A Reliably Weighted Collaborative Filtering System

Van-Doan Nguyen, Van-Nam Huynh

Japan Advanced Institute of Science and Technology (JAIST), Japan.

Abstract. In this paper, we develop a reliably weighted collaborative filtering system that first tries to predict all unprovided rating data by employing context information, and then exploits both predicted and provided rating data for generating suitable recommendations. Since the predicted rating data are not a hundred percent accurate, they are weighted weaker than the provided rating data when integrating both these kinds of rating data into the recommendation process. In order to flexibly represent rating data, Dempster-Shafer (DS) theory is used for data modelling in the system. The experimental results indicate that assigning weights to rating data is capable of improving the performance of the system.

1 Introduction

Research on collaborative filtering systems (CFSs) has focused on the sparsity problem, which is that the total number of items and users is very large while each user only rates a small number of items. The challenge in this problem is how to generate good recommendations when a small number of provided rating data is available. Until now, various methods have been developed for overcoming the problem. In [14], the author introduced a method that employs additional information about the users, e.g. gender, age, education, interests, or other available information that can help to classify users. Recently, Matrix Factorization methods [8, 10, 15, 18] have become well-known for combining good scalability with predictive accuracy; but they are not capable of tackling the data imperfection issue caused by some level of impreciseness and/or uncertainty in the measurements [9]. In [19], the authors proposed a new method that not only models rating data by using DS theory but also exploits context information of users for generating unprovided rating data. Further to the method developed in [19], the method in [12] employs community context information extracted from the social network for generating unprovided rating data. However, the methods in both [19] and [12] consider the role of the predicted rating data to be normally the same as that of the provided rating data, and they are not capable of predicting all unprovided rating data (see Example 1 in Section 4). In this paper, these two limitations will be overcome.

Additionally, over the years, management of data imperfection has become increasingly important; however, the existing recommendation techniques are rarely capable of dealing with this challenge [19]. So far, a number of mathematical theories have been developed for representing data imperfection, such as probability theory [4], fuzzy set theory [20], possibility theory [21], rough set theory [13], DS theory [3, 16]. Most of these approaches are capable of representing a specific aspect of data imperfection [9]. Importantly, among these, DS theory is considered to be the most general one in which different kinds of uncertainty can be represented [7, 19].

For CFSs, DS theory provides a flexible method for modeling information without requiring a probability to be assigned to each element in a set [11]. It is worth to know that different users can have different evaluations on the same item in that users' preferences are subjective and qualitative. Additionally, the existing recommender systems usually provide rating domains representing as finite sets, denoted by $\Theta = \{\theta_1, \theta_2, .., \theta_L\}$, where $\theta_i < \theta_j$ whenever $i < j$; these systems only allow users to evaluate an item as a hard rating value, known as a singleton, $\theta_i \in \Theta$. However, in some cases, users need to rate an item as a soft rating value, also referred to as a composite, representing by $A \subseteq \Theta$. For example, according to some aspects, a user intends to rate an item as θ_i, but regarding other aspects, the user would like to rate the item as θ_{i+1}; in this case, it is better to use a soft rating value as a set $A = \{\theta_i, \theta_{i+1}\}$. With DS theory, rating entries in the rating matrix can be represented as soft rating values. Besides, this theory supports not only modeling missing data by the vacuous mass structure but also generating both hard as well as soft decisions; here, hard and soft decisions can be known as the recommendations presented by singletons and composites, respectively. Specially, regarding DS theory, some pieces of evidence can be combined easily by using Dempster's rule of combination to form more valuable evidence. Under such an observation, DS theory is selected for modeling rating data in our system.

In short, the system in this paper is developed for not only dealing with the sparsity problem, but also overcoming the data imperfection issue. The main contributions of the paper include (1) a new method of computing user-user similarities which considers the significant role of the provided rating data to be higher than that of the predicted rating data, and (2) a solution for predicting all unprovided rating data using context information.

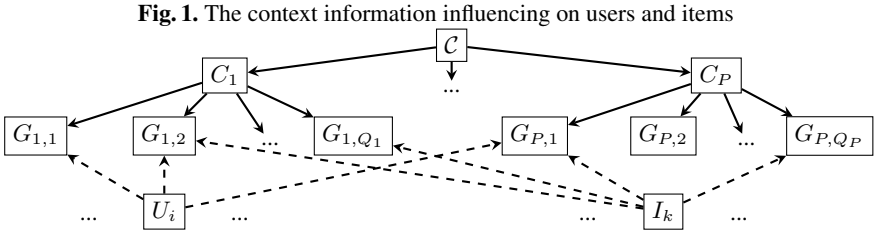
The remainder of the paper is organized as follows. In the next section, background information about DS theory is provided. Then, details of the methodology are described. After that, system implementation and discussions are represented. Finally, conclusions are illustrated in the last section.

