Wang X, Huang C, Yao L *et al.* A survey on expert recommendation in community question answering. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 33(4): 625–653 July 2018. DOI 10.1007/s11390-018-1845-0

A Survey on Expert Recommendation in Community Question Answering

Xianzhi Wang[1], *Member, ACM, IEEE*, Chaoran Huang[2], *Student Member, ACM, IEEE*
Lina Yao[2], *Member, ACM, IEEE*, Boualem Benatallah[2], *Member, IEEE*
and Manqing Dong[2], *Student Member, ACM, IEEE*

[1]*School of Software, University of Technology Sydney, Sydney, NSW 2007, Australia*
[2]*School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia*

E-mail: sandyawang@gmail.com; {chaoran.huang, lina.yao, b.benatallah, manqing.dong}@unsw.edu.au

Received January 21, 2018; revised June 11, 2018.

Abstract Community question answering (CQA) represents the type of Web applications where people can exchange knowledge via asking and answering questions. One significant challenge of most real-world CQA systems is the lack of effective matching between questions and the potential good answerers, which adversely affects the efficient knowledge acquisition and circulation. On the one hand, a requester might experience many low-quality answers without receiving a quality response in a brief time; on the other hand, an answerer might face numerous new questions without being able to identify the questions of interest quickly. Under this situation, expert recommendation emerges as a promising technique to address the above issues. Instead of passively waiting for users to browse and find their questions of interest, an expert recommendation method raises the attention of users to the appropriate questions actively and promptly. The past few years have witnessed considerable efforts that address the expert recommendation problem from different perspectives. These methods all have their issues that need to be resolved before the advantages of expert recommendation can be fully embraced. In this survey, we first present an overview of the research efforts and state-of-the-art techniques for the expert recommendation in CQA. We next summarize and compare the existing methods concerning their advantages and shortcomings, followed by discussing the open issues and future research directions.

Keywords community question answering, expert recommendation, challenge, solution, future direction

1 Introduction

The prosperity of crowdsourcing and web 2.0 has fostered numerous online communities featuring question answering (Q&A) activities. Such communities exist in various forms such as dedicated websites, online forums, and discussion boards. They provide a venue for people to share and obtain knowledge by asking and answering questions, known as community question answering (CQA)[1]. While traditional online information seeking approaches (e.g., search engines) retrieve information from existing information repositories based on keywords, they face several challenges. First, answers to some questions may not exist in the previously answered questions[2] and thus cannot be retrieved from existing repositories directly. Second, most real-world questions are written in complicated natural languages that require certain human intelligence to be understood. Third, some questions inherently seek people's opinions and can only be answered by humans. While machines find it difficult to handle the above cases, CQA can leverage the "wisdom of crowds" and obtain answers from multiple people simultaneously. Typical Q&A websites include Yahoo! Answers (answers.yahoo.com), Quora (www.quora.com), and Stack Overflow (stackoverflow.com). The first two websites cover a wide range of topics, while the last only focuses on the topic of computer programming.

Though there are advantages over the traditional information seeking approaches, CQA faces several unique challenges. First, a CQA website may have tens of thousands of questions posed every day, let alone millions of questions that have already existed on the web-

Survey
Special Section on Recommender Systems with Big Data

site. The huge volume of questions makes it difficult for a general answerer to find the appropriate questions to answer[3]. Second, answerers usually have varying interest and expertise in different topics and knowledge domains. Thus, they may give answers of varying quality to different questions. The time required for preparing answers[4] and the intention of answering also affect the quality of their responses. An extreme case is that answerers may give irrelevant answers that distract other users[5] without serious thinking. All the above situations cause additional efforts of an information seeker in obtaining good answers. Third, instead of receiving an answer instantly, users in CQA may need to wait a long time until a satisfactory answer appears. Previous studies[6] show that many questions on real-world CQA websites cannot be resolved adequately, meaning the requesters do not recognize the best answers to their questions within 24 hours.

Fortunately, several studies[7-9] have shown that some core answerers are the primary drivers of answer production in many communities. Recent work on Stack Overflow and Quora[10] further indicates that these sites consist of a set of highly dedicated domain experts who aim at satisfying requesters' queries but more importantly at providing answers with high lasting value to a broader audience. All these studies suggest the needs for recommending a small group of most competent answerers, or experts to answer the new questions. In fact, the long-tail phenomena in many real-world communities, from the statistic perspective, lay the ground of the rationale of expert recommendation in CQA[11], as most answers and knowledge in the communities come from only a minority of users[11-12]. As an effective means of addressing the practical challenges of traditional information seeking approaches, expert recommendation methods bring up the attention of only a small number of experts, i.e., the users who are most likely to provide high-quality answers, to answer a given question[13]. Since expert recommendation inherently encourages fast acquisition of higher-quality answers, it potentially increases the participation rates of users, improves the visibility of experts, and fosters stronger communities in CQA.

Given the advantages of expert recommendation and related topics such as question routing[6,14] and question recommendation[15] in the domains of natural language processing (NLP) and information retrieval (IR), we aim to present a comprehensive survey on the expert recommendation in CQA. On the one hand, considerable efforts have been conducted on the expert recommendation and have delivered fruitful results. Therefore, it is necessary to review the related methods and techniques to gain a timely and better understanding of the state of the art. On the other hand, despite the active research in CQA, expert recommendation remains a challenging task. For example, the sparsity of historical question and answer records, low participation rates of users, lack of personalization in recommendation results, the migration of users in or out of communities, and lack of comprehensive consideration of different clues in modeling users expertise are all regarded as challenging issues in literature. Given the diverse existing methods, it is crucial to develop a general framework to evaluate these methods and analyze their shortcomings, as well as to point out promising future research directions.

To the best of our knowledge, this is the first comprehensive survey that focuses on the expert recommendation issue in CQA. The remainder of the article is organized as follows. We overview the expert recommendation problem in Section 2 and its current applications in CQA in Section 3. In Section 4, we present the classification and introduction of state-of-the-art expert recommendation methods. In Section 5, we compare the investigated expert recommendation methods on various aspects and discuss their advantages and pitfalls. In Section 6, we highlight several promising research directions. Finally, we offer some concluding remarks in Section 7.

2 Expert Recommendation Problem

The expert recommendation issue is also known as the question routing or expert finding problem. The basic inputs of an expert recommendation problem include users (i.e., requesters and answerers) and user-generated content (i.e., the questions raised by requesters and the answers provided by answerers). More inputs might be available depending on the application scenarios. Typically, they include user profiles (e.g., badges, reputation scores, and links to external resources such as Web pages), users' feedback on questions and answers (e.g., textual comments and votings), and question details (e.g., the categories of questions and duplication relations among questions). The relationship among different types of inputs of an expert recommendation problem is described in the class diagram shown in Fig.1.

