Indonesian Journal of Electrical Engineering and Computer Science
Vol. 31, No. 3, September 2023, pp. 1696~1704
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v31.i3.pp1696-1704

Conceptual framework of recommendation system with hybrid method

Tammanoon Panyatip, Manasawee Kaenampornpan, Phatthanaphong Chomphuwiset
Department of Computer Science, Faculty of Informatics, Mahasarakham University, Maha Sarakham, Thailand

Article Info

Article history:

Received Oct 13, 2022
Revised Apr 25, 2023
Accepted May 6, 2023

Keywords:

Collaborative
Content based
Framework
Hybrid
Recommendation

ABSTRACT

Recommendation system relies on information of user preference and user behavior in order to recommend the useful information. The existing recommendation systems still have problems for new users and new items. This research proposes a new hybrid method to develop the conceptual framework of recommendation system that deals with new user and new movie data. The data used consists of a data from MovieLens and the internet movie database (IMDB). This work introduces a hybrid recommendation system which based on a combination of content-based filtering (CBF) and collaborative filtering (CF). Pre-filtering data is performed by finding an optimal number of clusters by calculating the total within cluster sum of square. In order to reduce the complexity of data and increase the relevance of the user-item ratings, the fuzzy c-mean (FCM) is employed. Then the similarity is calculated by using item-based method, the K-nearest neighbors and weight sum of the rating are applied. Finally, to recommend the movies, the research found that for new user data the precision is at 85% and mean absolute eror (MAE) value 2.1011. For new item data, the result of research obtains the precision at 87% and MAE value 2.0031. In conclusion, the new hybrid method developed can recommend movie efficiently.

Corresponding Author:

Tammanoon Panyatip
Department of Computer Science, Faculty of Informatics, Mahasarakham University
Kantarawichai District, Maha Sarakham 44150, Thailand
Email: tammanoon.pan@msu.ac.th

1. INTRODUCTION

Recommendation systems [1], [2] are one of the tools that are used in electronic customer relationship marketing: eCRM. They are implicitly implemented as a model to evaluate the consumption of products for marketing purposes. In 1990, the recommendation system is the introduction of real user experience data using information about people or other people who are similar to recommend the relevant information to the user at a suitable time. With the increasing amount of availability data about the users, the users can exploit the recommendation result more accurately and correctly. The recommendation systems use the data from the user's past experiences to filter the unnecessary information out, so that a particular user only gets the information that is related and important for him at the right time. For example, the recommendation systems analyzing the behavior history of the user such as purchased products or watched movies, it then recommends the most relevant products and movies to the user accordingly. As a result, with less time spent looking through the website and exploring the unrelated items, the user can then complete the shopping experience quickly and therefore the company makes more profit in a shorter period of time.

As mentioned, the recommendation systems analyze information about the user to filter and predict relevant data to present to the user at a particular situation. Therefore, the input data to be used by the recommendation system is one of the important factors in modeling the system. In general, modeling a recommendation system can be performed using; i) implicit [3] and ii) explicit input data [4]. The implicit-based method, on the contrary, uses more implicit information related to users such as the behaviors of users in consuming products and shopping history. However, with the new user, the recommendation systems cannot recommend the suitable products/services to the user as there is a lack of the valuable history of the user's behavior information to build the model of recommendation. The explicit-based method utilizes common and direct data from users such as user's rating information to model a recommendation system [5], [6]. With the new products or the products that have low ratings from users, the recommendation system suffers from the sparsity problem [7]. This is where insufficient transactional and feedback data are available for inferring specific user's similarities, which affects the accuracy and performance of the recommender system [8], [9]. However, many works [10]–[12] demonstrate that other related information, such as user's demographics, incomes, or education level, can be integrated into the recommendation system. Moreover, the information related to the products can also be consolidated to the systems. For example in the movie domain, the information about the movies, such as information about directors, main characters and genre can be exploited in modeling the recommendation systems. This increases the scalability of the system.

In order to increase the accuracy of the recommendations, both implicit and explicit input data are adopted in the recommendation systems. As a result, the recommendation systems are also dealing with big data issues [13]. As a result of increasing numbers of items in the data and exploiting more related information of items in the recommendation system, the multiple data dimensions occur in the recommendation system. Therefore, the computation efficiency becomes an issue and potentially degrades the performance of the recommendation systems. Data preparation is an applicable technique that has been used to resolve the problem. In addition, machine learning-based techniques have also been deployed to alleviate the problem such as classifying non-contextual information from the data or clustering the data before processing to the recommendation systems [14]. Clustering data in the preparation step can be one of the keys to improving the recommendation systems. The technique divides the data into k groups of similar items (regarding selected attributes). However, selecting a suitable k for the task is significant to achieving promising performances of the systems. This paper, therefore, aims to improve a recommendation system by investigating the data preparation technique for improving the hybrid models of recommendation systems. There are many categories of the recommendation systems. In this paper, we group the recommendation systems into three main categories based on the information that is used to model the system.

