A Community-Based Collaborative Filtering System Dealing with Sparsity Problem and Data Imperfections

Van-Doan Nguyen, Van-Nam Huynh

School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST).
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan.

Abstract. In this paper, we develop a collaborative filtering system for not only tackling the sparsity problem by exploiting community context information but for also dealing with data imperfections by means of Dempster-Shafer theory. The experimental results show that the proposed system achieves better performance when comparing it with a similar system, CoFiDS.

1 Introduction

In the research area of recommendation systems, collaborative filtering is considered to be the most widely implemented technique [12]. However, this technique also has its own limitations such as the new user issue, the new item issue, and the sparsity problem [1]. Among these limitations, the sparsity problem is known as a major drawback [9]. So far, various methods have been developed for overcoming the sparsity problem such as using latent factors [11] or context information [15] for generating unprovided rating data. This paper attempts to predict all unprovided rating data using community context information extracted from the social network consisting of all users for dealing with the sparsity problem.

Another, performances of collaborative filtering systems are usually limited by data imperfection issues. These issues are caused by the data affected by some level of impreciseness as well as uncertainty in the measurements [10]. Until now, a number of mathematical theories have been developed for representing data imperfections, such as Dempster-Shafer (DS) theory [4, 13], probability theory [5], possibility theory [17]. Among these, DS theory is considered to be the most general theory for representing imperfection data [8, 15]. With DS theory, rating entries in the rating matrix can be represented as soft rating values. Let us assume that the rating domain of a collaborative filtering system is a finite set $\Theta = \{1, 2, 3, 4, 5\}$. A hard rating value is represented as a proposition $\theta \in \Theta, 1 \leq \theta \leq 5$, referred to as a singleton; and a soft rating value is modeled as a set $A \subseteq \Theta$, known as a composite, e.g. $A = \{4, 5\}$. Similarly, the hard and soft decisions can be known as the recommendations presented by singletons and composites, respectively. Especially, with this theory, pieces of evidence can be combined for generating more valuable evidence [13]. Under such an observation, DS theory is used for representing rating data in our system.

The rest of this paper is organized as follows. Section 2 represents a brief introduction to DS theory. Then, details of proposed system are described in Section 3. After that, Section 4 represents the system implementation and discussion. Finally, some concluding remarks are depicted in Section 5.

2 Dempster-Shafer theory

Let us consider a problem domain is represented by a finite set $\Theta = \{\theta_1, \theta_2, ..., \theta_L\}$, called frame of discernment (FoD) [13]. A mass function, or basic probability assignment (BPA), $m : 2^\Theta \mapsto [0, 1]$ is the one satisfying $m(\emptyset) = 0$ and $\sum_{A \in 2^\Theta} m(A) = 1$. A function m is considered to be vacuous if $m(\Theta) = 1$ and $\forall A \neq \Theta, m(A) = 0$. A set $A \in 2^\Theta$ and $m(A) > 0$ is called a focal element of m. The belief function on Θ is defined as a mapping $Bl : 2^\Theta \mapsto [0, 1]$, where $Bl(A) = \sum_{B \subseteq A} m(B)$; and the plausibility function on Θ is defined as mapping $Pl : 2^\Theta \mapsto [0, 1]$, where $Pl(A) = 1 - Bl(\bar{A})$. A probability distribution Pr such that $Bl(A) \leq Pr(A) \leq Pl(A), \forall A \in 2^\Theta$ is said to be compatible with the mass function m and pignistic probability distribution Bp [14] is a typical one illustrated as follows: $Bp(\theta_i) = \sum_{\{A \in 2^\Theta | \theta_i \in A\}} (m(A)/ \mid A \mid)$.