2 Dempster-Shafer Theory

Let us consider that a problem domain is represented by a finite set, denoted as $\Theta = \{\theta_1, \theta_2, ..., \theta_L\}$, of mutually exclusive and exhaustive hypotheses, called frame of discernment [16]. A mass function, or basic probability assignment (BPA), $m : 2^\Theta \rightarrow [0, 1]$ is the one satisfying $m(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m(A) = 1$, where 2^Θ is the power set of Θ. The mass function m is called to be vacuous if $m(\Theta) = 1$ and $\forall A \subset \Theta, m(A) = 0$. A subset $A \subseteq \Theta$ with $m(A) > 0$ is called a focal element of m, and the set of all focal elements is called the focal set. If a source of information providing a mass function m has probability $\delta \in [0, 1]$ of trust, the discounting operation is used for creating new mass function m^δ, which takes this reliable probability into account. Formally, for $A \subset \Theta$, $m^\delta(A) = \delta \times m(A)$; and $m^\delta(\Theta) = \delta \times m(\Theta) + (1 - \delta)$.

Two evidential functions, known as belief and plausibility functions, are derived from the mass function m. The belief function on Θ is defined as a mapping $Bl : 2^\Theta \rightarrow [0, 1]$, where $A \subseteq \Theta$, $Bl(A) = \sum_{B \subseteq A} m(B)$; and the plausibility function on Θ is defined

Fig. 1. The context information influencing on users and items



as mapping $Pl : 2^{\Theta} \rightarrow [0, 1]$, where $Pl(A) = 1 - Bl(\bar{A})$. A probability distribution Pr satisfying $Bl(A) \leq Pr(A) \leq Pl(A), \forall A \subseteq \Theta$ is said to be compatible with the mass function m; and the pignistic probability distribution [17], denoted by Bp, is a typical one represented as $Bp(\theta_i) = \sum_{\{A \subseteq \Theta | \theta_i \in A\}} \frac{m(A)}{|A|}$. Additionally, a useful operation that plays an important role in the forming of two pieces of evidence into a single one is Dempster's rule of combination. Formally, this operation is used for aggregation of two mass function m_1 and m_2, denoted by $m = m_1 \oplus m_2$, in the following

$$m(A) = \frac{1}{1-K} \sum_{\{C,D \subseteq \Theta | C \cap D = A\}} m_1(C) \times m_2(D),$$

where $K = \sum_{\{C,D \subseteq \Theta | C \cap D = \emptyset\}} m_1(C) \times m_2(D) \neq 0$, and K represents the basic probability mass associated with conflict.

3 Methodology

3.1 Data Modeling

Let $\mathcal{U} = \{U_1, U_2, ..., U_M\}$ be the set of all users and let $\mathcal{I} = \{I_1, I_2, ..., I_N\}$ be the set of all items. Each user rating is defined as a preference mass function spanning over a finite, rank-order set of L preference labels $\Theta = \{\theta_1, \theta_2 ..., \theta_L\}$, where $\theta_i < \theta_j$ whenever $i < j$. The evaluations of all users are represented by a DS rating matrix created as $\mathcal{R} = \{r_{i,k}\}$, where $i = \overline{1, M}$, $k = \overline{1, N}$. For a provided rating entry regarding the evaluation of a user U_i on an item I_k, $r_{i,k} = m_{i,k}$, with $\sum_{A \subseteq \Theta} m_{i,k}(A) = 1$. Each unprovided rating entry is assigned the vacuous mass function; that means $r_{i,k} = m_{i,k}$, with $m_{i,k}(\Theta) = 1$ and $\forall A \subset \Theta$, $m_{i,k}(A) = 0$. All items rated by a user U_i, and all users rated an item I_k are denoted by ${}^{I}R_i = \{I_l \mid r_{i,l} \neq vacuous\}$, and ${}^{U}R_k = \{U_l \mid r_{l,k} \neq vacuous\}$, respectively.