Content-based filtering recommendation system normally performs based on the information that is contained in the system [15]–[17]. The technique uses the information of the product (features) and user preference to generate a recommendation for this user, which requires a machine learning algorithm to build the model. Applying the model, the recommender system can filter the suggestions that are suitable for each user. For example, Singla *et al.* [18] studied content-based movie recommender with internet movie database (IMDB) dataset. The product information that they adopted were movieID, plot, genre, title, votes, average rating score and year. The distributed bag of words version of paragraph vector (PV-DBOW) was used to extract feature words from the dataset. Then the term frequency-inverse document frequency technique was adopted for refining the preliminary recommendations. The final result showed the efficiency with a coverage value of 77.058%. Behera *et al.* [19] studied a movie recommendation system with MovieLens1M data. The product information that they adopted were year of release, genre, box office hit rate, descriptions of the movie. Then the user ratings were also used to compare the efficiency of the data mining techniques including KNN, restricted boltzmann machine (RBM), random techniques. The final result showed that the KNN method is the most efficient with a mean absolute eror (MAE) value of 0.72. The mae values of the RBM+KNN, RBM, RBM+random and random are 0.76, 0.86, 0.98, and 1.16 respectively. Even though, the content-based recommendation systems proposed by previous researchers seem to work well. But when there is a new user, the content-based recommendation system faces the difficulty of the cold start problem. This is because the system has no information about user preferences to generate a model from the user's interests.

Collaborative filtering method is based on the collection and analysis of similar user and other user data. This recommendations system based on user preferences are extracted first by considering user's ratings [20] item based collaborative filtering approach in movie recommendation system using different similarity measures either explicit rating from users to items, and implicit interactions or preferences of items are the basis of the collaborative filtering (CF) [21]. User neighbor search e.g. users with a rating history similar to current user [22]. The neighbor's ratings are then used to generate recommendations. The CF contains item-based and user-based, details are as follows; i) item-based is to compute the similarity between items using user ratings of those items and then to select the most similar items and ii) user-based is using that logic and recommends items by finding similar users to the active user, or used to predict the items that a user might like on the basis of ratings given to that item by the other users who have a similar taste with that of the target user. For example, Ochirbat *et al.* [10] studied a career recommendation system for students. The data that was adopted are the information about occupation, ratings or behavior of other users (such as the result from tracking of the student's scored answers in the questionnaire, checked/clicked occupational descriptions, rated content of the occupation searched occupation and ranked feedbacks ranging from 1 to 5). This work compared five types of similarities techniques including euclidean, intersection, jaccard, pearson and cosine techniques. It found that the pearson method is the most effective. Raghavendra and Srikantaiah [23] studied similarity based collaborative filtering model for a movie recommendation system with MovieLens data. The product information that they adopted were users ratings. Then the user ratings were also used to compare item-based CF and user-based CF. The accuracy values of the item-based CF and user-based CF are 0.84 and 0.76 respectively. This result is based on the 80:20 train and test data ratio. Yue *et al.* [24] studied an optimally weighted user-based and item-based collaborative filtering approach to predicting baseline data for friedreich's ataxia patients. The product information that they adopted were users ratings. Then the user ratings were also used to compare item-based and user-based. The MAE metric under density value of the item-based and user-based are 0.115, 0.120 at density 90%, and 0.120, 0.125 at density 80% respectively. In order to group the users and group the products accurately, the researchers have to collect a massive amount of features about the users and the products. Moreover, if there is not enough transactional or feedback data from the users, the sparsity problem occurs when available data is insufficient for identifying similar users (neighbors). This is a major issue that limits the quality of recommendations and the applicability of collaborative filtering in general.

Hybrid recommendation system is a method of combining the content-based and collaborative filtering methods to create a recommendation for a piece of information**,** or a product to the user to reduce the disadvantages and complement the advantages of both techniques of Afoudi *et al.* [25] studied the hybrid recommendation system by combining content-based filtering and collaborative prediction using artificial neural network. The MovieLens 100k dataset is used in this study. The product information that was chosen are gender, age, zip code, occupation and rating. This research suggested top-k at k=5 and k=10, where k is the number of suggested movies for all users. Content-based used TF-IDF and cosine similarity approach between the user profile and the items vector space model. The CF used the SVD approach and self-organizing map (SOM) method to improve the traditional CF with ANN and used k-mean clustering. The final result showed that the hybrid recommendation method is the most efficient with a f-measure (at k=10) with the value of 3.9274. The f-measure (at k=10) values of the hybrid, collaborative with k-means+SOM, collaborative, popularity and content-based methods are 2.5495, 1.8342, 0.9915, 0.6889, and 0.4834 respectively. This result is based on the 75:25 train and test data ratio. Sharma *et al.* [26] studied an automatic recommendation system based on a hybrid filtering algorithm with book recommendations. The product information that they adopted includes information about user, books and ratings. Content based filtering (CBF) that applied TF-IDF and cosine similarity to the information of user profile. The CF that applied cosine similarity with the rating vectors. Firstly, the hybrid method finds the users that are similar to the active user by looking at their profiles. The user's ratings are used to identify the similar users using the cosine. Then the related item for each neighbor is selected by obtaining vectors corresponding to the user profile (Vc) and item contents (Vm), by using the cosine. Lastly, the system can then recommend the items to the target user. In this study, the user ratings were also used to compare hybrid filtering, content based filtering and collaborative filtering. The final result showed that the hybrid filtering is the most efficient with a MAE value of 0.961. The MAE values of the content based filtering and collaborative filtering are 1.012 and 1.059 respectively.

