Performance Evaluation 77 (2014) 72–95

Contents lists available at ScienceDirect

Performance Evaluation

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

Mathematical modeling of group product recommendation with partial information: How many ratings do we need?

Hong Xie, John C.S. Lui *

Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 7 January 2013
Received in revised form 8 January 2014
Accepted 25 March 2014
Available online 12 April 2014

Keywords:
Group recommendation systems
Partial preference information
Randomized algorithm
Performance evaluation

ABSTRACT

Product recommendation is one of the most important services in the Internet. In this paper, we consider a product recommendation system which recommends products to a *group of users*. The recommendation system only has *partial preference information* on this group of users: a user only indicates his preference to a small *subset* of products in the form of ratings. This partial preference information makes it a challenge to produce an accurate recommendation. In this work, we explore a number of fundamental questions. What is the desired number of ratings per product so to guarantee an accurate recommendation? What are some effective voting rules in summarizing ratings? How users' misbehavior such as *cheating*, in product rating may affect the recommendation accuracy? What are some efficient rating schemes? To answer these questions, we present a formal mathematical model of a group recommendation system. We formally analyze the model. Through this analysis we gain the insight to develop a randomized algorithm which is both computationally efficient and asymptotically accurate in evaluating the recommendation accuracy under a very general setting. We propose a novel and efficient *heterogeneous rating scheme* which requires equal or less rating workload, but can improve over a homogeneous rating scheme by as much as 30%. We carry out experiments on both synthetic data and real-world data from TripAdvisor. Not only we validate our model, but also we obtain a number of interesting observations, i.e., a small of misbehaving users can decrease the recommendation accuracy remarkably. For TripAdvisor, one hundred ratings per product is sufficient to guarantee a high accuracy recommendation. We believe our model and methodology are important building blocks to refine and improve applications of group recommendation systems.

1. Introduction

Nowadays, we are living in the information age with information overload. To deal with such overload, *recommender systems* [1] were introduced which suggest products (hotels, books, songs, etc.) to a user by taking into account the preference of that user. Recommender systems have drawn a lot of attention from both commercial and academic communities over the last decade. We see a number of successful commercial recommender systems like Amazon.com [2], MovieLens [3], etc. A lot of research works have been done on investigating various algorithmic and complexity issues in designing recommender systems [1,4–7]. This type of recommender systems aim to make recommendations to one user and they are also called *classic recommender systems*.

* Corresponding author.
E-mail addresses: hxie@cse.cuhk.edu.hk (H. Xie), cslui@cse.cuhk.edu.hk (J.C.S. Lui).

http://dx.doi.org/10.1016/j.peva.2014.03.004
0166-5316/

However, when users operate in *groups, classic recommender systems* are not appropriate, because the system has to make recommendations by taking into account the preferences of all users within a group instead of one user. Examples of such contexts can be, recommending movies to a number of friends planning to watch together [8], recommending videos to an interest group on YouTube, etc. To deal with such contexts, *group recommendation systems* [9] were introduced. They aim to provide recommendations to a group of users maximizing the overall utility of that group. Recently, a number of successful commercial products of group recommendation systems have emerged [8,10–13]. In this paper, we consider group recommendation systems with *partial preference information*: there are a number of products and a number of users operates in a group, and each user only show his preference to a small *subset* of products in the form of ratings. The system applies some rating aggregation policies to summarize the ratings, and recommends a subset of products to a group users.

The *partial preference information* makes it a challenge to make an accurate recommendation. It is important for us to understand the *accuracy* and *effectiveness* of a group product recommendation system. However this is a challenging work, since a number of rating and human factors may affect the recommendation accuracy and effectiveness. Little attention has been made to this fundamental problem. In this paper, we explore a number of fundamental questions to fill in this void. *What is the desired number of ratings per product so to guarantee an accurate recommendation? What are some effective voting rules in summarizing ratings? How users' misbehavior such as cheating, in product rating may affect the recommendation accuracy? What are some efficient rating schemes?* To the best of our knowledge, this is the first paper which provides a formal model and analysis of such kind of systems. To summarize, our paper makes the following contributions:

- We propose a mathematical model to capture various factors which may influence the accuracy of a group product recommendation system under *partial preference information* settings.
- We formally analyze the model. Through this we gain the insight to develop a randomized algorithm to evaluate the recommendation accuracy under a general setting. We show that this algorithm is computationally efficient and also provides theoretical performance guarantees.
- We propose an efficient *two round heterogeneous rating scheme* which outperforms the homogeneous rating scheme by as much as 30% in recommendation accuracy with the same or less rating workload.
- We carry our experiments on both synthetic data and real-world data (rating data from TripAdvisor). We not only validate our model, but also examine various factors that may affect the recommendation accuracy. We find a number of interesting observations, for example, a small of misbehaving users can decrease the recommendation accuracy remarkably. For TripAdvisor, one hundred ratings per product is sufficient to guarantee a high recommendation accuracy.

This is the outline of our paper. In Section 2, we present the mathematical model of a group recommendation system. In Section 3, we present the formal analysis of the model. In Section 4, we present an efficient randomized algorithm with theoretical performance guarantees to evaluate the recommendation accuracy. In Section 5, we present the experimental results on synthetic data. In Section 6, we present the experimental results on a real-world dataset (from TripAdvisor). Related work is given in Sections 7 and 8 concludes.

2. Mathematical model

We consider a group product recommendation system which recommends k products from a finite set of N candidates denoted by $P_1, \ldots, P_N$, to a group of M users $\mathcal{U} = \{U_1, \ldots, U_M\}$, taking into account the collective preference of the whole user population with that group. Note that $1 \le k \le N$. Users show their preferences in the form of product rating. More concretely, a user only expresses ratings to a small *subset* of products on an m-level cardinal metric denoted by $\{1, \ldots, m\}$. Higher rating implies higher preference. For example, a 2-level (or binary) cardinal metric could be: $\{1 = \text{dislike}, 2 = \text{like}\}$. Ratings from different users are independent. We use the notation $\mathbf{r}_i = \{r_{i,1}, \ldots, r_{i,M}\}$ to denote a set of ratings for product P_i, where $r_{i,j} \in \{1, \ldots, m\}$ if user U_j rates product P_i, otherwise $r_{i,j} = 0$ denotes a missing rating. Let $n_i = |\{r_{i,j} \in \{1, \ldots, m\}, \forall j\}|$ denote the number of observed ratings for product P_i. We treat the observed ratings, say $r_{i,j} \in \{1, \ldots, m\}, \forall i, j$, as *partial preference information*. To decide whether a product should be recommended, the systems infers the collective preference of the user group $\mathcal{U}$ via evaluating ratings. The *partial preference information* makes it challenge to infer the collective preference accurately. There are a number of interesting questions to explore, i.e., how will the number of ratings per product affect the accuracy of the overall recommendation? To guarantee an accurate recommendation, what is the minimum number ratings per product? How users' misbehavior (such as *cheating*) in product rating may affect the final recommendation? *The objective of this work is to examine how various factors can influence the recommendation accuracy.*

To make a recommendation, the system applies a voting rule $\mathcal{V}$ to summarize ratings of a product. Many voting rules are possible. A simple and widely used voting rule is the *average score rule*. Let $\gamma_i = \mathcal{V}(\mathbf{r}^i)$ denote the aggregate rating of product P_i. For the *average score rule*, we compute the aggregate rating as $\gamma_i = \sum_j r_{i,j}/n_i$. There are a number of interesting questions to explore, i.e., what are some effective voting rules? Can one voting rule be more accurate than others?

However, specifying the voting rule is not enough. Recall that the system can only recommend k products. Suppose we rank products based on their aggregate ratings. It may happen that the aggregate rating of the kth ranked product, equals to that of the $(k+1)$th ranked product. In this case, we need to specify a *tie-breaking rule* to decide which product should be recommended. Let $\mathcal{T}$ denote a tie-breaking rule. In this work, we also explore whether the recommendation accuracy is sensitive to a particular tie-breaking rule.