A Clustering Approach for Collaborative Filtering Under the Belief Function Framework

Raoua Abdelkhalek(✉), Imen Boukhris, and Zied Elouedi

LARODEC, Institut Supérieur de Gestion de Tunis,
Université de Tunis, Tunis, Tunisia
abdelkhalek_raoua@live.fr, imen.boukhris@hotmail.com, zied.elouedi@gmx.fr

Abstract. Collaborative Filtering (CF) is one of the most successful approaches in Recommender Systems (RS). It exploits the ratings of similar users or similar items in order to predict the users' preferences. To do so, clustering CF approaches have been proposed to group items or users into different clusters. However, most of the existing approaches do not consider the impact of uncertainty involved during the clusters assignments. To tackle this issue, we propose in this paper a clustering approach for CF under the belief function theory. In our approach, we involve the Evidential C-Means to group the most similar items into different clusters and the predictions are then performed. Our approach tends to take into account the different memberships of the items clusters while maintaining a good scalability and recommendation performance. A comparative evaluation on a real world data set shows that the proposed method outperforms the previous evidential collaborative filtering.

Keywords: Collaborative filtering · Belief function theory · Clustering · Evidential C-Means

1 Introduction

During the last few years, Recommender Systems (RS) [1] have attracted considerable attention from several research communities and have reached a high level of popularity. The diversity of the information sources and the variety of domain applications gave birth to various recommendation approaches. According to the literature, CF is considered to be the most popular and the widely used approach in this area [1–3]. In order to provide recommendations, CF tends to predict the users' preferences based on the users or the items sharing similar ratings. To do so, this latter exploits the user-item matrix and computes the similarities between users (user-based [4]) or items (item-based [5]) in the system. Based on the computed similarities, the prediction process is then performed. CF has achieved widespread success in both academia and industry [2]. Despite its simplicity and efficiency, CF approach exhibits some limitations such as the scalability problems [6]. Actually, CF needs to search the whole user-item space in order to compute similarities. This computation increases with the number of items and users leading to poor scalability performance. To overcome the problem mentioned above, several recommendation approaches have

A. Antonucci et al. (Eds.): ECSQARU 2017, LNAI 10369, pp. 169–178, 2017.
DOI: 10.1007/978-3-319-61581-3_16

been proposed using different model-based techniques such as Bayesian network [7], Singular Value Decomposition (SVD) [8] and clustering techniques [6,9,10]. The common point of these approaches is to forecast pre-trained models using an item-user matrix. For instance, in clustering CF approaches, items can be assigned to clusters based on their historical ratings and recommendations are performed accordingly. However, an item may potentially belong to more than only one cluster. This concept is referred to as soft clustering. This imprecision may impact the relationship between the items and therefore the final prediction. Indeed, we show in a previous work [11] the relevance of handling uncertainty in CF throughout the prediction process. In this paper, we treat uncertainty involved in the clustering CF approaches where we consider the cluster membership of each item to be uncertain. To this end, we opt for the belief function theory (BFT) [12–14] which offers a rich representation about all situations ranging from complete knowledge to complete ignorance. Several clustering methods have been proposed under this theory. For example, the belief K-modes (BKM) has been proposed by [15] to deal with uncertainty in the attribute values. On the other hand, the Evidential C-Means (ECM) [16] has been conceived to handle uncertainty for objects' assignment. Since we are in particular interested in assessing the uncertainty in items cluster membership, we involve the Evidential C-Means method which is based on the concept of credal partition. Taking advantage of the BFT in particular the ECM technique, we propose an evidential clustering CF. The new approach allows us to assign the items to soft clusters whilst handling challenges imposed from the CF framework.

This paper is organized as follows: Sect. 2 recalls the basic concepts of the belief function theory and the Evidential C-Means. Section 3 presents briefly some related works on clustering CF as well as CF under the belief function framework. Our proposed recommendation approach is presented in Sect. 4. Section 5 exposes the experimental results conducted on a real world data set. Finally, the paper is concluded and some future works are depicted in Sect. 6.

