

Development of a recommender system based on a trust approach in social networks using the K-NN algorithm

Zohreh Mehrjooyan¹, Faraz Forootan^{2*}

Department of Computer Engineering, Islamic Azad University of dezfoul, dezfoul, Iran

Department of Computer Engineering, Islamic Azad University of dezfoul, dezfoul, Iran

*corresponding author: Faraz Forootan

Email: forootan@iaud.ac.ir

Abstract- In recent years, the increase in the volume of information has made it difficult to make accurate decisions about a specific matter. Recommender systems have been proposed as a suitable solution to handle this challenge. These systems use the processing of user comments and provide only useful information to them. This study aims to provide a trust-based recommender system in social networks using the K-nearest neighbor algorithm. Several operational tests have been performed on five datasets taken from GroupLens. Analysis of the results showed that the MAE value of the proposed method is better than the other three compared methods. This improvement is evident in all results from five data sets as well as the mean percentage of results (proposed method: 92.28, Bayesian: 84.45, CBR: 82.28, and RBF: 90.84). The average results show that the MAE of the proposed method is better compared to other methods. In general, the proposed method has better accuracy based on MAE than the other three methods.

Keywords- recommender systems, trust-based networks, social networks, collaborative filtering, K-nearest neighbor algorithm.

Introduction

Today, the growth of the E-commerce industry and its complexity has led to an increase in product data at an exponential rate. In such a situation, customers have been challenged to find optimal information about items or products. To help buyers in addressing this problem, large electronic trading companies provide their recommender systems to help customers in choosing the best products. This solution is a win-win trick. Recommendations provided by recommender systems can lead to two consequences. First, they help the user in making a decision (choosing the best item from a set of alternatives). Then they increase the user's awareness in his/her field of interest (while advising the user, new items and objects are introduced to him/her). Recommender systems are useful for both sides of a transaction (commercial or non-commercial) and provide many benefits. On the one hand, in business interaction, searching in the giant pool of information is facilitated and accelerated. That is, potential customers find it useful to use recommender systems. On the other hand, with the help of these systems, sellers can increase customer satisfaction and their sales rate.

In recent years, the unprecedented penetration of new Internet technology has created many applications in the field of e-commerce (Jiankun 2019). The attendance of applications in the field of B2C and B2B requires effective communication between machines.

One of the most important applications in this field is recommender systems. Recommender systems are a special type of information filtering system that extract items of interest to the user from a large pool of items and users. This system is an approach to dealing with the problems caused by the large and growing volume of information. It also helps the user caught in a huge amount of information to reach his objective faster. Some experts believe that the recommender system is a type of collaborative filtering (CF). This study aims for a comprehensive review of recommender systems. In general, the current generation of recommender systems is divided into four classes (Hill 1994).

The emergence of recommender systems in the mid-1990s provided an important research area in the field of collaborative filtering. In Greeshma Lingam (2017), a trust-based model with learning automata is proposed for recommending users in online social networks. For this purpose, the researchers used information on direct and indirect trust, the degree of association, and the proposed effect size. The proposed method uses a special parameter called Shannon's entropy information gain. Also, in Xiao (2017), researchers used collaborative filtering based on explicit trust/untrust clustering to develop a recommender system. By using the trust/untrust clustering technique, the proposed method can deal with the problem of sparse data and cold start. The

SVD algorithm was also used to discover trust communities. The sparse score complement algorithm was used to generate dense user score profiles and solve the cold start problem. Finally, the hybrid of these methods led to the improvement of the recommender system efficiency on the real data set. Massa et al. (2007) believe that trust information has a greater effect on self-starting recommender systems compared to item voting information. Trust comes in two forms, explicit and implicit. In explicit mode, trusted users can be manually entered into the system or extracted through social network links. In the implicit method, the final trust is obtained based on the history of users' votes. Both explicit trust (stated by the user) and implicit trust have been used in this field. However, explicit trust is more accurate than implicit trust.

