An Efficient Recommender System Based on Collaborative Filtering Recommendation and Cluster Ensemble

Hafed Zarzour[1], Faiz Maazouzi[1], Mohammad Al-Zinati[2], Yaser Jararweh[2], Thar Baker[3]
[1]*University of Souk Ahras, Souk Ahras, Algeria*
hafed.zarzour@gmail.com, mazouzi@labged.net
[2]*Jordan University of Science and Technology, Irbid, Jordan*
yijararweh@just.edu.jo, mhzinati@just.edu.jo
[3]*University of Sharjah, Sharjah, UAE*
tshamsa@sharjah.ac.ae

2021 Eighth International Conference on Social Network Analysis, Management and Security (SNAMS) | 978-1-6654-9495-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/SNAMS53716.2021.9732118

Abstract— In the last few years, cluster ensembles have emerged as powerful techniques that integrate multiple clustering methods into recommender systems. Such integration leads to improving the performance, quality and the accuracy of the generated recommendations. This paper proposes a novel recommender system based on a cluster ensemble technique for big data. The proposed system incorporates the collaborative filtering recommendation technique and the cluster ensemble to improve the system performance. Besides, it integrates the Expectation-Maximization method and the HyperGraph Partitioning Algorithm to generate new recommendations and enhance the overall accuracy. We use two real-world datasets to evaluate our system: TED Talks and MovieLens. The experimental results show that the proposed system outperforms the traditional methods that utilize single clustering techniques in terms of recommendation quality and predictive accuracy. Most importantly, the results indicate that the proposed system provides the highest precision, recall, accuracy, F1, and the lowest Root Mean Square Error regardless of the used similarity strategy.

Keywords—recommender system, collaborative filtering recommendation, recommender system for big data, cluster ensemble, Expectation Maximization, EM

I. Introduction

With the advent of big data, huge volume and high heterogeneity in the domains, sources, representations, and information structuring are notable characteristics of current Web information systems. Given these characteristics, tailoring Web information systems to adapt to the individual user needs and relevant preferences represents a challenge for constructing such systems. The continuous evolution of Web information makes the problem further complicated and harder to address.

Clustering-based Collaborative Filtering (CF) algorithms are used to create a set of clusters with highly related data from a given data set [1, 2]. These clusters are then used to calculate relevant recommendations. As opposed to using the entire data set, the use of the related data within clusters helps optimize the process of calculating recommendations. The online recommendations of such algorithms are based on a prediction model created and trained using a portion of data during the offline stage.

Expectation-Maximization (EM), k-means, and Self-Organizing Maps (SOM) are examples of single clustering algorithms that have been successfully applied to the CF domain.

Cluster ensemble has proven to be a compelling alternative for single clustering algorithms for enhancing the quality of predictions and recommendations. Due to its ability to combine recommendations of multiple clustering techniques, the cluster ensemble usually outperforms single clustering solutions in terms of robustness, consistency, novelty, and stability [3]. The cluster ensemble method is widely used in data mining [4-6]. However, it is still not commonly used in the recommender systems domain [7-9]. Furthermore, to the best of our knowledge, the effects of using different similarity functions on improving prediction accuracy are still not investigated in the field of recommender systems.

This paper proposes a recommender system based on a collaborative filtering recommendation and cluster ensemble model. The proposed system incorporates multiple similarity metrics to form clusters. Additionally, the proposed system utilizes the HyperGraph Partitioning Algorithm (HGPA) for combining the results of the Expectation-Maximization (EM) clustering strategy. We evaluate the proposed system using two real-world datasets, where the first is obtained from the TED website [3] that contains a repository of lecture recordings given by prominent speakers [10], whilst the second is MovieLens provided by the MovieLens Project. The experimental results show that the proposed system outperforms the traditional methods that utilize single clustering techniques in terms of recommendation quality and predictive accuracy.

The rest of this paper is structured as follows: Section 2 discusses the related work. Section 3 presents the research methodology along with the experimental process. In Section 4, we discuss our experimentation settings. Finally, we present the experimental results and findings and summarize our conclusions in Section 5.

