Decision Support Systems 83 (2015) 10–21

Contents lists available at ScienceDirect

Decision Support Systems

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

Prediction uncertainty in collaborative filtering: Enhancing personalized online product ranking

Mingyue Zhang, Xunhua Guo*, Guoqing Chen

Research Center for Contemporary Management, School of Economics and Management, Tsinghua University, Beijing100084, China

ARTICLE INFO

Article history:
Received 14 February 2015
Received in revised form 18 December 2015
Accepted 18 December 2015
Available online 30 December 2015

Keywords:
Online product ranking
Prediction uncertainty
Collaborative filtering
Posterior rating distribution
Confidence level

ABSTRACT

Personalized product ranking provides support to the decision making of online consumers and helps improve their satisfaction, since consumers always face a large volume of choices when they are shopping online. Recommender systems with collaborative filtering techniques are commonly used for this purpose, wherein products are ranked according to their predicted ratings. However, this kind of ranking approaches (namely, Ranking by Collaborative Filtering, *RCF* for short) have generally ignored the impacts of prediction uncertainty. This paper proposes a novel ranking approach called *RPU* (Ranking with Prediction Uncertainty), which utilizes posterior rating distribution and confidence level of prediction as two key factors for prediction uncertainty. Serving as a critical component of the generalized ranking framework, *RPU* aims to improve the accuracy of personalized product ranking through incorporating the uncertainty information. Experiments using real-world data of movie ratings show that *RPU* achieves higher ranking performance compared to traditional *RCF* and the results are robust in terms of sparse data.

1. Introduction

Online consumers often face large amounts of product choices that are beyond their cognitive capabilities to process information. For instance, a search for "woman party dress" on Amazon.com would return 164,170 results, as of January 2015. Due to time and energy constraints, most consumers are unable to inspect all available alternatives [1,2]. In response to the challenges brought by the overwhelming information, consumer decision support systems (CDSSs) have been extensively adopted by e-retailers and information intermediaries such as Amazon, Netflix, MovieLens, and PriceGrabber [1,3–5]. With the help of CDSSs, consumers can inspect alternatives that are sequentially sorted based on specific ranking principles, such as prices, release time, number of reviews, and average rating. Such ranking mechanisms typically focus on specific dimensions of general product features, whereas the heterogeneity of consumers is largely ignored [6]. Aiming to improve the services of CDSSs, personalized product ranking is being widely studied and actively developed [3,7–9].

The most popular strategy for personalized product ranking is known as collaborative filtering (*CF*) [7]. Although there exist various recommendation and display methods for *CF*-based systems, the general idea is to predict the preferences of potential customers by analyzing the opinions of large amounts of users [10]. Among them, rating-based recommendation represents a major portion of the existing literature on recommender systems [11]. The notion of *ratings* is commonly used as a representative of customer preferences. Different forms of input data can be utilized by *CF* techniques to predict the unknown ratings, including both explicit (i.e., existing ratings) and implicit feedback (such as purchase and click activities) of consumers [11]. For instance, Amazon.com makes use of purchase data as implicit feedback to estimate consumer preferences, which can also be seen as binary ratings with the values of either 0 (not purchased) or 1 (purchased). On Netflix, although they do not explicitly display the predicted ratings in the top pick list, rating data are used for calculating user preferences. After acquiring predicted ratings based on historical input data, the products are ordered according to their predicted ratings for a target consumer, from the highest to the lowest, resulting in the personalized top-*n* recommendation set [12–15]. Such mechanisms are designed to best fit consumers' personal tastes, in order to help them make wise purchase decisions within a shorter time. By adding value to consumers' shopping experiences, personalized product ranking has also been shown as beneficial to sellers [6].

Although personalized product ranking based on *CF* has been proved to be useful, it is argued that directly using predicted ratings as ranking scores may not be the best solution [8,16,17], because there is always uncertainty along with point prediction [18]. *CF* is a conventional point prediction technique which makes prediction for each individual rating and generates recommendations of

* Corresponding author.
E-mail address: guoxh@sem.tsinghua.edu.cn (X. Guo).

http://dx.doi.org/10.1016/j.dss.2015.12.004
0167-9236/

ranked products based on the predicted preferences [15]. Notably, existing research has not provided an effective method to consider the uncertainty in each individual prediction of product rating. For example, a movie website may recommend to a user the film 'The Hobbit' with a predicted rating of 5 and the film 'Interstellar' with a predicted rating of 4.5. Therefore, the user may decide based on the recommendation that 'The Hobbit' is likely to fit better with his/her preferences as compared with 'Interstellar'. However, in such a scenario the reliability of the two predictions are not incorporated [19]. Suppose that the recommendation for 'The Hobbit' is based on the data of 2 users whose preferences are similar to the target user, while the recommendation for 'Interstellar' is based on the data of 500 users of similar preferences, it would be reasonable to argue that the recommendation for 'Interstellar' is much more reliable and the user should be advised to choose it.

