Artificial Intelligence Review (2020) 53:843–874
https://doi.org/10.1007/s10462-018-09680-6

Expert finding in community question answering: a review

Sha Yuan[1] · Yu Zhang[2] · Jie Tang[1] · Wendy Hall[4] · Juan Bautista Cabotà[3]

Published online: 19 March 2019

Abstract
The rapid development of Community Question Answering (CQA) satisfies users' quest for professional and personal knowledge about anything. In CQA, one central issue is to find users with expertise and willingness to answer the given questions. Expert finding in CQA often exhibits very different challenges compared to traditional methods. The new features of CQA (such as huge volume, sparse data and crowdsourcing) violate fundamental assumptions of traditional recommendation systems. This paper focuses on reviewing and categorizing the current progress on expert finding in CQA. We classify the recent solutions into four different categories: matrix factorization based models (MF-based models), gradient boosting tree based models (GBT-based models), deep learning based models (DL-based models) and ranking based models (R-based models). We find that MF-based models outperform other categories of models in the crowdsourcing situation. Moreover, we use innovative diagrams to clarify several important concepts of ensemble learning, and find that ensemble models with several specific single models can further boost the performance. Further, we compare the performance of different models on different types of matching tasks, including *text vs. text*, *graph vs. text*, *audio vs. text* and *video vs. text*. The results will help the model selection of expert finding in practice. Finally, we explore some potential future issues in expert finding research in CQA.

Keywords Expert finding · Matrix factorization · Deep learning · Ensemble learning

1 Introduction

With the increasing demand of knowledge sharing services, Community Question Answering (CQA) websites, such as Quora, Toutiao and Zhihu, have already obtained the popularization use in reality. It is common to post questions and answers on CQA websites, where users' quest for professional and personal knowledge in various domains can be satisfied. The central task of CQA is to find appropriate users with willingness and relevant expertise to provide high-quality answers for given questions. This problem has been extensively studied in the past decade. Related researches include expert finding for community-based questions (Riahi et al. 2012; Zhao et al. 2016), expertise modeling (Han et al. 2016), and even a survey of

✉ Jie Tang
jietang@tsinghua.edu.cn

Extended author information available on the last page of the article

Artificial Intelligence Review (2020) 53:843–874
https://doi.org/10.1007/s10462-018-09680-6

专家在社区问题中找到回答：审查

Sha Yuan[1] · Yu Zhang[2] · Jie Tang[1] · Wendy Hall[4] · Juan Bautista Cabotà[3]

在线发布：2019年3月19日　Sp
ringer Nature B.V. 2019

摘要社区问题回答（CQA）的快速发展满足了用户对专业和个人知识的追求。在CQA中，一个核心问题是找到具有专业知识和愿意回答给定问题的用户。与传统方法相比，CQA的专家发现往往呈现出非常不同的挑战。
CQA的新功能（如巨大的批量，稀疏数据和众包）违反了传统推荐系统的根本假设。本文侧重于审查和对CQA中专家查找的当前进展进行分析。我们将最近的解决方案分为四种不同类别：基于矩阵分解的模型（基于MF的模型），基于梯度升压树的模型（基于GBT的模型），基于GBT的模型（基于GBT的模型）（基于DL的模型）和基于DL的模型）和基于排名的模型（R–基于模型）。我们发现基于MF的模型在众群情况下表现出其他类别的模型。此外，我们使用创新图来澄清合奏学习的几个重要概念，并发现具有多种特定单个型号的集合模型可以进一步提高性能。此外，我们将不同模型的性能进行比较在不同类型的匹配任务上，包括文本与文本，图表与文本，音频与文本和视频与文本。结果将有助于在实践中选择专家的选择。最后，我们在CQA中探索了一些潜在的未来问题。

关键词专家发现・矩阵分解・深度学习・集合学习

1 Introduction

随着知识共享服务的需求越来越多，社区问题应答（CQA）网站，如Quora，Toutiao和Zhihu，已经获得了现实的推广用途。在CQA网站上发布问题和答案是常见的，用户可以满足用户在各个领域中的Quest
ProfessionAnd的Quest。CQA的中央任务是为具有意愿和相关专业知识找到适当的用户，以便为给定问题提供高质量的答案。过去十年来，这个问题已被广泛研究过。相关研究包括社区问题的专家查找（Riahi等，2012;
Zhao等，2016），专业型号（Han等人2016），甚至是一项调查

✉ Jie Tang
jietang@tsinghua.edu.cn

在文章的最后一页上提供扩展作者信息

basic solutions (Balog et al. 2012). Though this problem has been studied before (Liu et al. 2005), the willingness of experts has been often ignored. This problem becomes more and more seriously – more than half of the questions on Quora only have one or even do not have any answers.[1]

Expert finding in CQA has generated huge impact to society. It provides a platform to connect questions with experts who can contribute quality answers. Questions about anything can be solved by crowdsourcing in CQA. For example, CQA can help to find a mathematician for a chef with a math problem. At the same time, cooking tips from the chef will be returned to the mathematician if necessary. However, it is often hard for CQA to establish such high-quality expert finding. How to match the questions with interested users' expertise? Can we predict who are the most likely to answer the given questions and what is the probability? Confronting these challenges, the focuses of expert finding in CQA have changed in practice.

Traditional expert finding problem focused on expert finding (Riahi et al. 2012) and expertise ranking (Zhao et al. 2016). The experts would be found for the given question based on text matching. In recent years, the core value of the problem is not finding expert, but solving problems by crowdsourcing. Moreover, expert finding in CQA often exhibits very different challenges compared to traditional methods. The characteristics of expert finding in CQA are summarised as follows.

First, *crowdsourcing*. The complex and intellective problems in CQA require considerable effort and quality contribution. Crowdsourcing considers the users' desire to solve a problem and then freely shares the answer with everyone. In CQA, the answer of the given question would be obtained by crowdsourcing from a large, relatively open and often rapidly-evolving group of interested experts.

Second, *sparse data*. The known question and answer pairs are rare compared to traditional expert finding applications. On one hand, seekers spend more time on finding the answer of their question. On the other hand, experts need to answer multiple versions of the same question. This also makes it challenging to directly use a supervised learning approach due to the lack of training samples.

Third, *new features*. The willingness of expert, the historical behavior of expert, and the quality of answer, all these new features have got more attention. They may contribute to further improve the rationality and effectiveness of expert finding in CQA. For example, the expert who often provides answers with high quality is more likely to answer the similar kinds of questions. How to use these features effectively is widely acknowledged as new challenge to improve the performance of expert finding in CQA.

Based on these observations, most well-known CQA websites and competitions, such as Quora, Toutiao and Kaggle are striving to match questions with interested users' expertise, that is, to find the best respondents to the questions. As for this study, we have got the labeled datasets of the competition ByteCup[2] organized by Toutiao, which is one of the most widely used information distribution platforms in China. We will take the datasets of *Toutiao Q&A* as an example to review the methodologies for expert finding in CQA in this paper.

In this paper, we firstly review the widely used expert finding solutions in CQA and classify all the solutions into different categories, including matrix factorization based models (MF-based models), gradient boosting tree based models (GBT-based models), deep learning based models (DL-based models) and ranking based models (R-based models). In addition, we illustrate the results of all the aforementioned categories of single models on the local validation dataset. The ensemble strategies of the Top 5 teams who won the competition are

[1] https://www.quora.com/What-percentage-of-questions-on-Quora-have-no-answers.

[2] https://biendata.com/competition/bytecup2016/.

844 S. Yuan et al.

基本解决方案（Balog等，2012）。虽然之前已经研究过这个问题（Liu等人，2005），专家的意愿经常被忽视。这个问题变得越来越认真 – Quora上的一半以上的问题只有一个甚至没有任何答案

CQA专家发现对社会产生了巨大影响。它提供了与可以贡献质量答案的专家来连接问题的平台。关于CQA中的众包可以解决关于任何事情的问题。例如，CQA可以帮助找到具有数学问题的厨师的数学家。与此同时，厨师的烹饪提示将在必要时返回数学家。然而，CQA通常很难建立如此高质量的专家发现。如何将问题与有兴趣的用户专业知识相匹配？我们可以预测谁最有可能回答给定问题，概率是多少？面对这些挑战，在CQA中专注的专注于实践中发生了变化。传统专家发现问题专注于专家查找（Riahi等，2012）和专业排名（Zhao等，2016）。专家将在基于文本匹配的情况下找到给定的问题。近年来，问题的核心价值不是找到专家，而是通过众包解决问题。此外，与传统方法相比，CQA的专家发现往往呈现出非常不同的挑战。
CQA中专家发现的特征总结如下。一，众包。
CQA中的复杂和智力问题需要相当大的努力和质量贡献。众群考虑用户解决问题的愿望，然后自由地与每个人分享答案。在CQA中，将通过从大型，相对开放，经常快速不断发展的感兴趣专家群体覆盖来获得给定问题的答案。第二，稀疏数据。与传统专家查找应用相比，已知的问题和答案对非常罕见。一方面，寻求者花更多的时间来寻找他们的问题的答案。另一方面，专家需要回答同一问题的多个版本。这也使得直接使用由于缺乏培训样本而直接使用监督的学习方法。三，新功能。专家的意愿，专家的历史行为，以及答案的质量，所有这些新功能都更加关注。他们可能有助于进一步提高CQA中专家发现的合理性和有效性。例如，经常提供高质量的答案的专家更有可能回答类似的问题。如何有效地使用这些功能被广泛被认为是提高CQA中专家发现性能的新挑战。基于这些观察，最着名的CQA网站和竞争，如Quora，Toutiao和Kaggle，正在努力将问题与有关用户的专业知识相匹配，即找到最佳受访者。至于本研究，我们已获得由Toutiao组织的Bytecup2竞争的标签数据集，这是中国使用最广泛的信息分布平台之一。我们将乘坐Toutiao问答的数据集作为审查CQA中专家在CQA中的方法的示例。在本文中，我们首先审查了CQA中广泛使用的专家查找解决方案，并将所有解决方案分类为不同的类别，包括基于矩阵分解的模型（基于MF的模型），基于渐变的升压树模型（基于GBT的模型），深度学习基于模型（基于DL的模型）和基于排名的模型（基于R的模型）。此外，我们说明了本地验证数据集上的所有上述单个模型类别的结果。赢得竞争的前5名球队的合奏策略是

[1] https://www.quora.com/What-percentage-of-questions-on-Quora-have-no-answers.

[2] https://biendata.com/competition/bytecup2016/.

Springer

also analyzed. What's more, we use innovative diagrams to clarify several important concepts of ensemble learning. This work will significantly help the correct understanding and proper use of ensemble learning in practice. Further, we investigate the performance of different models on different types of matching tasks. Finally, we statistically analyze the results of all expert finding solutions in CQA, and summarize the work of this paper.

The remainder of the paper is organized as follows. In the next section, we first give an overview of the related work. In Sect. 3, we present the problem definition, the widely used CQA datasets, and the categorization of the expert finding techniques. Sections 4, 5, 6 and 7 present the MF-based models, GBT-based models, DL-based models and R-based models, respectively. Section 8 specifies the details of ensemble learning. Sections 9, 10 and 11 present the results and the corresponding analysis. Finally, Sect. 12 concludes the paper.

2 Related work

2.1 Expert finding

Online services with a high-quality recommender system could help users to sift through the expanding and increasingly diverse content. There is a large body of research on recommendation algorithms, including collaborative filtering (Hu et al. 2008; Koren 2008), local focused models (Lee et al. 2013; Christakopoulou and Karypis 2016; Beutel et al. 2017), and more recently deep learning. It is important to choose the appropriate metric for the given recommendation task (Gunawardana and Shani 2009). The possible extensions that can improve recommendation capabilities (Adomavicius and Tuzhilin 2005) include an improvement of understanding users and items, incorporation of the contextual information into the recommendation process, and so on.

Inspired by recent advances in recommender system, expert finding has attracted a lot of attention in the information retrieval community (Li et al. 2015c; Dargahi Nobari et al. 2017; Boeva et al. 2017). The core task of expert finding is to identify persons with relevant expertise for the given topic. Massive efforts have been taken to improve the accuracy of expert finding (Wang et al. 2013). Most existing methods for expert finding can be classified into two groups, including the authority-based methods (Yeniterzi and Callan 2014; Zhu et al. 2014) and the topic-based methods (Deng et al. 2009; Daud et al. 2010; Hashemi et al. 2013). The authority-based methods are based on the link analysis of the past expert-topic activities (Bouguessa and Wang 2008; Liu et al. 2011). The topic-based methods are based on the latent topic modeling techniques (Momtazi and Naumann 2013; Liu et al. 2013b; Lin et al. 2013). Moreover, the emerging deep learning models are integrated with aforementioned methods to further improve the performance of expert finding (Wei et al. 2017; Li and Zheng 2017). They are capable of effectively learning high dimensional representations of expert information, topic information and expert-topic interactions (Ying et al. 2016).

Expert finding has been researched in various areas such as academic (Rani et al. 2015), organizations (Karimzadehgan et al. 2009), social networks (Bozzon et al. 2013; Li et al. 2013), and more recently question answering communities (Cheng et al. 2015). Finding experts with relevant expertise for a given topic has potential applications in these areas such as finding appropriate reviewers for a paper (Mimno and Mccallum 2007; Liang and de Rijke 2016), finding the right supervisor for a student in academic (Alarfaj et al. 2012) and finding the appropriate experts for the questions in CQA (Li et al. 2015a).

还分析了。更重要的是，我们使用创新图来澄清合奏学习的几个重要概念。这项工作将大大帮助正确理解和正确使用合奏学习。此外，我们调查不同模型对不同类型匹配任务的性能。最后，我们在统计上分析CQA中所有专家发现解决方案的结果，并总结了本文的工作。在本文的其余部分安排如下。在下一节中，我们首先概述了相关的工作。昆虫。3，我们介绍了问题定义，广泛使用的CQA数据集，以及专家发现技术的分类。第4,5,6和7节介绍了基于MF的模型，基于GBT的模型，基于DL的模型和基于R的模型。第8节规定了集合学习的详细信息。部分9,10和11呈现结果和相应的分析。最后，教派。12结束了这篇论文。

2 Related work

2.1 Expert finding

具有高质量推荐系统的在线服务可以帮助用户筛选扩展和越来越多的内容。关于辅助算法存在大量研究，包括协作过滤（胡等人。2008; koren
2008），当地重点模型（2013年Lee等; 2013;克里斯卡府和karypis 2016;
Beutel等，2017），和最近深入学习。为给定推荐任务（GunaWardana和Shani
2009）选择适当的公制非常重要。可以提高推荐功能（Adomavicius和Tuzhilin
2005）的可能扩展包括改进了解用户和项目，将上下文信息纳入推荐过程等。受到近期推荐系统的进步的启发，专家发现引起了信息检索社区的注意力（Li等人2015C; Dargahi Nobari等，2017;
Boeva等，2017）。专家查找的核心任务是识别给定主题的相关专业知识的人。已采取大规模努力提高专家发现的准确性（Wang等人。2013）。大多数现有专家查找方法可以分为两组，包括基于权限的方法（Yeniterzi和Callan 2014;
Zhu等，2014）和基于主题的方法（Deng等，2009; Daud等，2010;
Hashemi等。2013）。基于权威的方法是基于过去专家主题活动的链接分析（Bouguessa和Wang 2008;
Liu等人2011）。基于主题的方法基于潜在主题建模技术（MOMTAZI和NAUMANN
2013; LIU等人2013B;
LIN等人2013）。此外，新兴的深度学习模型与上述方法集成，以进一步提高专家发现的性能（Wei等人2017;李和郑2017）。它们能够有效地学习专家信息，主题信息和专家主题交互的高维度表示（Ying等，2016）。专家发现已经在学术界（Rani等，2015），组织（Karimzadehgan等，2009），社交网络（Bozzon等，2013;
Li等人2013）和更多最近的问题回答社区（Cheng et
al。2015）。在这些领域找到具有相关专业知识的专家在这些领域具有潜在的应用，例如为论文找到适当的审稿人（MIMNO和McCallum 2007; Liang And De Rijke
2016），为学术界寻找学生的权利主管（Alarfaj等人。
2012年）在CQA中找到了适当的专家（Li等人2015A）。

2.2 Expert finding in CQA

CQA websites, which provide users with a platform to share their experience and knowledge, are very popular in recent years. Successful CQA websites include general ones (such as Toutiao, Quora and Zhihu), and domain-specific ones (such as Stack Overflow). Finding users with relevant expertise for a specific question in CQA (Zhou et al. 2012; Liu et al. 2015) can increase the quality of answers. It further improves the crucial problems facing by CQA, such as the low participation rate of users, long waiting time for answers and low quality of answers (Neshati et al. 2017). Expert finding in CQA (Zhao et al. 2015a) is a challenging task due to the sparsity of the CQA data, and the emerging features. A great amount of studies have been conducted on expert finding in CQA (Riahi et al. 2012; Zhao et al. 2016). A survey about the early basic methods of expert recommendation in CQA is given in Lin et al. (2017), including query likelihood language (QLL), latent dirichlet allocation (LDA), PageRank, classification, collaborative filtering (CF) and their variants.

With the development of CQA, there are a large amount of advanced solutions of expert recommendation in CQA in recent years (Yang et al. 2013; Liu et al. 2013a; Zhou et al. 2014). Based on the matrix factorization approach, more efficient methods (Koren 2008; Chen et al. 2012; Rendle 2011) are proposed, including the singular-value decomposition (SVD), SVD++, bidirection SVD++ (also named SVD#), "Asymmetric-SVD" (ASVD) and so on. The problem of expert finding in CQA can be regarded as a classification problem when we classify the experts as a particular class of expert users from the other users (Lin et al. 2017). XGBoost (Chen and Guestrin 2016), which is a scalable open source system for the Gradient Boosted Decision Trees (GBDT), has shown its impact in a number of machine learning and data mining challenges. Recently, deep learning models have been widely exploited in various matching tasks with remarkable performance. Since the deep learning based methods have gained much attention, we review the related work in detail in the following subsection.

2.3 Deep learning for recommendation

As deep learning has grown in prominence for computer vision and natural language processing (NLP) tasks, there are a surge of recent works incorporating deep neural network (DNN) in recommender systems. Previous works have largely relied on applying the collaborative filtering intuition to neural networks, such as addressing collaborative filtering by applying a two layer Restricted Boltzmann Machine (Salakhutdinov et al. 2007), joint learning matrix factorization and a feedforward neural network (He et al. 2017), or replacing the traditional linear inner product as a nonlinear decomposition of the rating matrix in auto-encoder (Sedhain et al. 2015). Literature (Wu et al. 2016) utilizes the idea of denoising auto-encoders for top-N recommendation. AutoSVD++ (Zhang et al. 2017) extends the original SVD++ model with a contrastive auto-encoder to capture auxiliary item information.