2 Clustering in a Belief Function Framework

The BFT [12–14] represents a flexible and rich framework for reasoning under uncertainty. In this section, we provide an overview about its basic concepts and we recall the Evidential C-Means [16] as a clustering method under an uncertain framework.

2.1 Belief Function Theory

In the BFT, a problem domain is represented by the frame of discernment Θ. The belief committed to each element of Θ is expressed by a basic belief assignment (bba) which is a mapping function $m : 2^{\Theta} \rightarrow [0, 1]$ such that: $\sum_{A \subseteq \Theta} m(A) = 1$

Each mass $m(A)$ quantifies the degree of belief exactly assigned to the event A of Θ. The subsets A of Θ such as $m(A) > 0$ are called focal elements.

To make decisions, beliefs can be represented by pignistic probabilities defined as:

$$BetP(A) = \sum_{B \subseteq \Theta} \frac{|A \cap B|}{|B|} \frac{m(B)}{(1 - m(\varnothing))} \quad for\ all\ A \in \Theta \tag{1}$$

2.2 Evidential C-Means

The Evidential C-Means (ECM) [16] is a clustering technique based on the concept of credal partition. Given an object i, this method determines the mass m_{ij} representing partial knowledge regarding the cluster membership to any subset A_j of $\Theta = \{\omega_1, \omega_2, \ldots, \omega_n\}$ where n is the number of clusters. Every partition is represented by a center $v_k \in \mathbb{R}^p$ where p is the dimension of data. Each subset A_j of Θ is represented by the barycenter v_j of the centers v_k associated to the clusters composing A_j. The barycenter is computed as follows:

$$v_j = \frac{1}{c_j} \sum_{k=1}^{c} s_{kj} v_k \tag{2}$$

where $c_j = |A_j|$ denotes the cardinal of A_j and s_{kj} is defined as follows:

$$s_{kj} = \begin{cases} 1 & \text{if } \omega_k \in A_j \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The distance between an object i and any subset A_j of Θ is defined by:

$$d_{ij} = \|x_i - v_j\| \tag{4}$$

Finally, the credal partition is determined by minimizing the following objective function:

$$J_{ECM} = \sum_{i=1}^{n} \sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Theta\}} c_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^{\beta} \tag{5}$$

α, β and δ are the input parameters such that $\alpha \geq 0$ is a weighting exponent for cardinality. $\beta > 1$ is a weighting exponent controlling the hardness of the partition and δ represents the distance between all instances and the empty set. More details about parameters and credal partition process can be found in [16].

3 Related Work on Collaborative Filtering

CF has shown a great applicability in a wide variety of domains [2]. The key idea is that if two users rated some items similarly or had similar behaviors in the past then, they would rate or act on other items similarly. CF approaches are divided into two categories namely, memory-based and model-based. Memory-based CF approaches exploit the whole user-item matrix to find similar users or items and generate recommendations accordingly. In contrast, model-based algorithms rely

on the ratings matrix to infer a model which is then applied for predictions. The model building process can be performed using different methods. For example, Bayesian networks have been used in [7] for CF process. Clustering CF approaches that are based on a cluster model to reduce the time complexity have also been proposed [6,9,10,17]. In [6,17], authors have proposed a clustering approach for CF that classifies the users in different groups and neighborhood has been selected for each cluster. In [9], the users have been clustered from the views of both rating patterns and social trust relationships. Similarly, a CF approach has been implemented in [10] based on user's preferences clustering. All the clustering techniques mentioned above focus on user-based CF. In our work, we consider only item-based CF where items are clustered into groups rather than users. It is obvious that developing RSs that can quickly produce high quality recommendations have become more and more required in this area [6]. On the other hand, considering uncertainty during the recommendation process can be argued to be another important challenge in real-world problems [18]. The belief function theory [12–14] is among the most widely used ones for dealing with uncertainty. Recent studies have investigated the benefits of the adoption of such theory in RSs area. In fact, authors in [19] have represented the user's preferences through the BFT tools and integrate context information for predicting all unprovided ratings. Another approach developed in [20] relies on this theory to represent both user's preferences and community preferences extracted from social networks. The authors in [11] have proposed an evidential item-based CF where they considered the similar items as different pieces of evidence. They computed the similarities between the target item and the whole items in the system and the final prediction was an aggregation of the ratings corresponding to the similar items. However, a lot of heavy computations are needed in this case. This problem is referred to as the scalability problem which we tackle in our proposed recommendation approach.