Related research in collaborative filtering has discussed the confidence of individual prediction by exploring the process for calculating predictions [19]. Conceptually, the confidence of prediction informs how 'likely' the specific prediction is to be correct, reflecting uncertainty to some extent. However, how to make a balance between the prediction and its associated confidence remains an open question. In this paper, we focus on analyzing prediction uncertainty of collaborative filtering at the individual level. Based on the introduction of two key factors that can be derived from data characteristics for modeling the uncertainty of predictions made by *CF*s, we proposed a new personalized ranking approach that aims to take the uncertainty into account, which turns out to achieve higher ranking performance. This newly proposed ranking approach serves as a post-processing strategy and it can further improve the ranking accuracy after the *CF* algorithm being applied.

2. Related work

Prior work in the information systems and marketing literature has recognized the importance of product ranking in e-commerce consumers' decision making [6,20]. A considerable amount of research efforts have been devoted into developing business intelligence techniques that incorporate preferences of consumers with features of products in order to provide personalized product ranking [3,7–9,21]. Among the existing studies, recommender systems based on *CF* techniques are the most widely addressed [7] and have been proved to be advantageous largely due to its simplicity and effectiveness.

2.1. Collaborative filtering and prediction uncertainty

As an essential strategy of recommender systems, collaborative filtering (*CF*) aims to help consumers make better decisions by providing relevant items that they may be interested in. Specifically, *CF* techniques model consumer behavior through historical interaction data and make predictions for their preferences to products. They either predict the absolute values of ratings that consumers would give to the unseen items and then order them with descending predictions [22–24], or directly predict the relative preferences of consumers and output a ranked list of products [17,25]. Despite its success in many applications, *CF* approaches nevertheless have been reported to have several limitations [26], among which the sparsity problem attracts most attention. The sparsity problem occurs when historical feedback data is insufficient for identifying customers' preferences, which may severely limit the accuracy of recommendations [27]. Ever since, much research in the *CF* literature has been focused on improving the accuracy of rating predictions by proposing novel and sophisticated algorithms [28,29].

Theoretically, uncertainty exists in all predictions [30]. Since *CF* techniques compute predictions based on the models that are heuristic approximations of human processes and the data that are extremely sparse and incomplete [22], the predictions may occasionally be severely biased. Meanwhile, it has been empirically shown that a recommendation agent's system quality, including prediction uncertainty, has significant effects on users' decision making satisfaction [4]. Therefore, addressing the prediction uncertainty in *CF* should be of considerable research value.

At the aggregate level, uncertainty can be reflected by the overall accuracy of prediction, which is usually measured with the root-mean-square-error (RMSE) between all the predicted and actual values. It has been argued that high uncertainties are well correlated with larger prediction errors [31]. At the individual level, the uncertainty for each prediction depends on the process through which the prediction is generated. The next subsections present some related work dealing with the uncertainty issue from both the aggregation and the individual perspectives.

2.2. The sparsity problem from the aggregation perspective

As mentioned above, data sparsity is a major issue for the prediction uncertainty because *CF* cannot generate useful recommendation for the customer with insufficient previous ratings or purchases. Many researchers have attempted to alleviate this problem, trying to increase the overall prediction accuracy. This can be regarded as reducing the prediction uncertainty from aggregation level. Item-neighborhood based *CF* approach has been proposed to address the sparsity and scalability problems [17,23]. By calculating item–item similarities instead of user–user similarities, the item-based algorithms are much faster than the traditional user-neighborhood based *CF* and provide better recommendation performance. Another proposed approach, dimensionality reduction, aims to reduce the dimensionality of the consumer–product interaction matrix directly and addresses the sparsity problem by removing unrepresentative or insignificant consumers or products to condense the interaction matrix [27]. Some specific techniques can be applied to achieve this purpose, including Principle Component Analysis [32] and Singular Value Decomposition [26]. Empirical studies in [26] indicate that dimensionality reduction can alleviate the sparsity problem compared to the traditional neighborhood based methods. However, earlier systems relied on imputation to fill in missing ratings and make the rating matrix dense [26], which can be very expensive as it increases the amount of data and also have the risk of distorting the data considerably with inaccurate imputation [25]. Hence, more recent research has proposed to model directly the observed ratings only, avoiding overfitting through a regularized model [25,28]. For example, Salakhutdinov & Mnih (2007) [28] presented the Probabilistic Matrix Factorization (PMF) model which performs well on the large, sparse, and very imbalanced Netflix dataset and considerably outperforms standard SVD models. They further extended a fully Bayesian treatment of the PMF model (BPMF) in which model capacity is controlled automatically by integrating over all model parameters and hyperparameters [29] and the empirical results showed *BPMF* achieved even higher prediction accuracy than *PMF* model. In conclusion, existing research has mainly dealt with the sparsity problem from aggregation level, without analyzing the uncertainty associated with each individual prediction.

2.3. Measuring uncertainty from the individual perspective

In contrast with the approaches addressing the sparsity problem from the aggregation level, some studies has tried to measure the prediction uncertainty caused by the sparsity from the individual level, which inspired our relsearch. Firstly, research on the explanation of recommendations reckons that the uncertainty of individual prediction can be reflected by building an explanation facility into the recommendation [22,33–35]. For example, Herlocker et al. [22] explored the utility of explanations in collaborative