3.2 Predicting Unprovided Rating Data

As mentioned earlier, each unprovided rating entry in the rating matrix is modeled by the vacuous mass function. It can be seen that this function has high uncertainty. Thus, context information from different sources is used for the purpose of reducing the uncertainty introduced by the vacuous representation [19]. Here, context information, denoted by $\mathcal{C}$, is considered the concept for grouping users. Let us consider a

movie recommender system. In this system, characteristics such as user gender, user occupation, movie genre can be considered concepts because they may have significantly influenced user ratings. Each concept can consist of a number of groups, e.g. the movie genre might contain some groups such as drama, comedy, action, mystery, horror, animation. We assume that, in our system, there are P characteristics considered as concepts, and each concept $C_p \in \mathcal{C}$, consists of Q_p groups [12, 19], as shown in Fig. 1. Formally, the context information can be represented as follows

$$\mathcal{C} = \{C_1, C_2, ..., C_P\}; C_p = \{G_{p,1}, G_{p,2}, ..., G_{p,Q_p}\}, \text{ where } p = \overline{1, P}.$$

Simultaneously, a user U_i as well as an item I_k may belong to multiple groups from the same concept. For each $C_p \in \mathcal{C}$, the groups in which a user U_i is interested are identified by the mapping functions $f_p : \mathcal{U} \to 2^{C_p} : U_i \mapsto f_p(U_i) \subseteq C_p$; and the groups to which an item I_k belongs are determined by the mapping function $g_p : \mathcal{I} \to 2^{C_p} : I_k \mapsto g_p(I_k) \subseteq C_p$, where 2^{C_p} is the power set of C_p.

We also assume that the users belonging to a group can be expected to possess similar preferences. Based on this assumption, the unprovided rating entries are generated. For a concept $C_p \in \mathcal{C}$, let us consider an item I_k, the overall group preference of this item on each $G_{p,q} \in g_p(I_k)$, with $q = \overline{1, Q_p}$, is defined by the mass function ${}^{G}m_{p,q,k} : 2^{\Theta} \to [0, 1]$. This mass function is calculated by combining all the provided rating data of the users who are interested in $G_{p,q}$ and have already rated I_k, as below

$${}^{G}m_{p,q,k} = \bigoplus_{\{j | I_k \in {}^{I}R_j, G_{p,q} \in f_p(U_j) \cap g_p(I_k)\}} m_{j,k}. \tag{1}$$

If a user U_i has not rated an item I_k, the process for predicting the rating entry $r_{i,k}$ regarding the preference of user U_i on item I_k is performed as follows

- Firstly, the concept preferences corresponding to user U_i on item I_k, denoted by the mass functions ${}^{C}m_{p,i,k} : 2^{\Theta} \to [0, 1]$, with $p = \overline{1, P}$, are computed by combining the related group preferences of item I_k as follows

$${}^{C}m_{p,i,k} = \bigoplus_{\{q | G_{p,q} \in f_p(U_i) \cap g_p(I_k)\}} {}^{G}m_{p,q,k}. \tag{2}$$

- Secondly, the overall context preference corresponding to a user U_i on item I_k, denoted by the mass function ${}^{\mathcal{C}}m_{i,k} : 2^{\Theta} \to [0, 1]$, is achieved by combining all related concept mass functions as below

$${}^{\mathcal{C}}m_{i,k} = \bigoplus_{p=\overline{1,P}} {}^{C}m_{p,i,k}. \tag{3}$$

- Next, the unprovided rating entry $r_{i,k}$, which is vacuous, is replaced with its corresponding context mass function as follows

$$r_{i,k} = {}^{\mathcal{C}}m_{i,k}. \tag{4}$$

- Finally, in case the rating entry $r_{i,k}$ is still vacuous after replacing such as Example 1 in Section 4, we propose that this entry is assigned the evidence obtained by combining all preference mass functions of the users already rated item I_k as below

$$r_{i,k} = \bigoplus_{\{j | U_j \in {}^{U}R_k\}} m_{j,k}. \tag{5}$$

Please note that, at this point, all unprovided rating data are completely predicted.

3.3 Computing User-User Similarities

In the DS rating matrix, every rating entry $r_{i,k} = m_{i,k}$ represents user U_i's preference toward a single item I_k. Let us consider that the focal set of $m_{i,k}$ is defined by $F_{i,k} = \{A \in 2^{\Theta} | m_{i,k}(A) > 0\}$. The user U_i's preference toward all items as a whole can be defined over the cross-product $\boldsymbol{\Theta} = \Theta_1 \times \Theta_2 \times ... \times \Theta_N$, where $\Theta_i = \Theta, \forall i = \overline{1, N}$ [7, 19]. The cylindrical extension of the focal element $A \in F_{i,k}$ to the cross-product $\boldsymbol{\Theta}$ is $cyl_{\boldsymbol{\Theta}}(A) = [\Theta_1 ... \Theta_{i-1} A \Theta_{i+1} ... \Theta_N]$. The mapping $M_{i,k} : 2^{\boldsymbol{\Theta}} \rightarrow [0, 1]$ generates a valid mass function defined on $\boldsymbol{\Theta}$ by extending $r_{i,k}$; and if $B = cyl_{\boldsymbol{\Theta}}(A)$, $M_{i,k}(B) = m_{i,k}(A)$, otherwise $M_{i,k}(B) = 0$ [7].