Let us consider two evidences on the same frame Θ represented by two mass functions m_1 and m_2. Dempster's rule of combination operation, denoted by $\oplus$, is used for generating a new evidence. This operation is defined as follows: $(m_1 \oplus m_2)(\emptyset) = 0$; $(m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{\{B,C \in 2^\Theta | B \cap C = A\}} m_1(B) \times m_2(C)$, where $K = \sum_{\{B,C \in 2^\Theta | B \cap C = \emptyset\}} m_1(B) \times m_2(C) \neq 0$. The discounting operation is used when a source of information provides BPA m, but this source has probability δ of reliability. Then one may adopt $1 - \delta$ as a discount rate, resulting in a new BPA m^δ as follows

$$m^\delta = \begin{cases} \delta \times m(A), & \text{for } A \in 2^\Theta; \\ \delta \times m(\Theta) + (1 - \delta), & \text{for } A = \Theta. \end{cases}$$

3 Proposed system

3.1 Data modeling

Let $U = \{U_1, U_2, ..., U_M\}$ be the set of all users and let $I = \{I_1, I_2, ..., I_N\}$ be the set of all items. In our system, each user preference rating is defined as a mass function spanning over the FoD $\Theta = \{\theta_1, \theta_2, ..., \theta_L\}$, a rank-order set of L preference labels, where $\theta_j < \theta_l$ whenever $j < l$. The evaluations of all users are represented by a DS rating matrix $R = \{r_{ik}\}$, where $i = \overline{1, M}$, $k = \overline{1, N}$, and r_{ik} is the rating data of U_i on item I_k. Each unrated entry in the DS rating matrix is modeled by vacuous evidence. We obtain the method in [15] to incorporate context information from different sources for reducing the uncertainty introduced by vacuous evidence. Here, context information is represented as P concepts, and each of them consists of a number of groups. Formally, context information is modeled as follows

$$Context = \{Concept_1, Concept_2, ..., Concept_P\};$$

$$Concept_p = \{Group_{p1}, Group_{p2}, ..., Group_{pQ_p}\}, \text{where } p = \overline{1, P}.$$

We identify the groups to which a user belongs via mapping function $f^{(Concept_p)} : U \mapsto Concept_p$, where $p = \overline{1, P}$.

The social network is represented as an undirected graph $G = (U, E)$, with U is the set of all users (nodes) and E is the set of all friend relationships (edges). This graph is represented as an adjacency matrix $A = \{a_{ij}\}$, where $i = \overline{1, M}$, and $j = \overline{1, M}$. If there is an edge between two nodes U_i and U_j then $a_{ij} = 1$; otherwise $a_{ij} = 0$.

3.2 Identifying Communities

We adopt the SLPA algorithm [16] for overlapping community detection in the social network. Note that some detected communities might consist of a large number of, or a few users. Thus, we continue applying this algorithm to separate the large communities into several smaller communities (if possible). For each member in the small communities, we assign it to the community containing most of its neighbors. Since the SLPA algorithm allows to naturally uncover overlapping communities in the social network, the number of communities can only be known when uncovering task has already completed. After executing this algorithm, we assume that the social network is divided into K_C overlapping communities denoted by $c_1, c_2, ..., c_{K_C}$.

3.3 Separating the rating matrix

After identifying communities, the rating matrix R is divided into K_C sub-rating matrixes denoted by $R_1, R_2, ..., R_{K_C}$. Each sub-rating matrix R_t contains the rating entries of all users in community c_t, with $1 \leq t \leq K_C$.

3.4 Performing on each community

We employ the context information in community c_t for predicting unprovided rating data in sub-rating matrix R_t first; then, we use both predicted and provided rating data for computing user-user similarities. After that, we select neighborhoods and estimate the rating data for each active user in the community c_t. Note that these tasks, described in the rest of this section, are performed in each community c_t independently.