Question answering websites usually organize information in the form of threads. Each thread is led by a

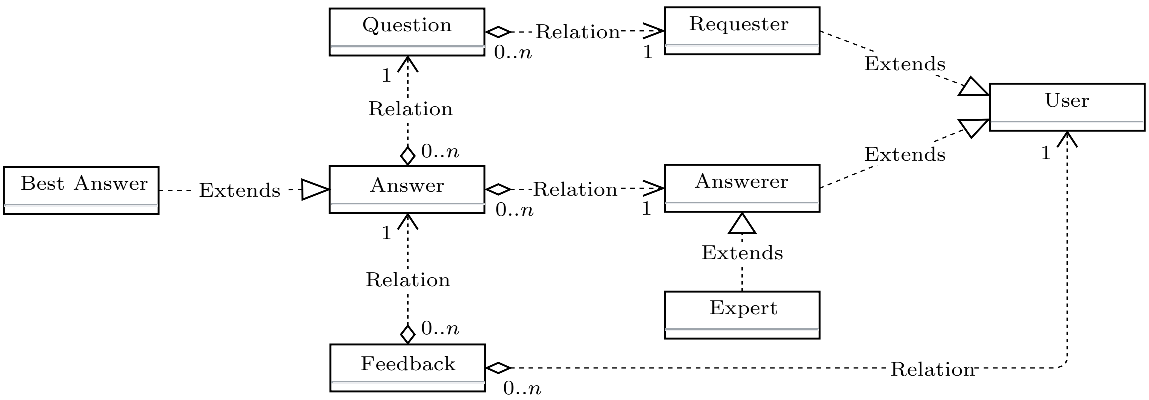

Fig.1. Elements of expert recommendation in CQA.

single question, which is replied to with none, one, or multiple answers. Each question or answer is provided by a single user, called a requester or an answerer, respectively. A requester may ask multiple questions, and each answerer may answer various questions. A user can be either a requester or an answerer, or both at the same time in the same CQA website, and all users are free to provide different types of feedback on the posted questions and answers. For example, in Stack Overflow, any registered user can comment and vote (by giving a thumb up or thumb down) on an answer posted for any question, and the requester has the authority to mark one from the posted answers as the best answer. In case that the requester has not designated the best answer within a specified period, the system will automatically mark the response that received the highest voting score as the best answer.

The objective of the expert recommendation problem is to raise the attention of experts, i.e., a small number of users who are most likely to provide high-quality answers, to the given question based on the above problem inputs. Despite various possible types of inputs, only a subset of them might be available in a specific application scenario. Therefore, researchers may define the expert recommendation problem differently according to the inputs. Besides, researchers may take into account different concerns and expect different types of outputs from their methods. Generally, topical relevance and expertise are the two most considered aspects of concerns by the existing research. While some researchers develop methods to find a group of high-quality answerers, other researchers aim to deliver a ranked list, where the users are ranked according to their potential to provide the best answer. We will elaborate the variations in the problem definition in Section 5.

Generally, it is only necessary to recommend experts when the new question is significantly different from any previous questions that have the best answer, meaning that no satisfactory answers are readily available within the archive of best answers to the earlier questions. Expert recommendation generally brings about the following advantages to CQA: 1) users usually prefer answers from experts, who are supposed to have sufficient motivation and knowledge to answer the given questions and therefore more likely to provide high-quality answers promptly; 2) expert recommendations can potentially reduce the waiting time of requesters in finding satisfactory answers as well as the time of experts in finding their questions of interests; 3) by bridging the gap between requesters and answerers, expert recommendations can potentially promote their participation rates and thus foster stronger communities. Since experts are recommended with questions that fit their expertise, their visibility is expected to be improved as well.

3 Current Applications in CQA

Currently, there exist various Q&A websites where expert recommendation techniques are applied or can be potentially applied. Due to the large number of Q&A websites that exist nowadays, we selectively list some typical Q&A websites by launch year in Table 1. In the following subsections, we will categorize and give further illustrations of several typical websites of each category.

3.1 Early CQA Services

Most early-stage Q&A services (e.g., the first four communities in Table 1) meet a requester's information needs by resorting to the opinions of experts rather than

Table 1. Some Popular Question Answering Communities

Community	Language	Specialized Domain	Launch Year	Still Active	Quality Guarantee
MedHelp	English	Medical	1994	Y	Y
Mad Scientist Netwok	English	Various	1995	Y	Y
WebMD	English	Medical	1996	Y	Y
Google Answers	Multiple	Various	2002	N	Y
Naver KiN	Korean	Various	2002	Y	N
WikiAnswers	English	Various	2002	Y	N
Answerbag	English	Various	2003	Y	N
IAsk	Chinese	Various	2005	Y	N
Baidu Knows	Chinese	Various	2005	Y	N
Live QnA	English	Various	2006	N	N
TurboTax Live Community	English	Tax	2007	Y	N
Sogou Wenwen	Chinese	Various	2007	Y	N
Stack Overflow	English	Programming	2008	Y	N
Quora	English	Various	2010	Y	N
Seasoned Advice	English	Cooking	2010	Y	N

the crowd. These experts are acknowledged by either the websites or third-party authorities and are often limited in number. They usually have rich knowledge and experience in some domains but require a payment for the answers they provide. We introduce two of these websites as examples as follows.

*Mad Scientist Network*①: a famous ask-a-scientist web service where people ask questions by filling forms and moderators take the responsibility of reviewing the questions and sending them to the appropriate members for answers. The moderators will also review the answers before making them public.

*Google Answers*②: a knowledge market service designed as an extension to Google's search service. There were a group of answerers called Google Answers Researchers who are officially approved to answer questions through an application process. Instead of passively waiting for other people to moderate or answer their questions, people can actively find the potential answerers by themselves and pay the answerers.

3.2 General-Purpose CQA Websites

The Q&A services that emerge in the past two decades are increasingly leveraging the "wisdom of the crowd" rather than a small number of experts to give answers. Websites following this philosophy allow any users to voluntarily answer any questions on their free will and most of them serve as general purpose platforms for knowledge sharing rather than domain focused ones. We overview some typical general purpose websites as follows.

Quora: one of the largest existing Q&A website where users can ask and answer questions, and rate and edit the answers posted by others.

*Zhihu*③: a Chinese Q&A website similar to Quora. It allows users to create and edit questions and answers, rate system, and tag questions. Also, users may also post blogs in Zhihu for sharing while others can view and comment on such posts.

*Naver KiN*④: a Korean CQA community, one of the earlier cases of expansion of search service using user-generated content.

*WikiAnswers*⑤: a wiki service that allows people to raise and answer questions, and edit existing answers to questions. It uses a so-called "alternates system" to automatically merge similar questions. Since an answer may be associated with multiple questions, duplicated entries can be avoided to some extent.

*Answerbag*⑥: a CQA community where users can ask and answer questions, give comments to answers,

①http://www.madsci.org/, May 2018.
②http://answers.google.com/, May 2018.
③http://www.zhihu.com/, May 2018.
④http://kin.naver.com/, May 2018.
⑤http://www.wikianswers.com/, May 2018.
⑥http://www.answerbag.com/, May 2018.

rate questions, rate answers, and suggest new categories.

*Live QnA*⑦: also known as MSN QnA, is part of Microsoft MSN group services. In this system, users can ask and answer questions, tag them to specific topics, and gain points and reputations by answering questions.

3.3 Domain-Focused CQA Websites

Compared with those general purpose websites, each domain-focused Q&A website only covers limited topics or knowledge domains. The Stack Exchange networks are probably the largest host of domain-focused Q&A websites nowadays. Some typical websites hosted by it include the followings.

*MathOverflow*⑧: a Q&A website focused on mathematical problems.

*AskUbuntu*⑨: a website supporting Q&A activities related to Ubuntu operation systems.

StackOverflow: a Q&A website focused on computer programming.

All these websites follow similar sets of styles and functions. Apart from the basic question answering features, they commonly use badges to recognize the achievement of answerers and grant badges to users based on their reputation points. Users can also unlock more privileges with higher reputation points.