Calculating the degree of similarity of the product with the user and the similarity of the user with the user is very important in recommender systems. The proposed methods for performing this calculation in recommender systems include a content-based approach, collaborative filtering, hybrid method, and knowledge-based method. The first two methods are basic techniques that have their limitations. As a result, most of the current recommender systems use the hybrid of these two methods in different ways. During the process of generating and presenting suggestions for a recommender system, the process of predicting the user's vote is considered a vital part of the system.

Onur et al. (2019) presented a method based on the fuzzy K-NN algorithm for suggesting the gender of anonymous users in social networks. The accuracy of the proposed method is better than kNN, naive Bayes, decision trees, and random forest methods. However, the solution considered by them is not based on trust. Adeniyi et al. (2016) presented a method for web usage data mining and web page recommender system development. Despite the desirable accuracy of the proposed method, their solution is not based on trust.

Previous recommender systems, especially systems based on collaborative filtering select similar neighbors for the active user with the help of similarity criteria and provide the desired suggestion based on the neighbors' opinions. The similarity criterion has a great effect on the performance of such systems. Choosing the right standard is one of the critical challenges in this field. Calculating implicit trust relationships of users through product scoring records, discovering such relationships through the K-NN algorithm to provide appropriate suggestions, and

at the same time not needing to find similar users can be introduced as the innovations of this study. In the proposed method, the history of the votes of all users even those different from the intended user can help the system in the process of providing the appropriate proposal. Therefore, the proposed system is completely independent of the specific relation.

Proposed method

To predict a user's vote for a specific item, you can use the estimation of similar users' votes for the new item. The default is that people with similar votes for items will most likely have similar votes in the future. As a result, finding similar users is the key challenge of this method. Users have voted for different numbers and types of items over time. Also, the overlap of different users (relative to different items) is different. As a result, it is difficult to find a similarity measure to properly account for all these features. Due to this challenge, collaborative filtering systems cannot achieve optimal efficiency.

A new approach in collaborative filtering is the use of a trust criterion. In general, trust is defined in two forms include explicit and implicit. In explicit mode, a person's trusted users are inserted manually. In the implicit method, trust is calculated based on the history of people's votes for items. In the proposed method, the efficiency of the collaborative refinement algorithm is improved by including the trust factor without considering the similarity of people. In this way, people who have different items overlaps with each other can help in predicting votes for items. The selected ranking algorithm is used to predict users' votes for the new item.

In recent years, ranking methods have been used in various applications, such as biology and high-scale satellite image processing and classification of web pages on the Internet. Developing efficient ranking methods in such fields has become a challenge. In this context, some classification methods such as the K-nearest neighbor algorithm (K-NN) are of special importance. Therefore, due to its practical importance, research in the field of inventing new methods is very active in this field. The combination of these applications with big data and various recommender systems has led to the definition of numerous projects in recent years. The proposed method of this research is based on the K-NN classification algorithm. The process of the proposed method is described in detail in the next section. Figure (1) represents the overall process of this study.