The discussion of the previous works mentioned in this section demonstrates that the researchers are trying to improve the accuracy of the system through different machine learning algorithms. However, the big data issues that occur in the recommendation system still raise an issue. The data preparation step becomes very important and could lead to a major improvement of the efficiency of the system. The important step of data preparation is the application of minimum sum of square error (SSE). It helps the acquisition of k-group values, as a result, researchers do not have to waste time finding the right k-clusters values. Before cluster using the fuzzy c-mean (FCM) method, it is a granular clustering method, for increased clustering accuracy, and the KNN determination within the grouped data group and use the data to calculate the approximate value using the item-based method for preparing the data before being processed for the information recommendation system.

Recently, in the big data, there was a problem in searching for information that was similar to user and item data, including a problem with the recommendation system, including new user and new item or cold start. If the data is estimated grouping with SSE and grouped by FCM and estimating the intra-group proximity first by using the KNN method, helping to reduce this problem. For example, Sitompul *et al.* [27] studied enhancement clustering with the method of determining centroid based on the minimum SSE and k-means algorithm. The product information that they adopted were birth and death rates, wholesale customers and seeds dataset. Then the datasets were also used to compare k-means and SSE+k-means. The final result showed that the SSE+k-means method is the most efficient with davies-bouldin index (DBI) value of 0.0378 or 3.78%. The k-means DBI value of 0.0301 or 3.01 %. Arora *et al.* [28] studied FCM clustering strategies: a review of distance measures. The product information that they adopted were iris, wine and breast cancer dataset. Then the datasets were also used to compare euclidean distance, mahalanobis, chebyshev and minkowski. The final result showed that the euclidean distance method is the most efficient, chebyshev distances and minkowski distances are equally suitable for clustering. Selvi and Sivasankar [29] movie recommendation system with MovieLens 100k data. The product information that they adopted were rating matrix. Then data point clustering using modified fuzzy c-means (MFCM) approach and optimization of grouped data points using modified cuckoo search (MCS) algorithm. Then the user ratings were also used to compare MCS, particle swarm optimization (PSO) and cuckoo search (CS). The final result showed that the MCS method is the most efficient with accuracy value of 79.65%. The accuracy value of the PSO and CS are 75.97% and 72.32% respectively. The rest of the paper will be organized as follows. The reviewing related literature regarding the method is discussed in section 2. The results and discussion in section 3. Finally, section 4 provides the conclusions.

2. METHOD

As mentioned in section 1, there are researchers who have performed well with movieLens and IMDB data, for example; Singla *et al.* [18], Behera *et al.* [19], Raghavendra and Srikantaiah [23]. As mentioned, solving the issue of big data could improve the accuracy of the recommendation system. Therefore this paper proposes a method that aims to improve a recommendation system by investigating the data preparation technique for improving the hybrid recommendation systems. The recommendation system, in this work, is implemented based on a hybrid model by integrating content-based filtering and collaborative filtering. This section is divided into 2 parts. Firstly, the datasets as the most important part of modeling the recommendation system which is discussed. Secondly, as a method.

2.1. Datasets

The experimental dataset was collected from MovieLens [29], [30] and IMDB dataset [31], [32] explainable reasoning over knowledge graphs for recommendation. MovieLens comprise with 100,000 records of the rating data of 943 users on 1,684 movies. IMDB contains the movie's details, comprising with 22,240 of actors, 10,942 of actresses and 861 of directors. Our dataset is used for 4 experiments. Firstly, we random 95% of this dataset as a training set, while the rest is used as the test set. The dataset in the second, third and fourth experiments is split into 90:10, 85:15 and 80:20 ratio respectively.

