Joining Items Clustering and Users Clustering for Evidential Collaborative Filtering

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. Recommender Systems (RSs) are supporting users to cope with the flood of information. Collaborative Filtering (CF) is one of the most well-known approaches that have achieved a widespread success in RSs. It consists in picking out the most similar users or the most similar items to provide recommendations. Clustering techniques can be adopted in CF for grouping these similar users or items into some clusters. Nevertheless, the uncertainty comprised throughout the clusters assignments as well as the final predictions should also be considered. Therefore, in this paper, we propose a CF recommendation approach that joins both users clustering strategy and items clustering strategy using the belief function theory. In our approach, we carry out an evidential clustering process to cluster both users and items based on past preferences and predictions are then performed accordingly. Joining users clustering and items clustering improves the scalability and the performance of the traditional neighborhood-based CF under an evidential framework.

Keywords: Recommender Systems · Collaborative Filtering · Uncertain reasoning · Belief function theory · Users clustering · Items clustering

1 Introduction

Collaborative Filtering (CF) is one of the most well-known approaches that have achieved a widespread success in Recommender Systems (RSs) [1]. It consists in finding out the most similar users (user-based) or the most similar items (item-based) to provide recommendations. Thus, the predictions of the users' preferences are performed based on the users or the items holding similar ratings. Although their simplicity and intuitiveness, these two CF methods disclose some limitations such as the scalability problem [2], since they require a lot of heavy computations to provide predictions. That is to say, in order to compute the user-user similarities and/or the item-item similarities, the entire ratings matrix needs to be searched which reveals a poor scalability performance. In this context, clustering techniques can be involved to first group users or items according to

H. Yin et al. (Eds.): IDEAL 2019, LNCS 11871, pp. 310–318, 2019.
https://doi.org/10.1007/978-3-030-33607-3_34

their available past ratings and predictions can be made independently within each cluster. Hence, clustering would be an ideal process that may increase the scalability of CF systems as well as maintaining a good prediction performance. While clustering users or items, they are likely to bear upon more than only one cluster which is known as soft clustering. Thus, the uncertainty spreading around the clusters assignment should be considered. In this paper, we embrace the belief function theory (BFT) [3–5] which is among the most used theories for representing and reasoning under uncertainty. We opt for the Evidential c-Means (ECM) [6], an efficient soft clustering method allowing to deal with the uncertainty at the clusters assignment level. When the uncertainty intervenes at clusters assignment, while using the belief function theory, the output is referred to as credal partition. The pertinence of reasoning under uncertainty in CF is also considered at the final prediction level. Indeed, we also tend to quantify and represent the uncertainty arising behind the provided predictions based on the Evidential k-Nearest Neighbors [7] formalism. Such evidential classifier improves the classification performance by allowing a credal classification of the objects. Therefore, the proposed approach allows us to provide more reliable and intelligible predictions by providing an evidential representation of the final results. This representation allows the users to get an overall information about their future preferences, which may increase their confidence and satisfaction towards the RS. Our aim then is not only to unify both user-based and item-based CF but also to represent and encapsulate the uncertainty appearing in both clusters assignment and prediction process under an evidential framework.

This paper is organized as follows: Sect. 2 recalls the concepts of the belief function theory. Section 3 introduces the related works on CF. Then, our proposed recommendation approach is emphasized in Sect. 4. Section 5 reports the experimental results. Finally, the paper is concluded in Sect. 6.

2 Belief Function Framework

The belief function theory [3–5], also referred to as evidence theory, represents a flexible framework for modeling and quantifying imperfect knowledge. Let Ω be the frame of discernment defined as a finite set of variables w. It refers to n elementary events such that: $\Omega = \{w_1, w_2, \cdots, w_n\}$. The basic belief assignment (*bba*) represents the belief committed to each element of Ω such that $m : 2^{\Omega} \rightarrow [0, 1]$ and $\sum\limits_{E \subseteq \Omega} m(E) = 1$. The mass $m(E)$ quantifies the degree of belief exactly assigned to an element E of Ω. Evidence may not be equally trustworthy. Hence, a discounting operation can be used to get the discounted *bba* m^{δ} such that: $m^{\delta}(E) = (1 - \delta) \cdot m(E), \forall E \subset \Omega \;\; and \;\; m^{\delta}(\Omega) = \delta + (1 - \delta) \cdot m(\Omega)$. where $\delta \in$ [0,1] is the discounting factor.

