Information Sciences 422 (2018) 21–50

Contents lists available at ScienceDirect

Information Sciences

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

A new confidence-based recommendation approach: Combining trust and certainty

Faezeh Sadat Gohari*, Fereidoon Shams Aliee, Hassan Haghighi

Faculty of Computer Science and Engineering, Shahid Beheshti University G.C., Tehran, Iran

ARTICLE INFO

Article history:
Received 22 December 2016
Revised 23 July 2017
Accepted 3 September 2017
Available online 5 September 2017

Keywords:
Recommender systems
Collaborative filtering
Trust
Certainty
Entropy
Confidence

ABSTRACT

Collaborative Filtering (CF) is one of the most successful recommendation techniques. Recently, implicit trust-based recommendation approaches have emerged that incorporate implicit trust information into CF in order to improve recommendation performance. Previous implicit trust models assume that all users have the same perception of ratings. However, although all users employ members of the same rating domain (e.g. ratings on a 1–5 scale), each individual has his own interpretations about a rating domain in order to express his preferences. Thus, it is reasonable that a user's rating vector has some degree of uncertainty, depending upon the rating usage trend of that user. In this paper, we present a new approach for confidence modeling in the context of recommender systems. The idea of this modeling is that confidence in a particular user depends not only on the trust in the opinions of that user but also on the certainty of these opinions. Based on this idea, we propose a new Confidence-Based Recommendation (CBR) approach. This approach employs four different confidence models that derive the users' and items' confidence values from both local and global perspectives. Experimental results on real-world data sets demonstrate the effectiveness of the proposed approach.

1. Introduction

With the rapid growth of the World Wide Web, people are faced with more information than they can process and understand. Recommender Systems (RSs) are developed to reduce this information overload problem by suggesting personalized services or products that are relevant to the needs and interests of users. These systems are applied in various domains such as news [4], digital libraries [45,46], movies [11], music [19], etc. There are two basic entities in an RS: user and item. The main purpose of RSs is to predict the ratings of new items for a user who is called the 'active user' . For this purpose, a filtering algorithm is applied to the observed ratings and a list of recommendations is generated for the active user.

Collaborative Filtering (CF) is the most known and widely used approach for providing accurate recommendations to users. The underlying assumption of the CF approach is that similar users share similar interests. Generally, CF can be classified into two categories [8]: memory-based CF (also known as neighborhood-based CF) and model-based CF. The heart of memory-based CF approaches is the measurement of similarity between users (user-based CF) [38] or items (item-based CF) [39]. The most commonly used similarity measures are based on Pearson correlation and cosine similarity [8,38,39]. Memory-based approaches use the entire rating matrix to compute the similarity between users or items. Thus, a major

* Corresponding author.
E-mail address: f_gohari@sbu.ac.ir (F.S. Gohari).

http://dx.doi.org/10.1016/j.ins.2017.09.001
0020-0255/

drawback of these approaches is their high computational complexity [44]. Model-based approaches learn a model from the observed data and use this model to predict unknown ratings. These models include Bayesian models [15], clustering models [11], fuzzy linguistic models [40], low-dimensional linear factor models such as Matrix Factorization (MF) [29], etc. Once the learning phase is finished, a model-based approach can quickly predict unknown ratings. Memory-based CF is more widely used than the model-based one due to its simplicity, justifiability, and stability [9,35].

Despite its wide application, CF suffers from some inherent problems such as *data sparsity* and *cold-start*. These problems have a serious impact on the accuracy of the generated recommendations. To alleviate these issues, hybrid approaches have been proposed that improve the performance of CF using additional information such as content information [4], sentiment information [1,33], trust information [12,34], semantic information [10,11], etc. Among the hybrid RSs, trust-based systems [6,9,12,25,42,48] have attracted much attention in recent years. In the context of RSs, trust can be defined as one's belief toward others in providing accurate ratings relative to the preferences of the active user [12]. Generally, there are two types of models to compute trust values: local trust and global trust models [28]. Global trust models compute the reputation of a user within the network [31,42], while local trust models calculate the trust between two users based on their past interactions [26,41]. In general, while local trust models can be more precise than global models [18,28], they suffer from the cold-start problem [14].