There has been a recent popularity in using recurrent neural networks (RNNs) for recommendation (Jing and Smola 2017; Tan et al. 2016; Wu et al. 2017). The sequential nature of RNNs (Hidasi et al. 2015) provides desirable properties for timeaware and session-based recommendation systems. More complex networks have been devised to incorporate context (Covington et al. 2016) or memory (Ebesu et al. 2018). The interactions among contextual features are handled through a linear portion (not the DNN portion) of the model (Cheng et al. 2016). Match-SRNN (Wan et al. 2016) applies text matching method to this task. The textual features (such as characters and words in the the expert and question descriptions) model

846 S. Yuan et al.

2.2 Expert finding in CQA

CQA网站为用户提供平台分享他们的经验和知识，近年来非常受欢迎。成功的CQA网站包括一般（例如Toutiao，Quora和Shihu），以及特定于域（例如堆栈溢出）。在CQA中寻找有关专业知识的用户（周等人2012;
Liu等人2015）可以提高答案的质量。它进一步提高了CQA面临的至关重要问题，例如用户的低参与率，答案等待时间和低答案质量低（Neshati等，2017）。
CQA（Zhao等人）专家发现是由于CQA数据的稀疏性以及新兴功能，这是一个具有挑战性的任务。在CQA（Riahi等人2012年）的专家发现，已经进行了大量研究;
Zhao等，2016）。关于CQA在CQA中的早期基本方法的调查在Lin等人中给出。
（2017），包括查询似然语言（QLL），潜在Dirichlet分配（LDA），PageRank，分类，协同过滤（CF）及其变体。随着CQA的发展，近年来CQA在CQA中有大量的专家推荐解决方案（杨等人2013;
Liu等人。2013A;周等人2014）。基于矩阵分解方法，更有效的方法（Koren 2008;
Chen等，2012; rendle 2011）被提出，包括奇异值分解（SVD），SVD ++，Bidirection SVD ++（也命名为SVD＃），“不对称－
SVD“（ASVD）等。当我们将专家作为来自其他用户的特定类别的专家用户归类时，CQA中专家发现的问题可以被视为分类问题（Lin等人。2017）。
XGBoost（Chen和Guestrin
2016），这是一个可扩展的渐变升降决策树（GBDT）的可扩展开源系统，已经对许多机器学习和数据挖掘挑战的影响显示了它。最近，深入学习模型已广泛利用各种匹配任务，具有显着性能。由于基于深度学习的方法遭受了很多关注，因此我们在以下小节中详细审查了相关的工作。

2.3深度学习推荐

随着深度学习的突出突出的计算机视觉和自然语言处理（NLP）任务，近期有兴趣在推荐系统中包含深神经网络（DNN）的作品。以前的作品在很大程度上依赖于将协同过滤直觉应用于神经网络，例如通过应用两层限制的Boltzmann机器（Salakhutdinov等，2007），联合学习矩阵分解和前馈神经网络（He等人）来解决协作滤波。
2017）或将传统的线性内部产品替换为自动编码器中评级矩阵的非线性分解（Sed−Hain等，2015）。文学（Wu等人。2016）利用去除自动编码器的概念，用于TOP−N推荐。
AutoSvd ++（Zhang等人2017）将原始SVD
++模型扩展到对比自动编码器以捕获辅助项目信息。近来，在使用反复性神经网络（RNNS）以获得常规的普及（Jing和Smola 2017; Tan等人2016; Wu等，2017）。
RNNS（HIDASI等，2015）的顺序性质为定日软件和基于会话的推荐系统提供了理想的属性。已经设计了更复杂的网络来包含Con文本（Covington等，2016）或记忆（Ebesu等，2018）。上下文特征之间的相互作用通过模型的线性部分（不是DNN部分）（Cheng等人。2016）。
Match−SRNN（WAN等人2016）将文本匹配方法应用于此任务。文本功能（例如专家和问题描述中的字符和单词）模型

the interaction information between texts. In the model, a neural tensor network is used to capture the character/word level interactions, and a spatial RNN is applied on the character/word interaction tensor to capture the global interactions. Literature (Beutel et al. 2018) bridges the contextual collaborative filtering literature and neural recommender literature. It demonstrates that making use of contextual data in deep neural recommenders (particularly in RNN models) could obtain a significant amount of information. Collaborative Memory Networks (CMN) (Ebesu et al. 2018) unifies the two classes of CF models capitalizing on the strengths of the global structure of latent factor model and local neighborhood-based structure in a nonlinear fashion. In the model, a memory component and a neural attention mechanism are fused as the neighborhood component.

3 Preliminaries

3.1 Problem definition

We first present required definitions and formulate the problem of expert finding in CQA. Our goal is to find experts for a given question in CQA in the way of crowdsourcing. More specifically, given certain question, one needs to find who are the most likely to (1) have the expertise to answer the question and (2) have the willingness to accept the invitation of answering the question.

Definition 1 *Expert* is the user with sufficient expertise for a certain *question* in CQA. The *expertise* are implied in relevant user documents, social interactions, past activities or personal information of each expert.

Given a set of M questions $Q = \{q_1, \ldots, q_M\}$, we need to predict which experts $E = \{e_1, \ldots, e_N\}$ are more likely to answer these questions. For simplicity, we reserve special indexing letters for distinguishing experts from questions, where u, v represent experts, and i, j represent questions.

Problem 1 For a given question i and its candidate expert $u \in E$, one needs to predict the *probability* $\hat{r}_{ui}$ of the expert u answering the question i.

The (u, i) pairs for which r_{ui} is known are stored in the set $Ł = \{(u, i)\,|r_{ui}\ is\ known\}$. The probability $r_{ui} \in [0, 1]$, high values mean stronger preference of the expert u to answer the question i. $\hat{r}_{ui}$ is the predicted probability that the question i will be answered by the expert u based on the labeled data. Here, it is a supervised learning problem to make prediction with the given labeled data. We need to infer a function from the labeled training examples, and then use the function to label the unknown data. In order to get the function, we need to reduce the error between $\hat{r}_{ui}$ and r_{ui}. Consequently, the objective optimization function is

$$L = \sum l(\hat{r}_{ui}, r_{ui}) \tag{1}$$

where l is the loss function.

Overfitting always happen. If we have too many features, the learned hypothesis may fit the training set very well, but fail to generalize the new examples. There are often two options to solve overfitting. The first is to reduce the number of features. The details is dependent on the specific problem. The second is regularization, which is used to reduce magnitude or values of each feature with parameter θ. It often works well when there are a lot of features, and each of them contributes a bit to the prediction $\hat{r}_{ui}$.

文本之间的交互信息。在该模型中，神经张传统网络用于捕获字符/字电平相互作用，并且在Charac−ter / Word交互Tensor上应用空间RNN以捕获全局交互。文献（Beutel等，2018）桥接上下文协同过滤文献和神经推荐文学。它展示了利用深神经建议者（特别是在RNN模型中）中的上下文数据可以获得大量信息。协作内存网络（CMN）（EBESU等人2018）统一两类CF模型，以非线性方式利用潜在因子模型的全球结构和基于地方邻域的结构的优势。在模型中，存储器组件和神经关注机制被融合为邻域组件。

3 Preliminaries

3.1 Problem definition

我们首先呈现必要的定义，并制定CQA中专家的问题。我们的目标是在众包中找到CQ A的给定问题的专家。更具体地说，鉴于某些问题，人们需要找到最有可能的（1）的人有专业知识，并且（2）愿意接受回答问题的邀请。

定义1专家是用户在CQA中具有足够专业知识的用户。各专家的相关性文件，社交互动，面包或个人信息中暗示了专业知识。

给定一组m问题q = {q1，。。。，QM}，我们需要预测哪个专家E = {E1，。。。，EN}更有可能回答这些问题。为简单起见，我们保护专业索引信件，以区分专家问题，其中U，V Ingublate和I，J代表问题。

问题1对于给定的问题，我及其候选专家U∈e需要预测专家你回答问题的概率rui。

已知RUI的（U，I）对存储在SET = {（U，I）| RUI是已知的}。概率rui∈[0,1]，高值意味着更强的专家你回答问题i的偏好。RUI是预测的概率，我将根据标记的数据由专家U回答问题。这里，通过给定标记的数据进行预测是一个监督的学习问题。我们需要从标记的训练示例中推断出函数，然后使用该函数来标记未知数据。为了获得功能，我们需要减少rui和rui之间的错误。因此，客观优化功能是

$$L = \quad l(\hat{r}_{ui}, r_{ui}) \tag{1}$$

其中l是损失功能。过度装备总是发生。如果我们有太多的功能，所学到的假设可能会非常适合培训设置，但未能概括新示例。通常有两个选项来解决过度装备。第一个是减少功能的数量。细节取决于具体问题。第二是正则化，其用于减少每个特征的幅度或值，其中具有参数θ。当存在大量功能时，它通常会很好地运行，并且它们中的每一个都贡献了预测rui。

For example, if we use L2-norm for regularization, the optimization problem is transformed into the following problem:

$$\mathbf{\Theta}^* = \underset{\Theta}{\arg\min} \sum \left(l(\hat{r}_{ui}, r_{ui}) + \sum_{\theta \in \Theta} \lambda_\theta \theta^2 \right) \tag{2}$$

where λ_θ is the regularization coefficient of parameter θ used in the hypothesis function. As it grows, regularization becomes heavier. Then, we need to find an appropriate optimization method to solve this optimization problem. In this way, we get the parameters of the prediction model, which can be used to label the unknown data.

Typical data in CQA implies large interaction between experts and questions. For instance, some experts prefer to answer than others, and some questions are more likely to be answered than others. In order to account for these affects, it is customary to adjust the data with baseline.

Definition 2 The *baseline* for the prediction $\hat{r}_{ui}$ is denoted by b_{ui}:

$$b_{ui} = \mu + b_u + b_i, \tag{3}$$

in which, the overall average probability is denoted by μ; the parameters b_u and b_i indicate the observed average deviations of expert u and question i, respectively. For example, suppose that we want to get a baseline for the probability of the question i answered by the expert u. The average probability over all questions $\mu = 0.6$. The expert u tends to answer question lower than the average with probability 0.3, so $b_u = 0.3 - 0.6 = -0.3$. The question i tends to be answered with probability 0.7, so $b_i = 0.7 - 0.6 = 0.1$. Thus, the baseline for question i answered by expert i is $b_{ui} = 0.6 - 0.3 + 0.1 = 0.4$.

3.2 CQA datasets

The early CQA systems (such as Google Answers,[3] which has been retired), provide services via allocating by the provider or finding by the user, rather than leveraging crowdsourcing. In this part, we overview some frequently used datasets of crowdsourcing-based CQA system. These datasets come from the real-world. They are applicable to evaluate the methods for the expert finding in CQA.

Quora[4] is one of the largest existing CQA websites where users can ask and answer questions, rate and edit the answers posted by others.

Yahoo! Answers[5] is the most popular and well studied dataset in CQA related researches. It is a large and diverse question answer community, acting not only as a medium for knowledge sharing, but also as a place to seek advice, gather opinions, and satisfy ones curiosity about things which may not have a single best answer (Zhao et al. 2013).

Stack Overflow[6] is a community question answering site focusing on technical topics such as programming languages, algorithms and operating systems (Cheng et al. 2015).

WikiAnswers[7] is a wiki service that allows people to raise and answer questions, as well as edit existing answers to questions. It uses a so-called alternate system to automatically merge

3 http://answers.google.com.
4 https://www.quora.com/.
5 https://answers.yahoo.com/.
6 https://stackoverflow.com/.
7 http://www.answers.com/.

848 S. Yuan et al.

例如，如果我们使用L2−Norm进行正常化，则优化问题会转换为以下问题：

$$\mathbf{\Theta}^* = \underset{\Theta}{\arg\min} \sum \left(l(\hat{r}_{ui}, r_{ui}) + \sum_{\theta \in \Theta} \lambda_\theta \theta^2 \right) \tag{2}$$

其中λ θ是假设函数中使用的参数θ的正则化系数。随着它的增长，正规化变得越来越重。然后，我们需要找到一个适当的优化方法来解决这个优化问题。通过这种方式，我们得到了预测模型的参数，可用于标记未知数据。CQA中的典型数据意味着专家和问题之间的较大互动。例如，一些专家们更喜欢回答他人，一些问题更有可能得到回答的胜利.InordertoAccountforthesefects，ItiscustomarytoadjustthedataWithbaseline。

定义2预测RUI的基线由BUI表示：

$$b_{ui} = \mu + b_u + b_i, \tag{3}$$

其中，总平均概率由μ表示;参数BU和BI分别表示专家U和问题I的观察到的平均偏差。例如，假设我们希望获得一个基线，以便我由专家U回答的问题的概率。所有问题的平均概率μ = 0.6。专家U倾向于回答低于概率0.3的平均值的问题，因此Bu = 0.3 − 0.6 = −0.3。我倾向于以概率0.7回答问题，因此双= 0.7−0.6 = 0.1。因此，我由专家I回答的问题的基线是Bui = 0.6 − 0.3 + 0.1 = 0.4。

3.2 CQA datasets

早期的CQA系统（例如已退休的Google Answers，如谷歌答案），通过提供者分配或由用户查找提供服务，而不是利用众包。在这部分中，我们概述了一些基于众包的CQA系统的常用数据集。这些数据集来自真实世界。它们适用于评估CQA中专家的方法。

Quora4是最大的现有CQA网站之一，用户可以询问和回答问题，评分和编辑其他人发布的答案。雅虎Answers5是CQA相关研究中最受欢迎，最良好的数据集。它是一个大型多样化的问题答案社区，不仅是知识分享的媒介，而且作为一个寻求建议，收集意见和满足关于可能没有单一最佳答案的事情的地方（Zhao等人）。2013）。堆栈overflow6是一个社区问题应答网站，专注于技术主题，如编程语言，算法和操作系统（Cheng等，2015）。

Wikianswers7是一个维基服务，允许人们提高和回答问题，以及编辑现有问题的答案。它使用所谓的备用系统来自动合并

3 http://answers.google.com.
4 https://www.quora.com/.
5 https://answers.yahoo.com/.
6 https://stackoverflow.com/.
7 http://www.answers.com/.

Springer

similar questions. Since an answer may be associated with multiple questions, duplicated entries can be avoided to some extent.

Zhihu[8] is a popular Chinese specialized CQA portal similar to Quora. It is able to provide questions with detailed and reliable answers that are voted by a large number of users. Users are also allowed to edit questions and answers, rate system, and tag questions.

Toutiao Q&A[9]: employs artificial intelligence technologies to deliver information with high efficiency and high quality. It aims to promote short-form content creation and user interaction on mobile devices in the format of question and answering.

Baidu Knows[10] is a popular Chinese general CQA, in which a user can put questions with bounty to promote others answering it.

Sogou Wenwen[11] is an interactive Chinese CQA with credit points and reputation points. Users can obtain points by asking or answering questions and use them as bounty.

We summarize the information of these CQA datasets in Table 1. *Quora*, *Yahoo! Answers*, *Stack Overflow* and *WikiAnswers* are in English. *Zhihu*, *Toutiao Q&A*, *Baidu Knows* and *Sogou Wenwen* are in Chinese. The "No. of QAP" is the number of question answering pairs in the given reference. It reflects the scale of the CQA datasets. In the "Available" column, we list the download address of the dataset. Toutiao is one of the most widely used information distribution platforms in China, so we will use the preprocessed datasets of *Toutiao Q&A* in the following parts of this paper.

3.3 Categorization of expert finding techniques

Based on the survey of the recent solutions, we categorize the techniques of expert finding in CQA under four subsettings, including MF-based models, GBT-based models, DL-based models and R-based models. As shown in the Table 2, we summarize the performance of these models on different types of matching tasks to explore the scope of application.[12] In the table, *text VS text* means to match text labels with text data, *graph VS text* means to match text labels with graph data, *audio VS text* is to match text labels with audio data; *video VS text* is to match text labels with video data.

We come to the conclusion that MF-based models achieve the best performance in the situation of *text VS text* with encoded text, while DL-based models are rarely used in these situations and not performing well due to the severe sparsity of the text datasets with less context information. In addition, R-based models have significant performance in the situation of *audio VS text*, DL-based models often achieve the best performance in the situation of both *graph VS text* and *video VS text*, which may due to their outstanding power of capturing high dimensional features from graphs and videos. We will discuss these four category solutions in detail in the following sections.

8 https://www.zhihu.com/.

9 https://www.wukong.com/.

10 https://zhidao.baidu.com/.

11 https://wenwen.sogou.com/.

12 More details of experiment results will be clarified in Sect. 10.

类似的问题。由于答案可能与多个问题相关联，因此可以在某种程度上避免重复的条目。

Zhihu8是一个类似于Quora的热门中国专业CQA门户网站。它能够提供由大量用户投票的详细且可靠的答案的问题。还允许用户编辑问题和答案，评分系统和标签问题。

Toutiao Q&A9：采用人工智能技术，以提供高效率和高品质的信息。它旨在以问答的格式促进移动设备上的短窗体内容创建和用户互动。百度知识10是一个受欢迎的中国一般CQA，其中用户可以用赏金提出问题来推广他人回答它。 Sogou Wenwen11是一个互动的中国CQA，具有信用点和声誉点。用户可以通过询问或回答问题来获得积分并将其用作赏金。我们总结了表1.

Quora，Yahoo!的这些CQA数据集的信息。答案，堆栈溢出和Wikianswers是英文的。

Zhihu，Toutiao Q&A，百度知道和Sogou Wenwen是中文的。 “不。

QAP“是给定参考中的问题答案对的数量。它反映了CQA数据集的比例。在“可用”列中，我们列出了数据集的下载地址。

Toutiao是中国使用的最广泛的信息分发平台之一，因此我们将在本文的以下部分中使用Toutiao Q&A的预处理数据集。

3.3专家发现技术的分类

基于对最近的解决方案的调查，我们将专家在CQA下的专家查找技术分析，包括基于MF的模型，基于MF的模型，基于DL的模型和基于R的模型。如表2所示，我们总结了这些模型在不同类型的匹配任务上的性能，以探索表中的应用范围.12在表中，文本与文本意味着将文本标签与文本数据相匹配，图形与文本匹配与图形数据的文本标签，音频VS文本是与音频数据匹配文本标签;视频VS文本是将文本标签与视频数据匹配。我们得出结论，基于MF的模型在文本与编码文本的情况下实现了最佳性能，而基于DL的模型很少在这些情况下使用并且由于文本数据集的严重伤口而不是良好的表现不佳较少的Context Information.inaddition，r基于r−priedelshaveSignificantPerianceIthings音频与文本，DL的模型通常在两个图表VS文本和视频与文本的情况下实现了最佳性能，这可能是由于它们从图形捕获高维特征的出色功能视频。我们将在以下部分详细讨论这四个类别解决方案。

8 https://www.zhihu.com/。9 https://www.wukong.com/。10 https://zhidao.baidu.com/。11 https://wenwen.sogou.com/。12篇在教派中澄清了实验结果的更多细节。10。

Table 1 CQA datasets

Datasets	Language	References	No. of QAPs	Available
Quora	English	Zhao et al. (2015b)	444,138	https://www.kaggle.com/quora/question-pairs-dataset
Yahoo! Answers		Qiu and Huang (2015)	312,000	https://webscope.sandbox.yahoo.com/
Stack Overflow		Riahi et al. (2012)	118,510	https://www.kaggle.com/stackoverflow/datasets
WikiAnswers		Bordes et al. (2014)	350,000	https://github.com/afader/oqa/tree/master/oqa-data
Zhihu	Chinese	Liu et al. (2015)	209,309	https://www.biendata.com/competition/CCIR2018/data/
Toutiao Q&A		Qian et al. (2018)	290,000	https://www.biendata.com/competition/bytecup2016/data/
Baidu Knows		Qiu and Huang (2015)	423,000	–
Sogou Wenwen		Li et al. (2015b)	291,304	–

850 S. Yuan et al.

Table 1 CQA datasets

数据集语言参考QAP的可用QAP否定数				
Quora	English	Zhao et al. (2015b)	444,138	https://www.kaggle.com/quora/question-pairs-dataset
Yahoo! Answers		Qiu and Huang (2015)	312,000	https://webscope.sandbox.yahoo.com/
堆栈溢出Riahi等人。(2012) 118,510 https://www.kaggle.com/stackoverflow/datasets				
WikiAnswers		Bordes et al. (2014)	350,000	https://github.com/afader/oqa/tree/master/oqa-data
Zhihu	Chinese	Liu et al. (2015)	209,309	https://www.biendata.com/competition/CCIR2018/data/
Toutiao Q&A		Qian et al. (2018)	290,000	https://www.biendata.com/competition/bytecup2016/data/
Baidu Knows		Qiu and Huang (2015)	423,000	–
Sogou Wenwen		Li et al. (2015b)	291,304	–

Springer

4 Matrix factorization based models

Matrix factorization (MF) (Koren et al. 2009), which is a common technique for collaborative filtering (CF) (Linden et al. 2003), covers a wide range of applications in recommender system with its variants. The *Problem* 1 can be modeled as recommendation problem solved by CF, because similar users may answer the similar questions. Therefore, MF can be applied to exploit latent information from data. In this part, we summarize the MF-based models, including MF, Singular Value Decomposition (SVD), SVD++, Bidirection SVD++, Bidirection Asymmetric-SVD (ASVD++) and Factorization Machine (FM).

4.1 MF

From the application point of view, MF can be used effectively to discover the latent features underlying the interactions between different kinds of entities. For example, several experts have answered same questions before as illustrated in Fig. 1. If some of them (the number is N) answer a new question, others may also answer the question (the probability is p). N is larger, p is larger.