Predicting unprovided rating data. We apply the method proposed in [15] for predicting unprovided rating data in sub-rating matrix $R_t = \{r_{ik}\}$, where $U_i \in c_t$, and $k = \overline{1,N}$. Let us consider that all items rated by a user U_i and all users who have already rated an item I_k are denoted by $R_i^{(user)} = \{I_l | r_{il} \neq \text{vacuous}\}$ and $R_k^{(item)} = \{U_l | r_{lk} \neq \text{vacuous}\}$, respectively. The predicting process in R_t is represented as below

1. Firstly, the group preference $m_k^{(Group_{pq})} : 2^{\Theta} \mapsto [0,1]$, with $U_i \in c_t, k = \overline{1,N}, p = \overline{1,P}$, and $q = \overline{1,Q_p}$, of each group $Group_{pq}$ of item I_k is computed as follows: $m_k^{(Group_{pq})} = \bigoplus_{\{i | U_i \in Group_{pq}, I_k \in R_i^{(user)}\}} m_{ik}$.
2. Then, the concept preference $m_{ik}^{(Concept_p)} : 2^{\Theta} \mapsto [0,1]$, with $k = \overline{1,N}, p = \overline{1,P}$, corresponding to user U_i and item I_k, is the obtained by combing these group preferences as follows: $m_{ik}^{(Concept_p)} = \bigoplus_{\{q | Group_{pq} \in f^{(Concept_p)}(U_i)\}} m_k^{(Group_{pq})}$.
3. Next, the overall context preference $m_{ik}^{(Context)} : 2^{\Theta} \mapsto [0,1]$ corresponding to a user U_i and an item I_k, is obtained by combining all the concept preferences as follows: $m_{ik}^{(Context)} = \bigoplus_{p=\overline{1,P}} m_{ik}^{(Concept_p)}$.
4. Finally, each unrated entry r_{ik} =vacuous is replaced with its corresponding context preference, that means $r_{ik} = m_{ik}^{(Context)}$.

Computing similarities. In the sub-rating matrix $R_t = \{r_{ik}\}$, each entry r_{ik} represents the user U_i's preference toward a single item I_k. The user U_i's preference toward all items as a whole can be represented by the cross-product FoD $\Theta_{all} = \Theta_1 \times \Theta_2 \times ... \times \Theta_N$, where $\Theta_i = \Theta, \forall i = \overline{1, N}$. The cylindrical extension of the focal element of r_{ik} to the cross-product Θ_{all} is $cyl_{\Theta_{all}}(A) = [\Theta_1 ... \Theta_{i-1} A \Theta_{i+1} ... \Theta_N]$. The mapping $M_{ik} : 2^{\Theta_{all}} \mapsto [0, 1]$, where $M_{ik}(B) = m_{ik}(A)$ for $B = cyl_{\Theta_{all}}(A)$ and 0 otherwise, generates a valid mass function defined on the FoD Θ_{all} [8].

For user U_i, consider $M_{ik}, k = \overline{1, N}$, generated by the extending r_{ik}. The mass function $M_i : 2^{\Theta_{all}} \mapsto [0, 1]$, where $M_i = \bigoplus_{k=1}^{N} M_{ik}$, is referred to as the user-BPA of user U_i. The pignistic probability of the singleton $\prod_{k=1}^{N} \theta_{i_k} = \theta_{i_1} \times ... \times \theta_{i_N} \in \Theta_{all}$, is $Bp_i\left(\prod_{k=1}^{N} \theta_{i_k}\right) = \prod_{k=1}^{N} Bp_{ik}(\theta_{i_k})$, where $\theta_{i_k} \in \Theta$, and Bp_i and Bp_{ik} are user U_i's pignistic probability distributions corresponding to the user-BPA and the preference mass function, respectively [15].

We apply the distance measure method in [3], denoted as $CD()$, to compute the distance among users. Let M_i and M_j denote the user-BPAs of users U_i and U_j respectively. The distance between U_i and U_j defined over the same cross-product FoD Θ_{all} is $D(M_i, M_j) = CD(Bp_i, Bp_j) = \sum_{k=1}^{N} CD(Bp_{ik}, Bp_{jk})$, where Bp_{ik} and Bp_{jk} refer to the pignistic probability distributions corresponding to BPAs of user U_i and U_j, respectively [15].