3.4 Summary

In summary, despite the prevalence of diverse types of Q&A websites, few of them have incorporated any effective expert recommendation techniques to bridge requesters and answers. To the best of our knowledge, currently, the only implementation of the idea of routing questions to the appropriate users in Q&A is called "Aardvark"[16]. However, the primary purpose of this system is to serve as an enhanced search engine, and the expert recommendation techniques it employs are still at a preliminary stage. Recently, Bayati[17] designed a framework for recommending security experts for software engineering projects. This framework offers more strength to facilitate expert recommendation by considering multiple aspects of users such as programming language, location, and social profiles on dominant programming Q&A websites like StackOverflow. Since the Q&A systems can be regarded as a type of crowdsourcing systems[18], the expert recommendation methods for a Q&A system can potentially be generalized and applied to general crowdsourcing systems as well.

4 Expert Recommendation Methods

As the major technique to facilitate effective CQA, considerable efforts have been contributed to the expert recommendation research from the information retrieval (IR), machine learning, and social computing perspectives, and have delivered fruitful results. We classify the state-of-the-art expert recommendation methods into eight categories and review the methods by category in the following subsections.

4.1 Simple Methods

One of the most critical tasks of expert recommendation is to evaluate users. Given a new question to be answered, some methods use simple metrics such as counts of positive/negative votes, proportions of best answers, and the similarity between the new question and users' previous answered questions to evaluate users' fitness to answer the questions. In the following, we introduce the methods that use the three metrics, respectively. For any of these methods, a higher score indicates a better answerer.

Votes. The method evaluates a user by the number of affirmative votes minus the number of negative votes, combined with the total percentage of affirmative votes that the user receives from other users averaged over all the answers the user has attempted.

Best Answer Proportion. This method ranks users by the fraction of the best answers among all the answers attempted by an answerer. The best answers are either awarded by the requester of questions or by the question answering platform when requesters designate no best answers.

Textual Similarity. The most famous method for measuring textual similarity is to compute the cosine similarity based on the term frequency-inverse document frequency (TF-IDF) model, a classic vector space model (VSM)[19] borrowed from the information retrieval domain. VSM is readily applicable to computing the similarity of an answerer's profile to a given question. Therefore, it can be directly used for the expert

⑦http://qna.live.com/, May 2018.
⑧http://mathoverflow.net/, May 2018.
⑨http://askubuntu.com/, May 2018.

recommendation by relating a new question to the answerers who have previously answered the most relevant questions to the given question.

4.2 Language Models

Despite the simplicity, VSM adopts the "bag-of-words" assumption and thus brings the high-dimension document representation issue. In contrast, language models use a generative approach to compute the word-based relevance of a user's previous activities to the given question, and in turn, to predict the possibility of a user answering the question. Such models can, to some extent, alleviate the high dimension issue. In a language model, the users whose profiles are most likely to generate the given question are believed to have to the highest probability to answer the given question. The model finally returns a ranked list of users according to their likelihood of answering the given question.

The language model-based methods include profile-based methods and document-based methods. The former[20] models the knowledge of each user with the associated documents and ranks the candidate experts for a given topic based on the relevance scores between their profiles and the given question. The latter[20] finds related documents for a given topic and ranks candidates based on mentions of the candidates in the related documents.

4.2.1 QLL and Basic Variants

Among the methods of this category, query likelihood language (QLL) model[21] is the most popular technique. QLL calculates a probability that user profiles will generate terms of the routed question. The traditional language models often suffer the mismatch between the question and user profiles caused by the co-occurrence of random words in user profiles or questions resulting from data sparseness. Translation models[22] overcome data sparseness by employing statistical machine translation and can differentiate between exact matched words and translated semantically related ones. A typical work[23] using this method views the problem as an IR problem. It considers the new question as a query and the expert profiles as documents. It next estimates an answerer's expertise by combining its previously answered questions, and regards experts as the users who answered the most similar questions in the past.

Besides the basic models, many variants of QLL have also emerged as alternatives or enhancements. For example, Lavrenko *et al.* proposed two variants of the basic language model, namely relevance-based language model[24] and cluster-based language model[25] to rank user profiles. Petkova and Croft[26] proposed a hierarchical language model which uses a finer-grained approach with a linear combination of the language models built on subcollections of documents.

4.2.2 Category-Sensitive QLL

Considering the availability of categories in many Q&A websites, Li *et al.*[27] proposed a category-sensitive QLL model to exploit the hierarchical category information presented with questions in Yahoo! Answers. Once a question gets categorized, the task is to find the users who are most likely to answer that question within its category. Their experiments over the Yahoo! Answers dataset show that taking categories into account improves the recommendation performance. A limitation of the category-sensitive model is that categories need to be well predefined and some questions might be closely related to multiple categories due to the existence of similar categories that share the same contexts. A possible solution to address this limitation is the transferred category-sensitive QLL model[27], which additionally builds and considers the relevance between categories.

4.2.3 Expertise-Aware QLL

Zheng *et al.*[28] linearly combined two aspects, user relevance (computed based on QLL) and answer quality (estimated using a maximum entropy model), using the simple weighted sum method to represent user expertise on a given question. Besides the relevance and quality aspects, Li and King[6] further considered the availability of users and used the weighted sum of the three aspects to represent user expertise on a given question. In particular, the relevance is estimated using the QLL model, the answer quality is estimated as the weighted average of previous answer quality incorporated with the Jelinek-Mercer smoothing[29] method, and users' availability to answer a given question during a given period is predicted by an autoregressive model. Compared with most existing methods, this method exploits not only time series availability information of users but also multiple metadata features such as answer length, question-answer length, the number of answers for this question, the answerer's total points, and the answerer's best answer ratio. These features have rarely been used by the existing research.

4.3 Topic Models

Since language models are based on exact word matching, they are most effective when they are used within the same topic. Besides, they are not able to capture more advanced semantics and solve the problem of the lexical gap between a question and user profiles. In contrast, topic models do not require the word to appear in the user profile, as it measures their relationship in the topic space rather than in the word space. It can, therefore, alleviate the lexical gap problem and previous experimental evaluations have confirmed the better performance of many topic models over language models[30-31]. Here, we focus on reviewing two most widely used topic models, Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA), as well as their variants and a few other models.

4.3.1 PLSA and Its Variants

Probabilistic Latent Semantic Analysis (PLSA) a.k.a. Probabilistic Latent Semantic Indexing (PLSI)[32] is developed based on Latent Semantic Indexing (LSI)[33], which uses Singular Value Decomposition to represent a document in a low-dimension space. Compared with LSI, which lacks semantic explanation, PLSA uses latent topics to represent documents and model the data generation process as a Bayesian network. In this way, it can leverage the semantic between words in documents to reduce the document representation space dimension. There are generally two classes of PLSA-based methods that model users directly and indirectly, respectively. We briefly review the two classes of methods as follows.

Direct User Model by PLSA. Methods of this class treat all the questions that a user accesses as one document. Then, PLSA is used directly to derive the topic information of the user using word distributions. A typical method of this class[15] would identify the underlying topics of questions to match users' interest and thereby help the capable users locate the right questions to answer. The expectation maximization (EM) algorithm is generally used to find a local maximum of the log-likelihood of the question collection and to learn model parameters.

Indirect User Model by PLSA. A typical method of this class is proposed in [34]. This work presents an incremental automatic expert recommendation framework based on PLSA. It considers both users' interests and feedback and takes questions as documents. It further gains a question's distribution on topics based on PLSA, followed by modelling a user using the average of topic distributions of all the questions that the user has previously accessed to facilitate recommendation.

A most important variant of PLSA is probably the Dual Role Model (DRM) proposed by Xu *et al.*[35] Instead of combining the consideration of a user as a requester and an answerer, DRM separately models users' roles as requesters and as answerers and derives the corresponding probabilistic models based on PLSA. Depending on the modeling approach of a user's role, DRM diverges into independent DRM, a type of method modeling user role indirectly, and dependent DRM, a method which learns the role model directly. In particular, the independent DRM assumes all users are independent of each other and models each user individually. In contrast, dependent DRM considers the dependence between users. Besides modeling users' topic distribution as requesters and answerers, it additionally models the relationship between answerers and requesters for better performance.