To make decisions, the *pignistic probability* denoted $BetP$ can be used, where $BetP$ is defined as follows: $BetP(E) = \sum_{F \subseteq \Omega} \frac{|E \cap F|}{|F|} \frac{m(F)}{(1-m(\varnothing))} \; for \; all \; E \subseteq \Omega.$

3 Related Works on Collaborative Filtering

Clustering-based CF approaches have been suggested in RSs to learn a cluster model and to provide predictions accordingly. For instance, the CF approach introduced in [2] consisted in a partition of the users in a CF system based on clustering and used the obtained partitions as neighborhoods. In the same way, authors in [8] have proposed a CF approach based on users' preferences clustering. In [9], a CF method has been presented where the training set has been used to generate users clusters and predictions are then generated. A different clustering based-CF has been proposed in [10], where, instead of users, authors proposed to group items into several clusters based on the k-METIS algorithm. These works mentioned above rely either on users clustering or items clustering to predict the users' preferences. In this work, we are rather interested in joining both users clustering and items clustering under the belief function theory [3–5]. In fact, a hybrid CF approach combining both user-based and item-based strategies under the belief function framework has been proposed in [11]. Such evidential approach showed that the fusion framework is effective in improving the prediction accuracy of single approaches (user-based and item-based) under certain and uncertain frameworks. However, it is not able to scale up to real data sets with large collections of items and users since a lot of heavy computations is required. This problem is referred to as the scalability problem which we tackle in our proposed approach.

4 E-HCBCF: Evidential Hybrid Clustering-Based CF

In this approach, we tend to cope with the scalability problem by performing a clustering process of both users and items where uncertainty is also handled.

4.1 Evidential Users Clustering

In this step, an evidential clustering process is performed for the users in the system using the Evidential c-Means (ECM) [6]. We define $\Omega_{clus} = \{w_1, w_2, \ldots, w_c\}$ where c is the numbers of clusters. Hence, the users clustering process bestows a credal partition that allows a given user to be associated to multiple clusters, or rather multiple partitions of clusters. We start by considering the ratings matrix to randomly initialize the cluster prototypes. Then, we compute the euclidean distance between the users and the non empty sets of Ω_{clus}. At the end, the convergence and the minimization of a given objective function [6] lead to the generation of the final credal partition. In order to make a final decision about the cluster of the current user, we transform each *bba* into a pignistic probability $BetP(w_k)$ based on the obtained credal partition of each user in the system. These values reflect the degrees of membership of the user u to the cluster w_k. A hard partition can be easily derived by assigning each user to the cluster holding the highest $BetP$ value.

4.2 Modeling Users' Neighborhood Ratings

According to the obtained users clusters, the set of the k-nearest neighbors of the active user U_a is extracted. We compute the distances between U_a and the other users in the same cluster as follows [7]: $dist(U_a, U_i) = \frac{1}{|I(a,i)|}\sqrt{\sum_{I \in I(a,i)} (r_{a,j} - r_{i,j})^2}$, where $|I(a,i)|$ is the number of items rated by the user U_a and the user U_i, $r_{a,j}$ and $r_{i,j}$ are respectively the ratings of the users U_a and U_i for the item I_j. The set of the k-most similar users, that we denote by Γ_k, is selected based on the computed distances. We define the frame of discernment $\Omega_{pref} = \{r_1, r_2, \cdots, r_L\}$ as a rank-order set of L preference labels (i.e ratings) where $r_p < r_l$ whenever $p < l$. Since we use the belief function theory in order to conveniently model uncertainty, we transform the rating of each user U_i belonging to Γ_k into a basic belief assignment m_{U_a,U_i} spanning over the frame of discernment based on the formalism in [7]. This *bba* reflects the evidence of U_i for the rating of U_a.