2.2. Process design

This recommendation system is implemented to improve the performance of recommendation systems by focusing on the data preparation step to solve the problem of new users and new data items (movies). The overall results of the proposed technique is shown in Figure 1. This recommendation system is implemented to improve the performance of recommendation. As mentioned, the main focus of this work is on the data preparation step. From Figure 1, the process is devided in to 4 main parts including input data, pre-filtering process parts and output. The Input data part consists of sex, age, occupation, genres from MovieLens [29], [30] and actors, actresses, directors from IMDB dataset [31], [32]. Information section userID, movieID, rating (1-5) make user rating matrix. The frequency values of genres, actors, actreses, directors are in the form of data matrix.

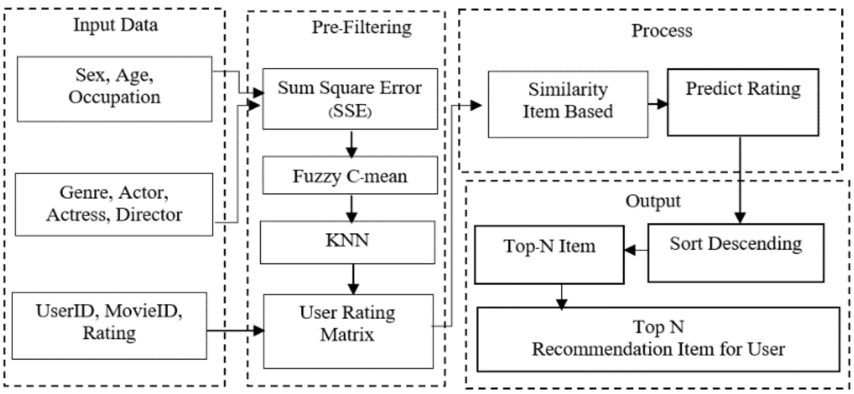

Figure 1. The overall process of the proposed technique

The aim of pre-filtering part is to reduce the size of the data and find the approximation of user and item data using the following methods. Firstly, the cluster estimation (k) is calculated by SSE [27]. As a result in this part we choose to use information about the user such as sex, age, occupation, genre, director, actor, actress to find SSE. The data is then converted to binary format and bring the data together in a matrix format. Before clustering, cluster k-values are required, wherein the investigators search for k-values by measuring the variance of the data within the cluster and define it as a line graph to see the refraction of the graph at the point corresponding to the smallest number of segments, when the value k is obtained, the user is grouped by information about the nature of the movie. That SSE is another technique for clustering validity and optimization cluster, which is positioned to indicate the appropriate number of clusters in the FCM clustering. Then the cluster with FCM [28], [29] part is applied after obtaining K-values from the group estimation using the SSE method, the researchers segment user demographic and movie feature as follows. Firstly, clustering demographic data contains information in regards to sex, age and occupation from MovieLens and correlation from userID attributes to help cluster the rating matrix data from MovieLens for a cold start solution. Then clustering movie feature data includes genre from MovieLens actor**s**, actresses, directors from IMDB and correlating from movieId attributes to help group the rating matrix data from MovieLens for solving cold start and sparsity problems.

After applying cluster with FCM, the KNN method [19] is the process of finding approximation from demographic and movie feature data. The new user's data includes sex, age, occupation to measure the distance from the center of the group (centroid) to allow the user to become a member of a particular group. When entering the prepared cluster, then the user's proximity within the group will be determined using the KNN technique, set the value k=12. The KNN method if the decision conditions are complex, this method can be used to create efficient models suitable for nominal data such as male, female and occupation. The new item data entry consists of genres, actors, actresses and directors. Its distance from the data center of the group (centroid) is measured so that the new item becomes a mumber of a particular group. When entering the prepared group, the user's proximity within the group is determined by using the KNN technique, set the value k=12. Then the user rating matrix is used to calculate proximity value method correlation-based similarity with Item-based [23] on the data containing userID, movieID and rating. All of the data is cleaned up. The reduction criteria for reducing the number of movies per user is 65, meaning that a movie must be watched by at least 65 people.

The last part is the process which is a process of collaborative filtering. It demonstrates a computation of similarity from the rating matrix data grouped. Calculate the approximation by Item-based method [23], corresponding to Yue *et al.* [24] approximately with the pearson method. Item-based on the correlation of the item-item score data, the results of the Raghavendra and Srikantaiah [23] and Yue *et al.* [24] experiments were more effective than the user-based approach and estimation of correlation. The pearson approximation, in which the results of Ochirbat and *et al.* [10] were more effective than approximation than the euclidean method. The last process is the output part. The output is a result of hybrid filtering which is the use of scores for each method. Content-based filtering and collaborative filtering: the weighted sum and the scores are sorted in descending order. Select the highest rating value of 10 movies (Top-N) to recommend movie information to users.

2.3. Evaluation

To evaluate the performance of the proposed method, this work applies MAE [19], [33] and precision [10]. We define the metrics for our pilot evaluation using the activities of subsection 2.1. We choose definite metrics for technical evaluation.

