Journal of Universal Computer Science, vol. 28, no. 10 (2022), 1003-1029
submitted: 25/12/2021, accepted: 28/7/2022, appeared: 28/10/2022

Towards more trustworthy predictions: A hybrid evidential movie recommender system

Raoua Abdelkhalek
(Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
https://orcid.org/0000-0002-3896-1718, raoua.abdelkhalek@gmail.com)

Imen Boukhris
(Université de La Manouba, Ecole Nationale des Sciences de l'Informatique
Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
https://orcid.org/0000-0003-4677-4311, imen.boukhris@hotmail.com)

Zied Elouedi
(Université de Tunis, Institut Supérieur de Gestion de Tunis, LARODEC, Tunisia
https://orcid.org/0000-0002-0476-4951, zied.elouedi@gmx.fr)

Abstract: Recommender Systems (RSs) are considered as popular tools that have revolutionized the e-commerce and digital marketing. Their main goal is predicting the users' future preferences and providing accessible and personalized recommendations. However, uncertainty can spread at any level throughout the recommendation process, which may affect the results. In fact, the ratings given by the users are often unreliable. The final provided predictions itself may also be pervaded with uncertainty and doubt. Obviously, the reliability of the predictions cannot be fully certain and trustworthy. For the system to be effective, recommendations must inspire trust in the system and provide reliable and credible recommendations. The user may speculate about the uncertainty pervaded behind the given recommendation. He could tend to a reliable recommendation offering him a global overview about his preferences rather than an inappropriate one that contradicts his activities and objectives. While such imperfection cannot be ignored, traditional RSs are rarely able to deal with the uncertainty spreading around the prediction process, which may affect the credibility, the transparency and the trustworthiness of the current RS. Thus, in this paper, we opt for the uncertain framework of the belief function theory (BFT), which allows us to represent, quantify and manage imperfect evidence. By using the BFT, the users' preferences and the interactions between the neighbors can be represented under uncertainty. Evidence from different information sources can then be combined leading to more reliable results. The proposed approach is a hybrid evidential movie RS that uses different data sources and delivers a personalized user-interface allowing a global overview of the possible future preferences. This representation would increase the users' confidence towards the system as well as their satisfaction. Experiments are performed on MovieLens and their additional features provided by the Internet Movie Database (IMDb) and Rotten Tomatoes. The new approach achieves promising results compared to traditional approaches in terms of MAE, NMAE and RMSE. It also reaches interesting Precision, Recall and F-measure values of respectively, 0.782, 0.792 and 0.787.

Keywords: Recommender Systems, Evidential predictions, Decision making, Uncertain reasoning, Belief Function Theory, Personalization
Categories: H.5.1, H.5.2, I.2.4, I.5.3
DOI: 10.3897/jucs.79777

1 Introduction

Recommender Systems (RSs) [Mohamed et al. (2019)] are a highly effective solution to cope with the information overload problem. Guiding the users and helping them in their decision making process are the main goals of these systems. Collaborative Filtering (CF) approaches [Ambulgekar et al. (2019)] have been drawing the attention of researchers in RSs. They predict the users' preferences towards items based on their past behaviors and provide accessible and personalized recommendations. An important class for this kind of RS is the neighborhood-based CF approaches categorized into user-based CF [Wang et al. (2021)] and item-based CF [Musa et al. (2020)]. Using the users' historical ratings, such approaches exploit the user-item ratings matrix and compute the similarities between users or items in the system. The predictions are then performed and the final recommendations are provided. Contrariwise, model-based approaches operate with these ratings to derive a predictive model, which is subsequently used in the recommendation [Behera et al. (2022)]. In our work, we are notably interested in item-based CF, which has shown a great applicability in many domains. For instance, movies RSs compute the similarities between movies based on their ratings. Then, they produce for the users personalized recommendations by predicting a likeness score for each movie not yet watched. While only the available ratings are used in these predictions,more additional information drawing out the user-item matrix can be exploited, such as items contents. Although some few works have extended CF approaches to integrate this additional side information, the impact of uncertainty involved in the recommendation process has not been considered. An effective RS ought to address ways to deal with uncertainty. As pointed out in [Nguyen et al. (2014)], CF techniques usually suffer from data imperfection issues. In this regard, uncertainty theories can be adopted to deal with this imperfection such as the belief function theory. The belief function theory (BFT) [Dempster (1968), Shafer (1976), Smets, (1998)], also known as Dempster-Shafer theory or evidence theory is considered as one of the powerful tools for dealing with imperfect information. Thanks to its flexibility, it represents and manages any forms of uncertainty, partial or even total ignorance. Likewise, the belief function framework is appropriate to handle uncertainty in classification problems, both supervised and unsupervised ones. For example, the Evidential k-Nearest Neighbors (EkNN) proposed in [Denoeux et al. (1995)] improves the classification performance by allowing a credal classification of the objects. That is to say, each object to be classified can belong to not only a single specific class. Another example would be the evidential clustering [Masson and Denoeux (2008), Masson and Denoeux (2009)] where an object may belong to more than only one cluster, which is commonly known as soft clustering. Two clustering techniques, Evidential c-Means (ECM) [Masson and Denoeux (2008)] and Relational Evidential c-Means (RECM) [Masson and Denoeux (2009)] have been proposed for this purpose. While the ECM has been designed to deal only with vectorial data, the RECM, which is a relational version of ECM, has been recently developed to deal also with pairwise proximity data.

Dealing with uncertainty is an important and challenging task in real world applications including RSs. Different kinds of uncertainty can be permeated at any level all over t the recommendation process, which follows to unreliable results. In other words, the final predictions delivered to the user cannot be certain and fully trusted. For the system to be effective, the provided recommendations must inspire trust, reliability and credibilityin the system. Therefore, we propose in this work a new evidential approach that extends traditional CF by using two sources of information in the prediction process namely users' ratings and items contents while considering uncertainty under the belief function framework. In fact, during the decision-making, the users may speculate about the uncertainty occurring behind the provided prediction. We assume that such uncertainty needs to be appropriately represented and processed to improve quality and reliability of RSs. Within the BFT, we can represent this prediction as an overview over all the possible cases of the user's future ratings rather than a single rating value. Actually, a user may get an item recommendation with a rating score ranging from 5 (Excellent) to 1 (Very bad), but one cannot expect the rating to be "hard" (i.e. "crisp", or "perfect"). The user may prefer a reliable prediction that gives him a complete overview about his preferences rather than a risky one that may contradict his activities and objectives

Under such an observation, the belief function theory is used in our approach for representing the user's final prediction under an uncertain context while taking into account the different memberships of the items clusters. Additionally, with this theory, pieces of evidence can be combined for generating more valuable evidence. Handling uncertainty that arises throughout the recommendation process may increase the intelligibility and the transparency of the predictions. This is a crucial challenge to improve the users' confidence towards the RS as well as their satisfaction [Ricci et al.(2015)]. Therefore, our aim in this paper is not only to integrate both items contents and users' ratings in item-based CF, but also to investigate the relevance of handling uncertainty pervaded throughout the prediction process. The new approach is inspired from neighborhood-based CF methods and uses exclusively the belief function theory tools. Overall, the main contributions of this work are as follows:

- An evidential hybrid framework is proposed for the neighbors' ratings modeling and the prediction process.
- Uncertainty pervaded in the users' ratings as well as the final predictions is represented and managed through the belief function theory.
- Trustworthiness of Recommender Systems is improved and proactive predictions are provided to users.
- Users' confidence is enhanced leading to a better chance for successful recommendations.
- A closer picture about the potentially future ratings is generated to the users, helping them for a better decision making.

The remainder of this paper is organized as follows: In Section 2, we highlight the motivation and the intuition behind our approach. Section 3 recalls the basic concepts of the belief function theory. Related works on Collaborative Filtering are provided in Section 4. Section 5 presents our proposed recommendation approach through a detailed description of the individual steps. Then, Section 6 details the experimentation conducted on two real world data sets. Finally, Section 7 concludes the paper and proposes some potential future works.

2 Problem statement and motivation

In the traditional item-based CF recommender, the users are provided with a prediction of their future evaluation of items not yet rated. For example, when a user is browsing a movie website in order to get an idea about the new releases, the system assigns for him a predicted rating for each movie. Suppose that the value of the rating assigned to the current user is 5 out of 5. That is to say, the system assumes that he will extremely like the suggested movie and that it would be an interesting and excellent one. In such case, the user would deeply trust the RS and make an immediate decision, which is logically watching the proposed movie. The question that arises here is how far the system can assume that the computed outputs are certain? In one way or another, the provided predictions are not perfect, and they involve uncertainty which should not be ignored. Indeed, this method does not take into account the uncertainty pertaining to the provided prediction. The recommended movie will not certainly fit users' preferences. Thus, they can be disappointed after watching it. To not give consideration and attention to the prediction's uncertainty may lead to unrepresentative results. In fact, the recommender in this case does not take into account all the possible cases of the user's future interests. Offering a display of confidence helps users in building trust since it can be considered as a sort of provided explanations [Shani et al. (2013)]. All these factors piled up together, explain the motivation behind this work, where we aim to perform better predictions by managing and representing the uncertainty under the framework of the belief function theory. Notations and concepts of the belief function theory fundamental for the understanding of our proposal are given in the next section.