4.3.2 LDA and Its Variants

The Latent Dirichlet Allocation (LDA) model[36] is probably the most widely used topic model among all existing topic models developed. In LDA, the topic mixture is drawn from a conjugate Dirichlet prior that remains the same for all users. More specifically, LDA assumes a certain generative process for data. To generate a user profile, LDA assumes that for each user profile a distribution over topics is sampled from a Dirichlet distribution. In the next step, for each word in the user profile, a single topic is chosen according to this topic distribution. Finally, each word is sampled from a multinomial distribution over words specific to the sampled topic. Here, we briefly review two important classes of LDA variants that have been applied for the expert recommendation in CQA.

Segmented Topic Model (STM)[37]. This is a topic model that discovers the hierarchical structure of topics by using the two-parameter Poisson Dirichlet process[38]. As a four-level probabilistic model, STM contains two levels of topic proportions and shows superiority over traditional models. Instead of grouping all the questions of a user under a single topic distribution, it allows each question to have a different and separate distribution over the topics. A user profile is considered as a document that contains questions (segments). The above distributions cover the expertise set of a user, the topics of each question in the profile, and the correlation between each profile and its questions.

TagLDA[39]. This method uses only tag information to infer users' topical interest. It is more efficiently, but the effectiveness is dependent on the accuracy and availability of tags.

4.3.3 Expertise-Aware LDA

The work in [40] considers both the topical interest and expertise of a user relevant to the topics of the given question. It also uses LDA to identify topical interest from previous answers of the user, but additionally compute the expertise level of users using collaborative voting mechanism. Sahu *et al.*[39] incorporated question tags and related voting information in LDA to compute user expertise, where user expertise is computed based on both the topical distribution of users and voting information under the same question tags.

4.3.4 Other Topic Models

Besides the famous QLL and LDA models, Zhou *et al.*[14] proposed a method that groups threads (i.e., a question and related answers) of similar content into clusters to build a cluster-based thread for each user. Each cluster represents a coherent topic and is associated with users to indicate the relevance relationship. The ranking score for a user is then computed based on the aggregation of the relevance of the user to all clusters given a new question. Guo *et al.*[3] proposed a user-centric and category-sensitive generative model for discovering topics, named User-Question-Answer (UQA). The work incorporates topics discovered by the UQA model with term-level matching methods to recommend experts and increase the participation rate of users in CQA. In this model, each user is considered as a pseudo-document which is a combination of all the questions the user has asked and all the answers the user has provided in reply to other users' questions. More methods can be derived based on this model as well as the combinations of these methods.

4.4 Network-Based Methods

The network-based methods evaluate users' authoritativeness in a user-user network formed by their asking-answering relations and recommend the most authoritative users as experts for a new question. The simplest network-based method uses Indegree[41] to rank and recommend users. In particular, an indegree score of a user equals the number of other users whose questions have been answered by the user, represented by an arrow from the requester to the answerer in the user-user network. Since frequent posters tend to have a significant interest in the topic and a larger degree of a node usually correlates with answer quality[41-42], this method regards the users with higher degrees as better answerers for the recommendation. The mainstream of this category includes three families of methods based on PageRank[43], HITS[44], and ExpertiseRank[41], respectively. We will also briefly introduce several other network-based methods to gain a comprehensive view of the related techniques.

4.4.1 PageRank and Its Variants

PageRank[45-46] uses nodes to represent users, and a directed edge to indicate one user (i.e., the source node) answers the questions of another user (i.e., the destination node). It estimates the likelihood that a random walk following links (and occasional random jumps) will visit a node. Each edge is associated with an affinity weight that measures the times that the answerer has replied to the requesters' questions. Two users are connected if their affinity weight is greater than 0, and the transition probabilities between nodes are obtained by normalizing the affinity weights. Now, the algorithm has been extended to bias the jump probability for particular topics[47] and many other static web ranking tasks. Choetkiertikul *et al.*[48] also used PageRank, but they measured the weights differently, i.e., evaluating the number of users' tags and activity times in common as weights between the users who have asking-answering relations.

The main variants of PageRank include SALSA[49], EntityRank[50], TwitterRank[51], and AuthorRank[52]. They differ from the above PageRank-based methods in focusing on some specific application domains. However, they still have the potential to be generalized to broader scenarios.

4.4.2 HITS and Its Variants

Different from PageRank, which does not distinguish between hub and authority nodes, the HITS algorithm is based on the observation that there are two types of nodes: 1) hubs, which link to authoritative nodes; 2) authorities, which provide useful information on the given topics. HITS assigns each node two scores: hub score and authority score. A hub score represents the quality of outgoing links from the nodes while an authority represents the quality of information located on these nodes. A typical work based on HITS was proposed by Jurczyk and Agichtein[45-46]. Instead of the times of the answerer replying to the requester's questions, this work models the weights of edges to indicate answer quality in HITS, based on users' explicit feedback, e.g., thumb-up/down from users, and whether the answer is the best answer. The results show that the HITS algorithm outperforms the methods based on simple graph measures such as in-degree.

An important variant of HITS was proposed by Shahriari *et al.*[53] Instead of considering the entire user-user network as a single community, this method regards the network as the combination of multiple user communities. Thereby, it detects overlapping communities[54] and differentiates the impact of intra-community and inter-community users to other users in the user-user network.

4.4.3 Expertise-Aware Network-Based Methods

Zhang *et al.*[41] proposed ExpertiseRank based on the intuition that an answerer who replies to a question usually has superior expertise compared with the requester on the specific topic. Their experimental results indicate that for a closed domain such as the Java developer forum, ExpertiseRank performs better than general graph-based algorithms like PageRank and HITS. ExpertiseRank considers not only how many other users a user has helped but also who the user has helped. It propagates the expertise scores of users through the question-answer relationship in a user-user network. The intuition behind ExpertiseRank is that one should get more credit for answering the questions of a user with higher expertise rather than the questions of a user with lower expertise.

The ExpertiseRank computation finally derives a score for each user, called z-score, based on which to quantify their authoritativeness. Z-score[41] combines users' asking-and-replying patterns to measure how many standard deviations above or below the expected "random" value a user lies. The more questions a user has answered and the fewer questions the user has asked, the higher the z-score of this user. Therefore, this method recommends users with the highest z-scores as experts. The experts obtained by this method should answer many questions and ask very few questions.

Besides ExpertiseRank, Jurczyk and Agichtein[45] incorporated users' authoritativeness on a given question (estimated by the tags of users' posts) and users' answer quality (predicted based on their past answer activities); the method in [55] differs in determining user quality by the number of the best answers they provide. Zhou *et al.*[14] used the post content of users to compute user expertise and reply frequencies of users computed by PageRank to re-rank users. They further used inverted indexes and threshold algorithm[56] to store and retrieve pre-computed intermediate scores to accelerate the computation. The final score of a user takes the product of the results of language model and the results of PageRank.

4.4.4 Other Network-Based Methods

Here, we review some typical enhancement or extension of the traditional link analysis methods as follows: Zhu *et al.*[57-58] additionally considered the category relevance of questions and ranked user authority in an extended category link graph; Liu *et al.*[59] comprehensively utilized multiple types of relationships between requesters and answerers, and between the best answerers and other answerers to find expert users; rather than leveraging the asking-answering interactions, Lai and Kao[60] employed the endorsement relationship among users to form user reputation graphs, based on which to recommend experts. Similarly, Lin and Shen[61] computed user reputation based on their trust relationship in a user-user network. By assuming each question has a known category and theme, they clustered users based on their reputation scores on each question theme and evaluated users based on their theme-specific reputation. Liu *et al.*[62] incorporated user-subject relevance (computed by cosine similarity) and user reputation with users' category-specific authoritativeness obtained from link analysis for the expert recommendation.