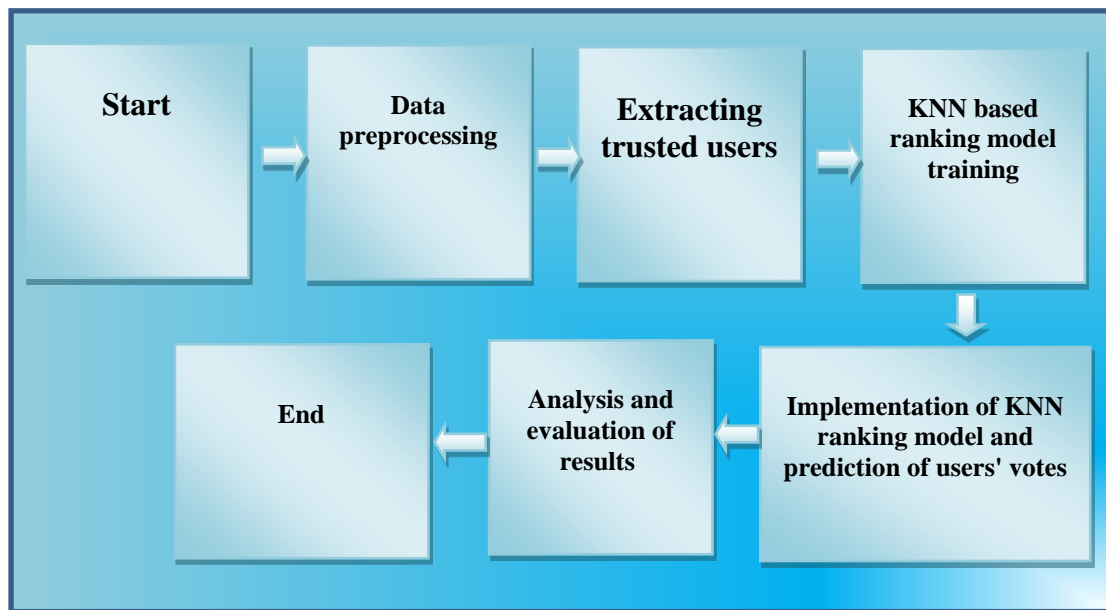


Figure 1: The overall research process

In the pre-processing section, the data input to the system is subjected to initial processing. First, the data are converted into a suitable form for prediction and the values of the features are normalized. Then the trusted users of a user are found. Based on their vote, the intended user's vote on a new item is predicted. Then the training of the system is done using the training data. Finally, the system learns the user's vote recognition pattern according to the input features. The developed model is the basis of the prediction process. In the next step, using the developed learning model, prediction is made according to the input data of users' votes about the new item, and the vote of a user about a new item is determined. In the last step, the output of the recommender system is evaluated and the correctness of the user's vote is determined by the system. Finally, the accuracy of the system is obtained based on the mean absolute error (MAE).

Data preprocessing

Before ranking, data must be preprocessed. The input data in the dataset are sparse. To be applied in ranking, these data must be converted into a complete matrix. Rows and columns represent users and items respectively. The value of each entry is equal to the value of the user's vote for the corresponding item. The output of this stage of pre-processing is shown in Figure (2).

	Item 1		Item j				Item m
User 1							
User i			Cell i,j				

User n									

Figure 2: User and vote matrix

Before ranking, data should be normalized. The data normalization method is described below. The set X with specific values is considered according to Relation (1).

$$X = \{ X_1 . X_2 . \dots . X_n \} \quad (1)$$

The minimum and maximum members of the set are determined according to the Relation (2).

$$Min(X) = r \mid r \in X \wedge \forall s \in X : r \leq s \quad (2)$$

$$Max(X) = r \mid r \in X \wedge \forall s \in X : r \geq s$$

The set of normalized values of each member of the set X is obtained between two values a and b according to equation (3).

$$Norm(X) = \{ a + (b - a) * (X_i - Min(X) / (Max(X) - Min(X)) \mid X_i \in X \wedge 1 \leq i \leq n \} \quad (3)$$

The output of this step is a set with well-defined data, which is better for comparisons and related analyses. At present, sufficient preparations have been made to rank and predict users' votes, the details of which are mentioned in the following sections.

Extracting trusted users

The meaning of "trust" between users in a recommender system is not exactly equivalent to the common meaning of this word. User A's trust in User B means that User A's opinions are similar to User B's. This similarity can be recognized by the user himself or implicitly recognized by the system based on the user's past behavior. In this research, the second approach is utilized.

In general, trust-aware systems generally use users' trust in each other, either explicitly or implicitly. An explicit trust is realized when the user explicitly expresses his trust to another user. In a social network, there are usually a set of explicit trusts named a web of trust. The criterion shown in Relation (4) is used to calculate the trust level.

$$Trust(U_i, U_j) \quad (4)$$

$$= \frac{n(\text{Score}(U_i) \cap \text{Score}(U_j))}{n(\text{Score}(U_i) \cup \text{Score}(U_j))}$$

where Score represents the set of items voted by a user, U_i and U_j represent two users whose trust level is measured, n represents the number of members of a set. In simpler terms, according to the above relation, trust between two users is equal to the ratio of the number of items voted by both users to the total number of items voted by one of the two users.