3 The uncertain framework of belief functions

The belief function theory [Dempster (1968), Shafer (1976), Smets (1998)] offers a particularly convenient framework for reasoning under uncertainty. It allows us to quantify uncertainty in data and handle it in a flexible way. Let Θ be the frame of discernment representing a finite set of n elementary events. Such set contains hypotheses concerning the given problem. It is defined as follows: $\Theta = \{\theta_1, \theta_2, \cdots, \theta_n\}$

The power set of Θ, denoted by 2^Θ, contains all the subsets of the frame of discernment, where $2^\Theta = \{E : E \subseteq \Theta\}$. It includes both the empty set $\emptyset$ and the entire set Θ, where the empty set is the unique set having no elements.

Knowledge in the belief function theory is modeled by a basic belief assignment (bba) which quantifies the impact of a piece of evidence on the different subsets of 2^Θ. A bba is a function that assigns a value in [0, 1] to every subset E of Θ such that:

$$m : 2^\Theta \to [0,1] \text{ and } \sum_{E \subseteq \Theta} m(E) = 1. \tag{1}$$

Each mass $m(E)$, called a basic belief mass (bbm), states the part of belief that exactly supports the event E of Θ. A subset $E \subseteq \Theta$, with $m(E) > 0$, is called a focal element.

The plausibility function, denoted *pl*, quantifies the maximum amount of belief that could be given to a subset E of Θ. It is defined as [Barnett (1991)]:

$$pl(E) = \sum_{E \cap F \neq \emptyset} m(F). \tag{2}$$

The case where the *bba* has at most only one focal element different from the frame of discernment Θ corresponds to the simple support function (*ssf*). It is defined as follows [Smets (1995)]:

$$\begin{cases} m(E) = 1 - \omega, \text{for some } E \subset \Theta. \\ m(\Theta) = \omega. \end{cases} \quad (3)$$

Where E is the focal element and $\omega \in [0, 1]$.
Note that w is the degree of support of the frame of discernment Θ and 1-ω is the degree of support of E.

The *bba* that models the state of the total ignorance is called vacuous *bba* and defined such that:

$$m(\Theta) = 1. \quad (4)$$

The fusion of two *bba*'s m_1 and m_2 derived from two reliable and distinct sources of evidence can be ensured using Dempster's rule of combination denoted by ⊕. It is defined as [Dempster (1967)]:

$$(m_1 \oplus m_2)(E) = \frac{\sum_{F \cap G = E} m_1(F) \cdot m_2(G)}{1 - \sum_{F \cap G = \emptyset} m_1(F) \cdot m_2(G)}, \forall F, G \subseteq \Theta. \quad (5)$$

Where F and G are the focal elements of m_1 and m_2, and $(m_1 \oplus m_2)(\emptyset) = 0$.

While holding beliefs tends to represent knowledge, such beliefs can be used to make a decision by transforming them into a probability distribution called the pignistic probability $BetP(E)$ as follows [Smets (1998)]:

$$\text{Bet}\,P(E) = \sum_{F \subseteq \Theta} \frac{|E \cap F|}{|F|} \frac{m(F)}{(1 - m(\emptyset))}, \text{for all } E \in \Theta. \quad (6)$$

To make decisions for a given problem, the hypothesis H having the highest pignistic probability has to be selected such as:

$$H = \text{argmax}_E(\text{Bet}\,P(E)) \text{ for all } E \in \Theta. \quad (7)$$

The maximum of plausibility is another possible solution to the decision making within the belief function theory where the hypothesis that has the highest value of the plausibility function *pl* can be chosen.

4 Related work

Quality improvement of predictions and recommendations has been the subject of several researches in RSs, notably in CF strategies. Basically, CF techniques, either user-based or item-based, rely exclusively on the users' ratings to provide predictions. They compute the users' similarities or the items' similarities based on the available ratings and provide predictions. Pearson and Cosine are the widely used similarity measures in neighborhood-based CF [Khojamli and Razmara (2021)]. However, new recommendation scenarios are emerging using additional information that goes beyond the user-item ratings matrix. Yet, there have already been a reasonable amount of researches in using attributes of items as background knowledge in CF systems. We recall that item attributes or item features reflect the properties characterizing the given item. For instance, [Pappas and Popescu (2015)] have proposed to combine items contents with users' preferences for non-fiction multimedia recommendations. They investigated the case of Technology, Entertainment and Design (TED) lectures where they exploited various TED metadata and users' ratings to predict the favorite videos talks to the active user. In [Liu et al, (2021), Abdelkhalek et al. (2018)], clustering techniques have been applied and items descriptions have been integrated into the CF framework. Otherwise, [Chow et al. (2014)] proposed a hybridization of content-based and user-based signals for mobile game recommendations. They built similarity graph by analyzing common meta data between games using the concepts of Page Rank algorithm. [Lu et al. (2015)] proposed a hybrid Content-based CF approach for the news topic recommendation in Bing where a piece of news could be interpreted by rich contexts. [Li et al. (2021)] predicted users who may be interested in new projects according to the matching of project characteristics and user characteristics scoring matrix. Similarly, [Messina et al. (2017)] presented a reputation-based model with measures of QoS and cost to help users in the selection of an optimal composed service in multi-cloud environments. Their predictions were based on system measures and reputation feedbacks provided by the customers. [Afoudi et al. (2021)] created a hybrid recommender framework that combines CF with Content Based Approach and Self-Organizing Map neural network technique. Another hybridization strategy has also been proposed by [Channarong et al. (2022)], where they integrated the bidirectional-encoder-representations-from-transformers (BERT) technique to both Content-based and CF to model user behavior sequences.

Getting suitable predictions and personalized recommendations is the main concern of RSs. However, the issue of trustworthiness in another fundamental challenge that can deeply affect the success of these systems [Abdelkhalek (2017)]. For instance, [Rrmoku et al. (2022)] have proposed to use the Naive Bayes classifier to enhance the process of recommendations as well as the trustworthiness confidence of the users. In such work, Social Network Analysis (SNA) metrics have been adopted and studied. In the same context, a trust-aware recommendation method based on deep sparse autoencoder has been proposed in [Ahmadian et al. (2022a)]. More specifically, they opted for a probabilistic model to evaluate the rating profiles of users. Their proposed mechanism consists in selecting the users' implicit ratings based on a reliability measure. The selected ratings and trust statements are then the input data of deep sparse autoencoder to generate the users' latent feature. Similarly, a new recommendation model is proposed in [Ahmadian el al, (2022b)] using deep neural networks where a sparse autoencoder is used to extract latent features from user-user trust relationships and user-tag matrices. The extracted latent features are then considered to calculate similarities between users and provide predictions accordingly. Later, [Ahmadian el al, (2022c)] incorporated temporal reliability and confidence measures to identify ineffective users from neighbors set and improve the recommendation performance.

We mention that these works rely on social information and implicit ratings such as the observation of the users' purchase history, comments or click streams. However, in our work, we rather focus on explicit users' ratings such as star rating expressed by the users.