For a user U_i, let us consider the mass functions $M_{i,k}$ defined over the cross-product $\boldsymbol{\Theta}$, with $k = \overline{1, N}$. The mass function $M_i : 2^{\boldsymbol{\Theta}} \rightarrow [0, 1]$, where $M_i = \bigoplus_{k=1}^{N} M_{i,k}$, is referred to as the user-BPA of user U_i.

Consider user U_i's user-BPA M_i and the rating mass functions $m_{i,k}, k = \overline{1, N}$, each defined over Θ. The pignistic probability of the singleton $\theta_{i_1} \times ... \times \theta_{i_N} \in \boldsymbol{\Theta}$, is $Bp_i(\theta_{i_1} \times ... \times \theta_{i_N}) = \prod_{k=1}^{N} Bp_{i,k}(\theta_{i_k})$, where $\theta_{i_k} \in \Theta$, and Bp_i and $Bp_{i,k}$ are user U_i's pignistic probability distributions corresponding to its user-BPA and preference rating of user U_i on item I_k , respectively [19].

For computing the distance among users, we adopt the distance measure method introduced in [2]. According to this method, the distance between two user-BPAs M_i and M_j defined over the same cross-product $\boldsymbol{\Theta}$ is $D(M_i, M_j) = CD(Bp_i, Bp_j)$, where CD refers to the Chan and Darwiche distance measure [2] represented as below

$$CD(Bp_i, Bp_j) = \ln \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{Bp_j(\boldsymbol{\theta})}{Bp_i(\boldsymbol{\theta})} - \ln \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{Bp_j(\boldsymbol{\theta})}{Bp_i(\boldsymbol{\theta})}.$$

In addition, $CD(Bp_i, Bp_j) = \sum_{k=1}^{N} CD(Bp_{i,k}, Bp_{j,k})$ [19]. Obviously, for each item I_k, it is easy to recognize as follows

- In case neither user U_i nor user U_j has rated item I_k, that means both $r_{i,k}$ and $r_{j,k}$ are predicted rating data. Since $Bp_{i,k}$ and $Bp_{j,k}$ are derived from entries $r_{i,k}$ and $r_{j,k}$, respectively, the value of the expression $CD(Bp_{i,k}, Bp_{j,k})$ is not fully reliable.
- The value of the expression $CD(Bp_{i,k}, Bp_{j,k})$ is also not fully reliable if either user U_i or user U_j has rated item I_k.
- The value of the expression $CD(Bp_{i,k}, Bp_{j,k})$ is only fully reliable if both user U_i and U_j have rated item I_k.

Under such an observation, in order to improve the accuracy of the distance measurement between two users, we propose a new method to compute the distance between two user-BPAs M_i and M_j, as shown below

$$\hat{D}(M_i, M_j) = \sum_{k=1}^{N} \mu(x_{i,k}, x_{j,k}) \times CD(Bp_{i,k}, Bp_{j,k}),$$

where $\mu(x_{i,k}, x_{j,k}) \in [0, 1]$ is a reliable function referring to the trust of the evaluation of both user U_i and user U_j on item I_k. $\forall (i, k), x_{i,k} \in \{0, 1\}$; $x_{i,k} = 1$ when $r_{i,k}$

is a provided rating entry, otherwise $r_{i,k}$ is a predicted rating one. Note that because of $\mu(x_{i,k}, x_{j,k}) \in [0, 1]$, the distinguishing of the provided and the predicted rating data does not destroy the elegance of the selected distance measure method [2]. When $\mu(x_{i,k}, x_{i,k}) < 1$ indicates that the distance between user U_i and user U_j is shorter than it actually is. That means user U_i has a high opportunity for being a member in user U_j's neighborhood set, and vice versa.