$$m_{U_a,U_i}(\{r_p\}) = \alpha_0 e^{-(\gamma_{r_p}^2 \times (dist(U_a,U_i))^2} \tag{1}$$

$$m_{U_a,U_i}(\Omega_{pref}) = 1 - \alpha_0 e^{-(\gamma_{r_p}^2 \times (dist(U_a,U_i))^2}$$

Where α_0 is fixed to the value of 0.95 and γ_{r_p} is the inverse of the average distance between each pair of users having the same ratings r_p. In order to evaluate the reliability of each neighbor, these *bba*'s are then discounted such that:

$$\begin{aligned} m^{\delta_u}_{U_a,U_i}(\{r_p\}) &= (1 - \delta_u) \cdot m_{U_a,U_i}(\{r_p\}) \\ m^{\delta_u}_{U_a,U_i}(\Omega_{pref}) &= \delta_u + (1 - \delta_u) \cdot m_{U_a,U_i}(\Omega_{pref}) \end{aligned} \tag{2}$$

The discounting factor $\delta_u = \frac{dist(U_a,U_i)}{max(dist)}$, where $max(dist)$ is the maximum value of the computed distances between the users.

4.3 Generating Users' Neighborhood Predictions

Once the $bba's$ of the similar users are generated, they are fused as follows [7]:

$$\begin{gathered} m^{\delta_u}(\{r_p\}) = \frac{1}{N}(1 - \prod_{U \in \Gamma_k}(1 - \alpha_{r_p})) \cdot \prod_{r_p \neq r_q} \prod_{U \in \Gamma_k}(1 - \alpha_{r_q}) \qquad \forall r_p \in \{r_1, \cdots, r_{Nb}\} \\ m^{\delta_u}(\Omega_{pref}) = \frac{1}{N}\prod_{p=1}^{Nb}(1 - \prod_{U \in \Gamma_k}(1 - \alpha_{r_p})) \end{gathered} \tag{3}$$

Nb is the number of the ratings provided by the similar users, α_{r_p} is the belief committed to the rating r_p, α_{r_q} is the belief committed to the rating r_q $\neq r_p$ and N is a normalized factor defined by [7]:

$$N = \sum_{p=1}^{Nb}(1 - \prod_{U \in \Gamma_k}(1 - \alpha_{r_p}) \prod_{r_q \neq r_p} \prod_{U \in \Gamma_k}(1 - \alpha_{r_q}) + \prod_{p=1}^{Nb}(\prod_{U \in \Gamma_k}(1 - \alpha_{r_q})))$$

4.4 Evidential Items Clustering

Until now, only the user-based assumption has been adopted. In what follows, the item-side would also be considered. In this phase, we aim also to carry out an evidential items clustering process for the corresponding items in the system. The uncertainty about items assignments to clusters is considered where each item in the user-item matrix can belong to all clusters with a degree of belief. The ECM [6] is applied once again and the ratings matrix is explored in order to generate the final credal partition of the corresponding items. Once the credal partition corresponding to each item in the system has been produced, a pignistic probability $BetP(w_k)$ is then generated based on the obtained $bba's$. Thereupon, we apply at this level these pignistic probabilities with the aim of assigning each item to its appropriate cluster through the consideration of the highest values.