2.3.1. MAE

This is an evaluation technique to determine the errors (differences) of the predicted results and the ground truth data. A small degree of MAE signifies the more similar between predicted data and ground truth. MAE is calculated as follows:

$$MAE = \left(\frac{1}{N}\right) \sum_{u,i,r \in R} | \hat{r}_{u,i} - r_{u,i} \qquad (1)$$

where N is the number of recommended movies, $\hat{r}_{u,i}$ is predicted score and $r_{u,i}$ is ground truth score.

2.3.2. Precision

this metric examins the degree of agreement of the recommended movies and the ground truth data. This is another value that we used to measure the accuratcy of our design. The precision is calculated as follows:

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

where *TP* true positive and *FP* fale positive.

3. RESULTS AND DISCUSSION

3.1. Experimental setup

Section 2 explains the method and the dataset used in this work. MovieLens and IMDB are used as the dataset for evaluating the performance of the proposed method. The dataset is divided into 3 groups, according to the method-explained in section 2, i.e.; i) input data: composing; user's demographic information, movie's details, and rating information, ii) pre-filtering data-used in the clustering process using SSE, fuzzy-c-means and KNN algorithm, and iii) testing data.

3.2. Experimental data preparation

The result of the K-value determination before the data grouping with SSE to determine the appropriate K-cluster value for the data. It was found that when considering the most appropriate data grouping results that the graph is refracted at number 2, so the K-value in the proper grouping is K=2. The overall of the proposed technique is shown in Figure 2. Results of data clustering by FCM method it is a characteristic of user data clustering, using FCM technique. The researcher set the value K=2, meaning to define 2 clusters of data. The overall of the proposed technique is shown in Figure 3. Experimental results with the KNN method within the user or movie data group and K=12 it was found that when using user data or movie data to find proximity to members within a group and assigns K=12 to get information that is close to new users or new movies. The overall results of the proposed technique is shown in Figure 4.