Recently, some works have been concentrated towards the belief function theory in RSs. For instance, a new fast combination method, called modified rigid coarsening (MRC) has been introduced in [Dong et al. (2018)] based on hierarchical decomposition (coarsening) of the frame of discernment. Another method for combining information about users' preferences based on the belief function theory has been proposed in [Nguyen et al. (2017)] to deal with highly conflicting mass functions. In [Troiano et al. (2015), Troiano et al.(2017)], authors used this theory to analyze the relationships between users and content they enjoy, looking at the items characteristics and the users' demographic features. In [Abdelkhalek et al. (2017)], authors have proposed a recommendation approach, which extends the standard user-based CF under the belief function framework. On the other hand, an item-based CF has been developed within the BFT in [Abdelkhalek et al. (2016)] where the selected similar items have been considered as different pieces of evidence contributing to the final prediction. Authors in [Abdelkhalek et al. (2017)] have adopted a discounting technique to the standard item-based CF using the BFT tools to quantify the reliability of each similar item. The Evidential c-Means technique has been adopted in [Abdelkhalek et al. (2017)] to cluster items based on their ratings. Given a target item, the rating prediction consists in the average of the ratings corresponding to the same clusters members. An extension of such work has been proposed in [Abdelkhalek et al. (2017)] to consider also the neighborhood formation based only on the users' past ratings. We mention that this proposed work is an extension of some preliminary works [Abdelkhalek et al. (2016), Abdelkhalek (2017), Abdelkhalek et al. (2019)] where we focused uniquely in individual CF approaches. Moreover, a very preliminary work has been performed in [Abdelkhalek et al. (2018)] where only some numerical features were studied. In the same spirit, [Bahri et al. (2022)] proposed a rule-based CF model dealing with evidential data and depending on the user's context and associated rules. [Ahmadian et al. (2020)] developed a social RS based on reliable implicit relationships among users. They created a connecting graph through the computation of implicit relationships using the belief function theory. A scalable friend recommendation framework in social networks has been conceived in [Cheng et al. (2019)] under the belief function theory. They performed a deep analysis on how the past observations affect the friend's selection in social networks and proposed to incorporate importance degree and reliability of evidence. [Vo et al. (2021)] integrated a recommendation framework combining the belief function theory, word embedding, and k-means clustering for user profiling problem. The proposed framework captures semantics of words in user corpus and particularly deals with uncertainty pervaded in the short user texts. A prediction of the potential high-entropy alloys (HEAs) under uncertainty has been performed by [Ha et al. (2021)] in the context of RSs based on the elemental substitution method. More recently, an evidential multi-criteria CF method for hotel recommendations has been proposed in [Le et al. (2022)]. Particularly, they integrated matrix factorization into a deep learning model to predict the multi-criteria ratings, which are aggregated using the belief function tools.

In contrast to these approaches, we propose in this paper, an evidential hybrid item-based CF for movies recommendations. The proposed approach explores the contents of items in addition to the users' ratings to improve the recommendation's performance while handling uncertainty pervaded throughout the recommendation process.

5 Evidential hybrid framework for ratings predictions

To generate personalized recommendations, different accessible data sources can be collected and exploited to deliver the final user-interface. Our aim in this work is to take advantage of both items features and users' ratings under the belief function framework to improve the quality of recommendations. We want to be able to infer more reliable and trustworthy predictions based on these two kinds of information sources. In this section, we describe our evidential hybrid approach, which extends state of the art CF methods and incorporates items contents and uncertainty at the same time.

5.1 Evidential CF based on items contents clustering (ECF-ICC)

Our hybrid CF approach under the belief function theory tends to join both users' ratings and items contents for the ratings predictions under uncertainty. First, an adapted evidential clustering model for the items is built based on their corresponding contents. Then, depending on each item's cluster, the k-nearest items are selected as the pieces of evidence contributing in the prediction process. Global evidence of the neighbors is finally aggregated to get an overall information about the user' future preferences. These predictions are presented on his own personalized user interface.

The whole process of our proposed recommendation approach is schematically given in Fig. 1.

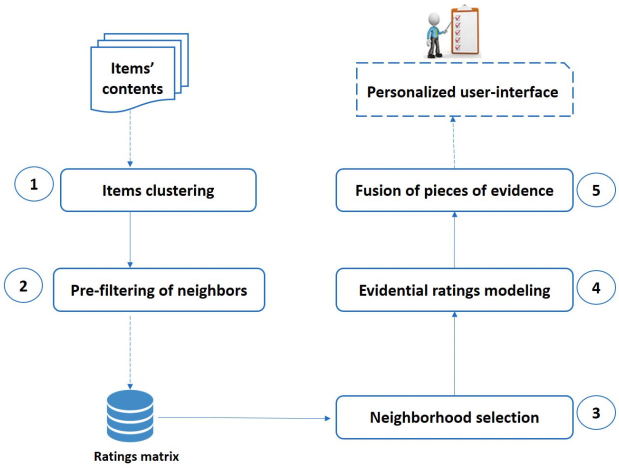

Figure 1: The evidential hybrid framework for ratings predictions

In the following, we describe the evidential hybrid CF framework, and we detail the different steps of the recommendation process.

5.1.1 Items Clustering

Exploiting auxiliary information aside from users' ratings would improve the prediction quality for CF. Therefore, we propose in this step to exploit the items contents to pick up a set of candidates from which neighbors should be selected. The key idea is to partition the items of a CF system based on their features while taking into account the uncertainty occurring during the clusters' assignment. Actually, when the uncertainty intervenes at clusters' assignment, while using the belief function theory, the output is referred to as credal partition. We assume that this additional flexibility considered during the clusters' assignment allows us to gain a deeper insight in the correlation between items, which may improve the quality of the produced predictions. Working under the belief function framework, we opt for the Relational Evidential c-Means (RECM) [Masson and Denoeux (2009)]. This unsupervised machine learning technique allows us to obtain a credal partition of the items. It enables a given item to be assigned to multiple clusters, or rather multiple partitions of clusters. Hence, this step is intended for evidential clustering due to its advantages and ability to cluster data under uncertainty. Before generating clusters, the similarities between items are derived from their information content. For many real domains, items are described by a mixture of numerical and categorical attributes. To this end, we propose to rely on the similarity measure proposed by Huang [Huang et al. (1997)] in order to handle mixed data sets. It consists of a combined similarity measure on both numerical and categorical attributes based on a combination coefficient σ. Such similarity measure is defined as follows:

$$D = \sigma D_{num} + (1 - \sigma) D_{cat} \tag{8}$$

Where D_{num} and D_{cat} are respectively the square Euclidean distance between numerical attributes and the simple matching coefficient between categorical attributes. For numerical attributes, we propose to employ the normalization scheme presented in [Witten et al.(2016)] before performing the similarity computation. For each item j, the snormalized values of their numerical attributes a are obtained as following:

$$a'_j = \frac{a_j - \underset{a_j}{min}}{\underset{a_j}{max} - \underset{a_j}{min}} \tag{9}$$

Once the distances between each pair of items are computed, we obtain a dissimilarity matrix containing the pairwise dissimilarities between these objects. Then, we adapt the RECM technique to group the items into different clusters. For the evidential cluster building, we define $\Theta_{clus} = \{C_1, C_2, \dots, C_n\}$ where n is the number of clusters. The choice of n can be ensured using cross-validation methods. At this level, each item can belong to all clusters with a degree of belief. In other words, based on the dissimilarity matrix already built, we allocate, for each item, a mass of belief to any subset of Θ_{clus}. First, the initial credal partition denoted by $M^{(0)}$ is randomly generated. Then, the final credal partition M is determined by minimizing a given objective function. For more details about the generation of the credal partition and the whole optimisation process within RECM, specific demonstrations can be found in [Masson and Denoeux (2009)]. Note that our goal until now is to express our beliefs regarding the class-membership of items, in the form of basic belief assignment. The resulting structure of the evidential clustering, which is the credal partition, represents the first level of the belief function theory where uncertainty is represented. More details about the evidential clustering can be found in [Denoeux and Orakanya (2016)].

In order to perform the items partition, we exploit the second level of the BFT, which is intended to make decisions. Thus, as each item is characterized by a *bba* on clusters, we apply the pignistic transformation *BetP* using Eq. (6) for each cluster $C_i \in \Theta_{clus}$. The cluster C_i holding the highest value of pignistic probability is then chosen and each item is assigned to its corresponding partition. Note that in some cases where the pignistic probability values are equal, the plausibility function *pl* in Eq. (2) can be computed and each item can be assigned to the cluster having the highest plausibility value.

5.1.2 Pre-filtering of neighbors

Until now, we have analyzed the items previously rated by the users and we have built a cluster model based on the characteristics of these items. After clustering has been carried out, we should pick out the set of candidates items from which neighbors should be extracted. In this phase, only the items belonging to the same cluster as the target item are extracted and selected as candidates neighbors. According to the obtained clusters, the user-item ratings matrix would be divided into n partitions.

This step can be considered as a global filtering step where the most likely candidates are kept for the neighborhood selection as illustrated in Fig. 2.

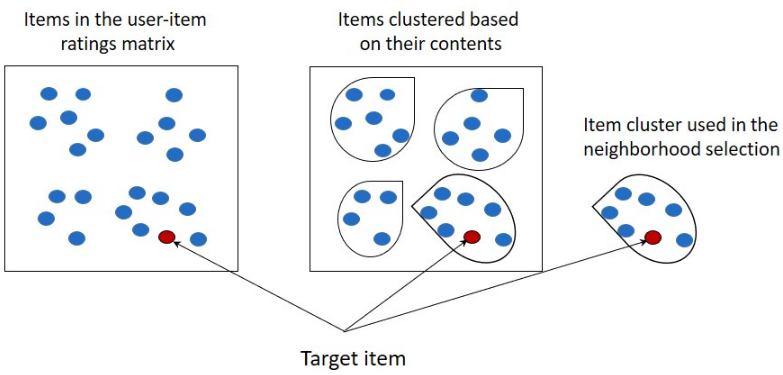

Figure 2: Pre-filtering of neighbors based on items contents

5.1.3 Neighborhood selection

After identifying the candidate sets of items based on their features similarities, the neighborhood selection is performed based on the ratings given by the users for these items. In other words, to select the *k*-nearest neighbors, we rely in this step on a subset of the user-item matrix which contains the ratings related to the candidates. In order to perform the selection, we compute the similarity between the target item and the other items in the same cluster. In our approach, we opt for the euclidean distance and we follow the similarity strategy proposed in [Sarwar et al.(2001)] by isolating the co-rated items as shown in Fig. 3.