System training and construction of the KNN ranking model

The recommendation generation stage is the most important step of a recommender system. The generation of recommendations includes collecting information inside or outside the system, processing and analyzing them to determine the characteristics of the items and the priorities of the users, and forming a set of items that are aligned and appropriate to the priorities of the target user. A target user or active user is a user who receives a set of items as a recommendation. In other words, the recommender system is always collecting and processing the target user's information (priorities and interests in previous items) to provide appropriate recommendations.

The formation of a set of recommendations is achieved by various methods. The most common solution is to predict the user's vote for items. In the user-centered collaborative filtering technique, the similarity of each other user (those who voted for the target item) with the target user is measured. Then the rates given to the target item are applied in the form of weighted values in predicting the target rate for the target item. The rated weight of each user shows his similarity

with the target user. To select similar users in this research, the trust approach was used, the details of which were mentioned in the previous section.

Implementation of the K-NN ranking model and user vote prediction

Based on a distance criterion, the K-NN algorithm finds the K number of the nearest neighbors of the input sample and makes a judgment about the input sample. This algorithm decides on test data based on k nearest neighbors. Each sample has n features that represent a point in n dimensional space. KNN searches the feature space to find k the closest training samples to the test sample. Subsequently, among the found neighbors, the class with the most votes is selected as the test sample class. The closeness of the data can be defined by "distance" criteria, such as the Euclidean distance between two points X_1 and X_2 , which is defined as follows.

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (5)$$

The value of K should be obtained experimentally according to the problem. Usually, the process starts with K=1 and its value gradually decreases based on the calculated diagnosis error. The K value with the lowest error is used as the final value.

The pseudo-code of the algorithm is as follows.

Let $v = \{V_1, \dots, V_n\}$ be a set of classes

Given a query instance, x_q to be classified

Let $x = \{x_1, \dots, x_n\}$ denote the k instances from Training that are nearest to x_q

For all $i: 1$ to n $vote_i = \{x \in X \mid \text{class}(x) = v_i\}$

Return v_i such that $|vote_i|$ is largest

Figure 3: K-NN algorithm pseudocode

After predicting the user's vote on new items, the items with the highest number of predicted votes are presented to the user as recommended items, and finally, the output of the recommender system is presented to the user.

Analysis and evaluation of results

In the analysis and evaluation step, the proposed method is implemented on the selected dataset and the accuracy is reported in terms of the mean absolute value of the difference between the predicted value of users' votes for a new item and the actual value of votes.

Results

The Grouplens dataset has been used to test the proposed method. In general, grouplens is one of the reliable datasets in the field of recommender systems, which can be accessed at

<https://grouplens.org/datasets/movielens/>. About 2000 data from this main data set have been randomly selected and compiled in the form of 5 subsets named data sets 1 to 5 and used in experiments. Also, the K-

NN algorithm has been implemented in RapidMiner as one of the most prominent data mining platforms. Figure (4) shows the implementation model of the KNN algorithm.