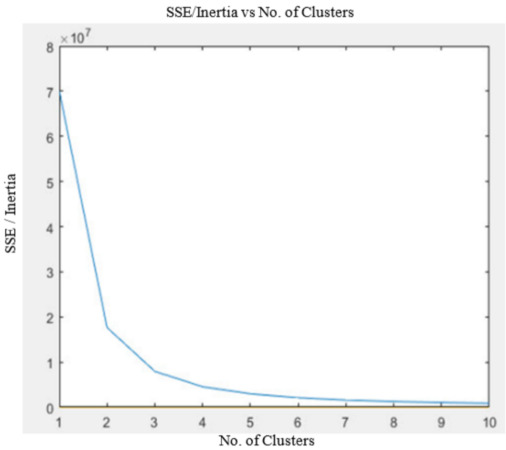

Figure 2. Clutering data with SSE

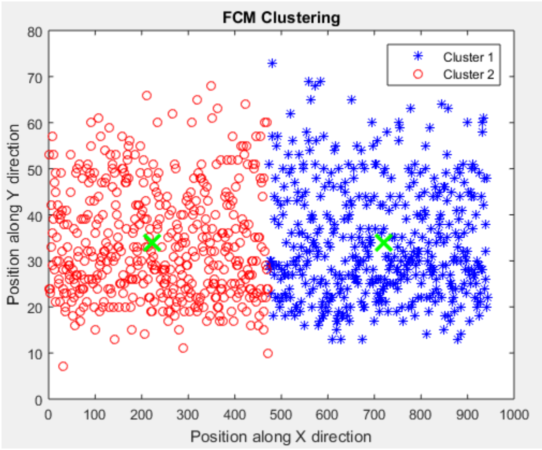

Figure 3. Clustering with fuzzy c-mean

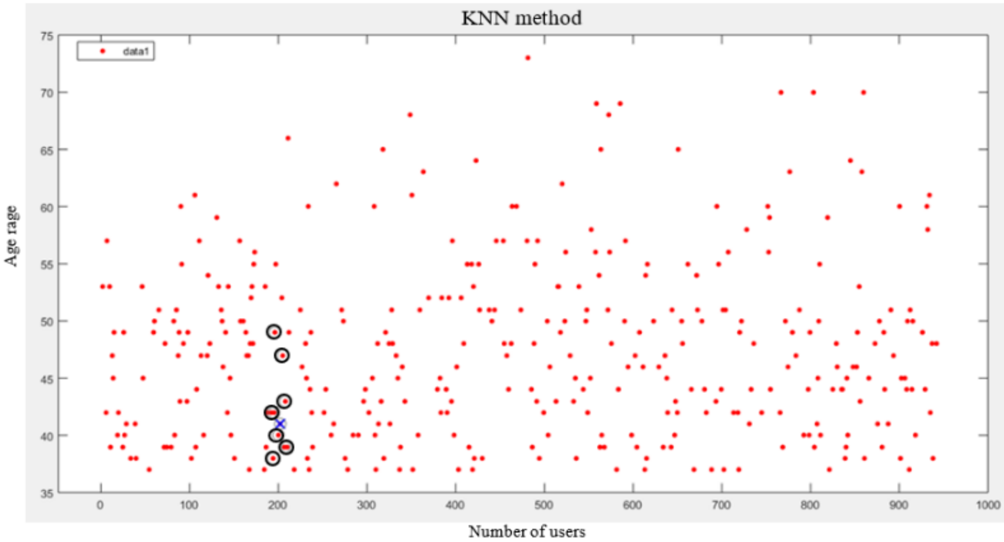

Figure 4. Finding proximity using KNN method

3.3. Experimental result and discussion
3.3.1. Experiment on new user

In this stage, a different number of users are tested, varying from 5%, 10%, 15%, and 20% of users from total data is selected. Each of the users are then experimented with the proposed method. In identifying a set of similar users, the set contains at most 12 similar users. The results from the experiment are shown in Table 1. From Table 1, the precision of the proposed method looks promising. Using 20% of the total users and selecting at most 5 movie titles, in the recommendation system provides the best result, achieving 85% of precision. This result is motivated with respect to the MAE value. The MAE for recommending 5 movie titles is less than recommending 10 movie titles. Therefore, the recommendation system suggests recommending 5 movie titles to a new user.

Table 1. Demonstrates the performance of the proposed method for new user issues

Dataset	Precision		MAE	
	Predicted 5 titles	Predicted 10 titles	Predicted 5 titles	Predicted 10 titles
5%	0.85	0.83	2.3312	2.3971
10%	0.78	0.80	2.0756	2.2649
15%	0.82	0.81	2.0778	2.2018
20%	0.85	0.84	2.1011	2.2070

3.3.2. Experiment on a new movie

1 experiment on new movies: in this stage, a different number of users were tested, varying from 5%, 10%, 15%, and 20% of movies from total data was selected. Each of the movies were then experimented with the proposed method. In identifying a set of similar movies, the set contains at most 12 similar titles. The results from the experiment are shown in Table 2. From Table 2, the precision of the proposed method looks promising. Using 20% of collected data from the total movie and selecting at most 5 movie titles in the recommendation system provides the best result, achieving 87% of precision. While, selecting at most 10 titles is at 86%. The MAE for recommending 5 users for a movie is less than recommending 10 users. Therefore, the recommendation system suggests recommending 5 users for a new movie.

Table 2. Demonstrates the performance of the proposed method for new movie issues

Dataset	Precision		MAE	
	Predicted 5 titles	Predicted 10 titles	Predicted 5 titles	Predicted 10 titles
5 %	0.87	0.85	2.2566	2.3721
10 %	0.82	0.81	2.1763	2.2430
15 %	0.84	0.84	2.0822	2.1892
20 %	0.87	0.86	2.0031	2.1104

For solving new user and new movie issues, the data that was used in the system composing of demographic information of the users (from MovieLens data), the rating data of the users and movies (from MovieLens) and the movie's details (from MIDB). The dataset is used in the preparation step. When a new user or a new movie is fed to the recommendation system, the new data will be assigned to a particular group. Within the group, a set of similar data items is selected using *K*NN algorithms, where K=12. The rating data and related information from KNN and fuzzy-c-means are used to calculate the weights. The experiment demonstrates that the proposed technique was outperformed when comparing to content-based filtering and collaborative filtering.

4. CONCLUSION

This work proposes a technique for recommending movies to users. This highlights the main aspects found in the recommendation systems, i.e., new user and new item. The proposed method focuses on the data preparation step by clustering the stored data into different groups, which will be used in a hybrid method for recommendation systems. The data used in the research consists of a data from MovieLens and the IMDB. The data from MovieLens contains 100,000 ratings from 943 users on 1,682 movies. As for the IMDB, the data includes information of actors, actresses and directors of the movies. In this work, a hybrid recommendation system with a combination of CBF and CF is introduced. Pre-filtering the data is performed by finding an optimal number of clusters to obtain optimal cluster centers. This is done by calculating the total within cluster sum of square. Then the FCM is implemented to reduce the complexity of data and increase the relevance of the user-item ratings. Then the similarity is calculated by using the item based method. Finally, the KNN and weight sum of the rating is applied to recommend the movies. The performance is measured with precision and MAE. The research found that for new user data the precision is at 85% and MAE value 2.1011. For new item data, the result of research obtains the precision at 87% and MAE value 2.0031. In conclusion, the new hybrid method developed, can recommend a movie efficiently. In future work, we aim to investigate a wide variety of dataset to experiment with the proposed method. In addition, a hybrid model for the recommendation system using deep learning methods is also one of the ideas that will be applied in the recommendation systems.

ACKNOWLEDGEMENTS

The authors are very grateful to the Faculty of Infomatics, Mahasarakham University and Kalasin University for supporting this research.