II. Background and Related Work

A. Expectation-Maximization (EM) Clustering

EM clustering is one of the essential model-based clustering strategies to find the unknown parameters within a statistical system [11, 12]. Typically, EM alternates between two main steps. The first step comprises producing an expectation to estimate unobserved data. The second step involves computing parameters that maximize the likelihood of complete data [13]. EM is considered an iterative technique and has the advantage of being simple, efficient, and easy to implement.

Researchers have investigated the application of EM clustering in the recommender systems domain. For example, Hofmann [14] developed an approach for mining user data based on a collaborative filtering method and EM clustering. The author evaluated the proposed model on the EachMovie dataset. The experimental results showed a substantial improvement in accuracy compared to other existing approaches.

Nilashi *et al.* [15] propose a new recommendation method using EM and regression to improve the accuracy of the multi-criteria recommendation systems. Their method employed the principal component analysis to reduce the data dimension and address multicollinearity problems. They used Yahoo! Movie and TripAdvisor datasets for their experiments. The results obtained from their proposed method showed a significant improvement in terms of the predictive accuracy related to multi-criteria collaborative filtering.

Later on, Nilashi *et al.* [16] investigate dimensionality reduction and prediction techniques for developing recommender systems. To this end, they proposed using multi-criteria collaborative filtering in the tourism domain. For this sake, they developed a hybrid recommendation model that includes EM clustering, Adaptive Neuro-Fuzzy Inference System, and the principal component analysis to improve the predictive accuracy. Their experimental results on the TripAdvisor dataset demonstrate the improvement in the predictive accuracy of tourism recommendations.

B. Cluster Ensemble

Clustering techniques aim at partitioning data into a set of groups that include the most similar elements. Although the literature is rich with many clustering solutions, none of these solutions proves to be generic and applies to all cases. Therefore, ensemble clustering has emerged as an alternative solution to achieve a better consensus clustering result by combining different clustering methods.

Tsai *et al.* [17] evaluated the applicability of the cluster ensemble methods in the context of collaborative filtering recommender systems. The proposed approach employed k-means and Self-Organizing Maps (SOM) as baseline clustering techniques. Besides, they used the HyperGraph Partitioning Algorithm (HGPA), the cluster-based similarity partitioning algorithm (CSPA), and majority voting as ensemble approaches. The authors evaluate their approach using MovieLens dataset. In [7], Zheng *et al.* propose a new system for recommending articles based on ensemble hierarchical clustering. Nilashi *et al.* [18] propose a multi-criteria collaborative filtering recommender system using prediction machine learning techniques and cluster ensembles. To improve the accuracy of online recommendations, they employed EM and SOM for data clustering and HGPA for ensemble clustering. The authors used the TripAdvisor dataset to evaluate their system. The experimental results for all of the studies mentioned above confirmed that the clustering ensemble methods provide better recommendation accuracy and precision in comparison to the baseline single clustering techniques.

III. The Recommender System Design

In this section, we discuss the design of our recommender system. We start by describing the datasets used to evaluate the model. Then, we provide a general overview of the proposed recommender system approach. After that, we discuss the EM clustering approach incorporated within our system. Finally, we present the used cluster ensemble method and the final recommendation system process.

A. Dataset Description

Our experiments were carried out on two real-world datasets, TED Talks, and MovieLens 100k. TED Talks dataset is a well-known and widely used dataset for recommendations collected by the NLP team at the Idiap Research Institute (www.idiap.ch/dataset/ted) [19]. The team obtained the metadata by crawling the official TED website (www.ted.com). The metadata contains two types of entry: users and talks (lectures). The users are the visitors of the TED website who have profiles and favorite lists of lectures [20]. The TED Talks is a collection of relatively short speeches covering a wide variety of topics, and for which high-quality manual transcriptions and translations into many languages are available [21]. The dataset contains threads that store a certain number of users' comments for each talk. The dataset includes 1,150 lectures given by 960 speakers and about 2,427 participants who made 12 or more ratings. The dataset MovieLens 100k is made public by Grouplens Cooperation (http://grouplens.org/datasets/movielens). The cooperation collected this dataset over various periods from the MovieLens website. The dataset contains 943 users, 1682 items, and 100.000 ratings ranging from 1 to 5. Users who had less than 20 ratings are removed from the dataset.

B. General Overview

We present the recommender system process using TED Talks dataset for clarity purposes. It is important to note that the same approach can be applied to other datasets representing users' choices.