Let us consider a monotonically deceasing function ψ: $[0, \infty] \mapsto [0, 1]$ satisfying $\psi(0) = 1$ and $\psi(\infty) = 0$. With this function, $s_{ij} = \psi(D(M_i, M_j))$ is referred to as the similarity between two users U_i and U_j. We adopt the function $\psi(x) = e^{-\gamma x}$, where $\gamma \in (0, \infty)$. The user-user similarity matrix is then generated as $S_t = \{s_{ij}\}$, where $U_i \in c_t$ and $U_j \in c_t$.

Selecting neighborhoods for active users. We adopt the method proposed in [6] for selecting neighborhoods. Formally, in order to select a neighborhood set $Nbhd_{ik}$ for an active user U_i, the users rated item I_k and whose similarity with user U_i is equal or greater than a threshold τ is extracted. Next, K users with highest similarity with user U_i is selected from extracted list. Note that, if the number of users who already rated item I_k is less than K, $Nbhd_{ik}$ is selected based on the space of all users in c_t.

Estimating rating data for active users. After obtaining $Nbhd_{ik}$, the rating r_{jk} of each neighbor $U_j \in Nbhd_{ik}$ is discounted by the user-user similarity $s_{ij} \in S_t$ between user U_i and U_j as follows

$$m_{jk}^{s_{ij}} = \begin{cases} s_{ij} \times m_{jk}(A), & \text{for } A \in 2^{\Theta}; \\ s_{ij} \times m_{jk}(\Theta) + (1 - s_{ij}), & \text{for } A = \Theta. \end{cases}$$

In community c_t, the estimated rating data for a user U_i on an unrated item I_k is represented as $\hat{r}_{ik}^{(c_t)} = \hat{m}_{ik}^{(c_t)} = m_{ik}^{(Nbhd)} \oplus m_{ik}$, where $m_{ik}^{(Nbhd)} = \bigoplus_{\{j | U_j \in Nbhd_{ik}\}} m_{jk}^{(disc)}$.

3.5 Generating recommendations

The suitable recommendation data for each active user is generated according to the number of communities to which the user belong. If an active user U_i is a member of only one community c_t, the finally estimated rating data of this user on an item I_k, denoted by $\hat{m}_{ik} = \hat{m}_{ik}^{(c_t)}$. In case the active user U_i belongs to a variety of communities simultaneously, the finally estimated rating data for this user on item I_k is achieved by using Dempster's rule of combination operation for fusing the estimated data on item I_k in the communities to which user U_i belong as follows: $\hat{m}_{ik} = \bigoplus_{\{i|U_i \in c_t, t=\overline{1,K_C}\}} \hat{m}_{ik}^{(c_t)}$.

For a hard decision on a singleton $\theta_i \in \Theta$, the pignistic probability is applied, and a singleton having the highest probability is selected as the preference label. For a soft decision, the maximum belief with overlapping interval strategy (maxBL) [2] is applied; in this case, the singleton preference label whose belief is greater than the plausibility of any other singleton is chose.

4 System implementation and discussion

We selected Flixster data set [1] consisting of friend relationships and hard rating data with the rating value from 0.5 to 5 with step size 0.5. Then, we enriched the data set by crawling the genres of movies. After crawling and cleaning, we achieved a new Flixster data set containing 49,410 friend relationships, 535,013 hard ratings from 3,827 users on 1210 movies. Additionally, each user has rated at least 15 movies and total of movies' genres is 19. Since the information about the genres to which a user belongs is not available, we also assume that the genres of a user U_i are assigned by the genres of all items rated by this user. For transforming each hard rating entry θ_l into soft rating entry r_{ik}, we applied the DS modeling function proposed in [15] as below