REFERENCES

[1] S. Rajarajeswari, S. Naik, S. Srikant, M. K. S. Prakash, and P. Uday, "Movie recommendation system," *Advances in Intelligent Systems and Computing*, vol. 882, pp. 329–340, 2019, doi: 10.1007/978-981-13-5953-8_28.

[2] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, and F. Xia, "Scientific paper recommendation: a survey," *IEEE Access*, vol. 7, pp. 9324–9339, 2019, doi: 10.1109/ACCESS.2018.2890388.

[3] B. Yi *et al.*, "Deep matrix factorization with implicit feedback embedding for recommendation system," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4591–4601, 2019, doi: 10.1109/tii.2019.2893714.

[4] J. Guerreiro and P. Rita, "How to predict explicit recommendations in online reviews using text mining and sentiment analysis," *Journal of Hospitality and Tourism Management*, vol. 43, pp. 269–272, 2020, doi: 10.1016/j.jhtm.2019.07.001.

[5] M. Karuppiah, H. R. Karimi, D. Malathi, V. Vijayakumar, R. Logesh, and V. Subramaniyaswamy, "Effective user preference mining-based personalised movie recommendation system," *International Journal of Computer Aided Engineering and Technology*, vol. 13, no. 3, p. 371, 2020, doi: 10.1504/ijcaet.2020.10029314.

[6] D. Datta, T. M. Navamani, and R. Deshmukh, "Products and movie recommendation system for social networking sites," *International Journal of Scientific & Technology Research*, vol. 9, pp. 10, pp. 262-270, 2020.

[7] N. Idrissi and A. Zellou, "A systematic literature review of sparsity issues in recommender systems," *Social Network Analysis and Mining*, vol. 10, no. 1, 2020, doi: 10.1007/s13278-020-0626-2.

[8] M. M. Reddy, R. S. Kanmani, and D. B. Surendiran, "Analysis of movie recommendation systems; with and without considering the low rated movies," *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, 2020, doi: 10.1109/ic-ETITE47903.2020.453.

[9] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data," *Expert Systems with Applications*, vol. 149, p. 113248, 2020, doi: 10.1016/j.eswa.2020.113248.

[10] A. Ochirbat *et al.*, "Hybrid occupation recommendation for adolescents on interest, profile, and behavior," *Telematics and Informatics*, vol. 35, no. 3, pp. 534–550, 2018, doi: 10.1016/j.tele.2017.02.002.

[11] B. K. Ye, Y. J. Tu, and T. P. Liang, "A hybrid system for personalized content recommendation," *Journal of Electronic Commerce Research*, vol. 20, no. 2, pp. 91–104, 2019.

[12] S. Kumar, K. De, and P. P. Roy, "Movie recommendation system using sentiment analysis from microblogging data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 915–923, 2020, doi: 10.1109/TCSS.2020.2993585.

[13] R. Agrawal and S. Prabakaran, "Big data in digital healthcare: lessons learnt and recommendations for general practice," *Heredity*, vol. 124, no. 4, pp. 525–534, Apr. 2020, doi: 10.1038/s41437-020-0303-2.

[14] T. Krismayer, M. Schedl, P. Knees, and R. Rabiser, "Predicting user demographics from music listening information," *Multimedia Tools and Applications*, vol. 78, no. 3, pp. 2897–2920, 2019, doi: 10.1007/s11042-018-5980-y.

[15] N. Pradeep, K. K. R. Mangalore, B. Rajpal, N. Prasad, and R. Shastri, "Content based movie recommendation system," *International Journal of Research in Industrial Engineering*, vol. 9, no. 4, pp. 337-348, 2020, doi: 10.22105/riej.2020.259302.1156.

[16] S. R. S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," *Smart Intelligent Computing and Applications*, pp. 391-397, 2019, doi: 10.1007/978-981-13-1927-3_42.

[17] P. Lops, D. Jannach, C. Musto, T. Bogers, and M. Koolen, "Trends in content-based recommendation: preface to the special issue on Recommender systems based on rich item descriptions," *User Modeling and User-Adapted Interaction*, vol. 29, no. 2, pp. 239–249, 2019, doi: 10.1007/s11257-019-09231-w.

[18] R. Singla, S. Gupta, A. Gupta, and D. K. Vishwakarma, "FLEX: a content based movie recommender," in *2020 International Conference for Emerging Technology, INCET 2020*, 2020, doi: 10.1109/INCET49848.2020.9154163.

[19] D. K. Behera, M. Das, S. Swetanisha, and P. K. Sethy, "Hybrid model for movie recommendation system using content K-nearest neighbors and restricted boltzmann machine," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 23, no. 1, pp. 445–452, 2021, doi: 10.11591/ijeecs.v23.i1.pp445-452.

