Electronic Commerce Research and Applications 26 (2017) 101–108

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

Using community preference for overcoming sparsity and cold-start problems in collaborative filtering system offering soft ratings ☆

Van-Doan Nguyen [a,]*, Songsak Sriboonchitta [b], Van-Nam Huynh [a]

[a] Japan Advanced Institute of Science and Technology, Japan
[b] Chiang Mai University, Thailand

ARTICLE INFO

Article history:
Received 23 January 2017
Received in revised form 1 October 2017
Accepted 6 October 2017
Available online 7 October 2017

Keywords:
Recommender systems
E-commerce
Soft ratings
Community preferences
Dempster-Shafer theory

ABSTRACT

This paper introduces a new collaborative filtering recommender system that is capable of offering soft ratings as well as integrating with a social network containing all users. Offering soft ratings is known as a new methodology for modeling subjective, qualitative, and imperfect information about user preferences, as well as a more realistic and flexible means for users to express their preferences on products and services. Additionally, in the system, community preferences that are extracted from the social network are employed for overcoming sparsity and cold-start problems. In the experiment, the new system is tested using a data set culled from Flixster, a social network focused on movies. The experiment's results show that this system is more effective than the selected baseline in terms of recommendation accuracy.

1. Introduction

Recommender systems (RSs) (Adomavicius and Tuzhilin, 2005; Hwang et al., 2010; Kim et al., 2002) have been developing rapidly since they were introduced in 1990s; in practice, RSs have been applied in a variety of e-commerce applications (Linden et al., 2003; Shambour and Lu, 2011; Kim et al., 2011). In general, RSs collect information about user preferences from multiple sources, estimate user preferences on unseen items, and then generate suitable recommendations based on the estimated data. Logically, the quality of the recommendations in an RS mainly depends on the representation of user preferences and the accuracy of estimations.

Conventional RSs described in the literature usually provide a rating domain defined as a totally ordered finite set of rating scores, and represent a user's preference for an item as a rating score (i.e., a hard rating). However, user preferences are naturally subjective and qualitative; therefore, representing user preferences as hard ratings is not appropriate in some cases (Nguyen and Huynh, 2015). For example, let us consider an RS offering hard ratings with a rating domain $\Theta = \{1, 2, 3, 4, 5\}$. Suppose that a user has rated two items, I_1 and I_3, with rating scores of 4 and 5, respectively; further, this user wants to rate item I_2 with a rating score indicating that it is better than I_1 but worse than I_3. By using a rating score, the user can evaluate item I_2 only as 4 or 5; as a consequence, this user may hesitate to express his/her preference for the item. In this scenario, it will be more comfortable for the user to use a combination of rating scores, such as $\{4, 5\}$ (i.e., a soft rating). Moreover, even though a user has evaluated an item by using a hard rating, this rating might encode imperfect information inside. For instance, when a user has rated an item with a rating score, such as 3, we cannot know exactly what this user thought about the item because the user probably wanted to evaluate the item as *at least 3* or *3 for 90%*. To addess such situations, the use of soft ratings in RSs has recently been studied and developed for the purpose of capturing subjective, qualitative, and imperfect information about user preferences (Wickramarathne et al., 2011; Nguyen and Huynh, 2014; Nguyen and Huynh, 2015). With a rating domain $\Theta = \{1, 2, 3, 4, 5\}$, a user can express his/her preference for an item by using a soft rating as follows:

- *3 for sure* ($\{3\}$ with a probability of 1.0)
- *At least 3* ($\{3, 4, 5\}$ with a probability of 1.0)
- *3 for 90%* ($\{3\}$ and Θ with probabilities of 0.9 and 0.1, respectively)
- *Less than 3* ($\{1, 2\}$ with a probability of 1.0)

According to the aforementioned studies, RSs offering soft ratings are developed based on the Dempster-Shafer theory (DST) (Dempster, 1967; Shafer, 1976), which is considered as one of

☆ This paper is an extended and revised version of the conference paper presented at CSoNet-2016 (Nguyen and Huynh, 2016)

* Corresponding author.
E-mail address: nvdoan@jaist.ac.jp (V.-D. Nguyen).

https://doi.org/10.1016/j.elerap.2017.10.002
1567-4223/

the most general theories for modeling imperfect information (Wickramarathne et al., 2011).

Furthermore, in the RS research area, recommendation techniques can be divided into three main categories: collaborative filtering, content-based, and hybrid (Adomavicius and Tuzhilin, 2005). Among these, collaborative filtering is regarded as the most well-known technique (Ricci et al., 2011; Jannach et al., 2012). To provide personalized recommendations to an active user, collaborative filtering RSs commonly try to find other users who are expected to have preferences similar to those of the user, and employ those users' ratings to estimate the original user's preferences for unseen items. However, collaborative filtering RSs also suffer from two fundamental limitations known as sparsity and cold-start problems (Adomavicius and Tuzhilin, 2005). The first problem is caused when each user only rates a very small number of items; as a result, the number of ratings is insufficient and recommendation performance is significantly affected (Huang et al., 2004). The second problem arises because of missing information about new items and new users.