Figure 1 illustrates the general process of the proposed recommender system. Initially, we construct a TED user-item matrix from two primary sources: 1) "Talks" or "Lectures", which include data about each talk such as identifier, title, short description, related tags, and the total number of views; 2) "Users" which contain data about each user, such as identifier and favorites list, representing the talks they like. We express the users' preferences in the TED user-item matrix using a binary rating. We use "1" to represent a favorite lecture and "0" for a lecture that is not liked or seen. After creating the TED user-item matrix, we compute the TED user similarity matrices using two standard metrics, namely Pearson Correlation Coefficient (PCC) and Cosine (COS) similarity. Then, we randomly split the dataset into two sets: the training set composed of (80%) of rated TED talks and the test set containing the remaining (20%). After that, we train and test the EM and k-means clustering techniques two times using the PCC metric and Cosine similarity.

Finally, we use the HGPA method to generate new recommendation results by combining multiple EM clusters for the clustering ensemble. Different ensemble sizes are considered to decide the clustering combination that produces the best recommendation results.

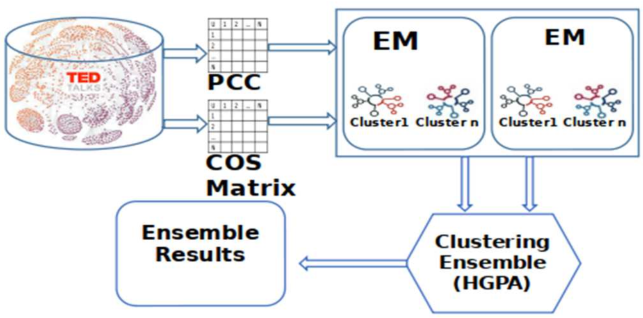

Fig. 1 A general overview of the recommender system process

C. *Expectation-Maximization (EM)*

The Expectation-Maximization EM algorithm is an iterative approach that relies on a parametric estimation technique within the general framework of maximum likelihood [22]. It describes the distribution of data in automatic clustering [23]. Also, it defines each cluster by a normal distribution law, parameterized by its center of gravity and its variance-covariance matrix. The EM algorithm is best known for estimating the parameters of the elementary distributions. It aims to maximize the log-likelihood for the data sample given a pre-defined cluster number. The Membership is used to define association with a group according to the parameterization $\emptyset$. In this case, the log-likelihood depends on the parameterization $\emptyset$ according to the association of the hidden variable Z.

Algorithm 1. EM clustering algorithm

Input: **k:** the number of clusters, **M:** matrix "User X User"

Output: **k** clusters

Repeat

E-STEP: For unobserved variables we calculate their expected values, we assume current parameter values

M-STEP: We calculate new parameter values for maximizing the data probability, both observed and estimated

Until a fixed point [of Q] is obtained

As illustrated in Algorithm 1, the EM algorithm iteratively cycles between two phases. The "Expectation" phase, often called "E-step", attempts to estimate the missing data using the observed data and the parameter value determined at the previous iteration. Therefore, during step E, the algorithm calculates the expectation of associations to groups $h_i^t \equiv P(G_i \vee x^t, \emptyset)$ with current $\emptyset$ parameterization.

The "Maximization" phase, often named "M-step", tries to maximize the likelihood function that is refined in each iteration by the E-step and updates the value of the parameter $\emptyset$ for the next iteration. In this step, we obtain a new parametrization $\emptyset'^{+1}$ to maximize the likelihood estimate $Q(\emptyset \vee \emptyset')$:

$$Q(\emptyset|\emptyset') = E[L(\emptyset \vee X, Z) \vee X, \emptyset'](1)$$

$$\emptyset'^{+1} = argmaxQ(\emptyset|\emptyset')(2)$$

D. *Cluster ensemble with HyperGraph Partitioning Algorithm (HGPA)*

The cluster ensemble approach combines multiple clustering baselines with the same elements into a unique consolidated clustering method [1]. Meta-Clustering Algorithm (MCLA), the cluster-based similarity partitioning algorithm (CSPA), and the HyperGraph Partitioning Algorithm (HGPA) are famous examples of cluster ensembles [24]. In our system, we use the HGPA as a cluster ensemble method for combining the results of Expectation-Maximization and k-means. To this effect, we use HMETIS, a hypergraph partitioning package, to decide the clustering combination that produces the best recommendation results.