Table 1. The values of the reliable function

$x_{i,k}$	$x_{j,k}$	$\mu(x_{i,k}, x_{j,k})$
0	0	1
0	1	$1 - w_1$
1	0	$1 - w_1$
1	1	$1 - 2 \times w_1 - w_2$

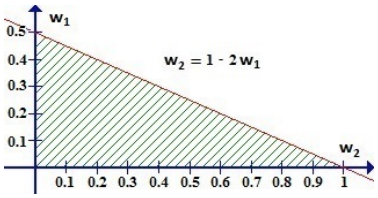


Fig. 2. The domains of w_1 and w_2

The reliable function $\mu(x_{i,k}, x_{j,k})$ can be selected according to specific application. In the general case, we suggest that $\mu(x_{i,k}, x_{j,k}) = 1 - w_1 \times (x_{i,k} + x_{j,k}) - w_2 \times x_{i,k} \times x_{j,k}$, where $w_1 \geq 0$ and $w_2 \geq 0$ are the reliable coefficients representing the state when a user has actually rated an item and two users together have rated an item, respectively. Because of $\forall(i, k), x_{i,k} \in \{0, 1\}$, the function $\mu(x_{i,k}, x_{j,k})$ has to belong to one of four cases as shown in Table 1. Under the condition $0 \leq \mu(x_{i,k}, x_{j,k}) \leq 1$, the domains of w_1 and w_2 must be in the parallel diagonal line shading area as illustrated in Fig. 2.

Consider a monotonically deceasing function ψ: $[0, \infty] \mapsto [0, 1]$ satisfying $\psi(0) = 1$ and $\psi(\infty) = 0$. Then, with respect to ψ, $s_{i,j} = \psi(D(M_i, M_j))$ is referred to as the user-user similarity between users U_i and U_j. We use the function $\psi(x) = e^{-\gamma \times x}$, where $\gamma \in (0, \infty)$. Consequently, the user-user similarity matrix is then generated as $S = \{s_{i,j}\}, i = \overline{1, M}, j = \overline{1, M}$.

3.4 Selecting Neighborhoods

The method of neighborhood selection proposed in [5] is an effective one because it prevents the recommendation result from the errors generated from very dissimilar users. This method, selected to apply in our system. Formally, we need to select a neighborhood set $\mathcal{N}_{i,k}$ for a user U_i. First, the users already rated item I_k and whose similarities with user U_i are equal or greater than a threshold τ are extracted. Then, K users with the highest similarity with user U_i are selected from the extracted list. The neighborhood is the largest set that satisfies $\mathcal{N}_{i,k} = \{U_j \in \mathcal{U} \mid I_k \in {}^{I}R_j, s_{i,j} \geq \max_{\forall U_l \notin \mathcal{N}_{i,k}}\{\tau, s_{i,l}\}\}$. Note that for a new user, the condition $I_k \in {}^{I}R_j$ is removed.

The estimated rating data for an unrated item I_k of a user U_i is presented as $\hat{r}_{i,k} = \hat{m}_{i,k}$, where $\hat{m}_{i,k} = \bar{m}_{i,k} \oplus m_{i,k}$. Here, $\bar{m}_{i,k}$ is the mass function corresponding to the neighborhood prediction ratings, as shown below

$$\bar{m}_{i,k} = \bigoplus_{\{j|U_j \in \mathcal{N}_{i,k}\}} m_{j,k}^{s_{i,j}}, \text{with } m_{j,k}^{s_{i,j}} = \begin{cases} s_{i,j} \times m_{j,k}(A), & \text{for } A \subset \Theta; \\ s_{i,j} \times m_{j,k}(\Theta) + (1 - s_{i,j}), & \text{for } A = \Theta. \end{cases}$$

3.5 Generating Recommendations

Our system supports both hard and soft decisions. For a hard decision, the pignistic probability is applied, and the singleton having the highest probability is selected as the preference label. If a soft decision is needed, the maximum belief with overlapping interval strategy (maxBL) [1] is applied, and the singleton whose belief is greater than the plausibility of any other singleton is selected; if such as class label does not exist, decision is made according to the favor of composite class label constituted of the singleton label that has the maximum belief and those singletons that have a higher plausibility.

4 Implementation and Discussions

Movielens data set[1], MovieLens 100k, was used in the experiment. This data set consists of 100,000 hard ratings from 943 users on 1682 movies with the rating value $\theta_l \in \Theta = \{1, 2, 3, 4, 5\}$, 5 is the highest value. Each user has rated at least 20 movies. Since our system requires a domain with soft ratings, each hard rating entry $\theta_l \in \Theta$ was transformed into the soft rating entry $r_{i,k}$ by the DS modeling function [19] as follows

$$r_{i,k} = \begin{cases} \alpha_{i,k} \times (1 - \sigma_{i,k}), & \text{for } A = \theta_l; \\ \alpha_{i,k} \times \sigma_{i,k}, & \text{for } A = B; \\ 1 - \alpha_{i,k}, & \text{for } A = \Theta; \\ 0, & \text{otherwise,} \end{cases} \text{with } B = \begin{cases} (\theta_1, \theta_2), & \text{if } l = 1; \\ (\theta_{L-1}, \theta_L), & \text{if } l = L; \\ (\theta_{l-1}, \theta_l, \theta_{l+1}), & \text{otherwise.} \end{cases}$$