The latest network-based method was proposed by Liu *et al.*[63] This work routs questions to potential answerers from the viewpoint of knowledge graph embedding and integrates topic representations with network structure into a unified question routing framework. The framework takes into account various types of relationships among users and questions. It is demonstrated to increase the answer rates of questions by using the recommended experts.

4.5 Classification Methods

When regarding experts as a particular class of users among all users, the problem of identifying the experts can be easily transformed into a classification problem that aims to distinguish such a particular class of expert users from the other users. Compared with the other methods, classification methods can easily apply multiple aspects of features from the perspective of the user, question, answer, or user-user interaction, to the expert recommendation problem. For example, Pal and

Konstan[64] used three classes of features and trained a binary classifier to distinguish experts from ordinary users. These features include question features (e.g., question length, word n-gram), user features (e.g., the number of answers and the number of the best answers provided by a user), and user feedback on answers (e.g., the user votes and comments to the answers).

Support Vector Machine (SVM) is the most used classification method for distinguishing experts from those non-experts. Besides diverse methods, the classification methods have used different features besides the basic question and user features, such as part-of-speech features, graph features to train their models. For example, Pal *et al.*[65] extracted features by modeling users' motivation and ability to help others and used SVM and C4.5 decision tree separately for expert detection. Zhou *et al.*[66] also used SVM but they defined both local and global features on questions, user history, and question-user relationship and additionally considered KL-divergence as a new feature. Ji and Wang[67] additionally used text similarities as features to train SVM and one of its variant, RankingSVM.

Besides, more methods such as random forests (RF)[48] and Naive Bayes[68] are used by existing studies. Some typical work includes: Le and Shah[69] considered more new features such as community features (e.g., average score of posts, average number of comments, average number of favorites marked), temporal features (e.g., time gaps between posts), and consistent features (e.g., scores of the posts, time gap between recent posts). They also tried some new methods like logistic regression and adaptive boosting in addition to decision trees and random forest for the classification. As an enhancement to decision trees, Dror *et al.*[70] proposed a representation model based on multi-channel vector space model, where users and questions are represented as vectors consisting of multi-dimensional features. Then, the matching degree between a user and a question is learned from their respective features using a binary classifier called Gradient Boosted Decision Trees (GBDT).

Instead of the conventional features used for identifying experts, Pal and Konstan[64] used a different criterion, question selection bias, for recommending experts, based on the assumption that experts prefer answering questions to which they bear a higher chance of making a valuable contribution. They used the probability of answering questions of different values as the feature vector and employed two methods, logistic regression and Gaussian Mixture Model, to solve the binary classification problem. In a later version, Pal *et al.*[71] used different equations to estimate the value of existing questions and to model the selection probabilities of questions by users. The method shows better performance of the Bagging metaclassifier over several single-version classification algorithms. The work also partially confirms that experts have some bias on question selection based on the existing value of answers to the questions.

Given the advantages of ranked recommendation results over unranked results, Ji and Wang[67] proposed RankingSVM, a ranking model based on SVM, for the expert recommendation. Burel *et al.*[72] extracted patterns from the question-selection behaviors of users in a Q&A community and then used Learning to Rank (LTR) models to identify the most relevant question to a user at any given time. They further employed Random Forests, LambdaRank[73], and ListNet[74] to derive a pointwise method, a pairwise method, and a listwise method, respectively, to gain ranked expert list along with the classification process. Similarly, Cheng *et al.*[75] also formalized expert finding as a learning-to-rank task. However, they leveraged users' voting information on answers as the "relevance" labels and utilized LambdaMART to learn ranking models which directly optimizes a rank-based evaluation metric, normalized discounted cumulative gain (nDCG). Logistic regression[76] has also been used recently to facilitate both ranking and classification.

4.6 Expertise Probabilistic Models

Dom and Paranjpe[77] proposed the only probabilistic model that focuses on user expertise for the expert recommendation in CQA. They used a Bayesian probabilistic model to obtain a posterior estimate of user credibility thereby recommending experts for a given question. This work assumes user expertise conforms to the Bernoulli distribution (or the mixture of two beta distributions) and uses maximum a posteriori (MAP) estimation to make predictions. It then ranks users according to their probabilities of providing the best answer to the given question, characterized by the probability of each user to be awarded the best answer on a question given the user's question-answering history. The work also discovered that Bayesian smoothing performs better than several other smoothing methods such as maximum a priori estimation, maximum likelihood (ML) estimation, and Laplace smoothing.

4.7 Collaborative Filtering Methods

Given the advantages such as flexibility and scalability of collaborative filtering (CF) methods, esp., the matrix factorization techniques[78] in the recommendation domain, some researchers seek to use CF for the expert recommendation in CQA. For example, Cho *et al.*[79] considered the case that only one expert would be designated to answer each question. They then captured users' behavioral features as potential answerers by matrix factorization[80]. Regularization terms are incorporated to represent user interest, user similarity, and users' probability to answer for better performance.

Instead of the basic matrix factorization model, Yang and Manandhar[81] employed probabilistic matrix factorization (PMF), which combines generative probabilistic model and matrix factorization, to address the expert recommendation problem. In particular, PMF learns the latent feature space of both users and tags to build a user-tag matrix[82], which is then used to recommend experts given a new question.

4.8 Hybrid Methods

To comprehensively take into account multiple aspects of clues, some researchers proposed hybrid methods that combine different aspects of concerns techniques and different techniques for better recommendation results. Here, we review the typical hybrid methods as follows.

4.8.1 Language Model + Topic Model

Liu *et al.*[30] combined QLL and LDA and showed the hybrid approach outperforms either of the original models.

4.8.2 Topic Model + Network-Based Method

Zhou *et al.*[83] identified authoritative users by considering both the link structure and topic information about users. They first applied LDA to obtain user-topic matrix and topic-word matrix, and then used PageRank and Topical PageRank for the expert recommendation. Liu *et al.*[84] proposed a topic-sensitive probabilistic model to estimate the user authority ranking for each question. The model is based on PageRank incorporating with topical similarities between users and questions. A very similar method named topic-sensitive link analysis was proposed by Yang *et al.*[85] Recently, Rao *et al.*[86] have proposed a similar approach that recommends experts based on users' topical relevance and authoritativeness given a new question. They also used LDA to discover topics but measure users' authoritativeness for each topic based on the "like" relationship among users in a social network.

Zhao *et al.*[87] used the TEL model to generate topic information, based on which to model experts over topics. TEL is based on LDA and is a unified model combining both graph-based link analysis and content-based semantic analysis for expert modeling in CQA. Instead of using a fixed user-answering graph to influence the prior of expert modeling, TEL highlights the causal relationship between topics and experts by using both user-user matrix and question-user matrix to represent a user's contribution to another user or a question, with best answers given higher weights. It is compared with two baseline topic modeling methods, one recommending experts based on requester-answerer interactions and the other recommending experts based on question-answerer interactions, which show its better performance.

Zhou *et al.*[88] extended PageRank with the topic-sensitive probabilistic model by considering topical similarity. In particular, the method improves the PageRank-based expert recommendation methods by running PageRank for each topic separately, with that each topic-specific PageRank prefers those users with high relevance to the corresponding topic. As a further step, Yang *et al.*[89] jointly modeled topics and expertise to find experts with both similar topical preference and superior topical expertise on the given question. The method integrates textual content model (GMM) and link structure analysis and leverages both tagging and voting information. It linearly incorporates the estimated user topical expertise score into the recursive PageRank score computation formula and extends PageRank to make the final recommendation.