From the mathematical point of view, MF is used to factorize a matrix obviously as its name suggesting. The original matrix can be represented by the multiply of two (or more) simple matrices with lower dimension. Let U and D be the set of experts and questions, respectively. Let $\mathbf{R}$ be the record matrix of the expert-question pairs. If we would like to discover k latent features, we need to find two matrices $\mathbf{P}$ (a $|U| \times k$ matrix) and $\mathbf{Q}$ (a $|D| \times k$ matrix) such that their product approximates $\mathbf{R}$:

$$\hat{\mathbf{R}} = \mathbf{P}^T \times \mathbf{Q} \approx \mathbf{R}. \tag{4}$$

Thus, matrix factorization maps experts and questions to a joint latent factor space of dimensionality k. Each row of $\mathbf{P}$ would represent the strength of the associations between an expert

Table 2 Performance of different categories of models on different types of matching tasks

Model category	text VS text	graph VS text	audio VS text	video VS text
MF-based models	√			
DL-based models		√		√
GBT-based models		√	√	
R-based models	√			

√Means that this category of models performs well

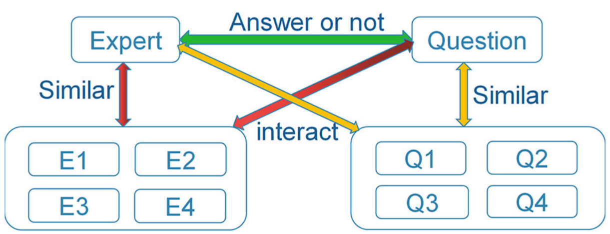


Fig. 1 Implied information

4基于矩阵分解的模型

矩阵分解（MF）（Koren等，2009），这是一个共同的合作滤波（CF）的常用技术（Linden等，2003），涵盖了具有其变体的推荐系统的广泛应用。问题1可以被建模为CF解决的推荐问题，因为类似的用户可以回答类似的问题。因此，可以应用MF来从数据中利用潜在信息。在这一部分中，我们总结了基于MF的模型，包括MF，奇异值分解（SVD），SVD ++，BIDirection SVD ++，BIDipecection非对称-SVD（ASVD ++）和分解机（FM）。

4.1 MF

从应用点来看，MF可以有效地用于发现不同类型实体之间的相互作用的潜在特征。例如，在图4中所示，几个专家已经回答了相同的问题。如果其中一些（数字是n）回答一个新问题，其他人也可能回答问题（概率是p）。n更大，p更大。从数学的角度来看，MF用于将矩阵分解，显然是其名称表明。原始矩阵可以由具有较低维度的两个（或更多）简单矩阵的乘法表示。让你和D分别成为专家和问题。让R是专家问题对的记录矩阵。如果我们想发现k潜在特征，我们需要找到两个矩阵P（a｜×k矩阵）和q（a｜d｜×k矩阵），使得它们的产品近似于R：

$$\hat{\mathbf{R}} = \mathbf{P}^T \times \mathbf{Q} \approx \mathbf{R}. \tag{4}$$

因此，矩阵分解地将专家和问题映射到维度k的联合潜在因子空间。每排P将代表专家之间协会的强度

表2不同类别的不同类型匹配任务的性能

模型类文本与文本图与文本音频与文本视频与文本				
MF-based models	√			
DL-based models		√		√
GBT-based models		√	√	
R-based models	√			

√这类模型的emeans表现良好

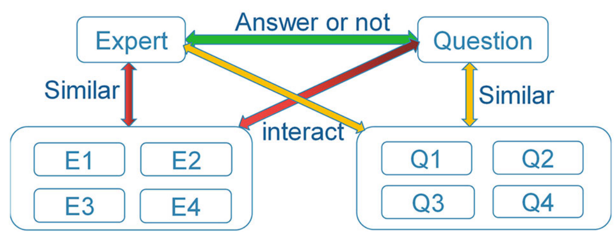


Fig. 1 Implied information

and the features. Similarly, each row of **Q** would represent the strength of the associations between a question and the features.

Matrix factorization maps experts and questions to a joint latent factor space of dimensionality k, such that expert-question interactions are modeled as inner products in that space. The resulting dot product $p_u^T q_i$ captures the interaction between expert u and question i.

$$\hat{r}_{ui} = p_u^T q_i. \tag{5}$$

Then we directly model the observed probabilities only, while avoiding over-fitting through a regularized model. To learn the factor vectors p_u and q_i, the system minimizes the regularized squared error on the set of known probabilities:

$$\min_{P,Q} \sum_{(u,i)\in Ł} (r_{ui} - q_i^T p_u)^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \tag{6}$$

where aforementioned Ł is the set of the (u, i) pairs for which r_{ui} is known.

4.2 SVD

One benefit of the matrix factorization approach to collaborative filtering is its flexibility in dealing with various data and other application-specific requirements. Eq. (5) tries to capture the interactions between users and questions without taking the baseline into consideration. Here we combine Eqs. (3) and (5) as follows:

$$\hat{r}_{ui} = b_{ui} + p_u^T q_i \tag{7}$$

The system learns by minimizing the squared error function, and avoids over-fitting through an adequate regularized model:

$$\min_{P,Q,B} \sum_{(u,i)\in Ł} (r_{ui} - \hat{r}_{ui})^2 + \lambda(\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2) \tag{8}$$

4.3 SVD++

MF and SVD models only consider explicit feedback which comes from the interaction between a user and a question. However, we can also obtain implicit feedback from the training data. For instance, a user prefers those questions that he answers in the past. Recommender systems can use implicit feedback to gain insight into user preferences. Indeed, we can gather behavioral information regardless of the user's willingness to provide explicit ratings. Here, we try to integrate both explicit feedback and implicit feedback. We could get more accurate results by a direct modification of Eq. (7):

$$\hat{r}_{ui} = b_{ui} + q_i^T \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \tag{9}$$

where $N(u)$ is the set of questions that user u has received invitation. A user u is modeled as $p_u + |N(u)|^{\frac{1}{2}} \sum_{j\in N(u)} y_j$. p_u is learnt from the given explicit ratings and $|N(u)|^{\frac{1}{2}} \sum_{j\in N(u)} y_j$ represents the perspective of implicit feedback. Here, a new set of item factors are necessary, where question j is associated with $y_j \in \mathbb{R}^f$. Model parameters are learnt by minimizing the squared error function.

852 S. Yuan et al.

和特征。同样，每行Q都将代表问题和特征之间的关联的强度。矩阵分解地图将专家和问题映射到Dimen−Sionality
K的联合潜在因子空间，使得专家问题交互被建模为该空间的内部产品。得到的点产品P
T U Qi捕获了专家U和问题的互动。

$$\hat{r}_{ui} = p_u^T q_i. \tag{5}$$

ThenWedirectlyModelTheobservedProbabilessonly，Whileavoidingover−Fittingthrough正则化模型。要了解因子向量PU和QI，系统最小化了已知概率集中的调节方案错误：

$$\min_{P,Q} \sum_{(u,i)\in Ł} (r_{ui} - q_i^T p_u)^2 + \lambda(\|p_u\|^2 + \|q_i\|^2) \tag{6}$$

在上述 是已知RUI的（U，I）对的集合。

4.2 SVD

矩阵分组方法的一个好处是协作滤波的一个好处是它在处理各种数据和其他特定于应用程序的要求中的灵活性。eq。（5）试图捕获用户与问题之间的相互作用而不考虑基线。我们在这里结合了eqs。（3）和（5）如下：

$$\hat{r}_{ui} = b_{ui} + p_u^T q_i \tag{7}$$

系统通过最大限度地减少平方误差功能，避免通过足够的正则化模型来拟合：

$$\min_{P,Q,B} \sum_{(u,i)\in Ł} (r_{ui} - \hat{r}_{ui})^2 + \lambda(\|p_u\|^2 + \|q_i\|^2 + b_u^2 + b_i^2) \tag{8}$$

4.3 SVD++

MF和SVD模型仅考虑来自用户与问题之间的交互的显式反馈。但是，我们还可以从培训数据获得隐含的反馈。例如，用户更喜欢他过去答案的问题。Rec−Olemender系统可以使用隐式反馈来获得用户偏好的洞察力。实际上，我们可以收集行为信息，无论用户愿意提供明确评级。在这里，我们尝试集成显式反馈和隐式反馈。我们可以通过直接修改EQ来获得更准确的结果。（7）：

$$\hat{r}_{ui} = b_{ui} + q_i^T \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \tag{9}$$

其中n（u）是用户您收到了邀请的一组问题。用户U被建模为PU + | N（U）|1
2 j∈n（u）y j。PU是从给定的明确评级中学到的

| n（u）|1 2 j∈n（u）y
j表示隐式反馈的视角。这里，需要一组新的项目因素，其中问题j与y
j∈rf相关联。通过最小化平方误差函数来学习模型参数。

$$\min_{P,Q,B,Y} \sum_{(u,i)\in Ł} (r_{ui} - \hat{r}_{ui})^2 + \lambda \|\theta\|^2 \tag{10}$$

where θ represents the parameters of the model. SVD++ (Koren 2008) does not offer the benefits of having less parameters and readily explainable results. This is because the model does abstract each user with a factors vector. However, SVD++ is clearly advantageous in terms of prediction accuracy than SVD.

4.4 Bidirection SVD++ (SVD#)

Appending another part of implicit feedback to the original SVD++ model, a new model named bidirection SVD++ model (also called SVD#) is built. The formula of this model turns to be:

$$\begin{aligned} \hat{r}_{ui} = b_{ui} &+ \left(q_i + |R(i)|^{-\frac{1}{2}} \sum_{j\in R(i)} x_j \right)^T \\ &\times \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \end{aligned} \tag{11}$$

$R(i)$ is the set of users who answer question i. Here, each question j is associated with $x_j, y_j \in \mathbb{R}^f$. The other parts of the formula are the same as the original SVD++ model.

This model shows the power of representing user/question embeddings using the neighborhood question/user embeddings. However, the embeddings here are static and indepent of time. When the time information is available, a more powerful proposed in Dai et al. (2016) will be helpful. This method incorporates the embedding co-evolving idea with time series models. The evolution of each user/question embedding depends not only on its old embeddings, but also the embeddings of question/user it interacting with.

4.5 Bidirection ASVD++

As mentioned in (Koren 2008, instead of providing an explicit parameterization for users, users can be represented through their preferred items. This model named "Asymmetric-SVD" (ASVD) offers several benefits: (1) fewer parameters; (2) handling new users; (3) explainability; (4) efficient integration of implicit feedback. Combining the "bidirection" strategy described in Sect. 4.4, there is a new model named bidirection ASVD++ model. The formula is listed as below:

$$\begin{aligned} \hat{r}_{ui} = b_{ui} &+ \left(|R(i)|^{-\frac{1}{2}} \sum_{j\in R(i)} x_j \right)^T \\ &\times \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \end{aligned} \tag{12}$$

4.6 Factorization machine

FM (Rendle 2011) is a generic approach based on matrix factorization to mimic most factorization models. libFM (Rendle 2012) proposed by Steffen Rendle is a software imple-

$$\min_{P,Q,B,Y} \sum_{(u,i)\in Ł} (r_{ui} - \hat{r}_{ui})^2 + \lambda \|\theta\|^2 \tag{10}$$

其中θ表示模型的参数。SVD ++（Koren 2008）没有提供较少参数和易于解释的结果的好处。这是因为模型对具有因素矢量的每个用户抽象。然而，在比SVD的预测精度方面，SVD ++显然是有利的。

4.4 Bidirection SVD++ (SVD#)

将隐式反馈的另一部分附加到原始SVD ++模型，构建了一个名为Bidirection SVD ++模型的新型号（也称为SVD＃）。该模型的公式变为：

$$\begin{aligned} \hat{r}_{ui} = b_{ui} &+ \left(q_i + |R(i)|^{-\frac{1}{2}} \sum_{j\in R(i)} x_j \right) \\ &\times \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \end{aligned} \tag{11}$$

R（i）是回答问题的用户集。这里，每个问题j与x j，yj∈Rf相关联。公式的其他部分与原始SVD ++模型相同。该模型显示使用邻语问题/用户嵌入式表示用户/问题嵌入的力量。但是，这里的嵌入品是静态和不可或缺的时间。当时间信息可用时，在Dai等人中提出了更强大的功能。（2016）会有所帮助。该方法包含与时间序列模型的嵌入共同发展的想法。每个用户/问题嵌入的演变不仅取决于它的旧嵌入，还取决于它与其交互的问题/用户的嵌入品。

4.5 Bidirection ASVD++

如（koren 2008，而不是为用户提供明确的参数化，用户可以通过他们的首选项目来表示。该型号名为“不对称-SVD”（ASVD）（ASVD）提供了多种优点：（1）参数更少;（2）处理新用户;（3）解释性;（4）有效地集成隐性反馈。组合“双向”策略。4.4，有一个名为B idirection ASVD ++模型的新型号。如下所示的公式如下所示：

$$\begin{aligned} \hat{r}_{ui} = b_{ui} &+ \left(|R(i)|^{-\frac{1}{2}} \sum_{j\in R(i)} x_j \right)^T \\ &\times \left(p_u + |N(u)|^{-\frac{1}{2}} \sum_{j\in N(u)} y_j \right) \end{aligned} \tag{12}$$

4.6 Factorization machine

FM（Rendle 2011）是一种基于矩阵分解的通用方法，以模拟大多数分解模型。Libfm（Rendle 2012）由Steffen Rendle提出的是一种软件

mentation for factorization machines. It combines the generality of feature engineering with the superiority of factorization models in estimating interactions between variables of large domain. FM model has the following advantages. Firstly, variable interactions are embedded in the FM model. Secondly, it is able to reliably estimate parameters under very high sparsity. Thirdly, the equation, which depends only on a linear number of parameters, can be computed in linear time. Forthly, it can be applied to a variety of prediction tasks, including regression, binary classification and ranking. In essence, FM model is a matrix factorization based machine learning model and it is similar to linear regression model. We all know the linear regression model has the following formula:

$$\hat{y}(x) = w_0 + w_1x_1 + \cdots + w_nx_n = w_0 + \sum_{i=1}^{n} w_ix_i. \tag{13}$$

where x_i is the feature and $\hat{y}$ is the predicted value.

On the basis of model above, if we consider the feature combination, the formula will be changed to the following form:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_ix_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w'_{ij}x_ix_j. \tag{14}$$

Because the sparsity of the feature, we find that many w'_{ij} will be zero after the training. Thus, in order to reduce the number of parameters, FM models the problem by the following formula:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_ix_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (V_i^TV_j)x_ix_j, \tag{15}$$

where V_i is the latant vector of the ith feature. We consider a maximum likelihood problem with Eq. (15). To avoid over-fitting, we add some regularization terms. That is, we solve the following optimization problem for FM model.

$$\min_{W,V} \sum_{i=1}^{n}(y_ilog(\sigma(\hat{y}_i)) + (1-y_i)log(1-\sigma(\hat{y}_i))) + \frac{\lambda}{2}\|\theta\|^2 \tag{16}$$

where θ represents the parameters of the model and $\sigma(x)$ is the sigmoid function. The learning algorithm of FM mainly contains (Rendle 2012): Stochastic Gradient Descent (SGD), Alternating Least Squares (ALS) and Markov Chain Monte Carlo (MCMC).

5 Gradient boosting tree based models

Tree ensemble methods are very widely used in practice. Gradient tree boosting is one of them that shines in many applications. The classic gradient boosting tree and its extension are described in Friedman (2001). XGBoost (Chen and Guestrin 2016) is a scalable open source system for tree boosting. The impact of the XGBoost has been widely recognized in a number of machine learning and data mining challenges. People often chooses XGBoost as the implementation of the Gradient Boosting Regression Trees (GBRT) in the application.

A tree ensemble model uses K additive functions to predict the output.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \mathcal{F}, \tag{17}$$

854 S. Yuan et al.

为分解机进行调炼。它将特征工程的一般性与分解模型的优势相结合，估计大域变量之间的相互作用。FM模型具有以下优点。首先，可变交互在FM模型中嵌入。其次，它能够在非常高的稀疏度下可靠地估计参数。第三，只能在线性时间计算仅取决于线性数量的等式。即，它可以应用于各种预测任务，包括回归，二进制分类和排名。实质上，FM模型是基于矩阵分解的机器学习模型，它类似于线性回归模型。我们都知道线性回归模型具有以下公式：

$$\hat{y}(x) = w_0 + w_1x_1 + \cdots + w_nx_n = w_0 + \sum_{i=1}^{n} w_ix_i. \tag{13}$$

其中xi是特征，y是预测值。在上面的模型的基础上，如果我们考虑特征组合，公式将改变为以下形式：

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_ix_i + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w'_{ij}x_ix_j. \tag{14}$$

因为特征的稀疏性，我们发现许多W'I
j在训练后将是零零。因此，为了减少参数的数量，FM通过以下公式模拟问题：

$$\hat{y}(x) = w_0 + \sum_{i=1} w_ix_i + \sum_{i=1}\sum_{j=i+1} (V_i^TV_j)x_ix_j, \tag{15}$$

其中VI是第i个功能的潜伏载体。我们考虑eq的最大可能性问题。（15）。为避免过度拟合，我们添加了一些正则化术语。也就是说，我们解决了FM模型的以下优化问题。

$$\min_{W,V} \sum_{i=1}^{n}(y_ilog(\sigma(\hat{y}_i)) + (1-y_i)log(1-\sigma(\hat{y}_i))) + \frac{\lambda}{2}\|\theta\|^2 \tag{16}$$

其中θ表示模型的参数和σ（x）是sigmoid函数。FM的学习算法主要包含（Rendle 2012）：随机梯度下降（SGD），交替的最小二乘（ALS）和马尔可夫链蒙特卡罗（MCMC）。

5基于梯度升压树的模型

树集合方法非常广泛地使用。渐变树增强是其中之一，其中许多应用程序都在闪耀。弗里德曼（2001年）描述了经典的梯度升压树及其扩展。XGBoost（Chen和Guestrin 2016）是一个可扩展的树升压开源系统。XGBoost对多种机器学习和数据采矿挑战的影响已被广泛认可。人们经常选择XGBoost作为应用程序中渐变升压回归树（GBRT）的实现。树集合模型使用K添加剂功能来预测输出。

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in \mathcal{F}, \tag{17}$$

where $\mathcal{F}$ is the space of regression trees (also known as CART). The regularized objective function is listed as follows:

$$\mathcal{L} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \tag{18}$$

where l is a loss function that measures the difference between the prediction $\hat{y}_i$ and the target y_i. The second term Ω penalizes the complexity of the model:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2. \tag{19}$$

T is the number of leaves in the tree. Each regression tree contains a continuous score on each leaf, ω_i is the score on i-th leaf.

Since the tree ensemble model in Eq.(18) includes functions as parameters but not just numerical vectors, it cannot be optimized using traditional optimization methods such as stochastic gradient descent (SGD) in Euclidean space. In XGBoost, Eq.(18) is trained in an additive manner.

$$\hat{y}_i^{(t)} = \sum_k f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), \tag{20}$$

where $\hat{y}_i^{(t)}$ is the prediction of the i-th instance at the t-th iteration. Then, the objective function is:

$$\mathcal{L} = \sum_i l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_k \Omega(f_k). \tag{21}$$

Considering the square loss and taking Taylor expansion approximation of the loss, we get:

$$\mathcal{L}^{(t)} \simeq \sum_i \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_k) + constant, \tag{22}$$

where

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}), \tag{23}$$

and

$$h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}). \tag{24}$$

Combining Eqs. (18) and (22), we remove constants and get:

$$\mathcal{L}^{(t)} \simeq \sum_i \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2}\lambda \sum_j \omega_j^2, \tag{25}$$

This is One Variable Quadratic Equation of ω_j. We can compute the optimal weight ω_j^* of leaf j by

$$\omega_j^* = -\frac{\sum_i g_i}{\sum_i h_i + \lambda}, \tag{26}$$

and calculate the corresponding optimal objective function value by

$$\tilde{\mathcal{L}}^{(t)} = -\frac{1}{2} \sum_j \frac{(\sum_i g_i)^2}{\sum_i h_i + \lambda} + \lambda T, \tag{27}$$

F是回归树的空间（也称为购物车）。正常的目标函数列出如下：

$$\mathcal{L} = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \tag{18}$$

其中L是损失函数，用于测量预测yi和目标yi之间的差异。第二项 惩罚模型的复杂性：

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2. \tag{19}$$

T是树中叶子的数量。每个回归树包含每个叶子上的连续分数，ωi是第i叶片上的分数。由于EQ中的树集合模型。（18）包括作为参数但不仅仅是数值矢量的功能，因此不能使用欧几里德空间中随机梯度下降（SGD）等传统优化方法进行优化。在XGBoost，EQ。（18）以添加方式培训。

$$\hat{y}_i^{(t)} = \sum_k f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), \tag{20}$$

其中y（t）i是在t−th迭代中预测第i个实例。然后，目标函数是：l =

考虑到平方损失并采取泰勒扩大近似值的损失，我们得到：

$$\mathcal{L}^{(t)} \simeq \sum_i \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_k) + constant, \tag{22}$$

where

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}), \tag{23}$$

and

$$h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}). \tag{24}$$

组合eqs。（18）和（22），我们删除常量并获得：

$$\mathcal{L}^{(t)} \simeq \sum_i \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2}\lambda \sum_j \omega_j^2, \tag{25}$$

这是ωj的一个可变二次方程。我们可以计算叶J的最佳权重ω⋆j

$$\omega_j^* = -\frac{{}_i\, g_i}{{}_i\, h_i + \lambda}, \tag{26}$$

并计算相应的最佳目标函数值

$$\tilde{\mathcal{L}}^{(t)} = -\frac{1}{2} \sum_j \frac{(\sum_i g_i)^2}{\sum_i h_i + \lambda} + \lambda T, \tag{27}$$

In practice, the greedy algorithm, that starts from a single leaf and iteratively adds branches to the tree, is usually used for evaluating the split candidates. It is impossible to efficiently do the exact greedy algorithm when the data does not fit entirely into memory. And then, the approximate algorithm for split finding is proposed in XGBoost. More details can be found in Chen and Guestrin (2016).

