Information Processing and Management 51 (2015) 163–178

Contents lists available at ScienceDirect

Information Processing and Management

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

Optimal answerer ranking for new questions in community question answering

Zhenlei Yan, Jie Zhou *

Tsinghua National Laboratory for Information Science and Technology (TNList), State Key Laboratory on Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 12 June 2014
Received in revised form 15 July 2014
Accepted 29 July 2014
Available online 24 August 2014

Keywords:
Answerer ranking
Learn to rank
Tensor model
AUC
Community question answering

ABSTRACT

Community question answering (CQA) services that enable users to ask and answer questions have become popular on the internet. However, lots of new questions usually cannot be resolved by appropriate answerers effectively. To address this question routing task, in this paper, we treat it as a ranking problem and rank the potential answerers by the probability that they are able to solve the given new question. We utilize tensor model and topic model simultaneously to extract latent semantic relations among asker, question and answerer. Then, we propose a learning procedure based on the above models to get optimal ranking of answerers for new questions by optimizing the multi-class AUC (Area Under the ROC Curve). Experimental results on two real-world CQA datasets show that the proposed method is able to predict appropriate answerers for new questions and outperforms other state-of-the-art approaches.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, Community Question Answering (CQA) services that allow users to ask and answer questions via the web have been successful on the internet. CQA portals such as Yahoo! Answers[1] and Tencent Wenwen[2] (which is one of the leading CQA service providers in China) provide online platforms for users to post their questions and share their knowledge by answering others' questions. Thousands of millions of questions have been resolved on these CQA portals. For example, in Tencent Wenwen, 207,469,570 questions have been resolved until October 30, 2012.[3] Although these CQA portals have brought significant benefits to users, there are still several drawbacks in current CQA systems. The most prominent problem is that lots of new questions cannot be solved by other users effectively. Previous survey (Furlan, Nikolic, & Milutinovic, 2012; Li & King, 2010) showed that lots of new questions actually cannot be answered efficiently (within 24 h). On the other hand, since the amount of new questions increases rapidly, it is inconvenient for the users who are usually expert in a few domains to find the questions, which they are able to answer. This leads to a low participation rate of users on CQA portals (Guo, Xu, Bao, & Yu, 2008).

To address this problem, several approaches have been proposed in both industry and academia. In the industry, for example Aardvark[4] is a social search engine which routes new questions to the person in the asker's extended social network

* Corresponding author. Tel.: +86 10 62782447.
E-mail address: jzhou@tsinghua.edu.cn (J. Zhou).

[1] http://answers.yahoo.com.
[2] http://www.wenwen.com.
[3] http://www.wenwen.com.
[4] http://vark.com/.

http://dx.doi.org/10.1016/j.ipm.2014.07.009
0306-4573/© 2014 Elsevier Ltd. All rights reserved.

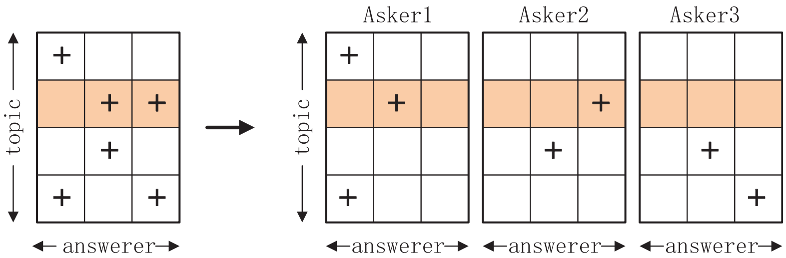

Fig. 1. Expand topic-answerer model in asker dimension. The symbol + means there is a relation between the answerer and topic (topics of questions are usually achieved by topic model tools) which represents that this answerer has given an answer in corresponding topic.

(e.g. Facebook,[5] LinkedIn[6]) who are most likely to be able to answer this question. The details of Aardvark are introduced in (Horowitz & Kamvar, 2010).