4.8.3 Language + Topic + Network-Based Model

Liu *et al.*[30] assumed that more authoritative answerers may give more accurate answers, and more active answerers may be more willing to answer new questions. They linearly combined QLL and LDA models to compute relevance, and additionally considered both user activity and authority information for the recommendation.

4.8.4 Topic Model + Classification Method

RankingSVM[67] employs LDA to calculate text similarity and use this similarity as a feature in a classification method for the expert recommendation. The experimental results show the resulting method achieves better performance than SVM.

4.8.5 Topic Model + Collaborative Filtering

Yan and Zhou[30] combined topic model and tensor factorization for the expert recommendation. They trained an LDA via Gibbs sampling with a manually defined topic number, followed by performing tensor factorization (TF) based on "requester-topic-answerer" triples via gradient descent to compute the recommendation scores of users.

4.8.6 Network-Based Method + Clustering

Following the similar idea of geo-social community discovery[91] in Point of Interest (POI) recommendation, Bouguessa *et al.*[55] incorporated clustering methods with network-based measures for the expert recommendation. They considered the number of best answers as an indicator of the authoritativeness of a user in a user-user network, where there is an edge from every requester to each of the corresponding best answers. Each edge is weighted by the number of the best answers in-between. In particular, they modeled the authority scores of users as a mixture of gamma distribution and used the Fuzzy C-Means algorithm to partition users into different numbers of clusters. They further used Bayesian Information Criteria (BIC) to estimate the appropriate number of mixtures. Finally, the users are classified into two classes, one representing authoritative users with high in-degrees and the other representing non-authoritative users with low in-degrees. In this way, the method can automatically surface the number of experts in a community rather than produce a ranked list of users.

5 Comparison and Discussion

To gain a better understanding of the state-of-the-art, we first summarize the existing expert recommendation methods concerning the used dataset, the required input & output, and the evaluation metric. We further compare and discuss the methods from three perspectives: the covered aspects of concern, reliance on sufficient data, and complexity, to identify their strengths and pitfalls. The three perspectives reflect the methods' capability in the recommendation power, applicability (robustness to cold start or sparse data), and easiness of usage (implementation difficulty), respectively.

5.1 Datasets

In this subsection, we list the most-used datasets by existing expert recommendation research. These datasets include both the real-world and synthetic ones, as well as those that do not belong to but are readily applicable to evaluating the methods for the expert recommendation in CQA. Among the real-world datasets, only the first two represent the dominant datasets used by most existing research while all the others are less used.

5.1.1 Yahoo! Answers

Yahoo! Answers[3,6,27,35,45,66,83,87,90,92] is perhaps the most popular and most studied dataset in Q&A related research. The characteristics of Yahoo! Answers (e.g., the diversity of questions and answers, the breadth of answering, and the quality of those answers) were first investigated by Adamic *et al.*[11] in 2008. In particular, they used user entropy to indicate a user's concentration or focus on different categories of topics. They further clustered questions and answers based on the content to understand users' activity on different topics in CQA. The results showed that the majority of users participated in a small number of topics. These features set the practical foundation for predicting answer quality by the amount of work and activities of users. Since each question has at most one best answer, the amount of ground truth might be sparse when only a part of the entire dataset is used in experiments. For this reason, some researchers set up their own criteria to determine whether an answer is a "good" answer or not, to expand the training and test set for their methods. For example, Li and King[6] labeled an answer a "good" answer either when it is selected as the best answer or when it obtains more than 50% of up-votes for all the answers of the question. Meanwhile, one answer is labeled as a "bad" answer if it receives more than 50% of rate-downs for all answers of the question.

5.1.2 Stack Overflow

Stack Overflow[31,37,39,48,64,69,71,75,93-94] involves over five million users and content about 11 053 469 questions, among which only 73% have received answers and closed and 55%, i.e., over six million questions, have known best answers (as of March 10, 2016). Like the Yahoo! Answers dataset, the records in the Stack Overflow dataset are massive, and most existing researches sampled a subset of the entire dataset for study. For example, Pal *et al.*[71] sampled a small dataset of 100 users and employed two expert coders to label the 100 users as either experts or non-experts. It turns out that the inter-rater agreement between the expert coders is 0.72 (Fleiss Kappa with 95%CI, $p \sim 0$), indicating the high agreement between the raters is not accidental. Out of the 100 users, 22 are labeled as experts and the rest as non-experts.

5.1.3 Other CQA Datasets

TurboTax Live Community (*TurboTax*)⑩[64-65,71]: a Q&A service related to the preparation of tax returns. TurboTax has employees that manually evaluate an expert candidate on factors, such as correctness and completeness of answers, politeness in responses, language and choice of words used. This dataset also has some labeled experts.

Quara[95-96]: a general and probably the world' largest Q&A website that covers various topics.

Java Developer Forum[41]: an online community where people come to ask questions about Java. It has 87 sub-forums that focus on various topics concerning Java programming. There is a broad diversity of users, ranging from students learning Java to the top Java experts. A typical sub-forum, e.g., "Java Forum", has accumulated hundreds of thousands of messages in tens of thousands of threads in the year of 2007.

Naver KnowledgeCiN: the largest question-answering online community in South Korea. Nam *et al.*[97] analyzed the characteristics of knowledge generation and user participation behavior in this website and found that altruism, learning, and competency are often the motivations for top answerers to participate.

*Baidu Knows*⑪: a Chinese CQA service, where an answer can receive a bounty after answering a question. Once an answer is accepted, it turns into the search result of relevant questions.

*Tripadvisor forums*⑫[14]: a travel-related website with user-generated content focusing on accommodation bookings. The service is free to users, who provide feedback and reviews to hotels, accommodation facilities, and other traveling-related issues.

Sogou Wenwen[90]: formerly known as Tencent Wenwen or Soso Wenwen, is similar to Quora and also run with credit points and reputation points. Users can obtain points by asking or answering questions and use them as bounty.

*Iask*⑬[30]: a leading web 2.0 site in China. The working mechanism of Zask is similar to that of Baidu Knows. The difference is that a requester in Iask can increase the bounty to keep the question open for another 15 days after obtaining the best answer.

Other Datasets on Stack Exchange: such as computer science⑭, fitness⑮[53], and cooking⑯. There are total 133 communities for knowledge sharing and question answering, covering enormous topics on Stack Exchange.

Estonian Nature Forum[53]: a Q&A website popular in Estonia.

*MedHelp*⑰[79]: a website which partners with doctors from hospitals and research facilities to provide online discussion and to satisfy users' medical information needs.

5.1.4 Synthetic Datasets

Generally, no single method outperforms all the others on all the datasets for two main reasons: first, online communities usually have different structural characteristics and lead to differences in the performance of methods[41]; second, the same users may behave differently in different communities due to various reasons such as the subjectivity and rewarding mechanism of a Q&A system. Given the lack of benchmarks to evaluate different methods, it has become a common practice to conduct controlled experiments with simulated datasets to test how a method performs under different scenarios. We will not give more introduction to the synthetic datasets due to the significant variances in the assumptions and conditions to generating these datasets.