6 Deep learning based models

As deep learning has grown in prominence for computer vision and natural language processing (NLP) tasks, there are many recent researches incorporating deep neural network (DNN) in recommender systems. Applying deep learning models into recommender system has been gaining momentum due to its state-of-the-art performances on popular benchmarks for recommender systems, such as MovieLens[13] and Netflix challenge datasets. Previous works have largely relied on applying the collaborative filtering intuition to neural networks, such as joint deep learning and CF models (Zheng et al. 2016; He et al. 2017), or auto-encoder (Sedhain et al. 2015). Further, to ensure generalization, denoising auto-encoders (DAEs) have been exploited to learn from corrupted inputs (Kawale 2015; Wu et al. 2016). Moreover, there has been a recent popularity in using recurrent neural networks (RNNs) for recommendation (Wan et al. 2016; Tan et al. 2016; Wu et al. 2017).

6.1 Autoencoder model

AutoRec (Sedhain et al. 2015) is an autoencoder based collaborative filtering model. Similar to traditional CF, AutoRec has two variants: an user-based autoencoder and an item-based autoencoder. They can respectively take user partial vectors and item partial vectors as input, project them into a hidden layer to learn the lower-dimensional representations, and further reconstruct them in the output layer to predict missing ratings for the purpose of recommendation.

While AutoRec is used in the *Problem* 1, experts are regarded as users, questions as items, and the question distribution data as rating matrix. The question distribution data indicates whether the expert answered the question (if answered, the tag is 1; otherwise 0). Then the AutoRec model is deployed to predict the ratings of the unknown expert-question pairs.

Both user-based AutoRec and an item-based AutoRec are exploited in expert finding in CQA. Experiment results show that item-based model performs better which may be due to the higher variance of user partial vectors. However, item-based AutoRec is not performing well than MF-based models as before in this task. The reason may be that the dataset of Toutiao is more sparse than the MovieLens dataset.

6.2 Collaborative denoising auto-encoder model

Denoising Auto-encoder (Vincent et al. 2008) is an extended auto-encoder model which aims to derive more robust features from the hidden layer. It reconstructs each data point from the corrupted version. The corrupted version of original inputs are usually derived from a conditional distribution. Common corruption choices contain the additive Gaussian noise and the multiplicative dropout noise. Utilizing the idea of DAEs, Collaborative Denoising

[13] https://grouplens.org/datasets/movielens/1m/.

在实践中，从单个叶子开始并迭代地将分支添加到树的贪婪算法通常用于评估分割候选者。当数据不完全符合内存时，不可能有效地进行精确的贪婪算法。然后，在XGBoost中提出了用于分流发现的近似算法。在陈和宾馆（2016年）可以找到更多细节。

基于深度学习的模型

由于深度学习在计算机视觉和自然语言的突出（NLP）任务中发展，最近有许多研究在推荐系统中包含深神经网络（DNN）。将深度学习模型应用于推荐系统，这一直是由于其在推荐系统的流行基准上的最先进的表演而获得势头，例如Movielens13和Netflix挑战数据集。以前的作品在很大程度上依赖于将协同过滤直觉应用于神经网络，例如联合深度学习和CF模型（Zheng等，2016;
He等人2017）或自动编码器（SEDHAIN等，2015）。此外，为了确保泛化，已经利用了去噪的自动编码器（Daes）从损坏的输入中学习（Kawale 2015;
Wu等人2016）。此外，近来，在使用经常性神经网络（RNNS）以供推荐有近来的普及（WAN等，2016; Tan等人2016; Wu等人2017）。

6.1 Autoencoder model

Autorec（Sedhain等，2015）是一种基于AutoEncoder的协作滤波模型。类似于传统的CF，Autorec有两个变体：基于用户的AutoEncoder和基于项目的AutoEncoder。它们可以分别使用用户部分向量和项目部分向量作为输入，将它们投入到隐藏层中以学习下方的表示，并进一步将它们重建在输出层中以预测推荐目的的缺失额定值。在问题1中使用Autorec时，专家被视为用户，问题作为项目，以及作为评级矩阵的问题分发数据。问题分发数据指示专家是否回答问题（如果回答，标签为1;否则为0）。然后部署Autorec模型以预测未知专家问题对的额定值。基于用户的Autorec和基于项目的Autorec在CQA中的专家中被利用。实验结果表明，基于项目的模型执行更好，这可能是由于用户部分向量的更高方差。但是，基于项目的Autorec不得在此任务中的基于MF的模型。原因可能是Toutiao的数据集比Movielens数据集更稀疏。

6.2 Collaborative denoising auto-encoder model

去噪自动编码器（VINCENT等，2008）是一个扩展的自动编码器模型，旨在从隐藏的层中获得更强大的功能。它从损坏的版本重建每个数据点。损坏的原始输入版本通常来自条件分布。常见的腐败选择包含添加高斯噪声和乘法丢失噪声。利用Daes的想法，协作去噪

[13] https://grouplens.org/datasets/movielens/1m/.

Auto-Encoder (CDAE) has been developed for recommendation tasks (Wu et al. 2016). The assumption of CDAE is that all the user-item interactions are a corrupted version of the user's full preference set.

Specifically, CDAE first learns latent representations from the corrupted version inputs. Then the latent representations are mapped back to the original input space to rebuild the input vectors. Parameters of CDAE are learned by minimizing the average reconstruction error. Finally, for recommendation, user's existing reference set (without corruption) is taken as input to predict recommendations for each user. While using CDAE for *Problem* 1, it regrades experts as items, questions as users, and the question distribution data as the users' preference set. A preference set is binary which only includes the information about whether the expert answered the question or not (if answered, the tag is 1; otherwise 0).

6.3 Neural autoregressive model

Inspired by the Restricted Boltzmann Machine (RBM) based CF model, an emerging Neural Autoregressive Distribution Estimator (NADE) based CF model named CF-NADE (Zheng et al. 2016) is proposed. It can model the distribution of expert ratings. CF-NADE with only one hidden layer can defeat all the previous state-of-the-art models in recommendation tasks upon the MovieLens 1M, MovieLens 10M and Netflix datasets. Furthermore, CF-NADE can be further extended to a deep model with more hidden layers which can further boost the performance.

CF-NADE, which is designed to model the ordering of the ratings, is a feed-forward and neural autoregressive architecture for CF tasks. Ideally, the order of items should follow the time-stamps of ratings. However, empirical study shows that random drawing permutations for each user also generates favourable performances. Since the expert IDs as well as the question IDs are anonymized and the descriptions of expert and questions in the dataset have been encoded into ID sequences, it is feasible to deploy CF-NADE to this competition without time-stamps information. While training the CF-NADE model, the experts and questions are considered as users and items, and the rating matrix is derived from question push notification records like in Sect. 6.1. Experiment results show that the performance of CF-NADE model in the *Problem* 1 is similar to the AutoRec model, in which item-based CF-NADE performs better than user-based CF-NADE but still not comparable to the matrix factorization based models such as SVD++ and ASVD++. Moreover, the CF-NADE model, though worth trying, is not integrated into any final ensemble models because it significantly reduces the performance when incorporated into ensemble models.

6.4 Neural network-based collaborative filtering

Recent studies on deep learning for recommendations usually employ deep learning methods to model auxiliary information such as textual descriptions of items and users. While for modeling the key factor of CF (the interaction between item and user features), they still rely on MF models and use an inner product on the latent features. The linear combination of the latent features' multiplication becomes a bottleneck in improving the performance. Replacing the inner product with a neural architecture, a general framework named neural network-based collaborative filtering (NCF) (He et al. 2017) has the ability to learn a non-linear userCitem interaction function from the implicit data.

NCF consists of a input layer, an embedding layer, several Neural CF layers and an output layer. The input layer consists of two feature vectors describing the user and the item

自动编码器（CDAE）已开发用于推荐任务（Wu等，2016）。CDAE的假设是所有用户项交互都是用户完整首选项集的损坏版本。具体来说，CDAE首先从损坏的版本输入中了解潜在的表示。然后将潜在表示映射回原始输入空间以重建输入向量。通过最小化平均重建误差来学习CDAE参数。最后，对于推荐，用户现有的参考集（不损坏）被视为输入以预测每个用户的建议。在使用CDAE进行问题1时，它将专家作为项目，问题作为用户，以及作为用户的首选项集的问题分发数据。偏好设置是二进制文件，它仅包括关于专家是否回答问题的信息（如果回答，标签为1;否则为0）。

6.3 Neural autoregressive model

由基于限制的Boltzmann机（RBM）的CF模型灵感，提出了一种名为CF-Nade（Zheng等人）的新兴神经自回转分布估算器（NADE）的CF型号（Zheng等人2016）。它可以模拟专家评级的分布。只有一个隐藏层的CF-Nade可以在Movielens 1M，Movielens 10M和Netflix数据集上打败所有以前的最先进的模型。此外，CF-NADE可以进一步扩展到具有更多隐藏层的深层模型，这可以进一步提高性能。
CF-NADE旨在模拟评级的排序，是用于CF任务的前馈和神经自动贸易架构。理想情况下，物品的顺序应遵循评级的时间戳。然而，实证研究表明，每个用户的随机绘制置换也产生有利的性能。由于专家ID以及问题ID是匿名的，并且数据集中的专家和问题的描述已被编码为ID序列，因此可以在没有时间戳信息的情况下部署CF-Nade到此竞争是可行的。在培训CF-NADE模型时，专家和问题被视为用户和项目，并且评级矩阵从问题推送通知记录中得出的额定矩阵。
6.1。实验结果表明，问题1中的CF-NADE模型的性能类似于Autorec模型，其中基于项目的CF-NADE比基于用户的CF-Nade更好，但仍然没有与基于矩阵分解的模型相当作为SVD ++和ASVD
++。此外，CF-NADE模型虽然值得尝试，但不会集成到任何最终的集合模型中，因为它在结合到合并模型中时显着降低了性能。

6.4基于神经网络的协同滤波

最近关于建议的深度学习的研究通常采用深度学习方法来模拟辅助信息，例如物品和用户的文本描述。同时用于建模CF的关键因素（项目和用户特征之间的交互），它们仍然依赖MF模型并在潜在功能上使用内部产品。潜在特征乘法的线性组合成为提高性能的瓶颈。用神经结构替换内部产品，是一个名为神经网络的协作滤波（NCF）的一般框架（He等人，2017）能够从隐式数据中学习非线性usercitem交互功能。NCF由输入层，嵌入层，若干神经CF层和输出层组成。输入层由描述用户和项目的两个特征向量组成

respectively. Then the embedding layer maps the sparse input vectors to dense vectors. They are regarded as the latent vectors for the user and the item. Finally, the embedding of user and item is fed into the neural CF layers to project the latent vectors to the final prediction scores. Each of the neural CF layers can be modified to learn specific latent structures of user-item interactions. While utilizing NCF for *Problem* 1, experts are items and questions are users. The expert tag data and the question data are regarded as the descriptions of experts and questions, and the question distribution data is regarded as the implicit expert-question interaction data.

6.5 Match-SRNN

Furthermore, the expert finding problem in CQA can also treated as a text matching problem. Thus, text matching methods can be applied to this task. It can take advantage of textual features such as characters and words in the the expert and question descriptions. For the *Problem* 1, a deep text matching model called Match-SRNN (Wan et al. 2016) is applied to model the interaction information between texts to further predict new expert-question pairs. The Match-SRNN model contains three parts: a neural tensor network to capture the character/word level interactions, a spatial recurrent neural network (spatial RNN) applied on the character/word interaction tensor to capture the global interactions recursively, and a linear scoring function to calculate the final matching score. The Match-SRNN model views the generation of the global interaction between two texts as a recursive process. It can not only obtain the interactions between nearby words, but also take advantage of long distant interactions.

7 Ranking based models

The evaluation criterion in this task is normalized discounted cumulative gain (NDCG), thus ranking based model is a natural fit for this target. There are two kinds of ranking based models appearing in the expert finding problem in CQA, including ranking based FM and ranking based SVM.

7.1 Ranking based FM

The basic idea of this model is coming from the FM method. We modify the objective function to optimize the pair-wise ranking loss. Let N^+ denotes the number of positive samples and N^- denotes the number of negative samples. Besides, x_i denotes the negative instances and x_j denotes positive instances. Then we solve the following optimization problem for ranking based FM.

$$\min_{W,V} \frac{1}{N^+ + N^-} \sum_{i=1}^{N^-} \sum_{j=1}^{N^+} log(1 + exp(\hat{y}(x_i) - \hat{y}(x_j))) + \frac{\lambda}{2} \|\theta\|^2 \tag{28}$$

where $\hat{y}(x)$ is the prediction in the Eq. (15). We expect that those positive samples have higher prediction score than those negative samples.

858 S. Yuan et al.

分别。然后嵌入层将稀疏输入向量映射到密集的矢量。它们被视为用户和项目的潜在矢量。最后，将用户和项目的嵌入馈送到神经CF层中，以将潜在的矢量投影到最终预测分数。可以修改每个神经CF层以了解用户项目交互的特定潜在结构.Wherizedncfforproblem 1，ExpersareItemsandQuestionsAreseUsers。专家标签数据和问题数据被认为是专家和问题的描述，并认为问题分配数据作为隐式专家问题交互数据。

6.5 Match-SRNN

此外，CQA中的专家发现问题也可以视为文本匹配问题。因此，可以将文本匹配方法应用于此任务。它可以利用专家和问题描述中的字符和单词等文本功能。对于问题1，应用于Match−SRNN的深文本匹配模型（WAN等人。2016）以模拟文本之间的交互信息，以进一步预测新的专家问题对。匹配−SRNN模型包含三个部分：神经张量网络，用于捕获字符/字交互张量的空间复发性神经网络（Spatial RNN），以递归捕获全局交互，以及一个线性得分障碍关键稳定性触发性核心.Thematch−srnnmodelviewsthegen−两个文本之间的全局交互作为递归过程。它不仅可以获得附近的单词之间的相互作用，而且还利用了长时间的相互作用。

7 Ranking based models

此任务中的评估标准是归一化的折扣累积增益（NDCG），因此基于排序的模型是对该目标的自然契合。CQA中专家发现问题出现了两种基于排名的模型，包括基于排名的FM和基于排名的SVM。

7.1 Ranking based FM

该模型的基本思想来自FM方法。我们修改目标函数以优化成对排名损失。让n +表示正样本的数量，n − 表示阴性样本的数量。此外，Xi表示负实例，x j表示正实例。然后我们解决基于排名的FM的以下优化问题。

$$\min_{W,V} \frac{1}{N^+ + N^-} \sum_{i=1}^{N} \sum_{j=1}^{N} log(1 + exp(\hat{y}(x_i) - \hat{y}(x_j))) + \frac{\lambda}{2} \|\theta\|^2 \tag{28}$$

其中y（x）是Eq中的预测。（15）。我们预计这些阳性样本具有比那些阴性样品更高的预测得分。

7.2 Ranking based SVM

Ranksvm (Joachims 2006), which is a linear pairwise ranking model, has also been used in the problem. Specifically, we first build the feature vectors for each user-question pair appeared in the training/test sets. Then those training pairs with same questions are organized together as a list. The pairwise constraints are then built within each list.

8 Ensemble learning

During the review of the ensemble learning solutions, we find that many contestants are obscure about the concept of ensemble learning, especially Stacking. These proper nouns are often inappropriately used in ensemble learning. Here, we comb through the relevant concepts of ensemble learning that are widely used in practice. In machine learning, ensemble learning [also called ensemble method (Bifet et al. 2009) before] is a proper noun. It is a method of using multiple learning algorithms to obtain better predictive performance than that could be obtained by any of the component learning algorithms alone. Ensemble learning can be used for classification problems, regression problems, feature selection, anomaly detection and so on. In the following part, we will use classification as an example.

If we use ensemble learning to improve the overall generalization ability of classifiers, the following two conditions should be satisfied. Firstly, differences exist between the base classifiers. The performance of the ensemble classifier will not be improved, if it is just an ensemble of the same kind of base classifiers. Secondly, the classification accuracy of every base classifier must be larger than 0.5. If the classification accuracy of the base classifier is less than 0.5, the classification accuracy of the ensemble classifier will decline with the increasing of ensemble size. If the two aforementioned conditions are satisfied, the classification accuracy of the ensemble classifier will edge up to 1 with the increasing of ensemble size. Generally, the classification accuracy of a weak classifier is just slightly better than random guess, while a strong classifier can make make very accurate predictions. The base classifiers are referred to as weak classifier.

There are two key points in ensemble learning. How to generate base classifiers with difference? How to combine the results of the base classifiers? We will introduce ensemble learning from these two aspects.

8.1 Types of ensemble learning

According to how the base classifiers are constructed, there are two paradigms of ensemble learning, the parallel ensemble learning and the sequential ensemble learning. In the parallel ensemble learning, the base classifiers are generated in parallel, with Bagging (Breiman 1996) as a representative. In the sequential ensemble learning, the base classifiers are generated sequentially, with Boosting (Friedman et al. 2000) as a representative.

8.1.1 Bagging

Bagging (Bootstrap aggregating) was proposed to improve classification accuracy by combining classifiers of randomly generated training sets. Fig. 2a illustrates the diagram of Bagging.

7.2 Ranking based SVM

Ranksvm（Joachims 2006），即线性成对排名模型也已用于问题。具体地，我们首先构建在训练/测试集中出现的每个用户问题对的特征向量。然后将那些具有相同问题的培训对作为列表组织。然后在每个列表中构建成对约束。

8 Ensemble learning

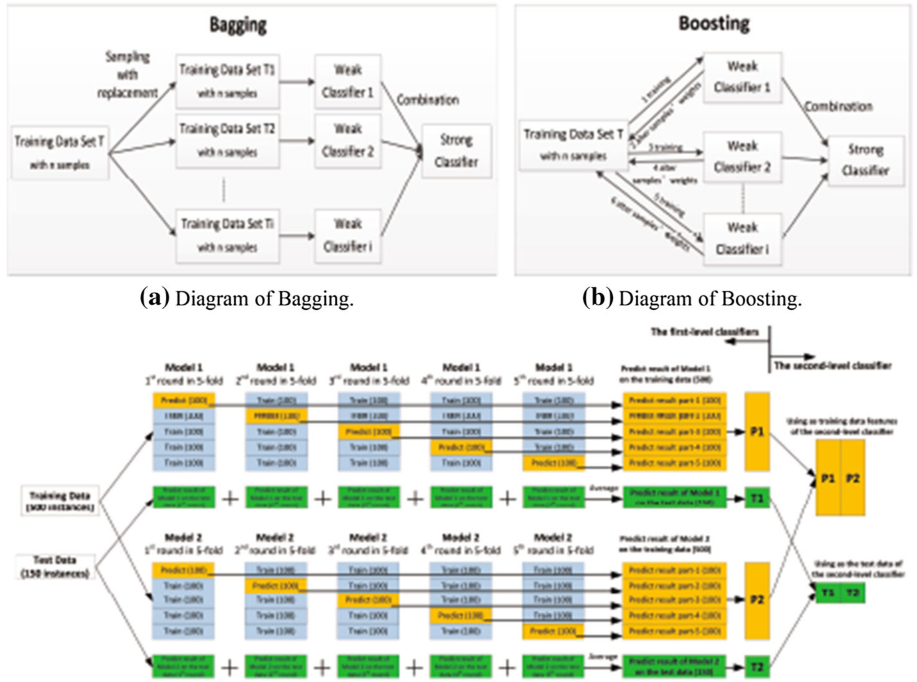
在审查集合学习解决方案期间，我们发现许多参赛者对集合学习的概念来说是模糊的，特别是堆叠。这些专有的名词通常在集合学习中使用不适当。在这里，我们通过在实践中广泛使用的集合学习的相关概念来梳理。在机器学习中，合奏学习[也称为集合方法（Bifet等，2009）之前]是一个合适的名词。它是使用多学习算法来获得比可以单独的任何组件学习算法获得的更好预测性能的方法。集合学习可用于分类问题，回归问题，特征选择，异常检测等。在以下部分中，我们将作为示例使用分类。如果我们使用集合学习来提高分类器的整体泛化能力，应满足以下两个条件。首先，基础分类器之间存在差异。如果只是同一类型基本分类器的集合，则不会改进集合分类器的性能。其次，每个基本分类器的分类准确性必须大于0.5。如果基本分类器的分类精度小于0.5，则集合分类器的分类准确性将随着集合大小的增加而下降。如果满足两个上述条件，则集合分类器的分类精度将随着合奏尺寸的增加而最高可达1。通常，弱分类器的分类准确性略高于随机猜测，而强大的分级器可以使得能够非常准确的预测。基本分类器称为弱分类器。集合学习中有两个关键点。如何生成带有差异的基本分类器？如何结合基本分类器的结果？我们将从这两个方面引入集合学习。

8.1种族学习类型

根据基本分类器的构建方式，共有学习的两个范式，并行集合学习和顺序集合学习。在并行集合学习中，基础分类器并行生成，袋装（Braing 1996）作为代表。在顺序集合学习中，基本分类器是顺序生成的，升压（Friedman等人）作为代表。

8.1.1 Bagging

提出了袋装（引导集合），通过随机生成的培训集的分类器来提高分类准确性。图。图2A示出了袋装的图。



(a) Diagram of Bagging. **(b)** Diagram of Boosting.