4 Evidential Clustering Approach for CF

In this section, we represent our evidential CF method based on items clustering. Figure 1 gives the overall flow of the proposed recommendation approach.

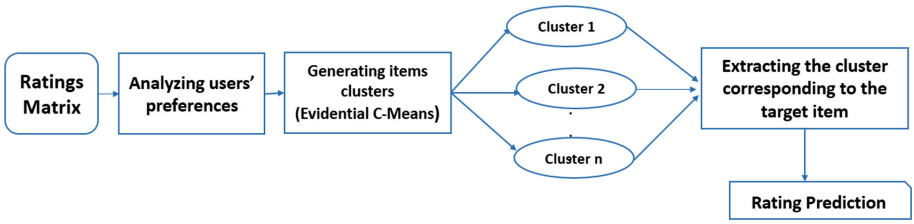

Fig. 1. A new clustering CF approach under the belief function theory

4.1 Items Clustering

Clustering is a crucial step in our approach since the predictions are then performed accordingly. The heart of this approach is to use the efficient soft clustering method, ECM [16] in order to provide a credal partition of the items. Hence, we allocate, for each item in the ratings matrix a mass of belief not only to single clusters, but also to any subsets of the frame of discernment Θ. Before performing the clustering process, we exploit the rating matrix and we randomly initialize the cluster centers commonly referred to as prototypes. Then, we compute the Euclidean distance between the items and the non empty subsets of Θ. We obtain the final credal partition when the objective function (Eq. 5) is minimized.

Example 1. *Let us consider the user-item matrix illustrated in Table 1.*

Table 1. User-item matrix

	Movie_1	Movie_2	Movie_3	Movie_4	Movie_5
User_1	3	?	4	1	2
User_2	4	4	2	?	?
User_3	3	2	4	3	2
User_4	?	1	5	2	3
User_5	5	2	0	2	5

Suppose that the number of clusters $c = 3$, the clustering process consists of providing a credal partition for the 5 movies. In other words, each movie in the system may belong to not only singleton clusters but also to disjunctions of clusters as represented in Table 2.

Table 2. The credal partition corresponding to the five movies

Movies	$\emptyset$	$\{C_1\}$	$\{C_2\}$	$\{C_1, C_2\}$	$\{C_3\}$	$\{C_1, C_3\}$	$\{C_2, C_3\}$	Θ
Movie_1	0.0025	0.9682	0.009	0.0078	0.0046	0.0043	0.0018	0.0017
Movie_2	0.0468	0.2946	0.2715	0.1106	0.1135	0.0731	0.0516	0.0382
Movie_3	0.0005	0.0010	0.0018	0.0004	0.9934	0.0009	0.0017	0.0004
Movie_4	0.0062	0.0212	0.8856	0.0174	0.0247	0.0107	0.0246	0.0097
Movie_5	0.0366	0.1484	0.4931	0.0909	0.0947	0.0479	0.0556	0.0327

4.2 Clusters Selection

In order to make a final decision about the cluster of the current item, we compute the pignistic probability $BetPi(C_k)$ (Eq. 1) induced by each *bba*. These

values are interpreted as the degree of membership of the item i to cluster k. Finally, a hard partition can be easily obtained by assigning each object to the cluster with the highest pignistic probability.