4.5 Modeling Items' Neighborhood Ratings

Given a target item I_t, we now consider the items having the same cluster membership as I_t. To this end, the distances between I_t and the whole items are computed. Formally, the distance between the target item I_t and each item I_j is computed as follows [7]: $dist(I_t, I_j) = \frac{1}{|U(t,j)|}\sqrt{\sum_{U\in U(t,j)} (r_{i,t} - r_{i,j})^2}$. $|U(t,j)|$ denotes the number of users that rated both the target item I_t and the item I_j where $r_{i,t}$ and $r_{i,j}$ correspond to the ratings of the user U_i for the target item I_t and for the item I_j. Finally, only the k items having the lowest distances are selected and the rating of each similar item is transformed into a bba defined as:

$$m_{I_t,I_j}(\{r_p\}) = \alpha_0 e^{-(\gamma_{r_p}^2 \times (dist(I_t,I_j))^2} \tag{4}$$

$$m_{I_t,I_j}(\mathit{\Theta}_{pref}) = 1 - \alpha_0 e^{-(\gamma_{r_p}^2 \times (dist(I_t,I_j))^2}$$

Following [7], α_0 is set to 0.95 and γ_{r_p} is defined as the inverse of the mean distance between each pair of items having the same ratings. Finally, these $bba's$ are discounted as follows:

$$m^{\delta_i}_{I_t,I_j}(\{r_p\}) = (1-\delta_i)\cdot m_{I_t,I_j}(\{r_p\}) \tag{5}$$

$$m^{\delta_i}_{I_t,I_j}(\mathit{\Theta}_{pref}) = \delta_i + (1-\delta_i)\cdot m_{I_t,I_j}(\mathit{\Theta}_{pref})$$

where δ_i is a discounting factor depending on the items' distances such as: $\delta_i = \frac{dist(I_t,I_j)}{max(dist)}$ where $max(dist)$ is the maximum value of the computed distances.

4.6 Generating Items' Neighborhood Predictions

The aggregation of the $bba's$ for each similar item is performed as follows:

$$m^{\delta_i}(\{r_p\}) = \frac{1}{Z}(1-\prod_{I\in\mathit{\Gamma}_k}(1-\alpha_{r_p}))\cdot \prod_{r_p\neq r_q}\prod_{I\in\mathit{\Gamma}_k}(1-\alpha_{r_q}) \qquad \forall r_p\in\{r_1,\cdots,r_{Nb}\} \tag{6}$$

$$m^{\delta_i}(\Theta_{pref}) = \frac{1}{Z}\prod_{p=1}^{Nb}(1 - \prod_{I\in\Gamma_k}(1-\alpha_{r_p}))$$

Note that Nb is the number of the ratings given by the similar items, α_{r_p} is the belief committed to the rating r_p, α_{r_q} is the belief committed to the rating $r_q \neq r_p$ and Z is a normalized factor defined by [7]:

$$Z = \sum_{p=1}^{Nb}(1-\prod_{I\in\Gamma_k}(1-\alpha_{r_p})\prod_{r_q\neq r_p}\prod_{I\in\Gamma_k}(1-\alpha_{r_q}) + \prod_{p=1}^{Nb}(\prod_{I\in\Gamma_k}(1-\alpha_{r_q})))$$

4.7 Final Predictions and Recommendations

After performing an evidential co-clustering process, predictions incorporating both items' aspects and users' aspects have been generated under the belief function theory. Thus, we obtain, on the one hand, *bba*'s from the k-similar items (m^{δ_i}) and, on the other hand, $bba's$ derived from the k-similar users (m^{δ_u}). The fusion of the $bba's$ corresponding to these two kinds of information sources can be ensured through an aggregation process using Dempster's rule of combination [3] such that: $m_{Final} = (m^{\delta_i} \oplus m^{\delta_u})$. The obtained predictions indicate the belief induced from the two pieces of evidence namely the similar users and the similar items.