(c) Diagram of Stacking.

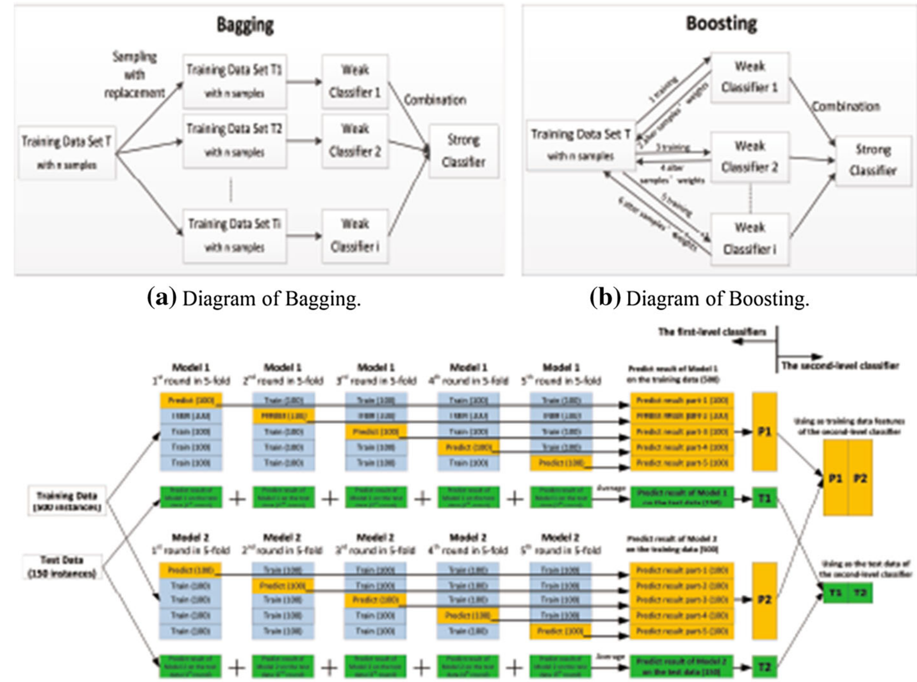
Fig. 2 Diagram of ensemble learning

Bagging applies bootstraping (Johnson 2001) to obtain the data subsets for training the base classifiers. In detail, given a training data set containing n training examples, a sample of n training examples will be generated by random sampling with replacement. Some original examples appear more than once, while some original examples are not present in the sample. If we need to train m number of base classifiers, this process will be applied m times. The combination methods used by Bagging are the most popular strategies, that is, voting for classification and averaging for regression. Here, the final classification results are determined by averaging on the respective results of these classifiers.

8.1.2 Boosting

Instead of resampling the training dataset as Bagging does, Boosting adjusts the distribution of the training dataset. Fig. 2b illustrates the diagram of Boosting. Boosting is an iterative process to generate base classifiers sequentially, where the later classifiers focus more on the mistakes of the earlier classifiers. In each round, the weight of the samples, which have been classified incorrectly, will be increased in the training dataset. The weight of the samples, which have been classified correctly, will be decreased in the training dataset. Finally, the ensemble classifier is a weighted combination of these weak classifiers.

860 S. Yuan et al.



(a) Diagram of Bagging. **(b)** Diagram of Boosting.

(c) Diagram of Stacking.

图2集合学习图

Bagging Apples Bootstrappt（Johnson 2001）以获取用于培训的数据子集基本分类器。详细地，给定包含N训练示例的训练数据集，将通过随机抽样进行N训练示例的样本。某些原始示例似乎不止一次，而示例中的某些原始示例不存在。如果我们需要培训M数量的基本分类器，则将应用此过程M次。袋装使用的组合方法是最流行的策略，即投票用于分类和对回归的平均。这里，最终分类结果是通过对这些分类器的各个结果的平均来确定的。

8.1.2 Boosting

AS装袋时，可以调整训练数据集的分发，而不是将训练数据集重新采样。图。图2B示出了升压的图。升级是一个迭代过程，用于顺序生成基础分类器，其中稍后的分类器更多地关注早期分类器的错误。在每一轮中，在训练数据集中将增加已经被错误分类的样本的重量。已正确归类的样本的重量将在训练数据集中减少。最后，集合分类器是这些弱分类器的加权组合。

8.2 Combination methods

The combination method plays a crucial role in ensemble learning. After generating a set of base classifiers, ensemble learning resorts to combination method to achieve an ensemble classifier with strong generalization ability, rather than trying to find a best single classifier. Generally, the most popular combination methods used in practice are Voting, Averaging and Learning. Voting and Averaging are the most popular and fundamental combination methods for nominal outputs and numeric outputs, respectively. These two methods are easy to understand and use. Here, we mainly focus on the Learning, with Stacking (stacked generalization) as a representative.

8.2.1 Stacking

Unlike Voting and Averaging, Stacking is a general combining procedure where the base classifiers are combined non-linearly in a serial model. In Stacking, the base classifiers are called the first-level classifiers, while the combiner is called the second-level classifier (or meta-classifier). The basic idea of Stacking is to train several first-level classifiers using the original training dataset. And then, a new dataset generated from the first-level classifier is used to train the second-level classifier, where the outputs of the first-level classifiers are regarded as the input features of the new training dataset, and the original labels are still the labels of the new training data.

In the training phase of Stacking, if all the instances in the training dataset are used to train the first-level classifiers, and the outputs of the first-level classifiers are used to train the second-level classifier, there will be a high risk of over-fitting. Therefore, the instances used for generating the input of the meta-classifier need to be excluded from the training instances of the first-level classifiers. Generally, a cross validation is used to avoid this problem.

Taking a Stacking model with 2 first-level classifiers and 5-fold cross validation as an example, Fig. 2c illustrates the diagram of Stacking. There are 500 instances in the training dataset. Using the Model 1 (the first-level classifier) in Fig. 2c as an example, in the 5-fold cross validation, the training dataset is divided into 5 parts, and each part has 100 instances. Four of them (with 400 instances in total) are used to train the Model 1. The remaining one part (with 100 instances) is used to do prediction. The prediction results (5 parts with 500 instances in total) are used as the features of the input of the second-level classifier. In every round in the 5-fold cross validation, the trained Model 1 makes prediction on the test dataset (with 150 instances). After 5 rounds, there are 5 parts of the prediction results on the test dataset. Making an average of these 5 parts, there are still 150 instances in the final prediction result of Model 1 on the test dataset.

Generally, Stacking can be viewed as a specific combination method of the Learning combination strategy. What's more, it can also be regarded as a general framework of many ensemble methods used in practice.

9 Results

In terms of the evaluation criteria, NDCG will be used. Specifically, we will rank the experts based on the forecasted probability for a certain question, and evaluate the $NDCG@5$ and $NDCG@10$ of ranking results. The final evaluation formula is: $NDCG@5 * 0.5 + NDCG@10 * 0.5$.

8.2 Combination methods

组合方法在集合学习中起着至关重要的作用。在生成一组基本分类器之后，组合方法来组合方法实现具有强大泛化能力的集合分类器，而不是尝试找到最佳单分类器。通常，在实践中使用的最流行的组合方法是投票，平均和学习。投票和平均分别是名义输出和数字输出最流行和最基本的组合方法。这两种方法很容易理解和使用。在这里，我们主要专注于学习，堆叠（堆叠泛化）作为代表。

8.2.1 Stacking

与投票和平均不同，堆叠是一般组合过程，其中基本分类器在串行模型中非线性地组合。在堆叠中，基本分类器称为第一级分类器，而组合器称为第二级分类器（或元分类器）。堆叠的基本概念是使用原始训练数据集训练几个第一级分类器。然后，从第一级分类器生成的新数据集用于训练第二级分类器，其中第一级分类器的输出被视为新训练数据集的输入特征，并且原始标签仍然是新培训数据的标签。在堆叠的训练阶段，如果训练数据集中的所有实例用于训练第一级分类器，并且使用第一级分类器的输出用于训练第二级分类器，则会有很高的风险过度拟合。因此，用于生成元分类输入的实例需要从第一级分类器的训练实例中排除。通常，交叉验证用于避免此问题。用2个第一级分类器和5倍交叉验证的堆叠模型作为示例，图2C示出了堆叠的图。训练数据集中有500个实例。使用图1中的型号1（第一级别分类器）。如图2C所示，在5倍交叉验证中，训练数据集分为5个部分，每个部分都有100个实例。其中四个（总共有400个实例）用于训练模型1.剩余的一个部分（有100个实例）用于进行预测。预测结果（总共500个零件）用作第二级分类器的输入的特征。在5倍交叉验证中的每一轮中，训练模型1对测试数据集进行预测（具有150个实例）。
5轮后，测试数据集有5个预测结果。平均这5个部分，在测试数据集上的模型1的最终预测结果中仍有150例。通常，可以将堆叠视为学习组合策略的特定组合方法。更重要的是，它也可以被视为在实践中使用的许多合奏方法的一般框架。

9 Results

就评估标准而言，将使用NDCG。具体而言，我们将基于某些问题的预测概率对专家进行排名，并评估排名结果的N DCG @ 5和N DCG @ 10。最终的评估公式是：n dcg @ 5 ⋆ 0.5 + n dcg @ 10 ⋆ 0.5。

9.1 Data analysis

In this paper, we analyze the problem of expert finding in CQA by taking the data of ByteCup competition as an example. The data provided for the competitors consisting of expert finding records in CQA with three types of information: expert tags, question data and question distribution data:

1. The expert tag data contains IDs of all expert users, their interest tags, and processed profile descriptions.
2. The question data contains IDs of all questions, processed question descriptions, question categories, total number of answers, total number of top quality answers, total number of upvotes.
3. The question distribution data contains 290,000 records of question push notification. Each contains the encrypted ID of the question, the encrypted ID of the expert user and whether the expert user answered the question (0 =ignored, 1 =answered).

The training set, validation set and test set are divided based on these records. The training set is used for the training of the model. Validation set is used for online real-time evaluation of the algorithm. Test set is used for the final evaluation.

All expert ID and question ID are encrypted to protect user privacy. Also for privacy protection purpose, the original descriptions of the questions and the experts are not provided. Instead, the ID sequence of the characters (each Chinese character will be assigned an ID) and the ID sequence of the words after segmentation (each word will be assigned an ID) are provided. Validation and testing labels have not been published. They are used for online evaluation and final evaluation only.

9.2 Feature extraction

We summarise all possible features in Table 3. The expert user tags *uTag* may be multiple tags, i.e., 18, 19 and 20 may represent baby, pregnancy and parenting, respectively. In the feature of *uwordIDseq*, user descriptions (excluding modal particles and punctuation) are first segmented, and then each word will be replaced by the Character ID, i.e., 284/42 may represent "Don't Panic". In the feature of *ucharIDseq*, user descriptions (excluding modal particles and punctuation) are first segmented, and then each character will be replaced by the Character ID, i.e., 284/42 may represent "BE". The question tag *qTag* may be a list of single tags, i.e., 2 may represent fitness. The feature *upvoteNum*, *ansNum* and *topAnsNum* may indicate the popularity of the question.

We also study the positive/negative contributions of each feature. As Table 3 illustrated, four features, including *uwordIDseq*, *ucharIDseq*, *qwordIDseq* and *qcharIDseq*, have negative impact on the model performance. The implicit features *imE* and *imQ*, which have strong positive influence on the model performance, are needed to be considered in the prediction model.

Table 4 illustrates the features used by the top 5 teams in the competition ByteCup. The four features including *uwordIDseq*, *ucharIDseq*, *qwordIDseq* and *qcharIDseq* that have negative impact on the model performance shown in Sect. 9.2, have not been used by any team. Therefore, we does't include them in Table 4. Although there are nine positive features, simply combining all of them will not lead to the best performance. All top 5 teams use the four features, including *uID*, *qID*, *imE* and *imQ*. The latent features *imE* and *imQ* underlying the interactions between different kinds of entities have important influence on the performance.

862 S. Yuan et al.

9.1 Data analysis

在本文中，我们通过作为一个例子的数据来分析CQA中专家发现的问题。为竞争对手提供的数据，由CQA中的专家查找记录组成，具有三种类型的信息：专家标签，问题数据和质疑分发数据：

1.专家标签数据包含所有专家用户的ID，其兴趣标记和已处理的配置文件描述。2.问题数据包含所有问题的ID，处理的问题描述，问题类别，答案总数，最高质量答案的总数，高度的总数。3.问题分发数据包含290,000条问题推送通知记录。每个都包含问题的加密ID，专家用户的加密ID以及专家用户是否回答问题（0 =忽略，1 =答案）。

培训集，验证集和测试集根据这些记录划分。培训集用于培训模型。验证集用于算法的在线实时评估。测试集用于最终评估。所有专家ID和问题ID都被加密以保护用户隐私。此外，对于隐私保护目的，未提供问题和专家的原始描述。相反，提供了字符的ID序列（每个汉字将被分配一个ID）和分割后的单词的ID序列（每个单词将被分配一个ID）。验证和测试标签尚未发布。它们仅用于在线评估和最终评估。

9.2 Feature extraction

我们总结了表3中的所有可能特征。专家用户标签utag可以是多个标签，即18,19和20分别代表婴儿，怀孕和育儿。在UWORDSEQ的特征中，首先分段，用户描述（不包括模态粒子和标点符号），然后每个单词将被字符ID替换，即，284/42可以表示“不恐慌”。在Ucharidseq的特征中，首先分段，用户描述（不包括模态粒子和标点符号），然后每个字符将由字符ID替换，即，284/42可以表示“be”。问题标记QTAG可以是单个标签的列表，即，2可以表示适合度。特征Upvotenum，Ansnum和Topansnum可能表示问题的普及。我们还研究了每个功能的正/负贡献。如表3所示，四个特征，包括UWORDSEQ，UCHarIDSEQ，QONDIDSEQ和QCharidseq，对模型性能产生了负影响。在预测模型中需要考虑对模型性能具有强烈积极影响的隐式功能IME和IMQ。表4说明了竞争对手的前5个团队使用的功能。包括uWORDIDSEQ，UCHARIDSEQ，QONDIDSEQ和QCharIdseq的四个功能，对SECT中显示的模型性能产生负面影响。
9.2，任何团队都没有使用过。因此，我们在表4中包含它们。虽然有九个正面功能，simplycombiningallofthemwillnotleadtothebestperformance.alltop5teamsusethefour，包括uid，qid，ime和imq。潜在特征IME和IMQ基础各种实体之间的相互作用对性能有重要影响。

Table 3 Designed features

Name	Notation	Description	Type	+/−
Anonymized expert user ID	*uID*	The unique identifier of each expert user	id	+
Expert user tags	*uTag*	The tag of user information	Category	+
Word ID sequence of user	*uwordIDseq*	Segmented user description. Each word is replaced by a unique wordID	Category	−
Character ID sequence of user	*ucharIDseq*	Segmented user description. Each character is replaced by a unique charID	Category	−
Anonymized question ID	*qID*	The unique identifier of each question	id	+
Question tag	*qTag*	The tag of each question	Category	+
Word ID sequence of question	*qwordIDseq*	Same as uwordIDseq instead of question description	Category	−
Character ID sequence of question	*qcharIDseq*	Same as ucharIDseq instead of question description	Category	−
Number of upvotes	*upvoteNum*	Number of upvotes of all answers to this question	Numeric	+
Number of answers	*ansNum*	Number of all answers to this question	Numeric	+
Number of top quality answers	*topAnsNum*	Number of top quality answers to this question.	numeric	+
Implicit expert	*imE*	Expert list with implicit relationship.	category	++
Implicit question	*imQ*	Question list with implicit relationship.	category	++

Table 4 Designed features

Team	*uID*	*uTag*	*qID*	*qTag*	*upvoteNum*	*ansNum*	*topAnsNum*	*imE*	*imQ*
Team-1	•	•	•	•	◦	◦	◦	•	•
Team-2	•	◦	•	◦	◦	◦	◦	•	•
Team-3	•	◦	•	◦	◦	◦	◦	•	•
Team-4	•	◦	•	◦	•	•	•	•	•
Team-5	•	◦	•	◦	◦	◦	◦	•	•

•Means that the feature is used. ◦Means that the feature is not used

9.3 Results of single models

SVDFeature (Chen et al. 2012) and Factorization Machine(libFM) (Rendle 2012) tools are used for MF-based models. XGBoost (Chen and Guestrin 2016) is used for GBT-based models. The code based on Theano framework is used for the DL-based models.

Table 3 Designed features

名称表示法描述类型+/−		
匿名的专家用户ID UID每个专家用户ID +的唯一标识符		
专家用户标记UTAG用户信息类别+		
用户UWORDSEQ分段用户描述的单词ID序列。每个单词都被唯一的WordID替换	Category	−
用户Ucharidseq分段用户描述的字符ID序列。每个角色都被一个独特的牧群代替	Category	−
匿名问题ID QID每个问题ID +的唯一标识符		
问题标签qtag每个问题类别+		
问题ID序列QWORDSEQ与UWORDIDSEQ相同，而不是质疑描述		
问题ID问题序列QCharIdSeq与Ucharidseq相同，而不是质疑描述	Category	−
Upvotes Upvotenum对这个问题数字+的所有答案的Upvotes的数量		
答案的数量ansnum这个问题数字+的所有答案的数量		
最高质量的数量答案Topansnum对此问题的最高质量答案的数量。数字+		
隐含专家IME具有隐式关系的专家列表。类别++		
隐式问题IMQ问题列表具有隐式关系。类别++		

Table 4 Designed features

Team	*uID*	*uTag*	*qID*	*qTag*	*upvoteNum*	*ansNum*	*topAnsNum*	*imE*	*imQ*
Team-1	•	•	•	•	◦	◦	◦	•	•
Team-2	•	◦	•	◦	◦	◦	◦	•	•
Team-3	•	◦	•	◦	◦	◦	◦	•	•
Team-4	•	◦	•	◦	•	•	•	•	•
Team-5	•	◦	•	◦	◦	◦	◦	•	•

表示使用该功能。 未使用该功能的emeans

9.3单一型号的结果

SVDFeature（Chen等，2012）和分解机（Libfm）（Rendle 2012）工具用于基于MF的模型。XGBoost（Chen和Guestrin 2016）用于基于GBT的模型。基于Theano框架的代码用于基于DL的模型。

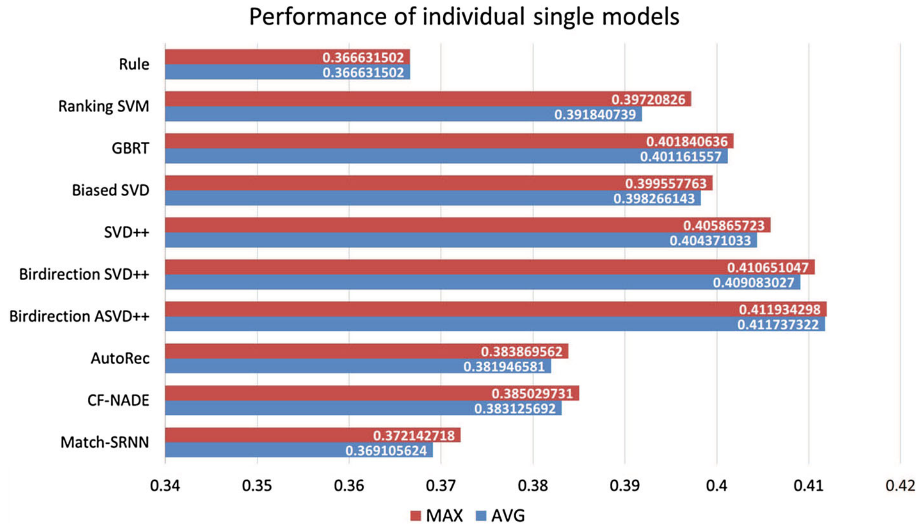


Fig. 3 Individual model performances on local validation dataset

The results of all aforementioned categories of single models on the local validation dataset is illustrated in Fig. 3. From the figure we can see that, some single models such as ASVD and bidirectional SVD++ make good performances. However, there are also weak models such as ranksvm and simple heuristic based method. In general, the MF-based models perform better than others including GBT-based models and DL-based models. The DL-based models don't perform well due to the sparse and encoded data in this task. We used different settings of parameters (max depth of each tree, number of trees, and boosting step size) to train several XGBoost models. Based on the experiments on local validation dataset, the performance of these models (refer to the performance of models starting with “GBRT” in Fig. 3) are reasonable, but not as good as MF-based models. Nevertheless, they do improve the performance of the final ensemble model. These models have quite different objective and underlying assumptions than MF-based methods. Therefore, a decent weak model will still improve the final ensemble results.