Here, $\alpha_{i,k} \in [0, 1]$ and $\sigma_{i,k}$ are a trust factor and a dispersion factor, respectively [19]. In the data set, context information is represented as below

$$\mathcal{C} = \{C_1\} = \{Genre\}; C_1 = \{G_{1,1}, G_{1,2}, ..., G_{1,19}\} = \{Unknown, Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film\text{-}Noir, Horror, Musical, Mystery, Romance, Sci\text{-}Fi, Thriller, War, Western\}.$$

Because the genres to which a user belongs is not available, we assume the genres of a user U_i are assigned by the genres of the movies rated by user U_i. Each unprovided rating entry was replaced with its corresponding context mass function predicted according to equations (1), (2), (3), (4), and (5). Note that if the context mass functions are fused by using the methods in [12, 19] (just applying equations (1), (2), (3) and (4)), some unprovided rating entries are still vacuous after replacing, as in Example 1.

Example 1. In the Movielens data set, let us consider a user U_c with $f_1(U_c) = \{G_{1,4}, G_{1,5}, G_{1,6}, G_{1,18}\} = \{Animation, Children's, Comedy, War\}$ and an item I_t with $g_1(I_t) = \{G_{1,17}\} = \{Thriller\}$. Assuming that user U_c has not rated item I_t and we need to predict the value for r_{ct}. The predicting process is as follows

- According to equation (1), ${}^{G}m_{1,17,t} = \bigoplus_{\{j | I_t \in {}^{U}R_j, G_{1,17} \in f_1(U_j)\}} m_{j,t}$; $\forall G_{1,q} \in C_1$ and $q \neq 17$, ${}^{G}m_{1,q,t} = vacuous$.
- Using equation (2), ${}^{C}m_{1,c,t} = \bigoplus_{\{q | G_{1,q} \in f_1(U_c) \cap g_1(I_t)\}} {}^{G}m_{1,q,t} = vacuous$.
- According to equation (3), ${}^{\mathcal{C}}m_{c,t} = {}^{C}m_{1,c,t} = vacuous$.
- Applying equation (4), $r_{c,t} = {}^{\mathcal{C}}m_{c,t} = vacuous$.

[1] http://grouplens.org/datasets/movielens/

Table 2. Overall MAE versus w_1 and w_2

w_2 \ w_1	0.0	0.1	0.2	0.3	0.4	0.5
0.0	0.8361	0.8366	0.8363	0.8350	0.8342	0.8334
0.1	0.8363	0.8363	0.8363	0.8347	0.8342	
0.2	0.8366	0.8363	0.8361	0.8342	0.8342	
0.3	0.8366	0.8363	0.8361	0.8339		
0.4	0.8363	0.8363	0.8358	0.8339		
0.5	0.8363	0.8363	0.8355			
0.6	0.8363	0.8361	0.8355			
0.7	0.8363	0.8361				
0.8	0.8361	0.8361				
0.9	0.8358					
1.0	0.8358					

Table 3. Overall DS-MAE versus w_1 and w_2

w_2 \ w_1	0.0	0.1	0.2	0.3	0.4	0.5
0.0	0.8406	0.8406	0.8405	0.8402	0.8399	0.8397
0.1	0.8406	0.8406	0.8405	0.8401	0.8399	
0.2	0.8406	0.8406	0.8404	0.8401	0.8400	
0.3	0.8406	0.8405	0.8404	0.8400		
0.4	0.8406	0.8405	0.8405	0.8400		
0.5	0.8406	0.8406	0.8404			
0.6	0.8406	0.8405	0.8404			
0.7	0.8406	0.8405				
0.8	0.8406	0.8405				
0.9	0.8405					
1.0	0.8406					