5 Experiments and Discussions

MovieLens[1], one of the commonly used real world data set in CF, has been adopted in our experiments. We perform a comparative evaluation over our proposed method (E-HCBCF) and the evidential hybrid neighborhood-based CF proposed in [11], denoted by (E-HNBCF). Such hybrid approach achieved better results than state of the art CF approaches both item-based and user-based ones. Compared to E-HNBCF, our new approach incorporates an evidential co-clustering process of the users as well as the items in the system and provides predictions accordingly. We followed the strategy conducted in [12] where movies are first ranked based on the number of their corresponding ratings. Different subsets are then extracted by progressively increasing the number of the missing rates. Hence, each subset will contain a specific number of ratings leading to different degrees of sparsity. For all the extracted subsets, we randomly extract 10% of the users as a testing data and the remaining ones were considered as a training data. We opt for three evaluation metrics to evaluate our proposal: the *Precision*, the *Distance criteron* (Dist_crit) [13] and the *elapsed time* such that: $Precision = \frac{IR}{IR+UR}$ *and* $Dist\_crit = \frac{\sum_{u,i}(\sum_{i=1}^{n}(BetP(\{p_{u,i}\})-\theta_i)^2)}{\|\widehat{p_{u,i}}\|}$. IR indicates that an interesting item has been correctly recommended while UR indicates that an uninteresting item has been incorrectly recommended. $\|\widehat{p_{u,i}}\|$ is the total number of the predicted ratings, n is the number of the possible ratings that

[1] http://movielens.org.

Table 1. Comparison results in terms of Precision and Dist_crit

Evaluation metrics	Sparsity degrees	E-HNBCF	E-HCBCF
Precision	53%	0.750	0.833
Dist_crit		0.946	0.905
Precision	56.83%	0.726	0.767
Dist_crit		0.933	1.020
Precision	59.8%	0.766	0.761
Dist_crit		0.940	0.938
Precision	62.7%	0.814	0.805
Dist_crit		0.932	0.928
Precision	68.72%	0.864	0.846
Dist_crit		0.961	0.844
Precision	72.5%	0.730	0.805
Dist_crit		0.936	0.927
Precision	75%	0.739	0.683
Dist_crit		0.872	0.972
Precision	80.8%	0.620	0.675
Dist_crit		1.080	1.003
Precision	87.4%	0.633	0.555
Dist_crit		0.665	0.651
Overall Precision		0.737	**0.747**
Overall Dist_crit		0.918	**0.909**

can be provided in the system. θ_i is equal to 1 if $p_{u,i}$ is equal to $\widehat{p_{u,i}}$ and 0 otherwise, where $p_{u,i}$ is the real rating for the user u on the item i and $\widehat{p_{u,i}}$ is the predicted value. Note that the lower the Dist_crit values are, the more accurate the predictions are while the highest values of the precision indicate a better recommendation quality. Experiments are conducted over the selected subsets by switching each time the number of clusters c. We used $c = 2$, $c = 3$, $c = 4$ and $c = 5$. For each selected cluster, different neighborhood sizes were tested and the average results were computed. Finally, the results obtained for the different numbers of clusters used in the experiments are also averaged. More specifically, we compute the precision and the Dist_crit measures for each value of c and we report the overall results. For the evidential co-clustering process, required parameters were set as follows: $\alpha = 2$, $\beta = 2$ and $\delta^2 = 10$ as invoked in [6]. For the $bba's$ generation, we remind that α_0 is fixed to the value 0.95 and γ_{r_p} is computed as the inverse of the mean distance between each pair of items and users sharing the same rating values. Considering different sparsity degrees, the results are displayed in Table 1. Compared to the evidential hybrid neighborhood-based CF, it can be seen that incorporating an evidential co-clustering process that joins both items clustering and users clustering provides a slightly better performance