IV. Experimental Results

A. *Evaluation Metrics*

To evaluate the quality of the predictions, we calculate the accuracy and Root Mean Square Error (RMSE). Accuracy aims to select high-quality items from all items set. RMSE puts more emphasis on more significant absolute errors. These metrics are defined as follows:

$$Accuracy = \frac{True_p + True_n}{True_p + False_p + False_n + True_n} (3)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(p_{u,i} - r_{u,i}\right)^2 (4)}$$

where the variable $p_{u,i}$ represents the predicted rating for a given user u on item i, the variable $r_{u,i}$represents the current rating, and N represents the total number of ratings on the set of items.

B. *Expectation-Maximization (EM) Clustering*

Table 1 presents the accuracy and the RMSE results for the proposed system using the PCC and COS similarity on TED Talks. Table 2 shows the same results using the MovieLens dataset. For the EM clustering algorithm, we use 3, 5, 7, 9, 11, 13, 15, 17, 19, and 21 clusters. We use PCC and COS methods to calculate the similarity for each dataset and each number of clusters.

As can be noticed by the given results, the EM approach performs better when using the COS similarity for most of the considered evaluation metrics on the two datasets. The improvement becomes more apparent when using a larger number of clusters. For instance, with TED Talks dataset, the accuracy for EM with PCC using 21 clusters is 0.9312. The accuracy for EM with COS using 21 clusters is 0.9388. For the same setting, the RMSE with PCC is 0.0959, while the RMSE with COS is 0.0891.

TABLE 1 PERFORMANCE RESULTS FOR EM BY PCC AND COS SIMILARITY (TED DATASET).

EM	Method	Accuracy	RMSE
3 clusters	PCC	0.7674	0.5973
	COS	0.7592	0.5901
5 clusters	PCC	0.7936	0.4534
	COS	0.7793	0.4544
7 clusters	PCC	0.8126	0.4478
	COS	0.8239	0.4323
9 clusters	PCC	0.8536	0.3909
	COS	0.8653	0.3782
11 clusters	PCC	0.8987	0.3023
	COS	0.9011	0.2878
	PCC	0.9128	0.2345

13 clusters	COS	0.9186	0.2233
15 clusters	PCC	0.9213	0.1802
	COS	0.9234	0.1792
17 clusters	PCC	0.9372	0.0937
	COS	0.9337	0.0967
19 clusters	PCC	0.9378	0.1029
	COS	0.9353	0.0987
21 clusters	PCC	0.9312	0.0959
	COS	0.9388	0.0891
Best	PCC	0.9372	0.0937
	COS	0.9337	0.0967

TABLE 2 PERFORMANCE RESULTS FOR EM BY PCC AND COS SIMILARITY (MOVIELENS DATASET).

EM	Method	Accuracy	RMSE
3 clusters	PCC	0.7934	0.5173
	COS	0.7934	0.5151
5 clusters	PCC	0.7956	0.345
	COS	0.7923	0.3578
7 clusters	PCC	0.8012	0.3569
	COS	0.8134	0.3356
9 clusters	PCC	0.8147	0.3467
	COS	0.8134	0.3234
11 clusters	PCC	0.8234	0.3133
	COS	0.8191	0.3213
13 clusters	PCC	0.8432	0.2213
	COS	0.8398	0.2486
15 clusters	PCC	0.8512	0.1734
	COS	0.8465	0.1713
17 clusters	PCC	0.8798	0.0932
	COS	0.8743	0.0944
19 clusters	PCC	0.8881	0.0829
	COS	0.8763	0.0863
21 clusters	PCC	0.9014	0.0755
	COS	0.8979	0.0769
Best	PCC	0.9014	0.0755
	COS	0.8979	0.0769

C. *EM Ensemble*

Tables 3 and 4 present the results of the HGPA-based EM ensemble by using PCC and COS similarities on the TED and MovieLens datasets, respectively. The used ensemble size ranges from 2 to 9, and the number of clusters varies between 3 and 19. As is noticed, the HGPA-based EM ensemble by using PCC obtains the highest values of accuracy (0.9683) and the lowest RMSE value (0.0802) in comparison to the HGPA-based EM ensemble using COS similarity in the case of TED dataset. We can also observe similar results when using the MovieLens dataset. Therefore, by PCC HGPA-based EM ensemble performs better than COS similarity.