To compute the similarity between two items, the first step consists in isolating the users who rated both of these items. Once common users are extracted, the similarity measure is applied. Formally, the distance between the target item I_t and each item I_j is computed as follows:

$$\mathrm{Dist}(I_t, I_j) = \frac{1}{|U(t,j)|}\sqrt{\sum_{U \in U(t,j)} \left(r_{i,t} - r_{i,j}\right)^2} \tag{10}$$

Where $|U(t,j)|$ denotes the number of users that rated both the target item I_t and the item I_j, $r_{i,t}$ 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 to be considered later in the next step.

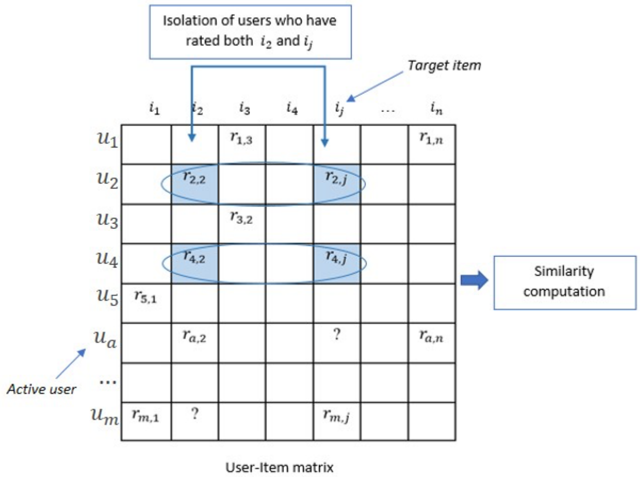

Figure 3: Similarity computation based on common users

5.1.4 Evidential ratings modeling

In this phase, we formalize our intuition behind the proposed evidential approach which covers uncertainty throughout the prediction process. Using the terminology of the belief function theory, we can define the frame of discernment corresponding to the ratings prediction situation as following: $\Theta_{pref} = \{\theta_1, \theta_2, \cdots, \theta_L\}$. It is defined as a rank-order set of L preference labels (i.e. ratings), where $\theta_p < \theta_l$ whenever $p < l$. CF predictions are usually generated by evidence gathered from the selected neighbors. Actually, despite the fact that the ratings of the selected neighbors can increase our belief about the most probable one, it does not imply that such knowledge is fully certain. That is why, we use the belief function theory in order to emphasize the presence of uncertainty in the neighbors' ratings. We assume that each similar item contributing to the final prediction can be considered as a distinct piece of evidence supporting a particular belief about the predicted rating. In view of such assumption, we start by exploring the ratings corresponding to these pieces of evidence and the related *bba*'s are then produced accordingly.

Inspired by the E*k*NN formalism [Denoeux et al. (1995)], these *bba*'s are generated overeachratingprovidedforthe*k*-similaritemsaswellasthewholeframeofdiscernment Θ_{pref}. Based on the distances computed during the items' neighborhood selection, we can represent this *bba* as a simple support function defined as:

$$m_{I_j} = \begin{cases} m_{I_j}(\{\theta_p\}) = \omega_{\theta_p} \\ m_{I_j}(\Theta_{pref}) = 1 - \omega_{\theta_p} \end{cases} \tag{11}$$

Where I_j is the neighbor of the target item I_t such that: $j=\{1,\cdots,k\}$ and $\omega_{\theta p}$ is the belief committed to the rating θ_p of I_j such that:

$$\omega_{\theta_p} = \alpha e^{-\left(\gamma_{\theta_p}^2 \times \mathrm{dist}(I_t, I_j)^2\right)} \tag{12}$$

Where $dist_{tj} = dist(I_t, I_j)$ is the distance between the items I_t and I_j already computed in the third phase, α and $\gamma_{\theta p}$ are two parameters such that $0 < \alpha < 1$ and $\gamma_{\theta p} > 0$. Based on [Denoeux et al.(1995)], we can define $\gamma_{\theta p}$ as the inverse of the mean distance between all the training patterns belonging to the rating θ_p. When it comes to the choice of α, it has been proven by Denoeux [Denoeux et al.(1995)] that a fixed value of 0.95 yields good classification results. The generated simple support system indicates that each neighbor of the target item has two possible hypotheses. The first one corresponds to the value of its provided rating while the rest of the committed belief is allocated to the frame of discernment Θ_{pref}. Therefore, the focal elements of the belief function are the sratings provided by the *k*-similar items and Θ_{pref}. By taking the *k*-most similar items as independent sources of evidence regarding the rating of I_t, each neighbor can be represented by a basic belief assignment. Thus, *k* different *bba*'s can be generated for each neighbor as illustrated in Fig. 4.

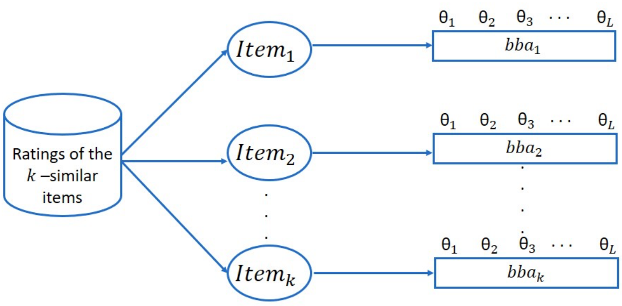

Figure 4: The evidential modeling of ratings for the k-similar items

5.1.5 Fusion of pieces of evidence

The aim of a RS is to generate the output interface to the users in terms of predictions. In this phase, we shed light on the prediction process of our contribution under the belief function framework. Traditionally, the user at this level gets a predicted rating that indicates a score of his future degree of satisfaction given an item. Nevertheless, in real world problems, the RS cannot draw any certain inference about the future rating which leads to very generic results and specifically not user-centric ones. However, recent studies found that putting users in control of their recommendations inspires trust in the system and results in more positive evaluations [Harper et al.(2015)]. This is the aim of this final step. Actually, in the previous step, we showed how to generate a *bba* for each similar item. Now, we describe how to aggregate these *bba*'s in order to synthesize the final belief about the rating of the target item. The *bba*'s of the *k*-similar items can be combined based on Dempster's rule of combination using Eq. (5) such that:

$$m_{I_t} = m_{I_1} \oplus m_{I_2} \oplus \cdots \oplus m_{I_k} \tag{13}$$

Thus, the induced final *bba* encodes the evidence of the *k*-nearest neighbors regarding the rating that should be provided to the target item. This aggregation of the contribution of each neighbor may lead to a more accurate and reliable prediction. Overall, this phase starts by exploring the *k* bba's already generated for each similar item and performs an aggregation of the different beliefs as illustrated in Fig. 5.

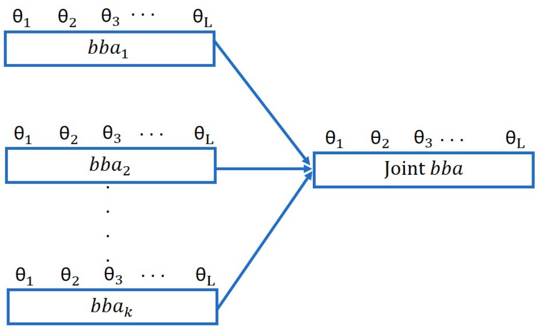

Figure 5: The evidential combination process for the k-similar items

The most important advantage of this prediction process modeling under the belief function theory is that it reflects more credible and intelligible results. To provide recommendations, we compute the pignistic probability and we take the rating having the greatest value, as it is more likely to be the potential future one. Furthermore, we generate for the active user his other possible preferences as illustrated in Fig. 6.

	★	★★	★★★	★★★★	★★★★★	Θ_{pref}
 Active User	-	0.114	-	-	0.846	0.040

Figure 6: Example of Evidential Predictions

According to the results illustrated in this example (Fig. 6), we note that the rating corresponding to the target movie has different possible alternatives to be shown to the active user. The given results predict for her that finding the target movie an interesting one is the most likely case where the evidence that she will extremely like it has a value of 0.846. However, these predictions support also some belief of 0.114 that she will be slightly interested in such movie. The rest of the belief is allocated to the frame of discernment Θ_{pref} which reflects the case of total ignorance.

6 Experimental study

In this section, we describe the experimental protocol and we present the experimental results of our proposed approach.