Example 2. *Based on the credal partition derived in the first step, the bba's can be transformed into pignistic probablities in order to select the corresponding cluster having the highest value as shown in Table 3.*

Table 3. The pignistic probabilities corresponding to the five movies

Movies	C_1	C_2	C_3	Selected cluster
Movie_1	0.9773	0.0144	0.0083	C_1
Movie_2	0.4188	0.3833	0.1979	C_1
Movie_3	0.0017	0.0029	0.9953	C_3
Movie_4	0.0387	0.9155	0.0458	C_2
Movie_5	0.2374	0.5992	0.1633	C_2

4.3 Ratings Prediction

The selected clusters are used to obtain knowledge about the items that should be considered in the rating prediction. In order to perform the prediction task, only the items belonging to the same cluster as the target item are extracted. The predicted rating consists of the average of the ratings corresponding to the same clusters members. Given a target item, the prediction is performed as follows:

$$\widehat{R}_{u,i} = \frac{\sum_{j \in C_i(u)} R_{uj}}{|C_i(u)|} \tag{6}$$

where $C_i(u)$ is the set of items belonging to the cluster of the target item i and that have been rated by the user u. R_{uj} is the rating given by user u to item j. $|C_i(u)|$ is the number of items in cluster C_i which have been rated by user u.

Example 3. *For instance, to predict the rating $\widehat{R}_{1,2}$ given by $User_1$ to $Movie_2$, we simply average the ratings of the items belonging to the same cluster and that have been rated by $User_1$. In our case, only $Movie_1 \in C_1$. Then $\widehat{R}_{1,2} = \frac{3}{1} = 3$.*

5 Experimental Evaluation

In order to evaluate our proposal, we test our approach using a real world data set which is widely used in CF and publicly available on the MovieLens[1] website. It contains 100.000 ratings collected from 943 users in 1682 movies.

[1] http://movielens.org.

We conducted our experiments by following the experimental protocol suggested by [7]. The movies rated by the 943 users are ranked according to the number of the ratings given by the users. Rating matrix do not have enough data for accurate predictions, which is known as sparsity. The experimentation strategy consists on increasing progressively the number of the missing rates leading to different sparsity degrees. Hence, we obtain 10 different subsets containing a specific number of ratings provided by the 943 users for 20 different movies. For each subset, we randomly extract 20% of the available ratings as a testing data and the remaining 80% were considered as a training data.

5.1 Evaluation Metrics

We assume that involving an evidential clustering approach for CF may lead to a better performance over the predicted ratings as well the consuming time.

Prediction and Recommendation
In order to assess the prediction accuracy and to evaluate the quality of recommendations provided to the active user, we opt for two evaluation metrics commonly used in CF: the *Mean Absolute Error* (MAE) which belongs in this case to $[0, 4]$ and the precision belonging to $[0, 1]$ defined by:

$$MAE = \frac{\sum_{u,i} |\widehat{R}_{u,i} - R_{u,i}|}{\|\widehat{R}_{u,i}\|} \tag{7}$$

$$Precision = \frac{IR}{IR + UR} \tag{8}$$

where $R_{u,i}$ is the real rating for the user u on the item i and $\widehat{R}_{u,i}$ is the predicted value. $\|\widehat{R}_{u,i}\|$ is the total number of the predicted ratings over all the users. IR indicates that an interesting item has been correctly recommended while UR indicates that an uninteresting item has been incorrectly recommended. The lower the MAE values are, the more accurate the predictions are. Otherwise, the highest values of the precision indicate a better recommendation quality.

Scalability
We also investigated the performance of our approach in terms of scalablity. We recall that the purpose of scalability refers to the ability of a method to be run quickly by handling the evolution regarding the number of items and users.

5.2 Experimental Results

We performed various experiments over the 10 selected subsets by varying each time the number of clusters c. We used $c = 2$, $c = 3$, $c = 4$ and $c = 5$. For each subset, the results corresponding to the different number of clusters used in the experiments are then averaged. In other words, we compute the MAE and the precision measure for each value of c and we note the overall results. For all our experiments, we used $\alpha = 2$, $\beta = 2$ and $\delta^2 = 10$ as invoked in [16].

