Expert Systems With Applications 58 (2016) 174–183

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Participation recommendation system for crowdsourcing contests[☆]

Yukino Baba[a,*], Kei Kinoshita[b], Hisashi Kashima[a]

[a] Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
[b] Lancers Inc., 3-10-13 Shibuya, Shibuya-ku, Tokyo, 150-0002, Japan

ARTICLE INFO

Article history:
Received 13 January 2016
Revised 17 March 2016
Accepted 7 April 2016
Available online 9 April 2016

Keywords:
Crowdsourcing
Contest
Recommendation system
Transfer learning
Feature-based matrix factorization

ABSTRACT

We propose a novel participation recommendation approach for crowdsourcing contests including probabilistic modeling of contest participation and winner determination. Our method estimates the winning and participation probability of each worker and offers ranked lists of recommended contests. Since there is only one winner in most contests, standard recommendation techniques fail to estimate the accurate winning probability using only the extremely sparse winning information of completed contests. Our solution is to utilize contest participation information and features of workers and contests as auxiliary information. We use the concept of a transfer learning method for matrices and a feature-based matrix factorization method. Experiments conducted using real crowdsourcing contest datasets show that the use of auxiliary information is crucial for improving the performance of contest recommendation, and also reveal several important common skills.

1. Introduction

Crowdsourcing contest is one of the popular embodiments of the idea of crowdsourcing, which is often employed for crowdsourcing complex and creative tasks including graphic design and software development projects. A crowdsourcing contest begins with requesting submissions from crowdsourcing workers. After a deadline, the requester determines a winner from among all the collected submissions, and the winner is paid.

In contrast to microtask crowdsourcing, which is another style of crowdsourcing used for small pieces of work that are usually done in several minutes where workers are usually expected to be paid as long as their submissions meet some threshold, whether or not a particular participant gets paid depends on the quality of submissions provided by other participants in crowdsourcing contests. It results in large fluctuations of the workers' actual incomes. Therefore, it is very important for workers to identify contests in which they have more chances to win. Adequate matching between workers and projects is also an issue for crowdsourcing platform providers, because participants get frustrated owing to the low chance of winning and leave the platforms, which might lead to future depopulation. According to our survey in Lancers,[1] which is one of the major general purpose crowdsourcing marketplaces in Japan, more than 80% of the workers participated in contests less than 10 times, and more than 90% of them have never won any contest. In addition, among the workers who have participated in more than 10 contests, the 25% of top workers occupy 90% of contest winners (Fig. 1).

We propose a statistical modeling of crowdsourcing contests and the result of a study toward the development of contest recommendation systems using two real crowdsourcing contest datasets, one with 602 workers for 12, 384 contests, and the other with 458 workers for 5703 contests. Our statistical model predicts the probability of each worker winning in each contest based on past contest results. The winning probability depends on the ability of workers and the affinity between workers and contests, and the winner is selected from among the participants of the contest. Model estimation is difficult because of the two important characteristics of contests; only one (or a few) participant(s) can be a winner(s) for a particular contest, and no winner is determined for the contests of current interest. This is a considerably extreme case of the "cold-start" situation in recommender systems where observations are too sparse to measure similarity among workers or contests, which results in severe degradation of predictive performance.

To tackle the intrinsic data sparsity problem in contest recommendation, we resort to the use of auxiliary information including participation histories and features of workers and contests. The participation history is a rich source of information for measuring the worker similarity because more than one worker participates in one contest (in contrast with the winner information). In addition, when the participants of the contest of current interest

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author.
E-mail addresses: baba@i.kyoto-u.ac.jp (Y. Baba), kinoshita.kei@lancers.co.jp (K. Kinoshita), kashima@i.kyoto-u.ac.jp (H. Kashima).

[1] https://www.lancers.jp.

http://dx.doi.org/10.1016/j.eswa.2016.04.010

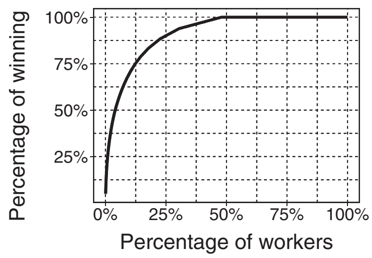

Fig. 1. Cumulative distribution of contest winners against cumulative number of workers (limited to workers who participated in more than 10 contests.) The workers were ordered by descending number of winnings.

are known, they can be used for finding past contests similar to the current contest. The features of workers and contests are also helpful, especially in measuring contest similarity. To utilize auxiliary information, we adopted a transfer learning method (Pan, Xiang, Liu, & Yang, 2010) that reuses the decomposition of the participation matrix, and a feature-based matrix factorization method (Chen, Zhang, Lu, Chen, Zheng & Yu, 2012).

We evaluated the predictive performance of our contest recommendation approach using the two datasets from the Lancers crowdsourcing marketplace. The experimental results showed that transfer learning from participation information is effective in winner prediction, especially with worker and contest features. Additionally, the use of features is considerably effective when the participation history is sparse.