5.1.5 Non-CQA Datasets

Plenty of datasets do not belong to the Q&A domain but are readily applicable to or have been used for the study of expert recommendation methods for CQA. A slight difference of the methods developed based on studying these datasets is that they most aim to rank and find the most best-skilled or authoritative users given an existing domain or topic instead of a new question. These datasets include

⑩http://ttlc.intuit.com/, May 2018.
⑪http://zhidao.baidu.com/, May 2018.
⑫http://www.tripadvisor.com/, May 2018.
⑬http://iask.sina.com.cn/, May 2018.
⑭http://cs.stackexchange.com/, May 2018.
⑮http://fitness.stackexchange.com/, May 2018.
⑯http://cooking.stackexchange.com/, May 2018.
⑰http://www.medhelp.org/, May 2018.

co-authorship network[52,98-99] such as DBLP[100-102], social networks[16,103-104], microblogs[105-107] such as Twitter[51], Email network[108-110], Internet forums[41], log data[111], e-Learning platform[112], Usenet newsgroups[7-8], Google Groups[9], general documents[113], and enterprise documents[20,26,114] such as Enterprise track of TREC[115-117].

5.2 Input and Output

To ease illustration, we first summarize the typical inputs and outputs of existing expert recommendation methods in Table 2. Here, we list five categories of commonly used inputs for expert recommendation methods. These inputs are either textual, numerical, or relational information, while the outputs, i.e., the recommended experts, are either ranked or unranked, depending on the methods adopted.

Based on the input/output list, we further present a comparison of the representative methods with respect to their inputs and outputs in Table 3. Some methods may use derived features from the original inputs as additional inputs. For example, classification methods may use the length of questions (implied by question content), the total question number of users (implied by users' question history), and the total answer number of users (implied by users' answer history) as additional features to train their models.

5.3 Evaluation Metrics

We summarize three categories of metrics used to evaluate expert recommendation methods for CQA, namely the basic, rank-based, and human-judgment-based metrics. The following subsections introduce the metrics of each category, respectively, where each metric is computed as (the mean of) the average of the metric values over a set of query questions or topics[14,27,53,79,83,90].

5.3.1 Basic Metrics

There are four set-based metrics to evaluate an expert recommendation method.

Precision[64-66,118]: the fraction of users who are true experts to the given questions, among all the users recommended by a method.

Recall[53,64-66,118]: the fraction of users who are recommended by a method and meanwhile turn out to be the real experts, among all the real experts to the given questions.

F_1-Score[64-66,68-69,118]: the harmonious average of the precision and the recall.

Accuracy[66,69-70,118]: the fraction of users who are correctly identified as either an expert or a non-expert by a method. The metric integrates the precision of the method in identifying the experts and non-experts.

5.3.2 Rank-Based Metrics

Precision at Top n ($P@n$)[14,27,79,83,87,90,92-93,119]: the percentage of the top-N candidate answers retrieved that are correct. It is also known as precision at top n ($P@n$)[87,90,93,119] or success at top N ($S@N$)[37]. A special case is $P@1$[95-96] when $n = 1$.

Recall at Top N ($R@N$)[90,92-93,96]: a natural expansion of the basic recall to rank-based scenario, similar to $P@n$.

Accuracy by Rank[96]: the ranking percentage of the best answerer among all answers. A similar metric using the best answerer's rank was proposed in [15] and [35].

Mean Reciprocal Rank (MRR)[6,14,27,53,72,83-84,90,94]: the mean of the reciprocal ranks of the first correct experts over a set of questions. It gives us an idea of how

Table 2. Typical Inputs and Outputs of Expert Recommendation Methods

Type	Category	ID	Input/Output Name	Input/Output Type
Input	Question profile	I0	Content (and category) of the given question	Textual
	User profile	I1	Users' question history	User-question mapping
		I2	Users' answer history	User-answer mapping
		I3	Users' historical viewing and answering activity	Multiple user-question mapping
		I4	Timestamps of users' answering activity	Numerical
	Historical questions & answers	I5	Question content	Textual
		I6	Question category information	Textual
		I7	Question tags	Textual
		I8	Qnswer content	Textual
		I9	Best answer information	Answer–{0,1} mapping
	Social profile	IA	Voting information	Numerical
		IB	User reputation	Numerical
	Network profile	IC	Question-answer relations among users	Directed user-user mapping
Output	Recommended experts	O1	Unranked group of experts	Set
		O2	Ranked list of experts	List

Table 3. Comparison of Inputs and Outputs of Representative Expert Recommendation Methods

Category	Representative Method	I0	I1	I2	I3	I4	I5	I6	I7	I8	I9	IA	IB	IC	O1	O2
Simple methods	Votes			√								√				√
	Best answer proportion			√							√					√
	Consine similarity based on TF-IDF[19]	√		√			√									√
Language models	QLL[23-25]	√		√			√									√
	Category-sensitive QLL[27]	√		√			√	√								√
	Expertise-aware QLL[6,28]	√		√			√				√					√
Topic models	PLSA[15,34], LDA[36], STM[37], UQA[3]	√		√			√									√
	DRM[35]	√	√	√			√									√
	TagLDA[39]	√		√					√							√
Network-based methods	Indegree[41], PageRank[47-48], HITS[46,53]													√		√
	z-score[41], Expertise-aware methods[14,41,45,55-56]			√							√			√		√
	Reputation-aware methods[60-61]												√	√		√
	Category-sensitive methods[57-58]	√						√						√		√
	Graph embedding method[63]	√	√	√			√			√	√			√		√
Classification	SVM[65-66], C4.5[65], RF[48], GBDT[70]	√		√	√		√		√		√	√	√	√	√	
methods	LTR[72]	√		√	√		√		√		√	√	√	√		√
Expertise probabilistic model	Bernoulli MAP model[77]			√							√					√
Collaborative filtering	MF[79]	√		√	√		√									√
	Tag-based PMF[81]	√		√	√				√		√					√
Hybrid methods	QLL+LDA[30], Topical PageRank[83], TEL[87], LDA+TF[90]	√		√			√									√
	Topical PageRank+Expertise[89]	√		√			√				√					√
	QLL+LDA+userActivity+Indegree[30]	√		√		√	√				√			√		√
	Indegree+Clustering[55]	√		√			√				√			√	√	

many users we generally need to check before finding the right expert from a ranked list.

Matching Set Count (*MSC*) *@n*[93-94]: the average number of the questions that were replied by any user ranked within top n recommended users.

Normalized Discounted Cumulative Gain (*nDCG*)[53,95]: a number between 0 and 1, measuring the performance of a recommendation system based on the graded relevance of the recommended items. A variant is *nDCG@k*, the division of the raw DCG by the ideal DCG, where k is the maximum number of items to be recommended.

Pearson Correlation Coefficient[45,120-121]: the correlation degree between the estimated ranking and the ranks of users according to the scores derived from the user feedback.

Area Under ROC Curve (*AUC*)[70]: the probability that an expert is scored higher than a non-expert.

5.3.3 *Human Judgment Based Metrics*

Correctness Percentage. Human judgment is necessary in the case where the ground truth is unavailable or hard to be determined automatically. In such cases, humans usually give either Yes/No answers[3] or ratings[41] to the recommended users. Then, the system calculates the percentage of correctly recommended users by investigating the agreement between the judgments made by different people[55,83,118].

5.4 Covered Aspects of Concern

To study the covered aspects of concern of different methods, we summarize the main aspects of concern and their indicators in an expert recommendation problem in Table 4. The inputs taken by each method directly reflect the method's covered aspects of concern. For example, to take user expertise into account, a method needs to consider at least one of the three aspects: user reputation, voting information to answers, and the best answer flags of answers. An indicator either belongs to or originates from the inputs. It falls into at least one of the two aspects: answer probability and expertise level of users.