6.1 Data sets

Two real world data sets have been adopted in order to evaluate our proposal. We opt for the well-known MovieLens[1] data set which contains 1682 movies rated by 943 users. This dataset was collected by the GroupLens Research Group [2] resulting in 100.000 ratings which are integer values between 1 ("bad") and 5 ("excellent"). Each user in the dataset has rated at least 20 movies. The movies involved in the MovieLens database are only described by their ids, titles and genres: (Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western). That is why, we rely also on a second dataset which extends the movies description by additional features provided by the Internet Movie Database - IMDb [3] and Rotten Tomatoe[4] in order to enrich the movies characteristics. Such dataset is released in the framework of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011) [5] at the 5th ACM Conference on Recommender Systems (RecSys 2011)[6]. It links the movies of MovieLens dataset with their corresponding web pages at these two movie review systems. The obtained dataset includes different attributes that can be either numerical (e.g. year, imdb-scores, rotten-scores, top critics, numbers of reviews) or categorical (e.g. genres, directors, actors, countries of origin, filming locations).

6.2 Experimental Process

We adopted the methodology proposed in [Su et al. (2008)] for conducting our experiments. First, the movies rated by the 943 users were ranked according to the number of their ratings such as:

$$Nb_{\text{user}}(\text{ movie }_1) \geq Nb_{\text{user}}(\text{ movie }_2) \geq \cdots \geq Nb_{\text{user}}(\text{ movie }_{1682})$$

where $Nb_{user}(movie_i)$ is the number of users who rated the $movie_i$.

Then, 10 different subsets are extracted by increasing progressively the number of the missing rates from 53.8 % to 95.9%. Finally, we obtain various subsets containing

1 http://movielens.org
2 https://grouplens.org
3 http://www.imdb.com/
4 http://www.rottentomatoes.com
5 http://ir.ii.uam.es/hetrec2011
6 http://recsys.acm.org/2011

a particular number of ratings provided by the 943 users for 20 different movies in the data set. We call these fractions the 20-movies subsets. Note that the more the available ratings are, the less sparse the data set is. For each subset, we randomly extract 20% of the available ratings as a test set and the remaining 80% are used as a training set.

6.3 Evaluation measures

In order to emphasize the performance of our evidential apporach, we propose to use two evaluation metrics which are commonly used in the RSs area: The *Mean Absolute Error* (MAE) which measures how close the predictions are to the user's ratings for each movie. This measure considers the average of the absolute deviation between each prediction and real rating for all held-out preference degrees of users in the testing set. Note that the MAE values are between 0 and 4 in our case and that the lower these values are the more accurately the recommendation engine predicts users' ratings. Mathematically:

$$MAE = \frac{1}{\parallel T \parallel} \sum_{r_{i,j} \in T} \left|\hat{r}_{i,j} - r_{i,j}\right| \tag{14}$$

Where $r_{i,j}$ corresponds to the actual rating for the user U_i on the item I_j and $\hat{r}_{i,j}$ corresponds to the predicted value. $\parallel T \parallel$ is the total number of the predicted ratings over all the users.

Since different numerical rating scales can be adopted in RSs, another common method is generally employed in the evaluation of RSs namely the *Normalized Mean Absolute Error (NMAE)*. It corresponds to a normalized version of the MAE aiming to define errors as percentages of full scale. Hence, a minimum value of 0 demonstrates a good prediction while a maximum value of 1 reflects a bad prediction result. It is computed as follows:

$$NMAE = \frac{\sum_{i,j} \in T\left|r_{i,j} - r_{i,j}\right|}{\parallel T \parallel (r_{max} - r_{min})} \tag{15}$$

where r_{max} is the upper bound of the ratings and r_{min} is the lower bound. Using NMAE may lead to an easier comparison error between different RSs.

In addition to MAE and NMAE, we have also used the *Root Mean Squared Error (RMSE)*. It is another evaluation method commonly used in RSs. In contrast to MAE and NMAE which correspond to a simple average over the differences between the real ratings and the predicted ones, the errors rate computed by RMSE are not treated on a similar way. That is to say, RMSE specifically assigns a relatively high weight to large errors unlike the MAE in which the errors are weighted equally. Formally, the RMSE is computed as follows:

$$RMSE = \sqrt{\frac{\sum_{(r_{i,j} \in T} \left|\hat{r}_{i,j} - r_{i,j}\right|)^2}{\parallel T \parallel}} \tag{16}$$

In addition to MAE, NMAE and RMSE, we are also interested in the precision measure which is another popular metric to evaluate the performance of RSs. It is used to evaluate how the recommendations help the active user in distinguishing good items from bad items. It is defined as follows:

$$\text{Precision } = \frac{TP}{TP+FP} \tag{17}$$

where TP denotes True Positive (a relevant item has been correctly recommended) and FP is False Positive (an irrelevant item has been incorrectly recommended). Using MoviesLens dataset, a common way to differentiate the relevant and irrelevant items is to mark the items with rating between 4 and 5 as relevant and those rated below 4 out of 5 as irrelevant to the user [Herlocker et al. (2004)]. In the same context, we have also opted for two other evaluation metrics namely, Recall and F-measure.

The Recall metric computes the portion of favored items that were suggested for the active user relative to the total number of the objects actually collected by him. It is defined as follows:

$$\text{Recall } = \frac{TP}{TP+TN} \tag{18}$$

The F-measure represents the weighted harmonic mean of the precision and recall of the test. It is defined as follows:

$$F - \text{ measure } = \frac{2 \cdot \text{ Precision } \cdot \text{ Recall}}{\text{Precision } + \text{ Recall}} \tag{19}$$

Obviously, the higher the Precision, Recall and F-measure are, the more effective the recommendation approach is.

6.4 Experimental results

We present, in this sub-section, the experimental results highlighting the effectiveness of our evidential CF based on items contents clustering that we denoted by ECF-ICC. We carry out several experiments over the extracted 20-movies subsets while switching each time some setting parameters. Authors in [Denoeux et al. (1995)], devoted that the parameter α used in the *bba*'s generation, does not have a great influence on the prediction performance and that a value of 0.95 leads to good results. Thus, as in [Denoeux et al. (1995)], we set α=0.95 over all our conducted experiments. The other required parameters are the similarity weight σ and the number of clusters n used during the items contents clustering. For the neighborhood size k, it may not be fixed in our clusteringbased approach since it fully depends on the number of items in the clusters. For instance, one cluster may contain 7 items while another one may have 15 items and so on. For the number of clusters, we recall that the choice is performed using cross-validation methods.

In order to find the optimal coefficient σ of the similarity computation of items contents in Eq. (8), we perform a series of experiments over the 10 extracted subsets by varying the combination coefficient from 0 to 1 with a step of 0.1. Fig. 7 and Fig. 8 show how the MAE and Precision results related to ECF-ICC vary with different values of the combination coefficient σ. We can observe through the two figures that using either MAE or Precision, the proposed approach has almost the same behavior over the different coefficient values. An optimal prediction and recommendation performance is achieved when σ attains a value of 0.5 (for both MAE and Precision). That is why, in all our experiments, we take σ=0.5 and α=0.95. Moreover, we observe that the worst results were obtained when σ = 0 and σ = 1. This can be interpreted by the fact that using only numerical features or, alternatively, only categorical features may badly affect the items clustering process, leading consequently to a poor predictions and recommendations quality.

Figure 7: MAE vs. coefficient σ

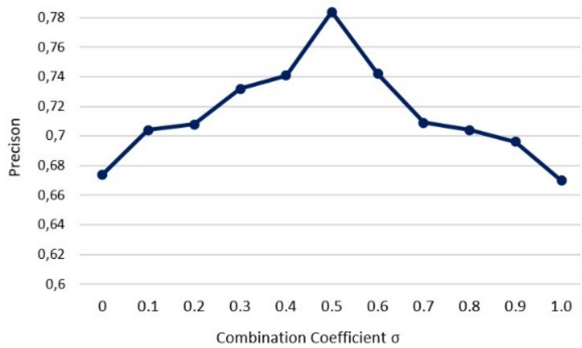

Figure 8: Precision vs. coefficient σ

Let us now look at the impact of the numbers of clusters *n* in the recommendation performance. As shown in Fig. 9, MAE varies with different values of *n* ranging from 2 to 5 clusters. For the first and fourth subsets (having respectively a sparsity degree of 53%, and 62.70%), the best results were obtained when using 3 clusters. On the other hand, most of the best results were achieved when *n*=2, such as the case of the subsets corresponding to sparsity of 72.70%, 80.8%, 87.4% and 95.90%. Using 4 and 5 clusters have also led to good results at sparsity levels of 56.83%, 59.80%, 68.72% and 75%. The suitable number of clusters depends on the sparsity degree of the current subset. Different behaviors are observed for different sparsity degrees. However, we can notice that in some cases, most of the best results were achieved when *n*=2 and *n*=3. This could be explained by the lack of available ratings in these subsets. Besides, increasing the number of clusters implies having small cluster sizes, which leads to a small number of neighbors contributing in the prediction process. Similarly, the precision results are reported in Fig. 10 to describe the behavior of the recommendation approach according to different sparsity degrees and different number of clusters *n*.