Trust information can be explicitly collected from users or implicitly inferred from a user's past ratings on items [12]. Although explicit trust tends to be more accurate than implicit trust, explicit trust statements require additional user effort and they may not always be available [50]. Hence, this work focuses on implicit trust.

Most of the implicit trust models use the prediction error on co-rated items as an indication of trust in the opinions of a user, but it is not known how much certainty can be placed in the inferred trust information. In the process of inferring implicit trust values, previous research works assume that all individuals have the same perception of ratings. However, each user has his own interpretations about a rating domain (e.g. ratings on a 1–5 scale) in order to express his preferences [21]. In other words, while all users employ the same rating domain to express their preferences, they are likely to have different rating usage trends. For example, a user may not regard all members of a 5-star rating scale to express his tastes on products, while another user may utilize each value of the rating scale in order to differentiate his preferences. In addition, a user might be an easily satisfied person who mostly employs a rating of 4 or 5, while another user might be an exacting person who mostly utilizes 1 or 2. Thus, due to the different perceptions of users, it is reasonable that a user's rating vector has some degree of uncertainty [21]. Such uncertainty can be estimated by calculating the entropy [43] of the rating vector. In the information theory, entropy measures the randomness or disorder of a set of random events. In general, the greater the disorder or randomness, the higher the information entropy is. Some previous studies [13,21,24,37,49] have shown positive outcomes of incorporating entropy values into the traditional CF approaches. In this paper, we aim to combine both trust and certainty measures in order to obtain a synthetic measure for identifying the most promising neighbors of a user.

Trust and certainty are two important concepts in RSs. To the best of our knowledge, previous studies only consider either trust or certainty as a separate model in an RS. These studies do not exploit the certainty of trust as an integrated model. To fill the current gap, we integrate these two concepts into a single confidence model and apply it to a new recommendation system. In other words, we define confidence as a combined model that integrates trust and certainty factors, where certainty is interpreted as how a user perceives the rating domain to distinguish his preferences. The idea of this hybrid model is that confidence in a particular user depends not only on the trust in the opinions of that user but also on the certainty of these opinions. Thus, confidence is a measure of the certainty of the inferred trust values, so that more certainty can be placed on the trust values between those who have similar rating usage trends. Such a hybrid model allows the system to identify the most confident neighbors of an active user. We note that this confidence model can also be easily employed on the selection of neighbors for an item. For this purpose, the confidence model must integrate trust information between items [42] and certainty information associated with the items' rating vectors [21].

Based on the presented confidence model, we propose a new recommendation approach called Confidence-Based Recommendation (CBR). In this approach, the user's neighborhood formation process relies on both trust values between users and their interpretations of ratings. The users' interpretations about the rating domain are compared by computing the entropy of their preference vectors. Therefore, the proposed approach can be viewed as a hybrid of entropy-based and trust-based CF systems.

For the sake of improving recommendation quality, CBR fuses user-level and item-level confidence models to exploit more of the available information. The user-level confidence model combines the implicit trust information between users and the certainty of their rating vectors to measure the degree of confidence between users. These confidence values help the system identify the most confident neighbors of the active user. The item-level confidence model combines the implicit trust information between items and certainty of the items' rating vectors to identify the best neighbors of the target item.

Moreover, in the proposed approach, the user-level and item-level confidence values are derived from both local and global perspectives. User-level local confidence is calculated based on: (1) local certainty, which compares the perceptions of two users about the rating domain; and (2) local trust, that is the trust between those users. User-level global confidence is calculated based on global certainty and global trust. The global certainty depends on the similarity between the perception of a specific user about the rating domain and that of other users. The global trust is the average opinion of the whole community about the trustworthiness of a user. Using such global confidence values, the system can provide reasonable recommendations for cold-start users who have rated only a few or even no items. Similarly, item-level local/global confidence is computed based on the local/global trust and certainty information regarding the items' rating vectors. The utilization