Unlike the evidential item-based CF (Evidential IB-CF) [11], the proposed evidential clustering item-based CF (Evidential Clustering IB-CF) relies on items clusters rather than the user-item matrix. Hence, we compare the two CF methods proposed under the BFT in order to evaluate the performance of our approach. Table 4 recapitulates the results of each evidential IB-CF considering different sparsity degrees.

Table 4. Comparison results in terms of MAE and precision

Evaluation metrics	Subsets	Sparsity degrees	Evidential IB-CF	Evidential clustering IB-CF
MAE	$Subset_1$	53%	0.751	0.749
Precision			0.79	0.792
MAE	$Subset_2$	56.83%	0.84	0.8
Precision			0.76	0.74
MAE	$Subset_3$	59.8%	0.761	0.747
Precision			0.77	0.785
MAE	$Subset_4$	62.7%	0.763	0.793
Precision			0.763	0.782
MAE	$Subset_5$	68.72%	0.831	0.845
Precision			0.735	0.752
MAE	$Subset_6$	72.5%	0.851	0.8
Precision			0.735	0.813
MAE	$Subset_7$	75%	0.744	0.733
Precision			0.78	0.805
MAE	$Subset_8$	80.8%	0.718	0.762
Precision			0.778	0.755
MAE	$Subset_9$	87.4%	0.840	0.873
Precision			0.707	0.73
MAE	$Subset_{10}$	95.9%	0.991	0.83
Precision			0.513	0.55
Overall MAE			0.809	**0.793**
Overall Precision			0,733	**0.75**

The proposed approach allows an improvement over the standard evidential item-based CF approach [11] by acquiring, in average the lowest error rates over the 10 subsets (0.793 compared to 0.809) as well as the highest overall precision (0.75 compared to 0.733). While the clustering CF proposed in [6] improves the scalability with a worse prediction quality compared to the traditional one, our evidential clustering CF outperforms the standard evidential CF in both cases.

Scalability Performance

We perform the scalability of our approach by varying the sparsity degree. We compare the results to the standard evidential CF as depicted in Fig. 2.

According to Fig. 2, the elapsed time corresponding to the clustering CF approach is substantially lower than the basic evidential CF. These results are

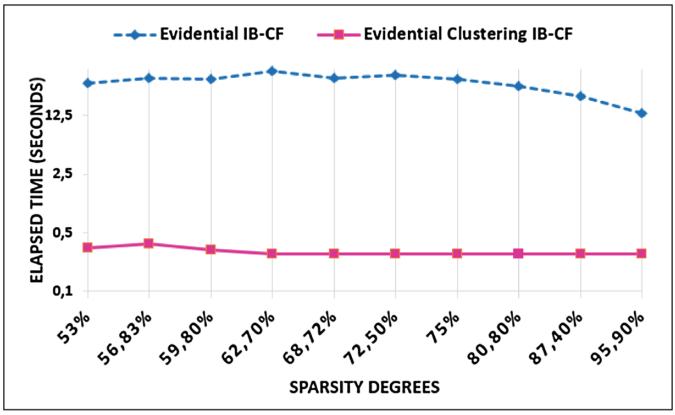

Fig. 2. Elapsed time of evidential clustering CF vs. evidential CF

explained by the fact that standard CF methods need to search the closest neighbors to the target item in the whole item space, which leads to huge computing amount.

6 Conclusion

In this paper, we have proposed a new clustering CF approach based on the Evidential C-Means method. Compared to a recent CF method under the belief function theory, elapsed time has been significantly improved, along with better prediction and recommendation performance. As future work, we intend to rely on the different $bba's$ corresponding to the different clusters rather than the most significant one.