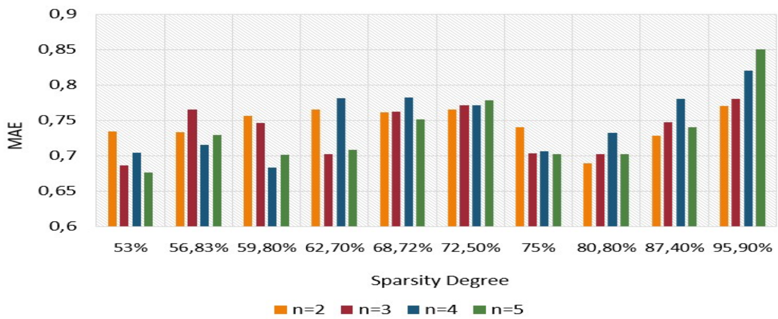

Figure 9: MAE vs. number of clusters n

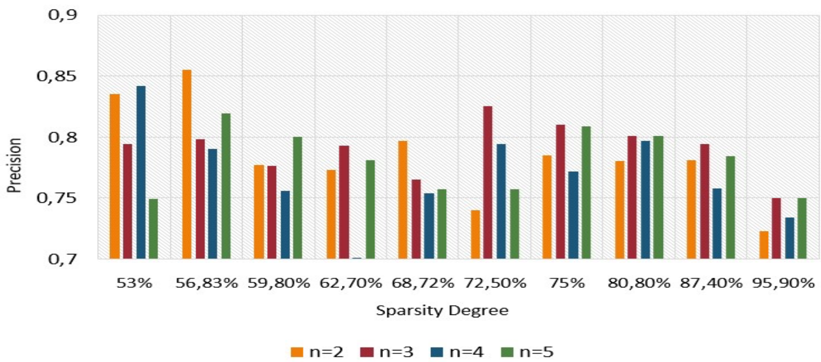

Figure 10: Precision vs. number of clusters

The performance of the proposed approach is compared against seven traditional item-based CF systems under certain and uncertain frameworks. First, we consider the two traditional neighborhood-based CF approaches commonly used in RSs: Pearson itembased CF (P-CF) and Cosine item-based CF (C-CF) approaches. Further, we consider the two neighborhood-based CF approaches under the belief function theory namely, the Evidential item-based CF (E-CF) [Abdelkhalek et al. (2016)] and the Evidential Discounting-based item-based CF (ED-CF) [Abdelkhalek et al. (2017)]. These two approaches are based on the Evidential *k*-Nearest Neighbors classifier. We also compare our method to the Evidential Model-based CF (EM-CF) [Abdelkhalek et al. (2017)] where items are clustered based on their ratings and predictions are performed as an average rating over the items in the same cluster. The Evidential CF approach joining Model-based and Neighborhood-based CF (EMN-CF) [Abdelkhalek et al. (2017)] is also considered in our experiments. All the CF approaches mentioned above rely exclusively on users' ratings to make predictions. That is why, we also compare the results to those obtained using ECF-CB, an evidential CF approach based on items' contents [Abdelkhalek et al. (2018)] where only numerical features have been considered.

While running our experiments, we use different numbers of clusters: n=2, n=3, n=4 and n=5. For each selected cluster, different neighborhood sizes were tested and the average results were computed. Finally, the results obtained for the different number of clusters used in the experiments are also averaged. More specifically, we compute the MAE and the Precision measures for each value of n and we report the overall results. Table 1 recapitulates the detailed results considering different sparsity degrees and the overall MAE, NMAE and RMSE values for all the compared methods. For the Precision, Recall and F-measures results, they are presented in Table 2.

Overall, on the 20-movies datasets, in terms of MAE, ECF-ICC achieves better results corresponding to a value of 0.740 compared to E-CF, ED-CF, EM-CF, EMN-CF, ECF-CB, C-CF and P-CF having respectively 0.809, 0.801, 0.793, 0.784, 0.788, 0.914 and 0.925. Besides, the proposed evidential approach allows an improvement over the five standard evidential item-based CF approaches by acquiring, in average, the highest overall precision over the 10 subsets (0.782 compared to 0.733, 0.731, 0.750, 0.755, and 0.757). Similarly, the new approach outperforms the traditional item-based CF working under a certain framework with a value of 0.782 compared to 0.706. These results show that integrating uncertainty throughout the prediction process while using both items' contents and users' ratings improve the recommendation performance under certain and uncertain frameworks. For all the item-based CF approaches, we observe a remarkable drop for the last subset corresponding to the highest sparsity degree (95.9%). These results can be explained by the lack of available ratings, which we tend to tackle in future works.

7 Conclusion and future works

In this paper, we have proposed a new evidential Collaborative Filtering approach for ratings predictions. The proposed approach integrates both users' ratings and items' contents under an uncertain framework. The use of the belief function theory allowed us to quantify both the belief regarding the ratings assigned to the similar items as well as the uncertainty pervaded in the final predictions. First, an evidential clustering process is proposed to group items based on their contents. Once the cluster model is built, the prediction task is ensured based on the ratings of the selected similar items. The neighboring items are considered as independent sources of information, their evidence is represented and fused to successfully predict the user's ratings. A main advantage of our method compared to the existing approaches is that it operates in an uncertain environment where all the information from the neighbors can be combined effectively leading to an evidential representation of the predictions. Our goal was to infer more reliable and trustworthy predictions to clarify and explain the recommendations provided to the active user. We assume that this intelligibility promotes the users confidence towards the system and provides for him a more convenient framework. This would be fully beneficial to the RSs field especially, nowadays where reliability becomes a crucial parameter to attend the user's satisfaction.

The proposed approach achieves a good prediction performance with a precision value of 0.782 and a Recall value of 0.792. However, the presented approach is not able to provide recommendations for new users added to the system. This problem is referred to as cold start problem, which we aim to handle in future works. Other interesting avenues for future works have to be mentioned. To start, it would be interesting to manage the uncertainty pervaded in the ratings assignment under a multi-criteria aspect. Moreover, we suggest handling uncertainty of all the users' ratings provided in the beginning of the recommendation process instead of only the more similar ones. We can also extend this research to include other types of users' ratings such as implicit feedbacks. It would also be fruitful to develop new evidential approaches by exploiting different kinds of additional information such as social networks and users' demographic in order to enhance the recommendation performance.

Table 1: Overall MAE, NMAE and RMSE

Metrics	Sparsity	P-CF	C-CF	E-CF	ED-CF	EM-CF	EMN-CF	ECF-CB	ECF-ICC
MAE		0.839	0.824	0.751	0.711	0.749	0.740	0.787	0.700
NMAE	53.00%	0.209	0.206	0.187	0.177	0.187	0.185	0.196	0.175
RMSE		1.231	1.158	1.089	1.060	1.021	0.930	1.012	0.991
MAE		0.936	0.870	0.840	0.802	0.800	0.851	0.810	0.735
NMAE	56.83%	0.234	0.217	0.210	0.200	0.200	0.212	0.202	0.183
RMSE		1.291	1.215	1.158	1.111	1.102	1.087	1.080	1.059
MAE		0.863	0.825	0.761	0.836	0.747	0.779	0.780	0.721
NMAE	59.80%	0.215	0.206	0.190	0.209	0.186	0.194	0.195	0.180
RMSE		1.256	1.198	1.135	1.121	1.108	1.188	1.181	1.010
MAE		0.905	0.876	0.763	0.743	0.793	0.750	0.748	0.739
NMAE	62.70%	0.226	0.219	0.190	0.185	0.198	0.187	0.187	0.184
RMSE		1.267	1.232	1.092	1.102	1.080	1.116	1.109	1.099
MAE		0.990	1.000	0.831	0.802	0.845	0.793	0.763	0.764
NMAE	68.72%	0.247	0.250	0.207	0.200	0.211	0.198	0.190	0.191
RMSE		1.367	1.351	1.184	1.121	1.097	1.244	1.102	1.016
MAE		0.976	0.917	0.851	0.843	0.800	0.845	0.785	0.771
NMAE	72.50%	0.244	0.229	0.212	0.210	0.200	0.211	0.196	0.192
RMSE		1.348	1.272	1.184	1.151	1.263	1.216	1.193	1.092
MAE		0.943	0.877	0.744	0.736	0.733	0.703	0.840	0.712
NMAE	75.00%	0.233	0.219	0.186	0.184	0.183	0.175	0.210	0.178
RMSE		1.270	1.212	1.187	1.191	1.182	1.025	1.189	1.099
MAE		0.927	0.848	0.718	0.723	0.762	0.711	0.745	0.706
NMAE	80.80%	0.231	0.212	0.179	0.180	0.190	0.177	0.186	0.176
RMSE		1.265	1.179	1.079	1.063	1.064	1.033	1.169	1.074
MAE		0.958	0.978	0.840	0.839	0.873	0.798	0.754	0.748
NMAE	87.40%	0.239	0.244	0.210	0.209	0.218	0.199	0.188	0.187
RMSE		1.309	1.334	1.180	1.163	1.133	1.093	1.086	1.074