[20] J. M. Musa and X. Zhihong, "Item based collaborative filtering approach in movie recommendation system using different similarity measures," *ACM International Conference Proceeding Series*. ACM, pp. 31–34, 2020, doi: 10.1145/3397125.3397148.

[21] J. Bobadilla, S. Alonso, and A. Hernando, "Deep learning architecture for collaborative filtering recommender systems," *Applied Sciences (Switzerland)*, vol. 10, no. 7, p. 2441, 2020, doi: 10.3390/app10072441.

[22] A. Gazdar and L. Hidri, "A new similarity measure for collaborative filtering based recommender systems," *Knowledge-Based Systems*, vol. 188, p. 105058, 2020, doi: 10.1016/j.knosys.2019.105058.

[23] C. K. Raghavendra and K. C. Srikantaiah, "Similarity based collaborative filtering model for movie recommendation systems," *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 2021, pp. 1143–1147, doi: 10.1109/ICICCS51141.2021.9432354.

[24] W. Yue, Z. Wang, W. Liu, B. Tian, S. Lauria, and X. Liu, "An optimally weighted user- and item-based collaborative filtering approach to predicting baseline data for Friedreich's Ataxia patients," *Neurocomputing*, vol. 419, pp. 287–294, 2021, doi: 10.1016/j.neucom.2020.08.031.

[25] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simulation Modelling Practice and Theory*, vol. 113, p. 102375, 2021, doi: 10.1016/j.simpat.2021.102375.
[26] S. Sharma, V. Rana, and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm," *Education and Information Technologies*, vol. 27, no. 2, pp. 1523–1538, 2022, doi: 10.1007/s10639-021-10643-8.
[27] B. J. D. Sitompul, O. S. Sitompul, and P. Sihombing, "Enhancement clustering evaluation result of davies-bouldin index with determining initial centroid of k-means algorithm," *Journal of Physics: Conference Series*, vol. 1235, no. 1, p. 12015, 2019, doi: 10.1088/1742-6596/1235/1/012015.
[28] J. Arora, K. Khatter, and M. Tushir, "Fuzzy c-means clustering strategies: a review of distance measures," *Advances in Intelligent Systems and Computing*, vol. 731, pp. 153–162, 2019, doi: 10.1007/978-981-10-8848-3_15.
[29] C. Selvi and E. Sivasankar, "A novel optimization algorithm for recommender system using modified fuzzy c-means clustering approach," *Soft Computing*, vol. 23, no. 6, pp. 1901–1916, 2019, doi: 10.1007/s00500-017-2899-6.
[30] R. Ahuja, A. Solanki, and A. Nayyar, "Movie recommender system using k-means clustering and k-nearest neighbor," in *Proceedings of the 9th International Conference on Cloud Computing, Data Science and Engineering, Confluence 2019*, 2019, pp. 263–268, doi: 10.1109/CONFLUENCE.2019.8776969.
[31] X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T. S. Chua, "Explainable reasoning over knowledge graphs for recommendation," in *Proceedings of the AAAI conference on artificial intelligence*, 2019, vol. 33, no. 01, pp. 5329–5336, doi: 10.1609/aaai.v33i01.33015329.
[32] A. Dash, A. Mukherjee, and S. Ghosh, "A network-centric framework for auditing recommendation systems," in *Proceedings - IEEE INFOCOM*, 2019, vol. 2019-April, pp. 1990–1998, doi: 10.1109/INFOCOM.2019.8737486.
[33] A. Ez-Zahout, H. Gueddah, A. Nasry, R. Madani, and F. Omary, "A hybrid big data movies recommendation model based k-nearest neighbors and matrix factorization," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 26, no. 1, pp. 434–441, 2022, doi: 10.11591/ijeecs.v26.i1.pp434-441.

BIOGRAPHIES OF AUTHORS

Tammanoon Panyatip is a Ph.D. student at Faculty of informatics, Mahasarakham University, Thailand. He is an assistant professor of information science at Kalasin University. His research includes recommendation systems and media for education. He can be contacted at email: tammanoon.pan@msu.ac.th.

Manasawee Kaenampornpan received Ph.D. degree in computer science from University of Bath, UK. She is an assistant professor of computer science at Mahasarakham University, Thailand. Her research interests are user experience design, context awareness, mobile and ubiquitous computing. She can be contacted at email: manasawee.k@msu.ac.th.

Phatthanaphong Chomphuwiset is a full-time lecturer at the department of computer science, Mahasarakham University, Thailan. He received a Ph.D. in computing from the University of Leeds, UK. His main research interests fall into a variety of areas, including medical image processing, computer vision applications, and machine learning. One of his current researches is to investigate techniques in machine learning and deep learning to process histopathological images. He tries to study a computational technique that can cope with the challenge of histopathology image analysis (HIA) where the images pose important challenges such as image size and resolution and the variations of the images due to the preparation process. He can be contacted at email: phatthanaphong.c@msu.ac.th.