In the MF-based models, the bidirection ASVD++ performs the best. What's more, if more implicit information is used, such as ratting action in online validation dataset or online test dataset, the model performance could be further improved. This phenomenon is reflected in Fig. 4. The accuracy of the bidirect ASVD++ is highest, followed by the bidirect ASVD++, the bidirect SVD++ and the bidirect SVD in the descending order.

Table 5 illustrates the parameters for the bidirection ASVD++ that achieves the best performance. Markov Chain Monte Carlo (MCMC) is used for the learning method in the model. Table 6 illustrates the best performance of the bidirection ASVD++ on the local validation dataset, the online validation dataset and the online test dataset. The results are 0.41193, 0.52412 and 0.50551, respectively.

9.4 Results of ensemble models

Taking the ensemble models of the Top 5 teams who won the competition ByteCup as the example, we analysis the results of the ensemble models.

864 S. Yuan et al.

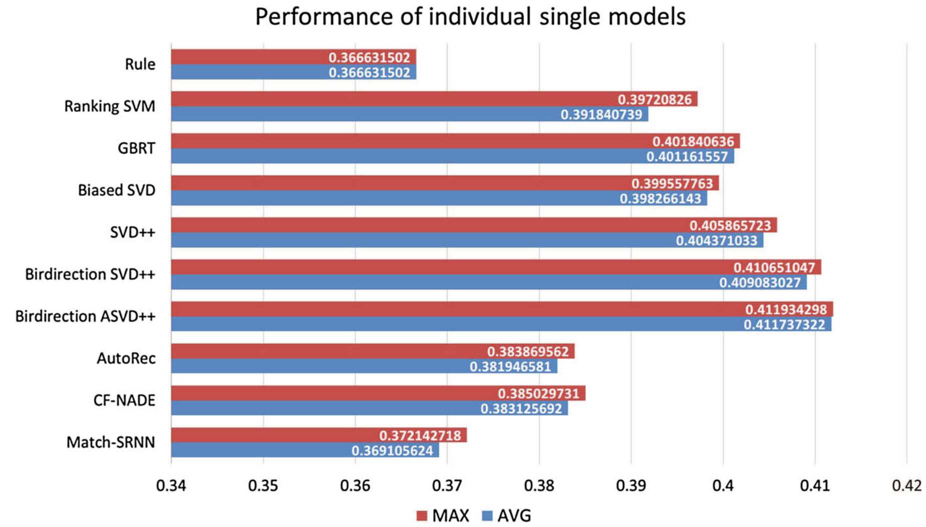


图3在本地验证数据集上的个人模型性能

所有上述本地验证的单一模型类别的结果
数据集在图3中示出。3.从图中我 们可以看出，某些单一型号，如ASVD和双向SVD ++具有良好的性能。但是，还有薄弱的模型，例如RankSVM和简单的启发式方法。通常，基于MF的模型比其他基于GBT的模型和基于DL的模型更好。基于DL的模型由于此任务中的稀疏和编码数据而不是良好的表现。我们使用了不同的参数设置（每棵树的最大深度，树木数量和升压步长），以训练几个XGBoost模型。基于本地验证数据集的实验，这些模型的性能（参考图3中的“GBRT”开始的模型的性能。3）是合理的，但不如基于MF的模型那么好。然而，他们确实改善了最终集合模型的性能。这些模型具有与基于MF的方法不同的目标和潜在的假设。因此，体面的弱模型仍将改善最终的集合结果。

在基于MF的模型中，BIDirection ASVD ++执行最佳。更重要的是，如果更多
使用隐式信息，例如在在线验证数据集或在线测试数据集中的roting动作，可以进一步提高模型性能。这种现象反映在图4中。Bidirect ASVD ++的准确性最高，然后是Bidirect ASVD ++，Bidirect SVD ++和Bidirect SVD的降序。

表5说明了Bidirection ASVD ++的参数，实现了最佳
表现。马尔可夫链蒙特卡罗（MCMC）用于模型中的学习方法。表6说明了本地验证数据集上的Bidirection ASVD
++的最佳性能，在线验证数据集和在线测试数据集。结果分别为0.41193,0.52412和0.50551。

9.4集合模型的结果

采用前5名队伍的集合模型作为竞争对手的竞争，我们分析了集合模型的结果。

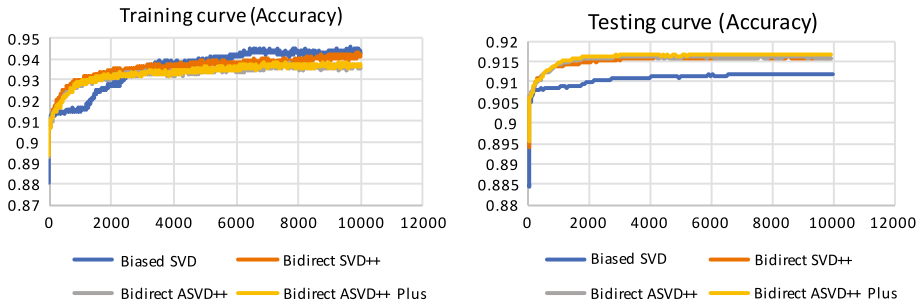


Fig. 4 MF-based models training/testing curve

Table 5 Parameters for the bidirection ASVD++

Parameters	Value
Learning method	MCMC
#Factor	8
#Iteration	10,000
Task	Binary classification
Stdev for init. of 2-way factors	0.1

Table 6 Performance of bidirection ASVD++

Test set	Performance (nDCG)
Local validation	0.41193
Online validation	0.52412
Online test	0.50551*

*Already rank first among all single models

9.4.1 Team-1

As shown in Table 7, Team-1 combines 45 models linearly with different settings (features, tools or hyper-parameters) using the linear ridge regression. Specifically, they do 5-fold cross validation on the local validation set. The final ensemble model is trained using local validation set. Note that, the predictions of local validation set are from those models trained on local training set. Thus the training set are not involved in the ensemble step. They also ensemble the predictions from same model with different parameters, such as different latent dimensions or different objective functions of matrix factorization models. The small variations make the single model more robust. To avoid the bias due to different scales, they do whitening for each model's prediction before ensemble.

Team-1 takes the predictions of each candidate model, and does a linear combination of those predicted values to make the final prediction. The score of these candidate models range from 0.367 to 0.412, they tune the weights of them based on the rating prediction on local validation set. The prediction ensemble of a set of base models further improves the performance. Finally, they get the score of 0.50812 on the final leaderboard. Team-1 has also tried to use nonlinear ensemble method, such as the gradient boosting tree, to do the ensemble. However, they found such tree models are very easy to over-fit the training set. It is also hard to regularize the model to get a good test performance.

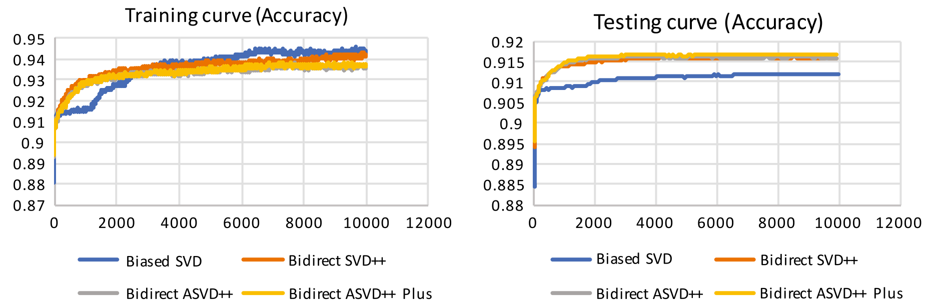


图4 MF的模型训练/测试曲线

表5 BDirection ASVD ++参数值的参数

Learning method	MCMC
#Factor	8
#Iteration	10,000
Task	Binary classification
stdev for init。双向因子0.1	

表6 Bidirection ASVD ++测试集性能的性能（NDCG）

Local validation	0.41193
Online validation	0.52412
Online test	0.50551*

*已经在所有单一模型中排名第一

9.4.1 Team-1

如表7所示，Team−1使用Linear Ridge回归与不同的设置（特性，工具或超参数）线性地结合了45个模型。具体而言，它们在本地验证集上执行5倍交叉验证。最终的集合模型使用本地验证集进行培训。注意，本地验证集的预测来自于在本地训练集上培训的模型。因此，训练集不参与集合步骤。它们还与不同参数的相同模型的预测合并，例如不同的潜在尺寸或矩阵分子化模型的不同客观函数。小变化使单一模型更加强大。为了避免由于不同的尺度而导致的偏差，它们会在合奏前进行每个模型的预测。

Team−1获取每个候选模型的预测，并且执行这些预测值的线性组合以进行最终预测。这些候选模型的得分范围为0.367至0.412，它们根据本地验证集的额定预测调整它们的权重。一组基础模型的预测集合进一步提高了性能。最后，他们在最终排行榜上获得0.50812的得分。

Team−1还试图使用非线性合奏方法，例如渐变升压树，进行合奏。但是，他们发现这种树模型非常容易过度拟合训练集。还难以正规化模型以获得良好的测试性能。

Table 7 Ensemble models used by the top 5 teams

Team	Details of the ensemble model	Final results	Compare with Team-1
Team-1	Linearly combine all models in Fig. 3	0.50812	0
Team-2	Use stacking strategy illustrated as Fig. 5	0.50307	− 1%
Team-3	FM + CF ∗	0.49905	− 1.82%
Team-4	MF+CF	0.49231	− 3.21%
Team-5	FM+RFM+(FM+RFM)+MF+SVD+(SVD++)	0.49003	− 3.69%

∗FM+CF represents the linear weighted sum of FM and CF

9.4.2 Team-2

For every expert, there is a list of questions that have been answered. Here, Team-2 regards the expert-question list as a document, and each question as a term. The TF-IDF of each question is calculated and used as the feature *imQ*. Similarly, The TF-IDF of each expert is calculated and used as the feature *imE*.

Team-2 uses the method of Stacking to integrate several single models. The Stacking strategy used by them is illustrated in Fig. 5. In the Stacking, FM, Logistic Regression (LR), XGBoost and Neural Network (NN) are the first-level classifiers. The results of them are used as inputs of the next layer, called meta features. SVD, TSNE (Pezzotti et al. 2017), NMF (Paatero and Tapper 1994) is used to get the dimension reduction features of the original features. Finally, the meta features and the dimension reduction features are combined to train the XGBoost.

The used NN has one hidden layer, in which the activation function is ReLu (Rectified Linear Units), the droupout rate is 0.75. Adam (Kingma and Ba 2014) is also used here to optimize the model. XGBoost is trained in the following steps. They uses the social graph to model the relationship between experts and questions $< E, Q >$. The experts and questions are regarded as nodes in an undirected graph. If a expert is invited to answer a question, there will be an undirected edge between them. DeepWalk (Perozzi et al. 2014) is used to convert $< E, Q >$ to work vector, which then be used to train XGBoost.

In addition, they find three implied CF messages based on the observation and analysis of the issues and data.

- If a expert has accepted most of the invitation for answering question, he will be more likely to accept the new invitation to answer question.
- Experts have answered some same questions. If some of them (assume the number is N) answer a new question, others may also answer the question (assume the probability is p). N is larger, p is larger.
- If questions $Q1$ and $Q2$ are given to the same user, $Q1$ and $Q2$ may be involved in the same field. If $Q1$ is answered by an expert, $Q2$ may be answered by the expert too.

And then, they combine the results of Stacking and CF by weight $2 : 1$. Finally, they get the score of 0.50307 on the final leaderboard. It is 1% less than Team-1.

9.4.3 Team-3

The weight of the question related to the expert *uid* is regarded as the feature *imQ* by Team-3. It is calculated as the reciprocal of the question numbers answered by the expert *uid*. The

表7前5支队伍使用的集合模型

团队细节的集合模型最终结果与团队−1相比			
Team−1线性结合了图3 0.50812 0的所有型号			
Team−2使用堆叠策略如图5 0.50307 −1%			
			%
Team-4	MF+CF	0.49231	−3.21%
Team-5	FM+RFM+(FM+RFM)+MF+SVD+(SVD++)	0.49003	−3.69%

⋆ FM + CF表示FM和CF的线性加权和

9.4.2 Team-2

对于每个专家，都有一个已解答的问题列表。在这里，Team−2将专家问题列表视为文件，每个问题都是一个术语。计算每个问题的TF−IDF，用作特征IMQ。类似地，计算每个专家的TF−IDF，并用作特征IME。
Team−2使用堆叠方法来集成几个单一型号。它们使用的堆叠策略在图5中示出。在堆叠，FM，逻辑回归（LR），XGBoost和神经网络（NN）中是第一级分类器。它们的结果用作下一层的输入，称为元特征。
SVD，TSNE（Pezzotti等，2017），NMF（Paatero和Tapper 1994）用于获得原始功能的尺寸减少功能。最后，元件特征和尺寸减少功能组合以培训XGBoost。使用的NN具有一个隐藏层，其中激活函数是Relu（整流线性单元），追踪速率为0.75。亚当（Kingma和Ba 2014）也用于优化模型。
XGBoost在以下步骤中受过培训。他们使用社会图来模拟专家与问题之间的关系<e，q>。专家和问题被视为一个无向图中的节点。如果邀请专家回答问题，则它们之间将有一个无向优势。
Deepwalk（Perozzi等人2014）用于将<e，q>转换为工作向量，然后使用它用于训练XGBoost。此外，它们根据问题和数据的观察和分析找到三个隐含的CF消息。

− 如果专家接受了大部分回答问题的邀请，他将更有可能接受新邀请来回答问题。 − 专家回答了一些同样的问题。如果其中一些（假设数字是n）回答一个新问题，其他人也可能回答这个问题（假设概率是p）。n更大，p更大。 − 如果对同一用户提供问题Q1和Q2，则Q1和Q2可以涉及相同的字段。如果Q1由专家回答，则可以由专家回答Q2。

然后，它们将堆叠的结果与重量2分2：1。最后，他们在最终排行榜上获得0.50307的得分。它比Team−1少1%。

9.4.3 Team-3

与专家UID相关的问题的重量被视为Team−3的特征IMQ。它计算为专家UID回答的问题编号的互惠。这

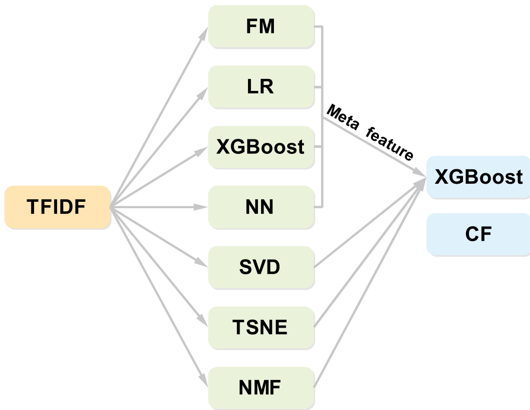


Fig. 5 Diagram of stacking used by Team 2

weight of the expert related to the question *qid* is regarded as the feature *imE*. It is calculated as the reciprocal of the expert numbers who answer the question *qid*. FM is achieve by libFM.

In CF, the probability of expert answering question is calculated as the weighted sum of the average similarity between experts and the average similarity between questions. The similarity between questions is calculated as the weighted difference between the positive similarity of the question and the negative similarity of the question. The positive similarity of question is the number of experts who have similar behavior on the specific question and answer the test question. The negative similarity of question is the number of experts who have similar behavior on the specific question and not answer the test question. The similarity between experts is calculated similarly as the similarity between questions.

As shown in Table 7, Team-3 combines the results of FM and CF with the linear weighted sum. Finally, they get the score of 0.49905 on the final leaderboard. It is 1.82% less than Team-1.

9.4.4 Team-4

As shown in Table 7, Team-4 combines the results of MF and CF with the linear weighted sum. In the scheme of CF, the prediction is calculated as the formula shown below:

$$pred(u,i) = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,i) * (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u,i)}, \tag{29}$$

where $sim(u,i)$ is calculated by

$$sim(u,i) = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_i (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_i (r_{v,i} - \bar{r}_v)^2}}. \tag{30}$$

$N(u)$ is the set of neighbors of the specific expert u. The number n of $N(u)$ is hyper parameter needed to be tune. They use $n = 5000$ in the final model.

Finally, they get the score of 0.49231 on the final leaderboard. It is 3.21% less than Team-1.

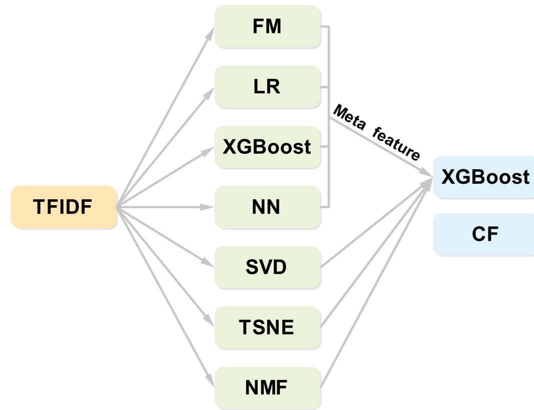


图5由团队使用的堆叠图

与问题相关的专家的重量被视为特征IME。它计算为回答问题Qid的专家号码的互惠。FM由Libfm实现。

在CF中，专家应答问题的概率计算为加权总和

专家之间的平均相似性和问题之间的平均相似性。问题之间的相似性被计算为问题的正相似性与问题的负相似性之间的加权差异。问题的积极相似之处是在特定问题上具有类似行为的专家数量并回答测试问题。问题的负面相似之值是在特定问题上具有类似行为的专家数量，而不是回答测试问题。专家之间的相似性与问题之间的相似性类似地计算。

如表7所示，Team-3将FM和CF的结果与线性加权结合在一起

和。最后，他们在最终排行榜上获得0.49905的分数。它比Team-1低1.82%。

9.4.4 Team-4

如表7所示，Team-4将MF和CF的结果与线性加权和相结合。在CF的方案中，将预测计算为下面所示的公式：

$$pred(u,i) = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,i) * (r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} sim(u,i)}, \tag{29}$$

其中sim（u，i）计算

$$sim(u,i) = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\ {}_i (r_{u,i} - \bar{r}_u)^2} \sqrt{\ {}_i (r_{v,i} - \bar{r}_v)^2}}. \tag{30}$$

n（u）是特定专家U的邻居集。n（u）的数字n是需要调整的Hyper参数。它们在最终模型中使用n =

5000。最后，他们在最终排行榜上获得0.49231的得分。它比Team-1小3.21%。

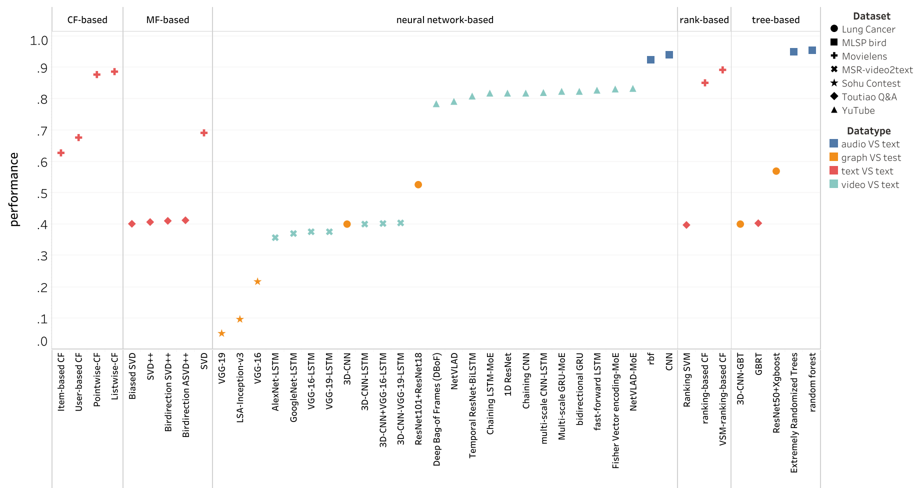


Fig. 6 Performances of diverse models on different type of datasets

9.4.5 Team-5

Team-5 combines the results of 6 individual models on the validation set, including FM, ranking based FM (RFM), the linear weighted sum of FM and RFM, three MF-based models (MF, SVD and SVD++). Assuming the predictions of the user-question pairs from the 6 individual models are $pred_1$, $pred_2$, $pred_3$, $pred_4$, $pred_5$, $pred_6$, respectively. A weight is assigned to every individual model and the final prediction of the user-question pairs is computed by the following formula:

$$\begin{aligned} pred = \alpha_1 pred_1 + \alpha_2 pred_2 + \alpha_3 pred_3 \\ + \alpha_4 pred_4 + \alpha_5 pred_5 + \alpha_6 pred_6 \end{aligned} \tag{31}$$

After the ensemble, the performance of the model turns out to be better.