MAE		0.913	1.130	0.991	0.978	0.830	0.870	0.870	0.805
NMAE	95.90%	0.228	0.282	0.247	0.244	0.207	0.217	0.217	0.201
RMSE		1.217	1.527	1.445	1.381	1.312	1.332	1.218	1.074
Overall MAE		0.925	0.914	0.8091	0.801	0.793	0.784	0.788	**0.740**
Overall NMAE		0.231	0.228	0.202	0.200	0.198	0.195	0.197	**0.185**
Overall RMSE		1.282	1.267	1.173	1.146	1.136	1.121	1.133	**1.058**

Table 2: Overall Precision, Recall and F-measure

Metrics	%	P-CF	C-CF	E-CF	ED-CF	EM-CF	EMN-CF	ECF-CB	ECF-ICC
Precision		0.737	0.739	0.760	0.748	0.740	0.755	0.736	0.815
Recall	53.00%	0.691	0.686	0.732	0.714	0.726	0.749	0.762	0.804
F-measure		0.730	0.729	0.759	0.742	0.757	0.787	0.783	0.805
Precision		0.737	0.739	0.760	0.748	0.740	0.755	0.736	0.815
Recall	56.83%	0.655	0.677	0.717	0.702	0.697	0.706	0.700	0.807
F-measure		0.693	0.706	0.737	0.724	0.717	0.729	0.717	0.810s
Precision		0.752	0.749	0.770	0.711	0.785	0.732	0.753	0.777
Recall	59.80%	0.726	0.718	0.725	0.698	0.731	0.759	0.781	0.811
F-measure		0.738	0.733	0.746	0.704	0.757	0.745	0.766	0.793
Precision		0.746	0.745	0.763	0.775	0.782	0.764	0.830	0.762
Recall	62.70%	0.705	0.709	0.770	0.772	0.775	0.775	0.777	0.787
F-measure		0.724	0.726	0.766	0.773	0.778	0.769	0.802	0.774
Precision		0.707	0.690	0.741	0.787	0.752	0.757	0.811	0.768
Recall	68.72%	0.657	0.699	0.764	0.778	0.747	0.781	0.793	0.803
F-measure		0.681	0.694	0.752	0.782	0.749	0.768	0.801	0.785
Precision		0.732	0.733	0.735	0.740	0.813	0.743	0.780	0.779
Recall	72.50%	0.639	0.665	0.680	0.687	0.784	0.700	0.748	0.786
F-measure		0.682	0.697	0.706	0.712	0.798	0.720	0.763	0.782

Precision		0.752	0.745	0.780	0.783	0.805	0.792	0.829	0.794
Recall	75.00%	0.737	0.698	0.718	0.718	0.751	0.853	0.802	0.779
F-measure		0.744	0.720	0.747	0.748	0.777	0.821	0.815	0.786
Precision		0.729	0.718	0.778	0.821	0.755	0.779	0.733	0.794
Recall	80.80%	0.666	0.673	0.696	0.782	0.746	0.750	0.756	0.785
F-measure		0.696	0.694	0.734	0.801	0.750	0.764	0.744	0.789
Precision		0.665	0.653	0.707	0.740	0.730	0.737	0.797	0.779
Recall	87.40%	0.590	0.737	0.747	0.763	0.744	0.590	0.696	0.779
F-measure		0.625	0.692	0.726	0.751	0.736	0.655	0.743	0.779
Precision		0.463	0.509	0.513	0.431	0.550	0.660	0.500	0.739
Recall	95.90%	0.600	0.685	0.716	0.684	0.661	0.690	0.605	0.779
F-measure		0.522	0.584	0.597	0.528	0.600	0.674	0.547	0.758
Overall Precision		0.706	0.706	0.733	0.743	0.750	0.755	0.757	**0.782**
Overall Recall		0.666	0.694	0.726	0.729	0.735	0.735	0.742	**0.792**
Overall F-measure		0.685	0.699	0.729	0.735	0.742	0.744	0.749	**0.787**

References

[Abdelkhalek (2017)] Abdelkhalek, R., 2017, August. Improving the trustworthiness of recommendations in collaborative filtering under the belief function framework. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 421–425).

[Abdelkhalek et al. (2019)] Abdelkhalek, R., Boukhris, I., & Elouedi, Z. (2019, November). Joining Items Clustering and Users Clustering for Evidential Collaborative Filtering. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 310–318). Springer, Cham.

[Abdelkhalek et al. (2016)] R. Abdelkhalek, I. Boukhris and Z. Elouedi, Evidential item-based collaborative filtering, *International Conference on Knowledge Science, Engineering and Management*, (Springer, 2016), pp. 628–639.

[Abdelkhalek et al. (2017)] R. Abdelkhalek, I. Boukhris and Z. Elouedi, A New User-Based Collaborative Filtering Under the Belief Function Theory, *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, (Springer, 2017), pp. 315–324.

[Abdelkhalek et al. (2017)] R. Abdelkhalek, I. Boukhris and Z. Elouedi, Assessing Items Reliability for Collaborative Filtering Within the Belief Function Framework, *International Conference on Digital Economy*, (Springer, 2017), pp. 208–217.

[Abdelkhalek et al. (2017)] R. Abdelkhalek, I. Boukhris and Z. Elouedi, A new evidential collaborative filtering: A hybrid memory- and model-based approach, *International Conference on Data Science and Knowledge Engineering for Sensing Decision Support*, (Springer, 2017), pp. 660–667.

[Abdelkhalek et al. (2017)] R.Abdelkhalek, I. Boukhris and Z. Elouedi,A ClusteringApproach for Collaborative Filtering Under the Belief Function Framework, *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, (Springer, 2017), pp. 169–178.

[Abdelkhalek et al. (2018)] R. Abdelkhalek, I. Boukhris and Z. Elouedi, An evidential collaborative filtering approach based on items contents clustering, *International Conference on Belief Functions*, (Springer, 2018), pp. 1–9.

[Afoudi et al. (2021)] Afoudi, Y., Lazaar, M. and Al Achhab, M., 2021. Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. Simulation Modelling Practice and Theory, 113, p.102–375.

[Ahmadian et al. (2022b)] Ahmadian, S., Ahmadian, M., & Jalili, M. (2022). A deep learning based trust-and tag-aware recommender system. Neurocomputing, 488, 557–571.

[Ahmadian et al. (2022c)] Ahmadian, S., Joorabloo, N., Jalili, M., & Ahmadian, M. (2022). Alleviating data sparsity problem in time-aware recommender systems using a reliable rating profile enrichment approach. Expert Systems with Applications, 187, 115849.

[Ahmadian et al. (2020)] Ahmadian, S., Joorabloo, N., Jalili, M., Ren, Y., Meghdadi, M. and Afsharchi, M., 2020. A social recommender system based on reliable implicit relationships. Knowledge-Based Systems, 192, p.105371.

[Ahmadian et al. (2022a)] Ahmadian, M., Ahmadi, M. and Ahmadian, S., 2022. A reliable deep representation learning to improve trust-aware recommendation systems. Expert Systems with Applications, 197, p.116697.

[Ambulgekar et al. (2019)] H. P.Ambulgekar, M. K. Pathak and M. B. Kokare,A Survey on Collaborative Filtering: Tasks, Approaches and Applications, *International Ethical Hacking Conference*, (Springer, 2019), pp. 289–300.

[Bahri et al. (2022)] Bahri, N., Bach Tobji, M.A. and Ben Yaghlane, B., 2022. ECFAR: A Rule-Based Collaborative Filtering System Dealing with Evidential Data. In International Conference on Intelligent Systems Design and Applications (pp. 944-955). Springer, Cham.

[Barnett 1991)] J. A. Barnett, Calculating Dempster-Shafer plausibility, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6 (IEEE, 1991), pp. 599–602.

[Behera et al. (2022)] Behera, G. and Nain, N., 2022. Trade-Off Between Memory and Model-Based Collaborative Filtering Recommender System. In Proceedings of the International Conference on Paradigms of Communication, Computing and Data Sciences (pp. 137-146). Springer, Singapore.