TABLE 3 PERFORMANCE RESULTS FOR EM ENSEMBLE BY HGPA USING PCC AND COS SIMILARITY (TED DATASET).

Ensemble size	Method	Ensemble Technique	Accuracy	RMSE
2 (k = 3, 5)	PCC	HGPA	0.8012	0.5234
	COS	HGPA	0.7986	0.5323
3 (k =3, 5, 7)	PCC	HGPA	0.8129	0.4094
	COS	HGPA	0.802	0.3947
4 (k =3, 5, 7, 9)	PCC	HGPA	0.8545	0.2877
	COS	HGPA	0.8492	0.2832
5 (k=3, 5, 7, 9, 11)	PCC	HGPA	0.8635	0.2023
	COS	HGPA	0.8843	0.2037
6 (k=3, 5, 7, 9, 11, 13)	PCC	HGPA	0.9268	0.1323
	COS	HGPA	0.9337	0.1324
7 (k = 3, 5, 7, 9, 11, 13, 15)	PCC	HGPA	0.968	0.0963
	COS	HGPA	0.9618	0.1056
8 (k = 3, 5, 7, 9, 11, 13, 15,17)	PCC	HGPA	0.9683	0.0802
	COS	HGPA	0.9612	0.0833
9 (k = 3, 5, 7, 9, 11, 13, 15, 17, 19)	PCC	HGPA	0.9556	0.0913
	COS	HGPA	0.9471	0.0928
Best	PCC	HGPA	0.9683	0.0802
	COS	HGPA	0.9655	0.0833

TABLE 4 PERFORMANCE RESULTS FOR EM ENSEMBLE BY HGPA USING PCC AND COS SIMILARITY (MOVIELENS DATASET).

Ensemble size	Method	Ensemble Technique	Accuracy	RMSE
2 (k = 3, 5)	PCC	HGPA	0.8022	0.4233
	COS	HGPA	0.8011	0.4134
3 (k =3, 5, 7)	PCC	HGPA	0.8234	0.3894
	COS	HGPA	0.8021	0.3927
4 (k =3, 5, 7, 9)	PCC	HGPA	0.8535	0.2671
	COS	HGPA	0.8431	0.2845
5 (k=3, 5, 7, 9, 11)	PCC	HGPA	0.8611	0.1934
	COS	HGPA	0.8822	0.1825
6 (k=3, 5, 7, 9, 11, 13)	PCC	HGPA	0.9168	0.1233
	COS	HGPA	0.9227	0.1124
7 (k = 3, 5, 7, 9, 11, 13, 15)	PCC	HGPA	0.9681	0.0911
	COS	HGPA	0.9567	0.1123
8 (k = 3, 5, 7, 9, 11, 13, 15,17)	PCC	HGPA	0.9693	0.0731
	COS	HGPA	0.9633	0.0815
9 (k = 3, 5, 7, 9, 11, 13, 15, 17, 19)	PCC	HGPA	0.9356	0.0934
	COS	HGPA	0.9411	0.0966
Best	PCC	HGPA	0.9681	0.0731
	COS	HGPA	0.9633	0.0815

D. Further Comparisons

Figures 2-6 depict the best values of precision, recall, F1, accuracy, and RMSE for all recommendation methods on TED Talks and MovieLens datasets. The figures show that HGPA-based EM and k-means ensembles outperform the baselines techniques regardless of the used similarity strategy. We can also notice that the HGPA-based EM ensemble using PCC provides the highest precision, recall, accuracy, F1, and the lowest value of RMSE on both datasets.