In summary, this paper makes three major contributions.

- We propose a statistical model of winner determination for crowdsourcing contests. To the best of our knowledge, this is the first attempt of predictive analysis of the winning and participation information of crowdsourcing contests.
- We use the concepts of transfer learning and feature-based matrix factorization to solve the intrinsic data sparsity problem in crowdsourcing contests.
- We perform experiments using real contest datasets from a crowdsourcing marketplace to show that the use of auxiliary information is crucial for contest recommendation systems. We also find several latent skills commonly required for a variety of types of crowdsourcing contests.

The rest of this paper is organized as follows: we next discuss related work in Section 2. In Section 3, we formalize our contest recommendation problem and introduce the details of the datasets. We describe our winner determination model and participation model in Section 4. The model estimation approach using the matrix transfer learning method and the feature-based matrix factorization method is discussed in Section 5, and application scenarios of our method are introduced in Section 6. In Section 7, we present the results of experiments using the real crowdsourcing datasets. We conclude with a brief summary and discuss future work in Section 8.

2. Related work

Task recommendation in microtask crowdsourcing has been studied. Ambati, Vogel, and Carbonell (2011), proposed a predictive model of worker preferences for tasks and used the model for task recommendation. Yuen, King, and Leung (2012), created a worker behavior matrix from several information such as the participation and the browsing histories of the crowdsourcing site, and used a matrix factorization method for task recommendation. Lin, Kamar, and Horvitz (2014), proposed a method incorporating positive and negative implicit feedback from workers to learn their preferences. Bozzon et al. (2012), and Difallah, Demartini, and Cudré-Mauroux (2013) developed task recommendation systems that were built on social networking services and used them to extract worker information. Ipeirotis and Gabrilovich (2014), proposed a crowdsourcing framework which used Internet advertising platforms to identify the expertise of workers and displayed suitable tasks to each worker. Task assignment in microtask crowdsourcing have also attracted considerable interest (Chen, Lin, & Zhou, 2013; Donmez, Carbonell, & Schneider, 2009; Ho, Jabbari, & Vaughan, 2013; Ho & Vaughan, 2012; Karger, Oh, & Shah, 2014; Tran-Thanh, Stein, Rogers, & Jennings, 2012; Yan, Rosales, Fung, & Dy, 2011), which aims to optimize the matching between tasks and workers.

In contrast with microtask crowdsourcing, crowdsourcing contests has not been extensively studied yet. Most of the existing research focus on theoretical analysis on designs such as reward allocation of contests (Archak & Sundararajan, 2009; Cavallo & Jain, 2012, 2013; Chawla, Hartline, & Sivan, 2012; DiPalantino & Vojnovic, 2009; Ghosh & McAfee, 2012). There are a few empirical studies on crowdsourcing contests. Araujo (2013) reported results of exploratory analysis of 38, 000 contests on 99 designs,[2] a marketplace focusing on design tasks such as logo, business card and web design. Archak (2010) investigated the players' behavior on TopCoder,[3] a platform for hosting software design and development contests. Boudreau, Lacetera, and Lakhani (2011), also conducted a study on TopCoder and discovered several factors effecting the quality of outcomes. The closest work to ours is presented in Daltayanni, de Alfaro, and Papadimitriou (2015), which proposed a link analysis approach for estimating worker abilities for supporting worker hiring decision. Their method provided a worker recommendation to requesters by using implicit feedback signals such as whether requesters adding a worker to their shortlist. In contrast to this work, we focus on the contest recommendation to workers and we only use explicit feedback such as winning and participation information.

In gaming environments, player rating systems based on winning probability estimation have been widely studied and already been in practical use. The Elo rating system provides a scheme for estimating player level based on performance in games (Elo, 1978). The system was developed for one-versus-one games such as chess and has been applied to various games and sports. Another popular rating system is TrueSkill, which was invented to estimate skill levels of individual players in a game where players are allowed to form a team and more than two teams compete (Dangauthier, Herbrich, Minka, & Graepel, 2007; Herbrich, Minka, & Graepel, 2006). The aim of such rating systems is to model the skill levels in order to optimize matchmaking in a single type of game. By contrast, we focus on crowdsourcing contests where various types of jobs are posted. Thus, we intend to model the skills over different jobs rather than a single type of job.

In educational data mining, the matrix representing the observed test outcomes between respondents and questions is decomposed into a skill matrix and a "Q-matrix" to find required skills for each question and to evaluate skill mastery of each respondent (Desmarais, 2012). However, the matrix is usually rather dense than the winner matrix we analyzed in this paper, and analysis of such extremely sparse winner matrix is one of the major challenges in this paper.

For community question answering sites such as Yahoo! Answers or Quora, a recommender system to suggest right questions to right answers has been developed (Dror, Koren, Maarek, & Szpektor, 2011). The system uses a multi-channel recommender

[2] http://99designs.com.
[3] http://www.topcoder.com.