[Channarong et al. (2022)] Channarong, C., Paosirikul, C., Maneeroj, S. and Takasu, A., 2022. HybridBERT4Rec: a hybrid (content-based filtering and collaborative filtering) recommender system based on BERT. IEEE Access.

[Cheng et al. (2019)] Cheng, S., Zhang, B., Zou, G., Huang, M. and Zhang, Z., 2019. Friend recommendation in social networks based on multi-source information fusion. International Journal of Machine Learning and Cybernetics, 10(5), pp.1003–1024.

[Chow et al. (2014)] A. Chow, M. H. N. Foo and G. Manai, HybridRank: A Hybrid ContentBased Approach To Mobile Game Recommendations, *Workshop on New Trends in Content-based Recommender Systems*, (2014), pp. 10–13.

[Dempster (1967)] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, Vol. 38 (Institute of Mathematical Statistics, 1967), pp. 325–339.

[Dempster (1968)] A. P. Dempster, A generalization of bayesian inference, *Journal of the Royal Statistical Society*, (Springer, 1968), pp. 205–247.

[Denoeux (1995)] T. Denoeux, A K-nearest neighbor classification rule based on DempsterShafer theory, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25 (IEEE, 1995), pp. 804–813.

[Denoeux and Orakanya (2016)] Denoeux, Thierry, & Orakanya Kanjanatarakul. "Evidential clustering: a review." International symposium on integrated uncertainty in knowledge modelling and decision making. Springer, Cham, 2016.

[Dong et al. (2018)] Y. Dong, X. Li, and Y. Liu, A fast combination method in DSmT and its application to recommender system, *PloS one*, Vol. 13 (Public Library of Science, 2018).

[Ha et al. (2021)] Ha, M.Q., Nguyen, D.N., Nguyen, V.C., Nagata, T., Chikyow, T., Kino, H., Miyake, T., Denœux, T., Huynh, V.N. and Dam, H.C., 2021. Evidence-based recommender system for high-entropy alloys. Nature Computational Science, 1(7), pp.470-478.

[Harper et al. (2015)] F. M. Harper, F. Xu, H. Kaur, K. Condiff, S. Chang and L. Terveen, Putting users in control of their recommendations, *ACM Conference on Recommender Systems*, (ACM, 2015), pp. 3–10.

[Herlocker et al. (2004)] J. L. Herlocker, J. A. Konstan, L. G. Terveen and J. T. Riedl, Evaluating collaborative filtering recommender systems, *ACM Transactions on Information Systems*, Vol. 22 (ACM, 2004), pp. 5–53.

[Huang (1997)] Z. Huang, Clustering large data sets with mixed numeric and categorical values, *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Vol. 64 (Wiley Online Library, 1997), pp. 21–34.

JointConferenceonArtificialIntelligence, Vol. 95 (Morgan Kaufmann Publishers, 1995), pp. 1896–1901.

[Khojamli and Razmara (2021)] Khojamli, H. and Razmara, J., 2021. Survey of similarity functions on neighborhood-based collaborative filtering. Expert Systems with Applications, 185, p.115–482.

[Le et al. (2022)] Le, Q.H., Mau, T.N., Tansuchat, R. and Huynh, V.N., 2022. A Multi-Criteria Collaborative Filtering Approach Using Deep Learning and Dempster-Shafer Theory for Hotel Recommendations. IEEE Access, 10, pp.37281–37293.

[Li et al. (2021)] Li, Lianhuan, Zhang, Zheng, and Zhang, Shaoda. Hybrid algorithm based on content and collaborative filtering in recommendation system optimization and simulation. Scientific Programming, 2021, vol. 2021.

[Liu et al. (2021)] Liu, J., Song, J., Li, C., Zhu, X. and Deng, R., 2021, April. A Hybrid News Recommendation Algorithm Based On K-means Clustering and Collaborative Filtering. In Journal of Physics: Conference Series (Vol. 1881, No. 3, p. 032050). IOP Publishing.

[Lu et al. (2015)] Z. Lu, Z. Dou, J. Lian, X. Xie, and Q.Yang, Content-based collaborative filtering for news topic recommendation, *AAAI conference on artificial intelligence*, (AAAI, 2015).

[Masson and Denoeux (2008)] M. H. Masson and T. Denoeux, ECM: An evidential version of the fuzzy c-means algorithm, *Pattern Recognition*, Vol. 4 (Elsevier, 2008), pp. 1384–1397.

[Masson and Denoeux (2009)] M. H. Masson and T. Denoeux, RECM: Relational evidential c-means algorithm, *Pattern Recognition Letters*, Vol. 11 (Elsevier, 2009), pp. 1015–1026.

[Messina et al (2017)] Messina, F., Pappalardo, G., Comi, A., Fotia, L., Rosaci, D. and Sarné, G.M., 2017. Combining reputation and QoS measures to improve cloud service composition. International Journal of Grid and Utility Computing, 8(2), pp.142-151.

[Mohamed et al. (2019)] Mohamed, M.H., Khafagy, M.H. and Ibrahim, M.H., 2019, February. Recommender systems challenges and solutions survey. In 2019 International Conference on Innovative Trends in Computer Engineering (ITCE) (pp. 149-155). IEEE.

[Musa et al. (2020)] Musa, J.M. and Zhihong, X., 2020, April. Item based collaborative filtering approach in movie recommendation system using different similarity measures. In Proceedings of the 2020 6th International Conference on Computer and Technology Applications (pp. 31-34).

[Nguyen et al. (2014)] V. D. Nguyen and V. N. Huynh, A community-based collaborative filtering system dealing with sparsity problem and data imperfections, *Pacific Rim International Conference on Artificial Intelligence*, (Springer, 2014), pp. 884–890.

[Nguyen et al. (2017)] V. D. Nguyen and V. N Huynh, Two-probabilities focused combination in recommender systems, *International Journal of Approximate Reasoning*, Vol. 80 (Elsevier, 2017), pp. 225–238.

[Pappas and Popescu (2015)] N. Pappas and A. Popescu-Belis, Combining content with user preferences for non-fiction multimedia recommendation: a study on TED lectures, *Multimedia Tools and Applications*, Vol .74 (Springer, 2015), pp. 1175–1197.

[Ricci et al. (2015)] F. Ricci, L. Rokach and B. Shapira, Recommender systems: introduction and challenges, *Recommender Systems Handbook*, (Springer, 2015), pp. 1–34.

[Rrmoku et al. (2022)] Rrmoku, K., Selimi, B. and Ahmedi, L., 2022. Application of Trust in Recommender Systems—Utilizing Naive Bayes Classifier. Computation, 10(1), p.6.

[Sarwar et al. (2001)] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Item-based collaborative filtering recommendation algorithms, *International Conference on World Wide Web*, (ACM, 2001), pp. 285–295.

[Shafer (1976)] G. Shafer, A mathematical theory of evidence, *Princeton university*, Vol. 1 (press Princeton, 1976).

[Shani et al. (2013)] G. Shani, L. Rokach, B. Shapira, S. Hadash and M. Tangi, Investigating confidence displays for top-N recommendations, *Journal of the American Society for Information Science and Technology*, Vol. 64 (Wiley Online Library, 2013), pp. 2548–2563.

[Smets (1995)] P. Smets, The canonical decomposition of a weighted belief, *International*
[Smets (1998)] P. Smets, The transferable belief model for quantified belief representation, *Quantified Representation of Uncertainty and Imprecision*, (Springer, 1998), pp. 267–301.

[Su et al.(2008)] X. Su and T. M. Khoshgoftaar, Collaborative filtering for multi-class data using bayesian networks, *International Journal on Artificial Intelligence Tools*, Vol. 17 (World Scientific, 2008), pp. 71–85.

[Troiano et al. (2015)] L.Troiano, L. J. Rodríguez-Muñiz and I. Díaz, Discovering user preferences using Dempster–-Shafer theory, *Fuzzy Sets and Systems*, Vol. 278 (Elsevier, 2015), pp. 98-117.

[Troiano et al. (2017)] L. Troiano, I. Díaz, and C. Gaglione, Matching media contents with user profiles by means of the Dempster–Shafer theory, *International Conference on Fuzzy Systems*, (IEEE, 2017), pp. 1–6.

[Vo et al. (2021)] Vo, D.V., Karnjana, J. and Huynh, V.N., 2021. An integrated framework of learning and evidential reasoning for user profiling using short texts. Information Fusion, 70, pp.27-42.

[Wang et al. (2021)] Wang, H., Shen, Z., Jiang, S., Sun, G. and Zhang, R.J., 2021. User-based Collaborative Filtering Algorithm Design and Implementation. In Journal of Physics: Conference Series (Vol. 1757, No. 1, p. 012168). IOP Publishing

[Witten et al. (2016)] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*, (Morgan Kaufmann, 2016).

[Yoshii et al. (2006)] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences, *International Conference on Music Information Retrieval*, Vol. 6 (2006)