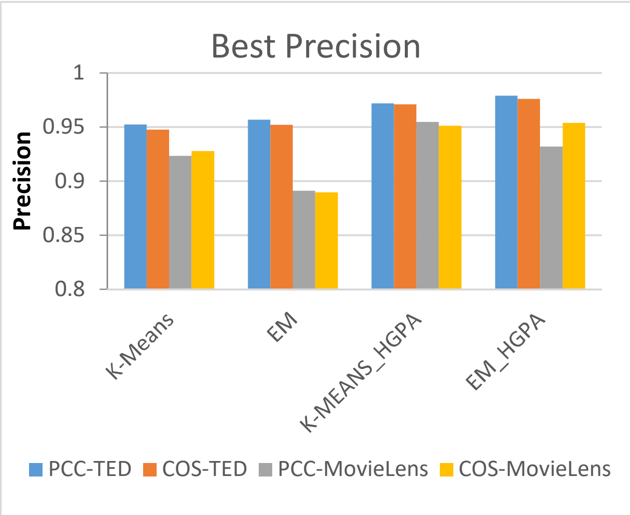

Fig. 2 Best precision results for all methods using PCC and COS similarity.

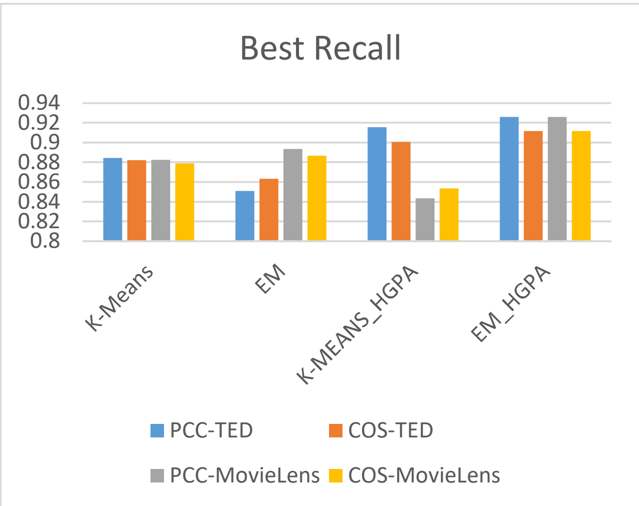

Fig. 3 Best recall results for all methods using PCC and COS similarity.

V. CONCLUSION AND DISCUSSION

Improving the performance of recommendations when building recommender systems is a challenging task. Hence, in this study, we attempted to consider this challenge to generate accurate recommendations efficiently. Instead of using a single clustering, we suggested using a cluster ensemble that combines multiple clustering solutions with two different similarity functions. Two datasets, MovieLens and TED lectures extracted from the TED website, were used to assess the effectiveness of the proposed method. Accordingly, we use precision, recall, and F1 score to evaluate the recommendation quality, while we use accuracy and RMSE to assess the prediction quality.

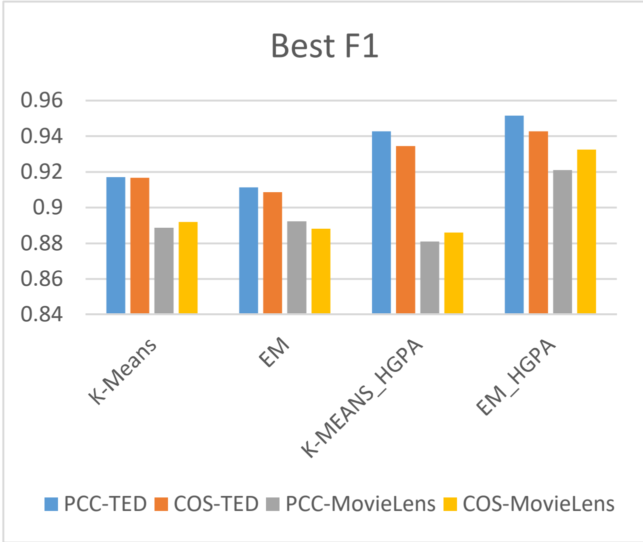

Fig. 4 Best F1 results for all methods using PCC and COS similarity.

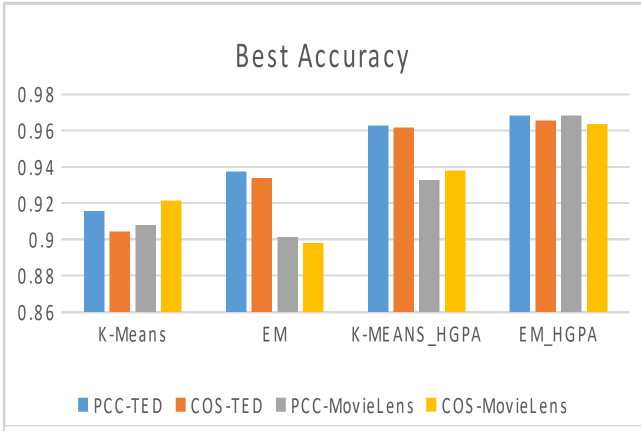

Fig. 5 Best accuracy results for all methods using PCC and COS similarity.