In academia, lots of methods (Dror, Koren, Maarek, & Szpektor, 2011; Guo et al., 2008; Li & King, 2010; Li, King, & Lyu, 2011; Liu, Liu, & Yang, 2010; Zhou, Cong, Cui, Jensen, & Yao, 2009; Zhou, Lyu, & King, 2012) have been proposed. Traditional methods often focused on semantic relations between questions and answerers via language model or topic model as shown in left side of Fig. 1. Although such approaches could describe the answerer's interests in questions or related topics, relations between asker and question, asker and answerer have not attracted enough attention. Actually, these relations indeed exist in online CQA portals. For asker and question, questions of the same asker are usually correlated. For example in Fig. 2 which is a snapshot of an asker's homepage on Yahoo! Answers, all questions were talking about *Chrysler* car, in categories of *Maintenance and Repair*, *Chrysler*. Thus, the information of asker is useful to understand questions. For asker and answerer, similar to Facebook where users can see their friends' news in their homepages, the asker's fans/followers are easier to find his/her new questions and more likely to answer them. A related example is shown in Fig. 3, which is a snapshot of a user's contacts/friends webpage on Yahoo! Answers. Then in this paper, we incorporate the asker dimension as traditional topic-answerer model and propose a novel asker-topic-answerer model which is a 3-order tensor (Kolda & Sun, 2008; Lu, Plataniotis, & Venetsanopoulos, 2011) as shown in the right side of Fig. 1.

Furthermore, CQA services allow askers to select the best one among all answers given by answerers. The corresponding answer is called the best answer. Traditional approaches (Li & King, 2010; Liu et al., 2010) usually consider only the best answers and answerers, ignoring the other answers and answerers, or treating them as the same. However, the information of other answers is also beneficial to solve questions and produces much more training data for learning procedure, e.g. more candidate answerers for predicting. Therefore, in our approach, we consider both the best and other answers as shown in Fig. 4. In the left side of Fig. 4 which represents the traditional approach, the blue grids show that asker2 was given two answers by the same answerer on two topics. There is no difference between these answers in traditional approach. However, actually the answerer achieved the best answer on the upper topic (symbol $\circ$) and other answers (symbol $\times$) on the below topic as shown in the right side of Fig. 4. It means this answerer may be more expert in the upper topic than the below one. Such useful information can be used to improve the ranking of answerers.

Thus our intuition is that if an answerer, u achieved many best answers in some topics of a specific asker (i.e. best answerer), u should rank higher than other answerers who gave non-best answers or no answers for new questions in this topic of the same asker. Furthermore, if u achieved many other answers, u should rank higher than the answerers who did not give answers (i.e. non-answerer). Based on this intuition and quality of answers, we classify users into three categories: best answerer, other answerer and non-answerer. As demonstrated in the previous work (Yan & Zhou, 2012), maximization of AUC (Area under the ROC curve) can be used to optimize ranking of two-class problem (answerer and no-answerer). Thus in this paper, it is reasonable to solve the ranking problem of three-class by maximizing multi-class AUC which has been studied in previous researches (Hand & Till, 2001; Marrocco, Duin, & Tortorella, 2008; Toh, Kim, & Lee, 2008). Now the problem mentioned above can be treated as answerer ranking problem.

In this paper, we formally describe the answerer ranking problem and propose a novel approach to learn optimal ranking for new questions effectively. The proposed approach consists of three stages: (a) learn topics of askers, answerers and construct asker-topic-answerer model; (b) optimize asker-topic-answerer model with tensor factorization by maximizing multi-class AUC; and (c) predict the final ranking of potential answerers for new questions. The main contributions can be summarized as below: (1) A novel model is proposed by combining tensor model and topic model to capture the latent semantic relations among asker, question and answerer simultaneously. (2) Based on this model, we propose a new learning approach with tensor factorization to perform the answerer ranking task, which outperforms previous related approaches on two real-world datasets.

The rest of this paper is organized as follows. In Section 2, we present an overview of related work. Next, we formally describe the answerer ranking problem in Section 3. In Section 4, we describe the details of the proposed framework. In

[5] http://www.facebook.com.
[6] http://www.linkedin.com.

Download English Version:

https://daneshyari.com/en/article/515383

Download Persian Version:

https://daneshyari.com/article/515383

Daneshyari.com