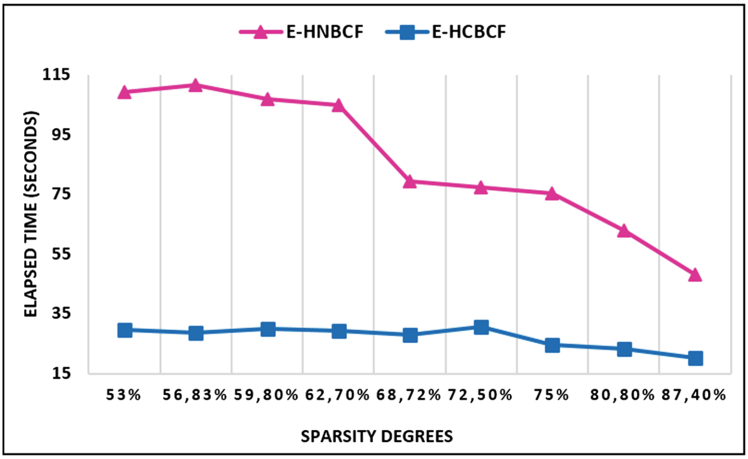

Fig. 1. Elapsed Time of E-HCBCF vs. E-HNBCF

than the other evidential approach. Indeed, the average precision of E-HCBCF is somewhat greater than the value obtained by the standard E-HNBCF (0.747 compared to 0.737). Besides, it acquires the lowest average value in terms of Dist_crit (0.909 compared to 0.918). These obtained results show that the evidential co-clustering process is effective to maintain a good prediction quality while improving also the scalability performance as depicted in Fig. 1. According to Fig. 1, the elapsed time corresponding to E-HCBCF is substantially lower than the basic E-HNBCF. This outcome is explained by the fact that the standard hybrid evidential CF method needs to search the closest neighbors of the active user as well as the similar neighbors of the target item by browsing the whole user-item ratings space, which results in a huge computing amount.

6 Conclusion

In this paper, we have proposed a new evidential co-clustering CF approach where both items' aspect and users' aspect come into play. In fact, a clustering model that joins both users clustering and items clustering has been built under the belief function theory based on the available past ratings. According to the obtained clusters, the k-nearest users and the k-nearest items are selected and predictions are then performed. Global evidence of the neighbors is finally aggregated to get an overall information about the provided predictions. Experiments on a real world CF data set proved the efficiency of the proposed approach where elapsed time has been significantly improved while maintaining a good recommendation performance.

References

1. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artif. Intell. **2009**, 1–19 (2009)
2. 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, Dhaka, Bangladesh. IEEE (2002)
3. Dempster, A.P.: A generalization of Bayesian inference. J. Roy. Stat. Soc. Ser. B (Methodol.) **30**, 205–247 (1968)
4. Shafer, G.: A Mathematical Theory of Evidence, 1st edn. Princeton University Press, Princeton (1976)
5. Smets, P.: The transferable belief model for quantified belief representation. In: Smets, P. (ed.) Quantified Representation of Uncertainty and Imprecision. HDRUMS, vol. 1, pp. 267–301. Springer, Dordrecht (1998). https://doi.org/10.1007/978-94-017-1735-9_9
6. Masson, M.H., Denoeux, T.: ECM: an evidential version of the fuzzy c-means algorithm. Pattern Recogn. **41**(4), 1384–1397 (2008)
7. Denoeux, T.: A K-nearest neighbor classification rule based on Dempster-Shafer theory. IEEE Trans. Syst. Man Cybern. **25**, 804–813 (1995)
8. 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). https://doi.org/10.1007/s10489-015-0756-9
9. Xue, G.R., et al.: Scalable collaborative filtering using cluster-based smoothing. In: ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 114–121. ACM (2005)
10. O'Connor, M., Herlocker, J.: Clustering items for collaborative filtering. In: ACM SIGIR Workshop on Recommender Systems, vol. 128. UC Berkeley (1999)
11. Abdelkhalek, R., Boukhris, I., Elouedi, Z.: Towards a hybrid user and item-based collaborative filtering under the belief function theory. In: Medina, J., et al. (eds.) IPMU 2018. CCIS, vol. 853, pp. 395–406. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91473-2_34
12. Su, X., Khoshgoftaar, T.M.: Collaborative filtering for multi-class data using Bayesian networks. Int. J. Artif. Intell. Tools **17**, 71–85 (2008)
13. Elouedi, Z., Mellouli, K., Smets, P.: Assessing sensor reliability for multisensor data fusion within the transferable belief model. IEEE Trans. Syst. Man Cybern. Part B Cybern. **34**, 782–787 (2004)