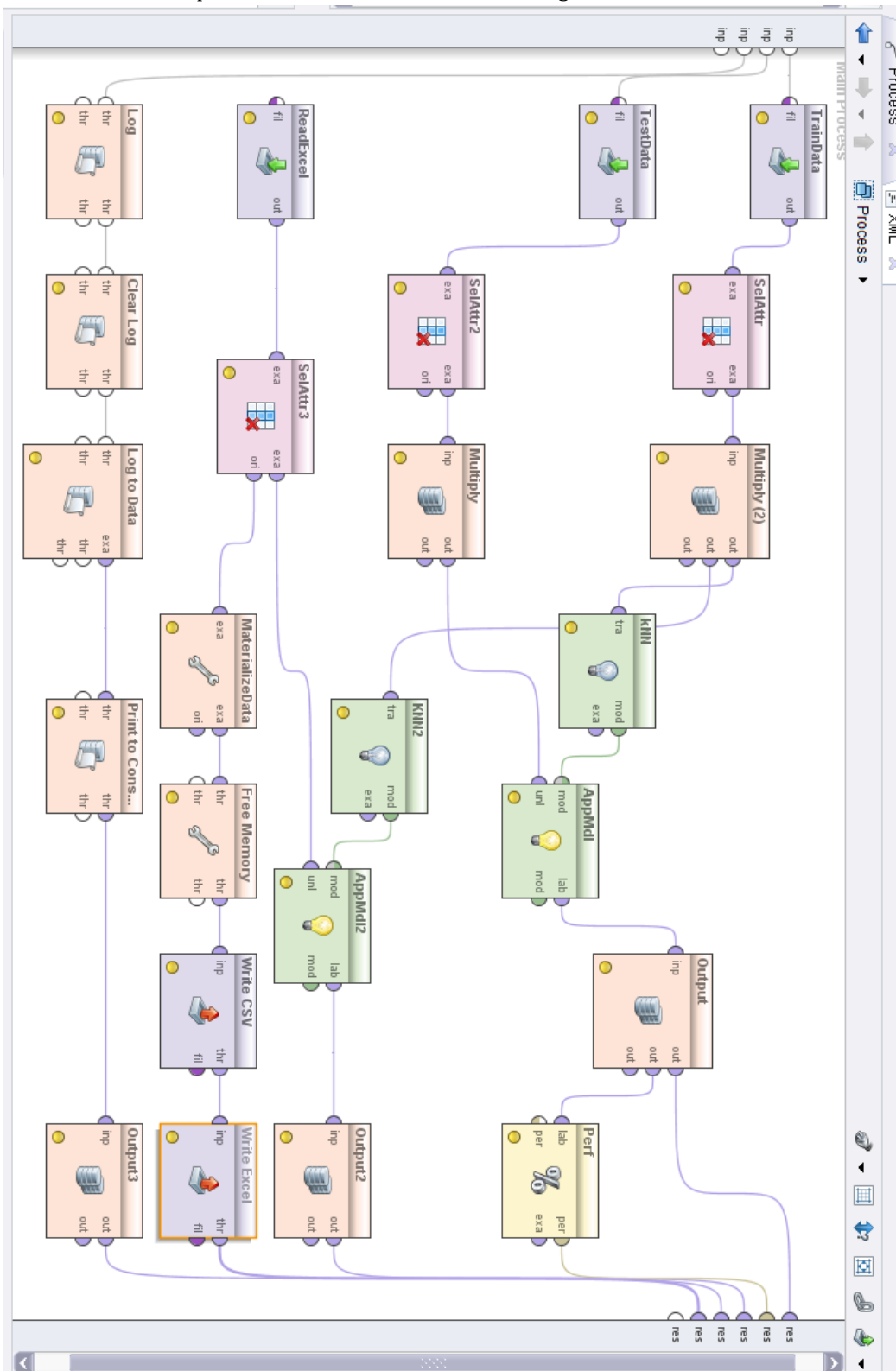


Figure 4: KNN algorithm model in Rapidminer

The matrix of votes is represented in Figure (5). Rows and columns represent users and items respectively. The value of each entry is equal to the value of the user's vote for the corresponding item. Here, the

columns of the matrix in Figure (6) are equivalent to the features and show a user's vote for an item. User votes for an item can be an integer between 1 and 10.

So here we have 10 distinct classes (integers in the range 1 to 10).

	I₁			I_j					I_m
U₁

U_i	Val _{i,j}

U_n

Figure 5: User and vote matrix
To analyze the resulting accuracy, the above table is analyzed according to Figure (6).

	I₁			I_j			I_i		I_m
U₁

U_i	Val _{i,j}

U_k

U_n

Figure 6: Accuracy matrix of the obtained results
In the matrix of Figure 6, columns I₁ to I_m correspond to new items voted by users. Users U_k to U_n are new users whose votes are predicted for new items. Let Val_{i,j} be the actual vote of user *i* to item *j*, Prd_{i,j} equal to the predicted vote of user *i* to item *j*, and the value of the vote varies between 1 and R. Relation (6) shows how to calculate the accuracy value for this particular prediction.

$$Accuracy_{i,j} = 1 - \frac{|Val_{i,j} - Prd_{i,j}|}{R} \quad (6)$$

Relation (7) shows the accuracy value for a specific user.

$$Accuracy_i = \left(\sum_{j=l}^m 1 - \frac{|Val_{i,j} - Prd_{i,j}|}{R} \right) / (m - l + 1) \quad (7)$$

Relation (8) shows the accuracy value for all users.

$$Accuracy = \left(\sum_{i=k}^n \sum_{j=l}^m 1 - \frac{|Val_{i,j} - Prd_{i,j}|}{R} \right) / (m - l + 1)(n - k + 1) \quad (8)$$

MAE is equal to the mean errors of observation. Also, Relation (8) represents the mean errors observed for all users' votes for all items. Therefore, Relation (8) exactly represents MAE. From now on, the resulting accuracy evaluation based on MAE is realized.