References

1. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. Knowl.-Based Syst. **46**, 109–132 (2013)
2. Park, Y., Park, S., Jung, W., Lee, S.G.: Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbor graph. Expert Syst. Appl. **42**(8), 4022–4028 (2015)
3. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artif. Intell. **2009**, 1–19 (2009)
4. Zhao, Z.D., Shang, M.S.: User-based collaborative-filtering recommendation algorithms on hadoop. In: Third International Conference on Knowledge Discovery and Data Mining, pp. 478–481. IEEE, Phuket (2010)
5. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: International Conference on World Wide Web, pp. 285–295. ACM, Hong Kong (2001)

6. Sarwar, B.M., Karypis, G., Konstan, J., Riedl, J.: Recommender systems for large-scale e-commerce: scalable neighborhood formation using clustering. In: International Conference on Computer and Information Technology. IEEE, Dhaka (2002)
7. Su, X., Khoshgoftaar, T.M.: Collaborative filtering for multi-class data using bayesian networks. Int. J. Artif. Intell. Tools **17**(01), 71–85 (2008)
8. Symeonidis, P.: Matrix and tensor decomposition in recommender systems. In: ACM Conference on Recommender Systems, pp. 429–430. ACM, Boston (2016)
9. Guo, G., Zhang, J., Yorke-Smith, N.: Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems. Knowl.-Based Syst. **74**, 14–27 (2015)
10. Zhang, J., Lin, Y., Lin, M., Liu, J.: An effective collaborative filtering algorithm based on user preference clustering. Appl. Intell. **45**(2), 230–240 (2016)
11. Abdelkhalek, R., Boukhris, I., Elouedi, Z.: Evidential item-based collaborative filtering. In: Lehner, F., Fteimi, N. (eds.) KSEM 2016. LNCS, vol. 9983, pp. 628–639. Springer, Cham (2016). doi:10.1007/978-3-319-47650-6_49
12. Dempster, A.P.: A generalization of bayesian inference. J. Roy. Stat. Soc. Series B (Methodological) **30**, 205–247 (1968)
13. Shafer, G.: A Mathematical Theory of Evidence, vol. 1. Princeton University Press, Princeton (1976)
14. Smets, P.: The transferable belief model for quantified belief representation. In: Smets, P. (ed.) Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, pp. 267–301. Springer, Dordrecht (1998)
15. Hariz, S., Elouedi, Z., Mellouli, K.: Clustering approach using belief function theory. In: Euzenat, J., Domingue, J. (eds.) AIMSA 2006. LNCS, vol. 4183, pp. 162–171. Springer, Heidelberg (2006). doi:10.1007/11861461_18
16. Masson, M.H., Denoeux, T.: ECM: An evidential version of the fuzzy c-means algorithm. Pattern Recogn. **41**(4), 1384–1397 (2008)
17. Xue, G.R., Lin, C., Yang, Q., Xi, W., Zeng, H.J., Yu, Y., Chen, Z.: Scalable collaborative filtering using cluster-based smoothing. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 114–121. ACM, Salvador (2005)
18. Nguyen, V.-D., Huynh, V.-N.: A community-based collaborative filtering system dealing with sparsity problem and data imperfections. In: Pham, D.-N., Park, S.-B. (eds.) PRICAI 2014. LNCS, vol. 8862, pp. 884–890. Springer, Cham (2014). doi:10.1007/978-3-319-13560-1_74
19. Nguyen, V.-D., Huynh, V.-N.: A reliably weighted collaborative filtering system. In: Destercke, S., Denoeux, T. (eds.) ECSQARU 2015. LNCS, vol. 9161, pp. 429–439. Springer, Cham (2015). doi:10.1007/978-3-319-20807-7_39
20. Nguyen, V.-D., Huynh, V.-N.: Integrating with social network to enhance recommender system based-on dempster-shafer theory. In: Nguyen, H.T.T., Snasel, V. (eds.) CSoNet 2016. LNCS, vol. 9795, pp. 170–181. Springer, Cham (2016). doi:10.1007/978-3-319-42345-6_15