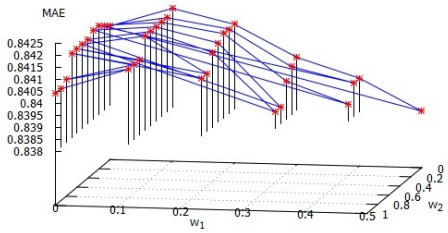


Fig. 3. Visualizing overall MAE

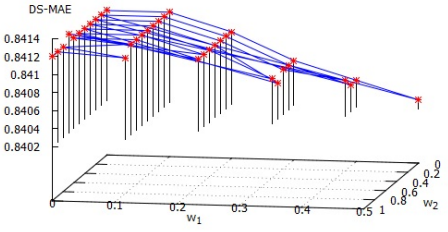


Fig. 4. Visualizing overall DS-MAE

Firstly, 10% of the users were randomly selected. Then, for each selected user, we accidentally withheld 5 ratings, the withheld ratings were used as testing data and the remaining ratings were considered as training data. Finally, recommendations were computed for the testing data. We repeated this process for 10 times, and the average results of 10 splits were represented in this section. Note that in all experiments, some parameters were selected as following: $\gamma = 10^{-4}$, $\beta = 1$, $\forall(i,k)\{\alpha_{i,k}, \sigma_{i,k}\} = \{0.9, 2/9\}$.

For recommender systems with hard decisions, the popular performance assessment methods are MAE, $Precision$, $Recall$, and F_β [6]. Recently, some new methods allowing to evaluate soft decisions are proposed, such as DS-$Precision$ and DS-$Recall$ [7]; DS-MAE and DS-F_β [19]. We adopted all these methods for evaluating the proposed system. Since the system is developed for aiming at extending CoFiDS [19], we also selected CoFiDS for performance comparison.

Table 2 and 3 show the overall MAE and DS-MAE criterion results computed by mean of these evaluation criteria with $K = 15, \tau = 0$ according to two reliable coefficients w_1 and w_2, respectively. The statistics in these tables indicate that the performance of the proposed system is almost linearly dependent on the value of w_1; this finding is the same for the other evaluation criteria. The coefficient w_2 just slightly influences the performance in hard decisions, but seems not to affect the performance in soft decisions; the reason is that, in the data set, when considering two users, the number of movies rated by these users is very small while the total of movies is large. Fig. 3 and 4 depict the same information as Table 2 and 3 in a visualization way.

For comparing with CoFiDS, we conducted the experiments with $w_1 = 0.5, w_2 = 0, \tau = 0$, and several values of K. Fig. 5 and 6 show the overall MAE and DS-MAE

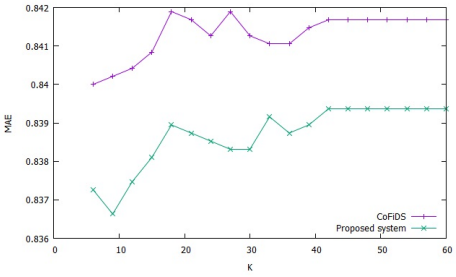


Fig. 5. Overall MAE versus K

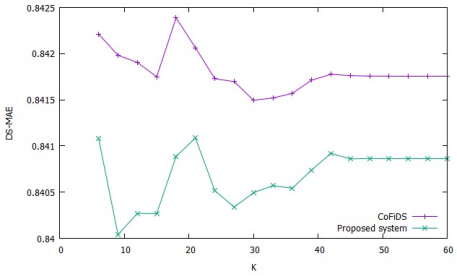


Fig. 6. Overall DS-MAE versus K

Table 4. The comparison in hard decisions

Metric	True Rating					Overall
	1	2	3	4	5	
Proposed system:						
$Precision$	**0.3201**	0.2210	**0.3188**	**0.4002**	**0.4179**	**0.3630**
$Recall$	**0.0906**	**0.0892**	**0.3179**	0.6413	**0.1885**	**0.3709**
MAE	**2.1368**	**1.4242**	**0.7790**	**0.4212**	**1.0175**	**0.8383**
F_1	**0.1205**	**0.1245**	**0.3170**	**0.4924**	**0.2571**	**0.3384**
CoFiDS:						
$Precision$	0.3118	**0.2151**	0.3177	0.3996	0.4171	0.3609
$Recall$	0.0873	0.0872	0.3157	**0.6418**	0.1866	0.3697
MAE	2.1435	1.4325	0.7813	0.4224	1.0202	0.8413
F_1	0.1148	0.1216	0.3152	0.4921	0.2551	0.3366