Analysis of the results

Table (1) shows the accuracy results of the proposed method compared to 3 other methods including RBF neural network, Case-Based Reasoning, and Bayesian classification. In general, 5 data sets have been used in this study. The data were randomly selected from the main data set including 2000 data (data sets 1 to 5) and were used in the experiments. For better readability, the values of the results between 0 and 1 have been multiplied by 100 and expressed as a percentage.

Table 1: Comparison of the values of the resulting accuracies (MAE)

	Proposed	Bayesian Classification	Case-Based Reasoning	RBF
DataSet1	91.5	89	88.2	92.9
DataSet2	92.6	90	90.5	92.1
DataSet3	91.8	87.1	84	92
DataSet4	93.4	82.3	80.5	91
DataSet5	92.1	74.5	68.2	86.2

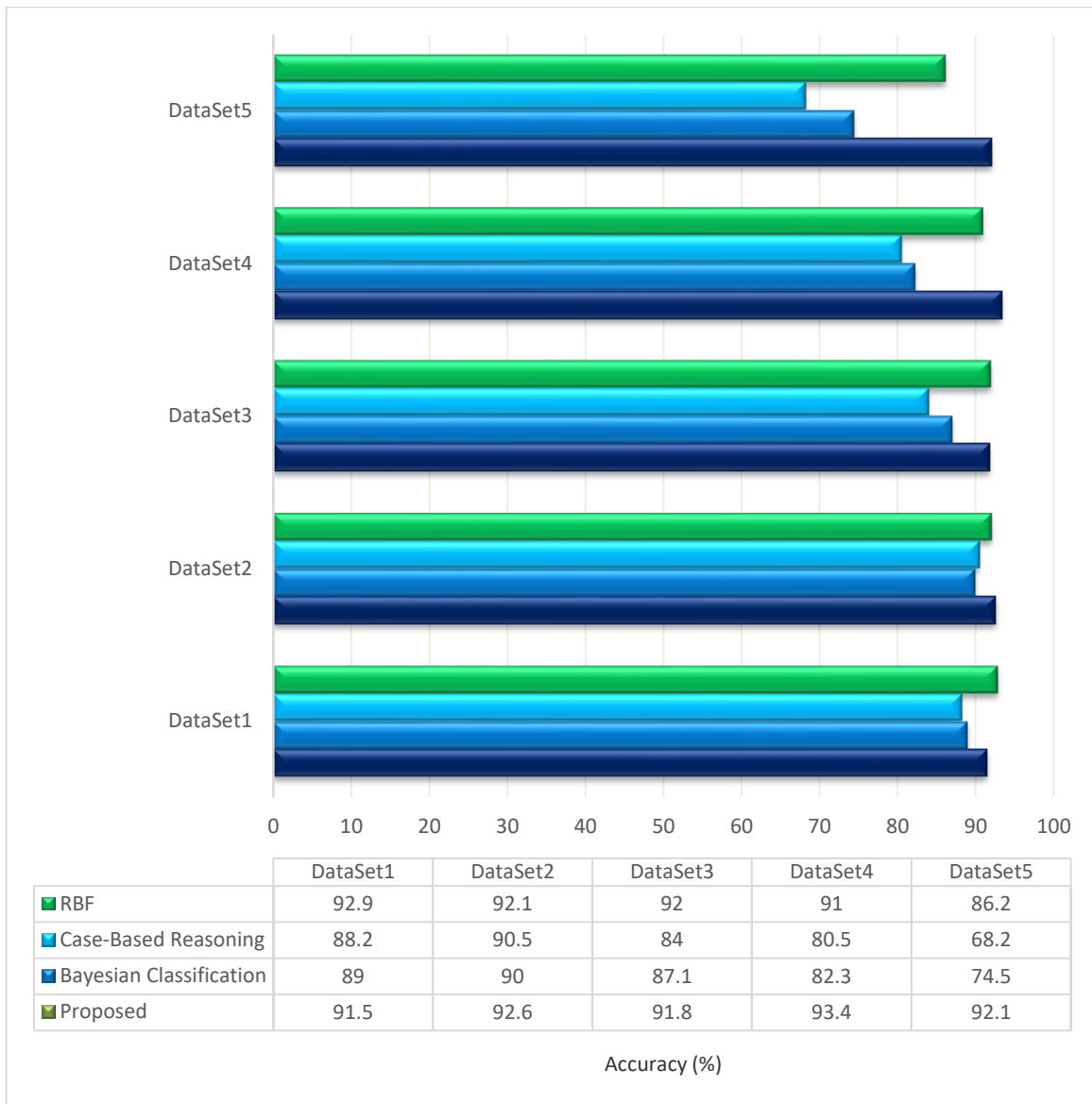


Figure 7: Comparative plot of the accuracies of the proposed method compared to other methods

Table 2: Comparison of the mean accuracies of the proposed method compared to other compared methods

	Proposed	Bayesian Classification	Case-Based Reasoning	RBF
Mean accuracy	92.28	84.58	82.28	90.84

As shown in the results, the performance of the proposed method and RBF is better than the other two methods. This issue is evident in each of the results obtained from each of the 5 data sets, as well as the mean results. The proposed method and RBF have a closer performance to each other. The RBF method has performed a little better than the proposed method in Data sets 1 and 3. However, the proposed method performs better on the other three datasets. The mean results show that the proposed method regarding MAE has performed better than other methods and the RBF method. Overall, the performance of the proposed method regarding MAE can be proven better than other methods.

Conclusion