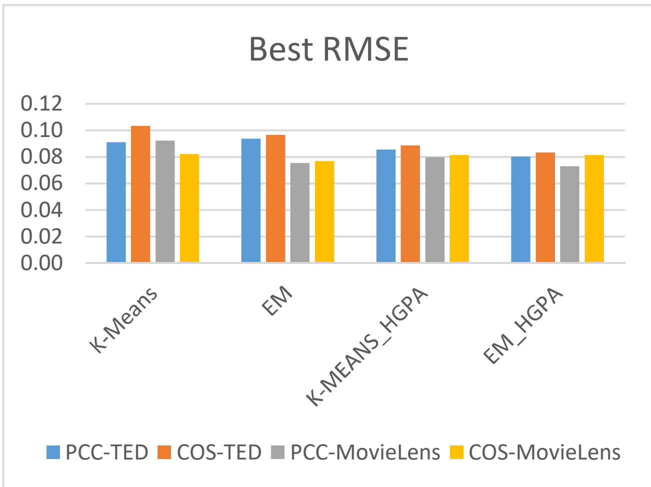

Fig. 6 Best RMSE results for all methods using PCC and COS similarity.

Concerning the quality of recommendations, our method significantly improves the quality of the generated recommendations compared to all traditional single clustering methods. Additionally, we found that using the PCC similarity in the HGPA-based EM ensemble results in the best values of the considered evaluation metrics. For the quality of prediction, our method also significantly enhances the predictive accuracy of the recommender system in comparison to other methods. Moreover, the results indicate that the prediction quality achieved its best value with a HGPA-based EM ensemble that applies PCC as a similarity measure. The HGPA-based k-means ensemble outperforms both single EM and k-means clustering methods, regardless of the used similarity measure. However, the HGPA-based EM ensemble outperformed all considered methods, including the HGPA-based k-means ensemble. Overall, the experimental results obtained in the present research are in agreement with the previous findings presented in [22, 25]. Both results indicate that the cluster ensembles in recommender systems are effective and can provide a better quality of recommendations and predictions than methods that rely on single clustering techniques.

REFERENCES

[1] H. Zarzour, Y. Jararweh, and Z. A. Al-Sharif, “An Effective Model-Based Trust Collaborative Filtering for Explainable Recommendations,” 2020 11th International Conference on Information and Communication Systems (ICICS), Apr. 2020.

[2] H. Zarzour, Al shboul Bashar, M. Al-Ayyoub, and Y. Jararweh, “A convolutional neural network-based reviews classification method for explainable recommendations,” 2020 Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), Dec. 2020.

[3] S. VEGA-PONS and J. RUIZ-SHULCLOPER, “A Survey Of Clustering Ensemble Algorithms,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 25, no. 03, pp. 337–372, May 2011.

[4] N. Iam-On and T. Boongoen, “Diversity-driven generation of link-based cluster ensemble and application to data classification,” Expert Systems with Applications, vol. 42, no. 21, pp. 8259–8273, Nov. 2015.

[5] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Y. Yu, and Z. Chen, “Scalable collaborative filtering using cluster-based smoothing,” Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’05, 2005.

[6] J. Hu, T. Li, C. Luo, H. Fujita, and Y. Yang, “Incremental fuzzy cluster ensemble learning based on rough set theory,” Knowledge-Based Systems, vol. 132, pp. 144–155, Sep. 2017.

[7] L. Zheng, L. Li, W. Hong, and T. Li, “PENETRATE: Personalized news recommendation using ensemble hierarchical clustering,” Expert Systems with Applications, vol. 40, no. 6, pp. 2127–2136, May 2013.

[8] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, X.-Z. Gao, and V. Indragandhi, “A hybrid quantum-induced swarm intelligence clustering for the urban trip recommendation in smart city,” Future Generation Computer Systems, vol. 83, pp. 653–673, Jun. 2018.