Table 5. The comparison in soft decisions

DS -Metric	True Rating					Overall
	1	2	3	4	5	
Proposed system:						
$Precision$	**0.3001**	**0.2035**	**0.3150**	0.3990	0.4016	**0.3551**
$Recall$	**0.0663**	0.0926	**0.3164**	0.6391	**0.1847**	**0.3680**
MAE	**2.1963**	**1.4313**	**0.7721**	0.4122	**1.0317**	**0.8405**
F_1	**0.1036**	**0.1248**	**0.3147**	0.4909	**0.2507**	**0.3349**
CoFiDS:						
$Precision$	0.2926	0.2032	0.3148	0.3990	**0.4020**	0.3547
$Recall$	0.0658	**0.0934**	0.3155	**0.6398**	0.1837	0.3679
MAE	2.1973	1.4323	0.7724	**0.4118**	1.0359	0.8415
F_1	0.1028	0.1255	0.3141	**0.4911**	0.2500	0.3347

criterion results of both CoFiDS and the proposed system change with the neighborhood size K. According to these features, the performances of two systems are fluctuated when $K < 42$, and then appear to stabilize with $K \geq 42$. In particular, both features show that the proposed system is more effective in all cases.

Table 4 and 5 show the summarized results of the performance comparisons between the proposed system and CoFiDS in hard and soft decisions with $K = 30, w_1 = 0.5, w_2 = 0, \tau_2 = 0$, respectively. In each category in these tables, every rating has its own column; and the bold values indicate the better performance, and underlined values illustrate equal performance. Importantly, the statistics in both tables show that, except for soft decisions with true rating value $\theta_4 = 4$, the proposed system achieves better performance in all selected measurement criteria. However, the absolute values of the performance of the proposed system are just slightly higher than those of CoFiDS. The reason is that the MovieLens data set contains a small number of provided rating data. In case more provided rating data are available, the proposed system can be much better than CoFiDS.

5 Conclusions

In summary, in this paper, we have developed a CFS that uses the DS theory for representing rating data, and integrates context information for predicting all unprovided rating data. Specially, after predicting all unprovided data, suitable recommendations are generated by employing both predicted and provided rating data with the stipulation that the provided rating data are more important than the predicted rating data.

References

1. Isabelle Bloch. Some aspects of dempster-shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17(8):905–919, 1996.
2. Hei Chan and Adnan Darwiche. A distance measure for bounding probabilistic belief change. *Int. J. Approx. Reasoning*, 38(2):149–174, 2005.
3. Arthur P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
4. William Feller. *An introduction to probability theory and its applications. Volume I*. J. Wiley & sons, 1968.
5. Jonathan L. Herlocker, Joseph A. Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR'99: Proceedings of the 22nd Annual International ACM SIGIR Conference*, pages 230–237, 1999.
6. Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.
7. K. K. Rohitha Hewawasam, Kamal Premaratne, and Mei-Ling Shyu. Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections. *IEEE Trans. Syst. Man Cybern., Part B*, 37(6):1446–1459, 2007.
8. Meng Jiang, Peng Cui, Fei Wang, Wenwu Zhu, and Shiqiang Yang. Scalable recommendation with social contextual information. *IEEE Trans. Knowl. Data Eng.*, 26(11):2789–2802, 2014.
9. Bahador Khaleghi, Alaa M. Khamis, Fakhreddine Karray, and Saiedeh N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28–44, 2013.
10. Yehuda Koren, Robert M. Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.
11. Andino Maseleno and Md. Mahmud Hasan. Dempster-shafer theory for move prediction in start kicking of the bicycle kick of sepak takraw game. *CoRR*, abs/1401.2483, 2014.
12. Van-Doan Nguyen and Van-Nam Huynh. A community-based collaborative filtering system dealing with sparsity problem and data imperfections. In *PRICAI 2014: Trends in Artificial Intelligence*, pages 884–890, 2014.
13. Zdzislaw Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, 1992.
14. Michael J. Pazzani. A framework for collaborative, content-based and demographic filtering. *AI Review*, 13:393–408, 1999.
15. Li Pu and Boi Faltings. Understanding and improving relational matrix factorization in recommender systems. In *Seventh ACM Conference on Recommender Systems*, pages 41–48, 2013.
16. Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
17. Philippe Smets. Practical uses of belief functions. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, UAI'99, pages 612–621. Morgan Kaufmann Publishers Inc., 1999.
18. John Z. Sun, Dhruv Parthasarathy, and Kush R. Varshney. Collaborative kalman filtering for dynamic matrix factorization. *IEEE Trans. Signal Processing*, 62(14):3499–3509, 2014.
19. Thanuka L. Wickramarathne, Kamal Premaratne, Miroslav Kubat, and Dushyantha T. Jayaweera. Cofids: A belief-theoretic approach for automated collaborative filtering. *IEEE Trans. Knowl. Data Eng.*, 23(2):175–189, 2011.
20. Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.
21. Lotfi A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.