Recommender systems in social networks are designed and developed to help users. The recommendations of these systems are based on the information and behavioral knowledge of users. This study has introduced a new application of recommender systems in the field of social networks. According to the experimental results, the performance of the proposed method is better than the three compared methods. This issue is evident in each of the results obtained from each of the 5 data sets, as well as the mean results. The mean results show that the proposed method regarding average MAE has better performance than other methods.

In line with the subject of the study, the following items can be suggested as future works. It will be desirable to develop and customize the proposed method for specific social networks such as social networks based on photos, audio, and video. Also, one of these developments can be realized in social networks based on streaming video. Also, adapting the proposed method to parallel and distributed processing systems to deal with big data can also lead to an efficient recommender system.

References

1. Adeniyi, Z. Wei, Y. Yongquan, Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, *Applied Computing and Informatics*, Volume 12, Issue 1, 2016, Pages 90-108, ISSN 2210-8327
2. Greeshma Lingam, Rashmi Ranjan Rout, 2017 DVLN Somayajulu, Learning automata-based trust model for user recommendations in online social networks, In *Computers & Electrical Engineering*.
3. Hill, L. Stead, M. Rosenstein, and G. Furnas, 1995 "Recommending and evaluating choices in a virtual community of use," presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA.
4. Jiankun Hu, Youfang Huang, Xiaoyan Zhao, 2019 Research on the slot allocation of container liner under E-commerce environment, *Computers & Industrial Engineering*, „ISSN 0360-8352
5. Massa, P. Avesani, 2007 "Trust-aware recommender systems," *Proceedings of ACM Recommender Systems*, Minneapolis, 2007 pp. 17-24.
6. Onur Dogan, Basar Oztaysi, Genders prediction from indoor customer paths by Levenshtein-based fuzzy kNN, *Expert Systems with Applications*, Volume 136, 2019, Pages 42-49, ISSN 0957-4174
7. Xiao Ma, Hongwei Lu, Zaobin Gan, Jiangfeng Zeng, 2017 An explicit trust and distrust clustering based collaborative filtering recommendation approach, In *Electronic Commerce Research and Applications*, Volume 25, Pages 29-39.