[9] J. Guan, R.-Y. Li, and J. Wang, “GRACE: A Graph-Based Cluster Ensemble Approach for Single-Cell RNA-Seq Data Clustering,” IEEE Access, vol. 8, pp. 166730–166741, 2020.

[10] C.-C. Hu and H.-L. Li, “Developing navigation graphs for TED talks,” Computers in Human Behavior, vol. 66, pp. 26–41, Jan. 2017.

[11] G. Feng, C. Lai, and N. C. Kar, “Expectation-Maximization Particle-Filter- and Kalman-Filter-Based Permanent Magnet Temperature Estimation for PMSM Condition Monitoring Using High-Frequency Signal Injection,” IEEE Transactions on Industrial Informatics, vol. 13, no. 3, pp. 1261–1270, Jun. 2017.

[12] H.-D. Qin, X. Yu, Z.-B. Zhu, and Z.-C. Deng, “An expectation-maximization based single-beacon underwater navigation method with unknown ESV,” Neurocomputing, vol. 378, pp. 295–303, Feb. 2020.

[13] G. D. Evangelidis and R. Horaud, “Joint Alignment of Multiple Point Sets with Batch and Incremental Expectation-Maximization,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1397–1410, Jun. 2018.

[14] T. Hofmann, “Latent semantic models for collaborative filtering,” ACM Transactions on Information Systems, vol. 22, no. 1, pp. 89–115, Jan. 2004.

[15] Nilashi M., Esfahani M. D, Roudbaraki M. Z, Ramayah T, Ibrahim O (2016) A multicriteria collaborative filtering recommender system using clustering and regression

techniques. Journal of Soft Computing and Decision Support Systems: 3(5): 24-30.

[16] M. Nilashi, O. bin Ibrahim, N. Ithnin, and N. H. Sarmin, “A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS,” Electronic Commerce Research and Applications, vol. 14, no. 6, pp. 542–562, Oct. 2015.

[17] C.-F. Tsai and C. Hung, “Cluster ensembles in collaborative filtering recommendation,” Applied Soft Computing, vol. 12, no. 4, pp. 1417–1425, Apr. 2012.

[18] M. Nilashi, K. Bagherifard, M. Rahmani, and V. Rafe, “A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques,” Computers & Industrial Engineering, vol. 109, pp. 357–368, Jul. 2017.

[19] N. Pappas and A. Popescu-Belis, “Combining content with user preferences for TED lecture recommendation,” 2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI), Jun. 2013.

[20] N. Pappas and A. Popescu-Belis, “Sentiment analysis of user comments for one-class collaborative filtering over ted talks,” Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, Jul. 2013.

[21] F. Maazouzi, H. Zarzour, and Y. Jararweh, “An Effective Recommender System Based on Clustering Technique for TED Talks,” International Journal of Information Technology and Web Engineering, vol. 15, no. 1, pp. 35–51, Jan. 2020.

[22] C. LIU, “MAXIMUM LIKELIHOOD ESTIMATION FROM INCOMPLETE DATA VIA EM-TYPE ALGORITHMS,” Advanced Medical Statistics, pp. 1051–1071, Nov. 2003.

[23] D. Castillo-Barnes, F. J. Martinez-Murcia, J. Ramírez, J. M. Górriz, and D. Salas-Gonzalez, “Expectation–Maximization algorithm for finite mixture of α-stable distributions,” Neurocomputing, vol. 413, pp. 210–216, Nov. 2020.

[24] A. Strehl, J. Ghosh, "Cluster ensembles---a knowledge reuse framework for combining multiple partitions", Journal of machine learning research, vol. 3, Dec, p. 583-617, 2002.

[25] Faqeeh, Mosab, Nawaf Abdulla, Mahmoud Al-Ayyoub, Yaser Jararweh, and Muhannad Quwaider. "Cross-lingual short-text document classification for Facebook comments." In 2014 International Conference on Future Internet of Things and Cloud, pp. 573-578. IEEE, 2014.

[26] Obaidat, I., Mohawesh, R., Al-Ayyoub, M., Mohammad, A.S. and Jararweh, Y., 2015, November. Enhancing the determination of aspect categories and their polarities in arabic reviews using lexicon-based approaches. In 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1-6). IEEE.