What's more, Team-5 finds a rule in the training set, and it can be used in the validation set to improve the model performance. In the training set, a certain user-question pair only appears once or twice and a user answers the question once at most. Therefore, they assume that expert won't answer the same question twice and it is consistent with the reality. When the user-question pair appears in the validation set and it also appears in the training set where the user answers the question, they predict that user won't answer the question again. This rule helps to boost the performance on the validation set again.

Finally, they get the score of 0.49003 on the final leaderboard. It is 3.69% less than Team-1.

10 Diverse models on different types of matching tasks

In this section, we compare the performance of diverse models on different types of matching tasks to explore the difference among the models on different matching tasks (Fig. 6). Totally, seven matching tasks were involved in the study including:

1. Toutiao: The evaluation metric of ByteCup is $NDCG@5 * 0.5 + NDCG@10 * 0.5$;

868 S. Yuan et al.

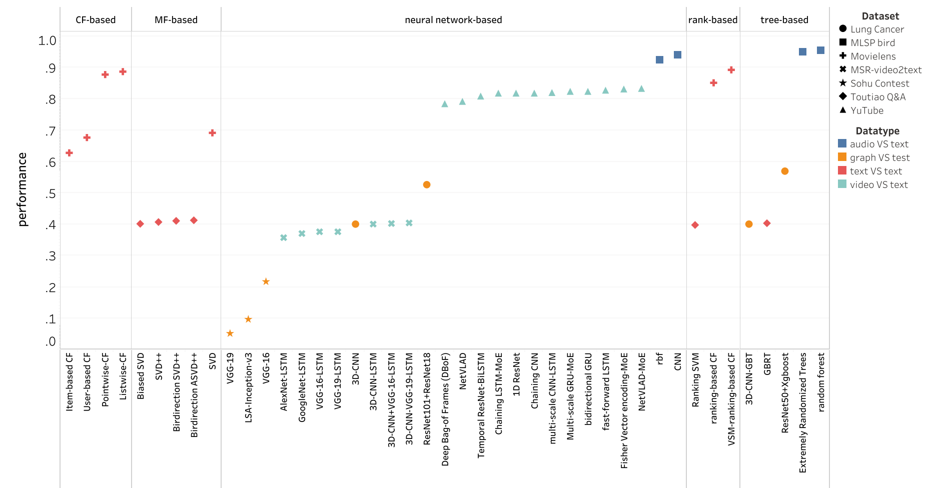


图6不同类型的数据集不同模型的性能

9.4.5 Team-5

Team−5将6个单独模型的结果组合在验证集中，包括FM，基于排名的FM（RFM），FM和RFM的线性加权和，基于三种MF的模型（MF，SVD和SVD++）。假设来自6个单独模型的用户问题对的预测分别是pred1，pred2，pred3，pred4，pred5，pred6。将重量分配给每个单独的模型，并且通过以下公式计算用户问题对的最终预测：

$$\begin{aligned} pred = \alpha_1 pred_1 + \alpha_2 pred_2 + \alpha_3 pred_3 \\ + \alpha_4 pred_4 + \alpha_5 pred_5 + \alpha_6 pred_6 \end{aligned} \tag{31}$$

在集合之后，模型的性能结果变得更好。

更重要的是，Team−5在培训集中找到规则，可以在验证中使用设置以提高模型性能。在培训集中，某些用户问题对仅出现一次或两次，用户最多都回答问题。因此，他们认为专家不会两次回答相同的问题，并且它与现实一致。当用户问题对中出现在验证集中时，它还显示在用户回答问题的培训集中，他们预测用户不会再回答问题。此规则有助于再次提升验证上的性能。

最后，他们在最终排行榜上获得0.49003的得分。比团队−1小3.69%。

10种不同类型的匹配任务的不同模型

在本节中，我们可以比较不同类型的不同类型匹配任务的性能，以探索不同匹配任务的模型之间的差异（图6）。完全涉及七项匹配任务，包括：

1. Toutiao：Bytecup的评估度量是N DCG @ 5 ⋆ 0.5 + N DCG @ 10 ⋆ 0.5;

2. Movielens: Movie recommendation on MovieLens data with evaluation metric $NDCG@10$;
3. Sohu Contest: Sohu Programming Contest[14] on news pictures data with evaluation metric average NDCG;
4. Lung Cancer: Data Science Bowl 2017[15] on Lung CT images data with evaluation metric LogLoss;
5. MLSP bird: MLSP 2013 Bird Classification Challenge[16] on bird sounds audio data with evaluation metric micro-AUC;
6. YouTube: Google Cloud & YouTube-8M Video Understanding Challenge[17] on YouTube videos data with evaluation metric Global Average Precision@20;
7. MSR-video2text: Video to Language Challenge[18] on MSR-video2text data with evaluation metric $BLEU@4$.

Based on the data type of the tasks, we classified the seven tasks into 4 categories. There are : (1) *text vs. text*, which means to match text labels with text data, includes ByteCup and Movie recommendation; (2) *graph vs. text*, which means to match text labels with graph data, contains Sohu Programming Contest and Data Science Bowl 2017; (3) *audio vs. text*, which aims to match text labels with audio data, includes MLSP 2013 Bird Classification Challenge; (4) *video vs. text*, which is to match text labels with video data, includes Google Cloud & YouTube-8M Video Understanding Challenge and Video to Language Challenge.

The models used in the seven tasks are also classified into four categories including MF-based models, GBT-based models, R-based models and DL-based models. As shown in Fig. 6, MF-based models and rank-based models are used only in *text vs. text* category of matching tasks, while DL-based models are not employed in these tasks since they are not performing well (which may due to the severe sparsity of the datasets). MF-based models usually achieve the best performance in *text vs. text* category of matching tasks. In addition, DL-based models achieve the best performance in *graph vs. text* and *video vs. text* categories, which may due to their outstanding power of capturing high dimensional features from graph and video, and they are also utilized in the *audio vs. text* category. Finally, GBT-based models have significant performance in the *audio vs. text* category.

11 Discussion

In this article, we statistically analyze all the existing solutions for the expert finding problem in CQA. We summarise the results analysis and the learned lessons in this part.

11.1 Results analysis

We describe the different individual methods used in the task, and also introduce several types of ensemble learning. And then, we present the results of both of them. It is worth noting that the different individual methods get scores from 0.3665 to 0.4119 when used independently. The results of ensemble learning range from a score of 0.49003 to a score of

[14] https://biendata.com/competition/luckydata/.
[15] https://www.kaggle.com/c/data-science-bowl-2017.
[16] https://www.kaggle.com/c/mlsp-2013-birds.
[17] https://www.kaggle.com/c/youtube8m.
[18] http://ms-multimedia-challenge.com/2016/challenge.

2. movielens：movierecommendationonmovielensdatawithevalionmetric n dcg @ 10;3.搜狐竞赛：搜狐编程竞赛14关于新闻图片数据评估度量平均NDCG;4.肺癌：数据科学碗201715在肺CT图像数据上评价度量标准Logloss;5. MLSP鸟：MLSP 2013鸟类分类挑战16在鸟类声音和评估度量微AUC的音频数据;6. YouTube：Google Cloud＆YouTube−8M视频了解挑战17关于YouTube视频数据，评估度量全球平均平均精度@ 20;7. MSR−Video2Text：视频到MSR−Video2Text数据上的语言挑战18，具有评估度量BLEU @ 4。

根据任务的数据类型，我们将七个任务分为4类。有：（1）文本与文本，这意味着将文本标签与文本数据匹配，包括ByTecup和Movie推荐;
（2）图表与文本，这意味着匹配具有图形数据的文本标签，包含Sohu编程比赛和2017年数据科学碗;（3）音频与文本，旨在将文本标签与音频数据匹配，包括MLSP 2013鸟类分类挑战;（4）视频与文本，用于将文本标签与视频数据匹配，包括Google Cloud＆YouTube−8M视频了解挑战和视频到语言挑战。七项任务中使用的模型也分为四类，包括基于MF的模型，基于GBT的模型，基于R基的模型和基于DL的模型。如图6所示，基于MF的模型和基于秩的模型仅在文本与文本类别中使用匹配任务，而基于DL的模型不会在这些任务中采用，因为它们不顺利（可能是由于到数据集的严重稀疏性）。基于MF的模型通常在匹配任务中的文本与文本类别中实现最佳性能。此外，基于DL的模型在图表和视频与视频与视频与视频与视频与视频与文本类别中实现了最佳性能，这可能是由于它们从图形和视频捕获高维特征的出色功能，并且它们也在音频与音频上使用。文本类别。最后，基于GBT的模型在音频与文本类别中具有显着性能。

11 Discussion

在本文中，我们统计分析CQA中专家发现问题的所有现有解决方案。我们总结了这一部分的结果分析和学习课程。

11.1 Results analysis

我们描述了任务中使用的不同单独方法，并介绍了几种类型的集合学习。然后，我们介绍了它们两个的结果。值得注意的是，不同的单独方法在独立使用时从0.3665到0.4119获得分数。合奏学习的结果从分数为0.49003到分数

[14] https://biendata.com/competition/luckydata/.
[15] https://www.kaggle.com/c/data-science-bowl-2017.
[16] https://www.kaggle.com/c/mlsp-2013-birds.
[17] https://www.kaggle.com/c/youtube8m.
[18] http://ms-multimedia-challenge.com/2016/challenge.

0.50812. Since the data used in the task is the real data from Toutiao with about 580 million users, even minor improvements can affect millions of users.

Based on the analysis of the solutions and the observation of the results, we find that the ensemble methods outperform any of the single models when they were used independently. That is, ensemble learning really outperforms every single component model, if the two conditions mentioned in Sect. 8 are both satisfied. Although there are some model with poor performance, the use of them with other different kind of models leads to a considerable improvement of the prediction. YES! A weak model in combination with other different kind of models can still improve the performance of the final ensemble model. In general, the combination of different kinds of models even with a weak model[19] leads to significant performance improvements over every single component model.

11.2 Important lessons

As known from the No Free Lunch Theorem, none of the algorithms is better than a random one. In the field of machine learning, there isn't an almighty algorithm that is applicable to all situations. Different data sets and different problems have different best algorithms respectively. In previous years, XGBoost shows its absolute advantage in the structured data. However, it puts up a poor show than MF-based models in this task. It is a reasonable explanation that the dataset here is more sparse than movie rating datasets used in previous tasks.

As noticed, a single model won't win. This shows that, as expected, the field of machine learning is getting stronger. This paper witnesses the advantage of ensemble learning applied to the combination of different learning models. In addition, many mobile social platforms in China, such as WeChat, Sina Weibo, Toutiao and so on, have hundreds of million users. Even minor improvements of the solution results can affect millions of users.

Moreover, from the survey of the performance of different models on different types of matching tasks, we learned that MF-based models and rank-based models are more suitable for *text vs. text* matching tasks, DL-based models and GBT-based models achieve the best results for *audio vs. text* matching tasks. DL-based models are appropriate for both *video vs. text* and *audio vs. text* matching tasks.

12 Conclusion

This survey paper focuses on the expert finding problem in CQA. Given certain question, one needs to find who are the most likely to (1) have the expertise to answer the question and (2) have the willingness to accept the invitation of answering the question. We have reviewed the most recent solutions and classified them to four different categories: MF-based models, GBT-based models, DL-based models and R-based models. Experimental results demonstrate the effectiveness and efficiency of the MF-based models in the expert finding problem in the crowdsourcing situation.

In the future, several important research issues need to be addressed. First, how to efficiently integrate the implicit feedback is an open problem. Obviously, implicit feedback becomes increasingly important in practical application, because users provide much more implicit feedback than explicit one. In addition, explainability is usually ignored in the research. The existing methods face real difficulties to explain predictions. Finally, how

19 Its accuracy is larger than 0.5.

870 S. Yuan et al.

0.50812。由于任务中使用的数据是来自大约580万用户的Toutiao的实际数据，即使是小的改进也会影响数百万用户。基于对解决方案的分析和结果的观察，我们发现集合方法独立使用时越优于任何单一型号。也就是说，如果在教派中提到的两个条件，集合学习真的胜过每个单个组件模型。8都满意。虽然性能差有一些型号，但与其他不同类型的模型一起使用它们导致预测的相当大改善。是的！与其他不同类型的模型相结合的弱模型仍可提高最终集合模型的性能。通常，即使具有弱模型的不同类型的模型的组合也会导致每个单个组件模型的显着性能改进。

11.2 Important lessons

从免费的午餐定理中已知，算法都没有比随机的更好。在机器学习领域，没有一种适用于所有情况的全能算法。不同的数据集和不同的问题分别具有不同的最佳算法。在过去几年中，XGBoost在结构化数据中显示了其绝对优势。但是，它在这项任务中展现了比基于MF的模型的差。这是一个合理的解释，即这里的数据集比以前任务中使用的电影额定值数据集更稀疏。如注意，单个型号将无法获胜。这表明，正如所预期的那样，机器学习领域越来越强烈。本文见证了合并学习应用于不同学习模型组合的优势。此外，许多中国的移动社交平台，如微信，新浪微博，Toutiao等，拥有数亿用户。即使对解决方案结果的轻微改进也可能影响数百万用户。此外，从对不同类型的匹配任务的不同模型的表现调查中，我们了解到基于MF的模型和基于秩的模型更适合文本与文本匹配任务，基于DL的模型和基于GBT的模型。达到音频与文本匹配任务的最佳效果。基于DL的模型适用于视频与文本和音频与文本匹配任务。

12 Conclusion

本调查纸重点介绍CQA中的专家发现问题。考虑到某些问题，人们需要找到最有可能（1）有谁有的专业知识来回答问题，并且（2）愿意接受回答问题的邀请。我们已审查了最新的解决方案并将其分类为四种不同的类别：基于MF的模型，基于MF的模型，基于Modelsandr−BasiaModels.preisherSumentStemontriseStemontristresstriseSumpriseStemontrists，在众包中专家发现问题中的基于MF的模型的有效性和效率。将来，需要解决几个重要的研究问题。首先，如何有效地整合隐式反馈是一个打开的问题。显然，隐式的反馈在实际应用中变得越来越重要，因为用户提供比明确的反馈更加隐含的反馈。此外，在研究中通常忽略解释性。现有方法面临着解释预测的真正困难。最后，怎么做

19其精度大于0.5。

Springer

to make sure that the established model is no needed to be retrained is a crucial issue in expert finding in CQA. We hope that the overview presented in this paper will advance the discussion in the expert finding technologies in CQA.

Acknowledgements This work is supported by the NSFC for Distinguished Young Scholar (61825602), National Natural Science Foundation of China (61806111), and the National High Technology Research and Development Program of China (863 Program) (2015AA124102).

References

Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 6:734–749

Alarfaj F, Kruschwitz U, Hunter D, Fox C (2012) Finding the right supervisor: expert-finding in a university domain. In: Association for Computational Linguistics, pp 1–6

Balog K, Fang Y, De Rijke M, Serdyukov P, Si L (2012) Expertise retrieval. Found Trends Inf Retr 6(23):127–256

Beutel A, Chi EH, Cheng Z, Pham H, Anderson J (2017) Beyond globally optimal: focused learning for improved recommendations. In: International conference on world wide web, pp 203–212

Beutel A, Covington P, Jain S, Xu C, Li J, Gatto V, Chi EH (2018) Latent cross: making use of context in recurrent recommender systems. In: International conference on web search and data mining, pp 46–54

Bifet A, Holmes G, Pfahringer B, Kirkby R, Gavaldà R (2009) New ensemble methods for evolving data streams. In: International conference on knowledge discovery and data mining, pp 139–148

Boeva V, Angelova M, Tsiporkova E (2017) Data-driven techniques for expert finding. In: International conference on agents and artificial intelligence, pp 535–542

Bordes A, Chopra S, Weston J (2014) Question answering with subgraph embeddings. arXiv preprint arXiv:1406.3676

Bouguessa M, Wang S (2008) Identifying authoritative actors in question-answering forums: the case of yahoo! answers. In: International conference on knowledge discovery and data mining, pp 866–874

Bozzon A, Brambilla M, Ceri S, Silvestri M, Vesci G (2013) Choosing the right crowd: expert finding in social networks. In: International conference on extending database technology, pp 637–648

Breiman L (1996) Bagging predictors. Mach Learn 26(2):123–140

Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: International conference on knowledge discovery and data mining, pp 785–794

Chen T, Zhang W, Lu Q, Chen K, Zheng Z, Yu Y (2012) SVDFeature: a toolkit for feature-based collaborative filtering. J Mach Learn Res 13:3619–3622

Cheng X, Zhu S, Chen G, Su S (2015) Exploiting user feedback for expert finding in community question answering. In: International conference on data mining, pp 295–302

Cheng HT, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M et al (2016) Wide and deep learning for recommender systems. In: Workshop on deep learning for recommender systems, pp 7–10

Christakopoulou E, Karypis G (2016) Local item-item models for top-n recommendation. In: ACM conference on recommender systems, pp 67–74

Covington P, Adams J, Sargin E (2016) Deep neural networks for youtube recommendations. In: Conference on recommender systems, pp 191–198

Dai H, Wang Y, Trivedi R, Song L (2016) Recurrent coevolutionary latent feature processes for continuous-time recommendation. In: Recsys workshop on deep learning for recommendation systems, pp 29–34

Dargahi Nobari A, Sotudeh Gharebagh S, Neshati M (2017) Skill translation models in expert finding. In: International ACM SIGIR conference on research and development in information retrieval, pp 1057–1060

Daud A, Li J, Zhou L, Muhammad F (2010) Temporal expert finding through generalized time topic modeling. Knowl Based Syst 23(6):615–625

Deng H, King I, Lyu MR (2009) Formal models for expert finding on dblp bibliography data. In: International conference on data mining, pp 163–172

Ebesu T, Shen B, Fang Y (2018) Collaborative memory network for recommendation systems. In: International ACM SIGIR conference on research and development in information retrieval

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

为了确保既定的模式是不需要培训，是CQA中专家发现的重要问题。我们希望本文提出的概述将推进CQA专家发现技术的讨论。

致谢这项工作得到了中国国家自然科学基金（61806111），国家自然科学基金（61806111）和中国高科技研发计划（863计划）（2015AA124102）的支持。

References

Adomavicius G，Tuzhilin A（2005）朝着下一代推荐系统：对最先进的和可能的扩展调查。 IEEE Trans Katabl Data Eng 6：734−749 Alarfaj F，Kruschwitz U，Hunter D，Fox
C（2012）找到了右主管：在大学域名的专家调查。 in：计算语言学协会，第1−6 Palog K，Fang Y，De Rijke M，Serdyukov P，Si L（2012）专业知识检索。发现趋势Inf Retr 6（23）：127−256
Beutel A，Chi Eh，Cheng Z，Pham H，Anderson
J（2017）超出全球最优：专注于改善建议的学习。在：世界范围内的国际会议，PP 203−212 Beutel A，Covington P，Jain S，XU C，Li J，Gatto V，Chi
EH（2018）潜在的十字：在经常性推荐系统中使用上下文。
in：Web搜索和数据挖掘国际会议，第46−54页Bifet A，Holmes G，Pfahringer B，Kirkby R，Gavald à R（2009）新的Ensemble方法，用于不断发展数据流。
in：国际知识发现和数据挖掘会议，第139−148页Boeva V，Angelova M，Tsiphova
E（2017）专家发现的数据驱动技术。在：国际代理商和人工智能国际会议，PP 535−542 BORDES A，Chopra S，Weston J（2014）问题用子图嵌入回答。 arxiv预印刷品

arXiv:1406.3676
Bouguessa M，Wang
S（2008）在问答论坛中识别权威行动者：雅虎的案例！答案。在：国际知识发现和数据挖掘会议，PP 866−874 Bozzon A，Brambilla M，Ceri S，Silvestri M，Vesci
G（2013）选择合适的人群：专家在社交网络中找到的专家。在：延长数据库技术国际会议，PP 637−648 Breiman L（1996）装袋预测因子。 MOCH学习26（2）：123−140 Chen T，Guestrin
C（2016）XGBoost：可伸缩的树升压系统。在：国际知识发现和数据挖掘会议，PP 785−794陈T，张W，Lu Q，Chen K，Zheng Z，Yu
Y（2012）SVDFeature：一种用于基于功能的协作过滤的工具包。 J Mach Learn Res 13：3619−3622
Cheng X，朱S，陈G，SU S（2015）利用社区问题的专家发现的用户反馈。
in：国际数据挖掘会议，第295−302页成本，康涅尔州，哈科，哈姆森j，sh sh t，chandra t，aradhye h，安德森g，conrado g，chai w，ispir
m等（2016）广泛和深度学习适用于推荐系统。在：建议系统的深度学习研讨会，PP 7−10
Crestikopoulou E，Karypis G（2016）本地物品 − 项目项目适用于TOP−N建议书。
in：ACM会议推荐系统，PP 67−74 Covington P，Adams J，Sargin
E（2016）为YouTube建议深神经网络。在：推荐人员会议，PP 191−198戴H，王Y，Trivedi R，Song L（2016）经常性的共同协调潜在功能，用于连续推荐。
in：Recsys关于建议系统的深度学习研讨会，第29−34 PP 29−34 Dargahi Nobari A，Sotudeh Gharebagh S，Neshati M（2017）技能翻译模型。 in：International ACM
Sigir关于信息检索的研究和开发会议，PP 1057−1060 Daud A，Li J，周L，Muhammad
F（2010）时间专家通过广义时间主题建模。基于Knowl的SYST
23（6）：615−625邓H，王，王，LYU先生（2009年）正式模型用于DBLP参考书目的专家查找。
in：数据挖掘国际会议，PP 163−172 EBESU T，Shen B，Fang
Y（2018）建议系统的协作记忆网络。 in：International ACM
Sigir关于信息检索弗里德曼的研发大会jh（2001）贪婪函数近似：梯度升压机。 ANN STAT 29（5）：1189−1232

