power Comput. Technol. (ICCPCT), 2017 Int. Conf., IEEE; 2017, p. 1–5.

[27] SHUVAYAN DAS. Beginners Guide to learn about Content Based Recommender Engines. Analyticsvidhya 2015:1.

[28] Isinkaye FO, Folajimi YO, Ojokoh BA. Recommendation systems: Principles, methods and evaluation. Egypt Informatics J 2015;16:261–73.

[29] Lee HC, Lee SJ, Chung YJ. A study on the improved collaborative filtering algorithm for recommender system. Softw. Eng. Res. Manag. Appl. 2007. SERA 2007. 5th ACIS Int. Conf., IEEE; 2007, p. 297–304.

28