Social networks are growing very fast and play a significant role on the Internet. Additionally, communication and collaboration in social networks have become more and more convenient and frequent. In a social network, users naturally form into various communities whose members interact frequently with one another (Tang and Liu, 2010). These social relations might influence individual behaviors and decisions, including those related to buying items. Commonly, when consulting for advice before buying new items, most people tend to believe in recommendations from friends in the same community rather than in recommendations from outside users. In a community, each member has his/her own preference for an item, and the overall preference of all members is called the community preference for that item. In practice, community preferences can be used for better understanding of user behaviors and ratings; thus, these preferences can potentially be used for predicting the missing information about user preferences. In other words, community preferences might be useful for overcoming the limitations in collaborative filtering RSs.

On the basis of this observation, this paper proposes a new collaborative filtering RS that is capable of (1) representing user preferences as soft ratings and (2) exploiting community preferences derived from a social network containing all users, which helps address the aforementioned sparsity and cold-start problems. In the system, community preferences are employed for predicting all unprovided user ratings on items (including new users and new items); subsequently, active users receive personalized recommendations based mainly on the provided and predicted ratings.

The remainder of this paper is organized as follows. In Section 2, background information about DST is introduced. In Section 3, related work is discussed. The proposed system is described in Section 4. In Section 5, experiments are presented, and their results are discussed. Finally, conclusions and suggestions for future research are presented in Section 6.

2. Dempster-Shafer theory

DST (Dempster, 1967; Shafer, 1976), also called evidence theory or the theory of belief functions, is a well-known theory for modeling uncertain, imprecise, and incomplete information. In the context of this theory, let us consider a problem domain represented by a finite set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_L\}$ of mutually exclusive and exhaustive hypotheses (Shafer, 1976). A function $m : 2^\Theta \to [0, 1]$ is called a mass function if it satisfies

$$m(\emptyset) = 0 \text{ and } \sum_{A \subseteq \Theta} m(A) = 1. \tag{1}$$

A subset $A \subseteq \Theta$ with $m(A) > 0$ is called a focal element. In addition, mass function m is vacuous if $m(\Theta) = 1$ and $\forall A \neq \Theta,\ m(A) = 0$.

If the information source that provides mass function m has probability δ of reliability, this function needs to be discounted by a rate of $1 - \delta \in [0, 1]$, as follows:

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

Based on mass function m, belief function $Bel : 2^\Theta \to [0, 1]$ and plausibility function $Pl : 2^\Theta \to [0, 1]$ can be derived, as shown below:

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \text{ and } Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \tag{3}$$

According to Smets' two-level view in the so-called transferable belief model (Smets and Kennes, 1994; Smets, 1999), when a decision needs to be made, mass function m should be transformed into a probability distribution called pignistic probability function $Bp : \Theta \to [0, 1]$ defined by:

$$Bp(\theta_i) = \sum_{\{A \subseteq \Theta | \theta_i \in A\}} \frac{m(A)}{|A|}. \tag{4}$$

Suppose that we need to combine two mass functions m_1 and m_2 defined on Θ. In this case, we can use Dempster's rule of combination (Shafer, 1976), denoted by $m = m_1 \oplus m_2$, as follows:

$$\begin{aligned} & m(\emptyset) = 0; \\ & m(A) = \frac{1}{1 - \mathcal{K}} \sum_{\{B,C \subseteq \Theta | B \cap C = A\}} m_1(B) m_2(C), \end{aligned} \tag{5}$$

where $\mathcal{K} = \sum_{\{B,C \subseteq \Theta | B \cap C = \emptyset\}} m_1(B) m_2(C)$, and $\mathcal{K}$ represents the mass associated with the conflict.

3. Related work

Over the years, many researchers have focused on tackling the two fundamental problems in collaborative filtering RSs, and various methods have been developed for addressing these problems. Regarding the previous studies, matrix factorization (Barjasteh et al., 2015; Bauer and Nanopoulos, 2014; Guo et al., 2015), which exploits latent factors for predicting all unprovided ratings, is known as a popular method. In addition, some authors proposed combining collaborative filtering with content-based techniques (Jiang et al., 2015) or employing information from other sources, such as demographic information (Kim et al., 2010; Lika et al., 2014) or implicit preferences derived from users' behaviors (Grcar et al., 2005). Further, integrating RSs with social networks has emerged as a major research topic because these networks contain a large amount of information that should be valuable for improving the quality of recommendations (Konstas et al., 2009; Sun et al., 2015). Additionally, a variety of collaborative filtering RSs employ social trust for overcoming the problems (Guo et al., 2014; Li et al., 2013; Wu et al., 2016). However, the matrix factorization method and the other methods were developed for collaborative filtering RSs offering hard ratings only.

As mentioned previously, RSs using soft ratings have recently been studied and developed. The authors in (Wickramarathne et al., 2011) introduced a collaborative filtering RS that offers soft ratings and overcomes the sparsity problem by employing context information for generating unprovided ratings. However, when context information does not influence a user on an item, the corresponding unprovided rating cannot be generated (Nguyen and Huynh, 2015). Moreover, this system is not capable of addressing the cold-start problem or integrating with social networks. In