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. Ann Stat 28(2):337–374

Gunawardana A, Shani G (2009) A survey of accuracy evaluation metrics of recommendation tasks. J Mach Learn Res 10(12):2935–2962

Han F, Tan S, Sun H, Srivatsa M, Cai D, Yan X (2016) Distributed representations of expertise. In: International conference on data mining, pp 531–539

Hashemi SH, Neshati M, Beigy H (2013) Expertise retrieval in bibliographic network: a topic dominance learning approach. In: International conference on information and knowledge management, pp 1117–1126

He X, Liao L, Zhang H, Nie L, Hu X, Chua TS (2017) Neural collaborative filtering. In: International conference on world wide web, pp 173–182

Hidasi B, Karatzoglou A, Baltrunas L, Tikk D (2015) Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939

Hu Y, Koren Y, Volinsky C (2008) Collaborative filtering for implicit feedback datasets. In: ieee international conference on data mining, pp 263–272

Jing H, Smola AJ (2017) Neural survival recommender. In: International conference on web search and data mining, pp 515–524

Joachims T (2006) Training linear SVMS in linear time. In: International conference on knowledge discovery and data mining, pp 217–226

Johnson RW (2001) An introduction to the bootstrap. Teach Stat 23(2):49C54

Karimzadehgan M, White RW, Richardson M (2009) Enhancing expert finding using organizational hierarchies. In: European conference on information retrieval, pp 177–188

Kawale J, Fu Y (2015) Deep collaborative filtering via marginalized denoising auto-encoder. In: International on conference on information and knowledge management, pp 811–820

Kingma DP, Ba J (2014) A method for stochastic optimization. In: International conference on learning representations, pp 1–15

Koren Y (2008) Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: International conference on knowledge discovery and data mining, pp 426–434

Koren Y, Bell R, Volinsky C et al (2009) Matrix factorization techniques for recommender systems. Computer 42(8):30–37

Lee J, Kim S, Lebanon G, Singer Y (2013) Local low-rank matrix approximation. In: International conference on machine learning, pp 82–90

Li Q, Zheng X (2017) Deep collaborative autoencoder for recommender systems: a unified framework for explicit and implicit feedback. arXiv preprint arXiv:1712.09043

Li X, Ma J, Yang Y, Wang D (2013) A service mode of expert finding in social network. In: International conference on service sciences, pp 220–223

Li H, Jin S, Shudong L (2015a) A hybrid model for experts finding in community question answering. In: International conference on cyber-enabled distributed computing and knowledge discovery, pp 176–185

Li X, Liu Y, Zhang M, Ma S, Zhu X, Sun J (2015b) Detecting promotion campaigns in community question answering. In: International joint conference on artificial intelligence, pp 2348–2354

Li Y, Ma S, Huang R (2015c) Social context analysis for topic-specific expert finding in online learning communities. Smart Learn Environ 5(1):57–74

Liang S, de Rijke M (2016) Formal language models for finding groups of experts. Inf Process Manag 52(4):529–549

Lin L, Xu Z, Ding Y, Liu X (2013) Finding topic-level experts in scholarly networks. Scientometrics 97(3):797–819

Lin S, Hong W, Wang D, Li T (2017) A survey on expert finding techniques. J Intell Inf Syst 49(2):255–279

Linden G, Smith B, York J (2003) Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Comput 7(1):76–80

Liu X, Koll M, Koll M (2005) Finding experts in community-based question-answering services. In: International conference on information and knowledge management, pp 315–316

Liu J, Song YI, Lin CY (2011) Competition-based user expertise score estimation. In: ACM SIGIR conference on research and development in information retrieval, pp 425–434

Liu DR, Chen YH, Kao WC, Wang HW (2013a) Integrating expert profile, reputation and link analysis for expert finding in question-answering websites. Inf Process Manag 49(1):312–329

Liu J, Qi LI, Liu B, Zhang Y (2013b) An expert finding method based on topic model. J Natl Univ Def Technol 35(2):127–131

Liu X, Ye S, Li X, Luo Y, Rao Y (2015) Zhihurank: a topic-sensitive expert finding algorithm in community question answering websites. In: International conference on web based learning, pp 165–173

872 S. Yuan et al.

弗里德曼J，Hastie T，Tibshirani r（2000）添加原始回归：升压的统计视图。ANN STAT 28（2）：337−374 Gunawardana A，Shani G（2009）推荐任务的准确性评估度量调查。J Mach Seart Res 10（12）：2935−2962韩F，Tan S，Sun H，Srivatsa M，Cai D，Yan X（2016）专业知识的分布式表示。in：国际数据挖掘国际会议，PP 531−539 Hashemi Sh，Neshati M，Beigy
H（2013）教专业知识网络中的专业知识：主导地位学习方法。：国际信息和知识管理会议，PP 1117−1126
Hex，Liaol，Zhangh，Niel，Hux，Chuats（2017）NeultCollaborativeFiltering.in：在万维网上，PP 173−182 Hidasi B，Karatzoglou A，Baltrunas L，TIKK
D（2015）基于会议的与经常性神经网络的建议。ARXIV预印亚克日期：1511.06939

胡y，koren y，Volinsky
C（2008）隐式反馈数据集的协同过滤。在：IEEE数据挖掘国际会议，PP 263−272 Jing H，Smola AJ（2017年）神经生存推荐。 in：Web搜索和数据挖掘国际会议，PP 515−524 Joachims T（2006）在线性时间训练线性SVM。
in：国际知识发现和数据挖掘会议，第217−226页，约翰逊rw（2001）介绍了引导。教学统计23（2）：49C54 Karimzadehgan M，White RW，Richardson
M（2009）使用组织Hierar−chies增强专家寻找。在：欧洲信息检索会议，第177−188页，PP 177−188 KAWALE J，FU Y（2015）通过边缘化去噪自动编码器深度协作滤波。
in：国际关于信息和知识管理会议，PP 811−820 Kingma DP，BA
J（2014）一种用于随机优化的方法。在：学习陈述国际会议，PP
1−15克伦Y（2008）分解符合附近：多方面的协作滤波模型。
in：全国知识发现会议和数据挖掘，PP 426−434 Koren Y，Bell R，Volinsky
C等人（2009）推荐系统的矩阵分解技术。电脑42（8）：30−37 Lee J，Kim S，黎巴嫩G，歌手Y（2013）本地低秩矩阵近似。 in：International Machine Learning会议，PP 82−90 Li Q，Zheng
X（2017）深度协作AutoEncoder为推荐系统：一个统一的明确反馈框架。
Arxiv预印亚克日期：1712.09043

李X，MA J，Yang Y，Wang D（2013）社交网络专家的服务模式。 in：国际服务科学会议，PP 220−223李H，金斯，尚荣L（2015A）一个用于社区问题的专家的混合模型。
in：关于网络的分布式计算和知识发现的国际会议，第176−185 Pi X，Liu Y，Zhang M，Ma S，Zhu X，Sun J（2015B）检测社区问题的促销活动。在：国际人工智能联席会议，第2348−2354 PP 2348−2354 Li Y，Huang
R（2015C）在线学习社区中专门专题专家的社会背景分析。智能学习环境5（1）：57−74梁S，De Rijke M（2016）查找专家组的正式语言模型。 inf Process Monm 52（4）：529−549 Lin L，Xu Z，Ding Y，Liu X（2013）在学术网络中寻找主题专家。科技97（3）：797−819林S，洪W，Wang D，Li T（2017）专家发现技术调查。 J INTEM SYST 49（2）：255−279 Linden G，史密斯B，York J（2003）Amazon.com建议：项目到项目协同过滤。 IEEE Internet Comput 7（1）：76−80 Liu X，Koll M，Koll M（2005）在社区的问答服务中寻找专家。
在信息和知识管理中的国际会议，第315−316 PP
315−316刘继，宋毅，林CY（2011）基于竞争的用户专业知识分数估计。 in：ACM Sigir关于信息检索的研究和开发会议，PP
425−434刘博士，陈毅，高WC，王HW（2013A）集成专家概况，声誉和联系分析，以了解答案答案网站。 inf Process Monm
49（1）：312−329刘家，齐丽，刘B，张y（2013b）基于主题模型的专家发现方法。 J Natl Univ Def Technol 35（2）：127−131 Liu X，Ye S，Li X，Luo Y，Rao
Y（2015）Zhihurank：一个主题敏感的专家查找算法在社区问题应答网站中。在：基于Web的学习国际会议，PP 165−173

Springer

Mimno D, Mccallum A (2007) Expertise modeling for matching papers with reviewers. In: International conference on knowledge discovery and data mining, pp 500–509
Momtazi S, Naumann F (2013) Topic modeling for expert finding using latent Dirichlet allocation. Wiley Interdiscip Rev Data Min Knowl Discov 3(5):346C353
Neshati M, Fallahnejad Z, Beigy H (2017) On dynamicity of expert finding in community question answering. Inf Process Manag 53(5):1026–1042
Paatero P, Tapper U (1994) Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2):111–126
Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: online learning of social representations. In: International conference on knowledge discovery and data mining, pp 701–710
Pezzotti N, Lelieveldt B, Maaten LVD, Hollt T, Eisemann E, Vilanova A (2017) Approximated and user steerable tsne for progressive visual analytics. IEEE Trans Vis Comput Graph 23(7):1739–1752
Qian Y, Tang J, Wu K (2018) Weakly learning to match experts in online community. In: International joint conference on artificial intelligence, pp 3841–3847
Qiu X, Huang X (2015) Convolutional neural tensor network architecture for community-based question answering. In: International joint conference on artificial intelligence, pp 1305–1311
Rani SK, Raju K, Kumari VV (2015) Expert finding system using latent effort ranking in academic social networks. Int J Inf Technol Comput Sci 7(2):21–27
Rendle S (2011) Factorization machines. In: International conference on data mining, pp 995–1000
Rendle S (2012) Factorization machines with libfm. Trans Intell Syst Technol 3(57):1–22
Riahi F, Zolaktaf Z, Shafiei M, Milios E (2012) Finding expert users in community question answering. In: International conference on world wide web, pp 791–798
Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: International conference on machine learning, pp 791–798
Sedhain S, Menon AK, Sanner S, Xie L (2015) Autorec: autoencoders meet collaborative filtering. In: International conference on world wide web, pp 111–112
Tan YK, Xu X, Liu Y (2016) Improved recurrent neural networks for session-based recommendations. In: Workshop on deep learning for recommender systems, pp 17–22
Vincent P, Larochelle H, Bengio Y, Manzagol P (2008) Extracting and composing robust features with denoising autoencoders. In: International conference on machine learning, p 1096–1103
Wan S, Lan Y, Guo J, Xu J, Pang L, Cheng X (2016) Match-srnn: modeling the recursive matching structure with spatial rnn. In: International joint conference on artificial intelligence, pp 2922–2928
Wang GA, Jiao J, Abrahams AS, Fan W, Zhang Z (2013) Expertrank: a topic-aware expert finding algorithm for online knowledge communities. Decis Support Syst 54(3):1442–1451
Wei J, He J, Chen K, Zhou Y, Tang Z (2017) Collaborative filtering and deep learning based recommendation system for cold start items. Expert Syst Appl 69:29–39
Wu Y, DuBois C, Zheng AX, Ester M (2016) Collaborative denoising auto-encoders for top-n recommender systems. In: ACM international conference on web search and data mining, pp 153–162
Wu CY, Ahmed A, Beutel A, Smola AJ, Jing H (2017) Recurrent recommender networks. In: International conference on web search and data mining, pp 495–503
Yang L, Qiu M, Gottipati S, Zhu F, Jiang J (2013) Cqarank: jointly model topics and expertise in community question answering. In: International conference on information and knowledge management, pp 99–108
Yeniterzi R, Callan J (2014) Constructing effective and efficient topic-specific authority networks for expert finding in social media. In: International workshop on social media retrieval and analysis, pp 45–50
Ying H, Chen L, Xiong Y, Wu J (2016) Collaborative deep ranking: a hybrid pair-wise recommendation algorithm with implicit feedback. In: Pacific-asia conference on knowledge discovery and data mining, pp 555–567
Zhang S, Yao L, Xu X (2017) Autosvd++: an efficient hybrid collaborative filtering model via contractive auto-encoders. In: SIGIR conference on research and development in information retrieval, pp 957–960
Zhao T, Bian N, Li C, Li M (2013) Topic-level expert modeling in community question answering. In: International conference on data mining, pp 776–784
Zhao Z, Wei F, Zhou M, Ng W (2015a) Cold-start expert finding in community question answering via graph regularization. In: International conference on database systems for advanced applications, pp 21–38
Zhao Z, Zhang L, He X, Ng W (2015b) Expert finding for question answering via graph regularized matrix completion. IEEE Trans Knowl Data Eng 27(4):993–1004
Zhao Z, Yang Q, Cai D, He X, Zhuang Y (2016) Expert finding for community-based question answering via ranking metric network learning. In: International joint conference on artificial intelligence, pp 3000–3006
Zheng Y, Tang B, Ding W, Zhou H (2016) A neural autoregressive approach to collaborative filtering. arXiv preprint arXiv:1605.09477

Mimno D, Mccallum A (2007) Expertise modeling for matching papers with reviewers. In: International conference on knowledge discovery and data mining, pp 500–509
Momtazi S, Naumann F (2013) Topic modeling for expert finding using latent Dirichlet allocation. Wiley Interdiscip Rev Data Min Knowl Discov 3(5):346C353
Neshati M, Fallahnejad Z, Beigy H (2017) On dynamicity of expert finding in community question answering. Inf Process Manag 53(5):1026–1042
Paatero P, Tapper U (1994) Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values. Environmetrics 5(2):111–126
Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: online learning of social representations. In: International conference on knowledge discovery and data mining, pp 701–710
Pezzotti N, Lelieveldt B, Maaten LVD, Hollt T, Eisemann E, Vilanova A (2017) Approximated and user steerable tsne for progressive visual analytics. IEEE Trans Vis Comput Graph 23(7):1739–1752
Qian Y, Tang J, Wu K (2018) Weakly learning to match experts in online community. In: International joint conference on artificial intelligence, pp 3841–3847
Qiu X, Huang X (2015) Convolutional neural tensor network architecture for community-based question answering. In: International joint conference on artificial intelligence, pp 1305–1311
Rani SK, Raju K, Kumari VV (2015) Expert finding system using latent effort ranking in academic social networks. Int J Inf Technol Comput Sci 7(2):21–27
Rendle S (2011) Factorization machines. In: International conference on data mining, pp 995–1000
Rendle S (2012) Factorization machines with libfm. Trans Intell Syst Technol 3(57):1–22
Riahi F, Zolaktaf Z, Shafiei M, Milios E (2012) Finding expert users in community question answering. In: International conference on world wide web, pp 791–798
Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: International conference on machine learning, pp 791–798
Sedhain S, Menon AK, Sanner S, Xie L (2015) Autorec: autoencoders meet collaborative filtering. In: International conference on world wide web, pp 111–112
Tan YK, Xu X, Liu Y (2016) Improved recurrent neural networks for session-based recommendations. In: Workshop on deep learning for recommender systems, pp 17–22
Vincent P, Larochelle H, Bengio Y, Manzagol P (2008) Extracting and composing robust features with denoising autoencoders. In: International conference on machine learning, p 1096–1103
Wan S, Lan Y, Guo J, Xu J, Pang L, Cheng X (2016) Match-srnn: modeling the recursive matching structure with spatial rnn. In: International joint conference on artificial intelligence, pp 2922–2928
Wang GA, Jiao J, Abrahams AS, Fan W, Zhang Z (2013) Expertrank: a topic-aware expert finding algorithm for online knowledge communities. Decis Support Syst 54(3):1442–1451
Wei J, He J, Chen K, Zhou Y, Tang Z (2017) Collaborative filtering and deep learning based recommendation system for cold start items. Expert Syst Appl 69:29–39
Wu Y, DuBois C, Zheng AX, Ester M (2016) Collaborative denoising auto-encoders for top-n recommender systems. In: ACM international conference on web search and data mining, pp 153–162
Wu CY, Ahmed A, Beutel A, Smola AJ, Jing H (2017) Recurrent recommender networks. In: International conference on web search and data mining, pp 495–503
Yang L, Qiu M, Gottipati S, Zhu F, Jiang J (2013) Cqarank: jointly model topics and expertise in community question answering. In: International conference on information and knowledge management, pp 99–108
Yeniterzi R, Callan J (2014) Constructing effective and efficient topic-specific authority networks for expert finding in social media. In: International workshop on social media retrieval and analysis, pp 45–50
Ying H, Chen L, Xiong Y, Wu J (2016) Collaborative deep ranking: a hybrid pair-wise recommendation algorithm with implicit feedback. In: Pacific-asia conference on knowledge discovery and data mining, pp 555–567
Zhang S, Yao L, Xu X (2017) Autosvd++: an efficient hybrid collaborative filtering model via contractive auto-encoders. In: SIGIR conference on research and development in information retrieval, pp 957–960
Zhao T, Bian N, Li C, Li M (2013) Topic-level expert modeling in community question answering. In: International conference on data mining, pp 776–784
Zhao Z, Wei F, Zhou M, Ng W (2015a) Cold-start expert finding in community question answering via graph regularization. In: International conference on database systems for advanced applications, pp 21–38
Zhao Z, Zhang L, He X, Ng W (2015b) Expert finding for question answering via graph regularized matrix completion. IEEE Trans Knowl Data Eng 27(4):993–1004
Zhao Z, Yang Q, Cai D, He X, Zhuang Y (2016) Expert finding for community-based question answering via ranking metric network learning. In: International joint conference on artificial intelligence, pp 3000–3006
Zheng Y, Tang B, Ding W, Zhou H (2016) A neural autoregressive approach to collaborative filtering. arXiv preprint arXiv:1605.09477

Springer

Zhou G, Lai S, Liu K, Zhao J (2012) Topic-sensitive probabilistic model for expert finding in question answer communities. In: International conference on information and knowledge management, pp 1662–1666

Zhou G, Zhao J, He T, Wu W (2014) An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities. Knowl Based Syst 66(9):136–145

Zhu H, Chen E, Xiong H, Cao H, Tian J (2014) Ranking user authority with relevant knowledge categories for expert finding. World Wide Web 17(5):1081–1107

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sha Yuan[1] · Yu Zhang[2] · Jie Tang[1] · Wendy Hall[4] · Juan Bautista Cabotà[3]

Sha Yuan
yuansha@mail.tsinghua.edu.cn

Yu Zhang
zhang.yu@imicams.ac.cn

Wendy Hall
wh@ecs.soton.ac.uk

Juan Bautista Cabotà
jcabota@gmail.com

1 Knowledge Engineering Lab, Department of Computer Science and Technology, Tsinghua University, Beijing, China

2 Institute of Medical Information, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China

3 Computer Science Department, University of Valencia, Valencia, Spain

4 Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

874 S. Yuan et al.

周G，Lai S，Liu K，Zhao
J（2012）关于问题答案社区专家查找的主题敏感概率模型。in：国际信息和知识管理会议，PP
1662–1666周G，赵家，何T，吴W（2014）关于问题答案社区专家查找主题敏感概率模型的实证研究。基于Knowl的SYST 66（9）：136–145 Zhu H，Chen E，Xiong H，Cao H，Tian
J（2014）与相关知识类别的用户权限排名为专家发现。万维网17（5）：1081–1107

出版商的说明Springer
Nature仍然是关于发表地图和机构附属机构的司法管辖权索赔中立。

Affiliations

Sha Yuan[1] · Yu Zhang[2] · Jie Tang[1] · Wendy Hall[4] · Juan Bautista Cabotà[3]

Sha Yuan
yuansha@mail.tsinghua.edu.cn

Yu Zhang
zhang.yu@imicams.ac.cn

Wendy Hall
wh@ecs.soton.ac.uk

Juan Bautista Cabotà
jcabota@gmail.com

1个知识工程实验室，清华大学计算机科学与技术系，北京

2中国医学科学院北京联盟医学院医学信息研究所，北京

3，巴伦西亚大学，巴伦西亚大学，西班牙

4电子与计算机科学，南安普敦大学，英国南安普顿