$$r_{ik}{=}m_{ik} = \begin{cases} \alpha_{ik}(1-\sigma_{ik}), & \text{for } A=\theta_l; \\ \frac{2}{5}\alpha_{ik}\sigma_{ik}, & \text{for } A=B; \\ \frac{3}{5}\alpha_{ik}\sigma_{ik}, & \text{for } A=C; \\ 1-\alpha_{ik}, & \text{for } A=\Theta; \\ 0, & \text{otherwise,} \end{cases} \text{ with } B = \begin{cases} (\theta_1,\theta_2,\theta_3), & \text{if } l=1; \\ (\theta_1,\theta_2,\theta_3,\theta_4), & \text{if } l=2; \\ (\theta_{L-3},\theta_{L-2},\theta_{L-1},\theta_L), & \text{if } l=L-1; \\ (\theta_{L-2},\theta_{L-1},\theta_L), & \text{if } l=L; \\ (\theta_{l-2},\theta_{l-1},\theta_l,\theta_{l+1},\theta_{l+2}), & \text{otherwise,} \end{cases}$$

and $C = \begin{cases} (\theta_1,\theta_2), & \text{if } l=1; \\ (\theta_{L-1},\theta_L), & \text{if } l=L; \\ (\theta_{l-1},\theta_l,\theta_{l+1}), & \text{otherwise,} \end{cases}$ where $\alpha_{ik} \in [0,1]$ is a trust factor and $\sigma_{ik} \in [0,1]$ is a dispersion factor.

For each user in the data set, we withheld accidentally 5 ratings. These withheld ratings were used as the testing data; and the remaining ratings were considered as the training data. CoFiDS [15] with context information for rating refinement was selected for performance comparison using the following assessment methods: MAE, $Precision$, $Recall$, F_β [7]; $DS\text{-}Precision$, $DS\text{-}Recall$ [8]; $DS\text{-}MAE$, $DS\text{-}F_\beta$ [15].

In the experiment, we selected the parameters as follows: $\gamma = 10^{-5}$, $\beta = 1$, $K = 20$, $\tau = 0.9$, $\forall(i,k)\{\alpha_{ik},\sigma_{ik}\} = \{0.9, 2/9\}$, and $z = 0.1$. After detecting communities, we achieved 7 overlapping communities. Table 1 and Table 2 show summarized results of the performance comparisons between the proposed system and CoFiDS in soft and hard recommendations, respectively. In these tables, every rating value has its

[1] http://www.cs.ubc.ca/∼jamalim/datasets/

Table 1. The comparison in soft recommendations

$DS\text{-}Metric$	True rating value									
	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Proposed system:										
MAE	3.2795	2.8546	2.3471	1.8341	1.3714	0.8807	0.4626	0.1501	0.5632	0.9573
$Precision$	0.8138	0	**0**	0.1270	0.2143	0.1981	0.1799	0.2031	0.1733	0.3874
$Recall$	0.0223	**0**	**0**	0.0008	0.0019	0.0628	0.1569	0.7784	0.0141	0.0906
F_1	0.0434	**0**	**0**	0.0015	0.0038	0.0954	0.1676	0.3221	0.0261	0.1468
CoFiDS:										
MAE	3.3278	2.8982	2.4152	1.8932	1.4068	0.8977	0.4796	0.1244	0.5714	0.9995
$Precision$	0.8631	0.0058	**0**	0	0.0289	0.2027	0.1742	0.2004	0.1116	0.3945
$Recall$	0.0221	**0**	**0**	0	0	0.0631	0.1219	0.8214	0.0014	0.0696
F_1	0.0431	**0**	**0**	0	0	0.0962	0.1434	0.3222	0.0028	0.1183

Table 2. The comparison in hard recommendations

$Metric$	True rating value									
	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Proposed system:										
MAE	3.3107	2.8694	2.3846	1.8681	1.3871	0.8948	0.4703	0.1540	0.5781	0.9801
$Precision$	0.8750	*0*	*0*	*0.1667*	*0.2727*	0.1958	0.1812	0.2030	0.1742	0.3870
$Recall$	0.0223	**0**	**0**	0.0015	0.0031	0.0639	0.1585	0.7776	0.0137	0.0893
F_1	0.0435	*N/A*	*N/A*	*0.0030*	*0.0061*	0.0964	0.1691	0.3220	0.0254	0.1451
CoFiDS:										
MAE	3.3265	2.8976	2.4141	1.8936	1.4073	0.8980	0.4798	0.1242	0.5712	0.9990
$Precision$	0.7778	*N/A*	*N/A*	*N/A*	*N/A*	0.2017	0.1736	0.2004	0.1034	0.3953
$Recall$	0.0221	**0**	**0**	0	0	0.0629	0.1210	0.8215	0.0013	0.0698
F_1	0.0429	*N/A*	*N/A*	*N/A*	*N/A*	0.0959	0.1426	0.3222	0.0026	0.1186

own column; underlined values indicate the better performance, bold values illustrate equal performances, and italic values mention that they are incomparable for comparison. Since, in the data set, the number users rated as 1.0, 1.5, 2.0 or 2.5 is very small compared to the number of people rated as higher values, the column regarding rating value ranging from 1.0 to 2.5 contains some values as 0 or N/A (Not applicable). As we have seen from the statistics in both Table 1 and Table 2, the proposed system achieves better performance in all selected assessment criteria in most of true rating value categories. However, the absolute values of the performance of the proposed system are just slightly higher than those of CoFiDS. If we identify communities in the social network by using another information such as the number of messages, emails, comments, tags. maybe the different absolute values will be greater.

5 Conclusion

In this paper, we have developed a community-based collaborative filtering system dealing with data imperfections based on DS theory, and integrating the community context information extracted from the social network into the purpose of tackling the sparsity problem. In the experiment, we selected Flixster data set for evaluating our system. Additionally, we already enriched this data set by crawling the movies genres. Regarding the experimental results, our system gains better performances in both hard and soft decisions compared with CoFiDS.

References

1. G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 17(6):734–749, 2005.
2. I. 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.
3. H. Chan and A. Darwiche. A distance measure for bounding probabilistic belief change. *Int. J. Approx. Reasoning*, 38(2):149–174, 2005.
4. A.P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
5. H.F. Durrant-Whyte and T.C. Henderson. Multisensor data fusion. In Bruno Siciliano and Oussama Khatib, editors, *Springer Handbook of Robotics*, pages 585–610. Springer, 2008.
6. J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *SIGIR '99: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 15-19, 1999, Berkeley, CA, USA*, pages 230–237. ACM, 1999.
7. J.L. Herlocker, J.A. Konstan, L.G. Terveen, and J. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53, 2004.
8. K.K.R. Hewawasam, K. Premaratne, and M.L. Shyu. Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(6):1446–1459, 2007.
9. Z. Huang, H. Chen, and D. Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems*, 22(1):116–142, January 2004.
10. B. Khaleghi, A. Khamis, F.O. Karray, and S.N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28 – 44, 2013.
11. Y. Koren, R.M. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.
12. F. Ricci, L. Rokach, B. Shapira, and P.B. B. Kantor, editors. *Recommender Systems Handbook*. Springer, 2011.
13. G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
14. P. 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.
15. T.L. Wickramarathne, K. Premaratne, M. Kubat, and D.T. Jayaweera. Cofids: A belief-theoretic approach for automated collaborative filtering. *IEEE Trans. Knowl. Data Eng.*, 23(2):175–189, 2011.
16. J. Xie and B.K. Szymanski. Towards linear time overlapping community detection in social networks. In *PAKDD (2)*, pages 25–36, 2012.
17. L.A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.