The inputs and outputs only give some intuitive

Table 4. Aspects of Concern and Their Indicators Covered by Representative Expert Recommendation Methods

Category	Representative Method	Answer Probability	Expertise Level
Simple methods	Votes		Vote counts of answers
	Best answer proportion		Best answer ratio
	Consine similarity of TF-IDF[19]	Textual relevance	
Language models	QLL[23-25]	Textual relevance	
	Category-sensitive QLL[27]	Question category Textual relevance	
	Expertise-aware QLL[6,28]	Textual relevance	Best answer ratio
Topic models	PLSA[15,34], LDA[36], STM[37], UQA[3]	Topical relevance	
	DRM[35]	Topical relevance	
	TagLDA[39]	Topical relevance of tags	
Network-based methods	Indegree[41]	a# of users	
	PageRank[47-48], HITS[46,53]	a# & q# of users	a# & q# of users
	z-score[41]	a# & q# of users	a# & q# of users, best answer number
	Expertise-aware methods[14,41,45,55-56]	a# & q# of users	a# & q# of users, best answer number
	Reputation-aware methods[60-61]	a# & q# of users	a# & q# of users, user reputation
	Category-sensitive methods[57-58]	a# & q# of users, category relevance	a# & q# of users
	Graph embedding method[63]	a# & q# of users, textual relevance	a# & q# of users, best answer number
Classification methods	SVM[65-66], C4.5[65], RF[48], GBDT[70], LTR[72]	Textual relevance, question features, user features (e.g., a#), metrics (e.g., z-score)	Best answer number
Expertise probability model	Bernoulli MAP model[77]		Best answer ratio
Collaborative filtering	MF[79]	Textual relevance	
	Tag-based PMF[81]	Textual relevance of tags	Best answer ratio
Hybrid methods	QLL+LDA[30], TEL[87]	Textual & topical relevance	
	LDA+TF[90]	Textual & topical relevance	
	Indegree+Clustering[55]	a# & q# of users	Best answer number
	Topical PageRank[83]	Topical relevance, a# & q# of users	a# & q# of users
	Topical PageRank+Expertise[89]	Topical relevance, a# & q# of users	a# & q# of users, best answer number
	QLL+LDA+userActivity+Indegree[30]	Textual/topical relevance, a#, q#, active time of users	Best answer number

Note: q# and a# denote the number of questions asked and answered by a user, respectively.

clues of how powerful each method might be. In fact, it is not only the inputs but also the ways of using these inputs and the underlying techniques that determine a method's capability in adequately addressing the expert recommendation problem. In the following, we elaborate each aspect of concern and make a comparative discussion of the investigated methods according to their covered aspects of concern.

5.4.1 *Answer Probability*

A large body of the research on the expert recommendation in CQA focuses merely on ranking the relevance of the given question with users' previously answered questions. The underlying assumption is that users are more likely to answer those questions that are similar to their previously answered ones. This type covers all the methods that consider users' answer probability yet not users' expertise level in Table 4 and range from the simple Cosine similarity method to QLL, LDA, MF, and further to hybrid methods that combine the above techniques like TEL.

Similar to the general recommendation approach, it is reasonable to assume a user prefers to answer the questions that are similar to his/her already-answered questions. On the other hand, there is no evidence to show that the stronger relevance of a user to a given question also indicates a higher expertise level of the user in answering that question. Therefore, such methods may not be able to distinguish users' different levels of expertise in either a question or a topic, and the recommended experts may not provide quality answers to the given question.

Other issues with this type of methods are related to the consideration of categories or topics. While considering the category or topic information enables a method to provide personalized recommendations customized to the given question, it raises additional issues to the method. First, the category-sensitive methods highly rely on the availability and proper definition of categories. For example, suppose a user has answered a lot of questions about "c++" covered by the category of "Programming," and is deemed as a potential answerer for questions of this category. Given a question related to another programming language like "Python", which is also covered by the category of "Programming", recommending this user as an expert may not be appropriate as the user may not be familiar with "Python" as well. The topic-sensitive approach is more reasonable as topics are usually not predefined but dynamically constructed by algorithms, and therefore they can adapt to the ever-increasing questions and answers in a Q&A community.

Second, an inevitable issue with considering categories or topics is that a user may have an expertise in multiple categories or topics. For category-sensitive methods, although the correlation among categories can be incorporated explicitly, it could be difficult to designate a question to a single category when the question is related to multiple categories. For the topic-sensitive methods, they mostly discover topics based on probabilistic models such as LDA. Since the probabilistic models distribute the total probability of 1 among all the topics for each user, having a higher probability on one topic will discount the probability on other topics. However, the fact is that a user could be more relevant to multiple topics than another user simultaneously. This situation has not been sufficiently taken into account by the existing research.

5.4.2 User Expertise

There are a few methods that take into account user expertise while neglecting to consider the answer probability of users in the expert recommendation in CQA. These methods typically include simple techniques (e.g., votes, the best answer proportion) and the expertise probabilistic model in Table 4. A limitation of these methods is that they only consider the probability of a user giving the best answer under the condition that the user has decided to answer the given question. The fact is that a user with the appropriate expertise may not answer the question in the first place due to various reasons such as lack of interest or unavailability. Therefore, as far as answer probability is concerned, the recommended experts may not have a good chance of giving a quality answer.

Another issue with the methods of this type is that they commonly compute a global expertise score for each user while neglecting the fact that user expertise may also be topic-sensitive, similar to answer probability. Consequently, the recommendation results are independent of the specific problem scenario: given a new question, the recommended experts may perform generally well but unfortunately perform poorly on the given question. The global expertise model is most suitable for scenarios where the same topic covers all questions.

5.4.3 Both Aspects

Given the importance of both aspects of concern, many existing methods, especially the latest ones, combine the two above aspects for the better expert recommendation. These methods typically include expertise-aware QLL models, all variants of PageRank/HITS, classification methods, collaborative filtering methods, and hybrid methods that combine two or more of the above methods. The straightforward way of integrating the two aspects is to compute a score for each aspect separately and then combine the two scores into a weighted sum, which would be used as the criterion for ranking users and deriving the recommendation results. The other methods, including network-based methods, classification methods, and collaborative filtering methods, combine the two aspects of consideration more naturally.

For example, the network-based methods naturally incorporate the two aspects to compute a single criterion, user authoritativeness, for the expert recommendation — the users who have provided many questions yet asked few questions are considered to be authoritative; such users are generally believed to have both higher probabilities to answer and better expertise in a user-user network. An alternative is to use the related user number to replace question number in the above computation. This replacement slight changes the computation method but does not affect the authoritativeness definition. Instead of using a single metric such as authoritativeness to recommend experts, the classification methods directly take various factors as features, without explicitly distinguishing the two aspects, to train a classification model. A limitation with classification methods is that they generally deliver a set of users as experts without further differentiating them. Therefore, they lack the flexibility to decide how many users to recommend as experts on the fly.

To explicitly utilize the best answer information in a network-based method, a simple approach is to replace the "requester-answerer" relationship in a user-user network into the "requester-best answerer" relationship among users. In this case, a link is drawn from one user to another user only when the second user has provided the best answer to the first user's questions. The indegrees and other metrics (e.g., authoritativeness) of users derived from the modified model can directly be used to recommend users who are both active and have the right expertise. In this way, the recommended experts are those who have provided the best answers to the largest numbers of other users and have been answered by the fewest other users on their raised questions.

Currently, almost all the hybrid methods that cover both aspects involve network-based methods as a component. These methods therefore still share some drawbacks of the basic network-based methods. First, the experts recommended by such methods are specific to a user-user network rather than a particular topic or a question. Intuitively, both the transitivity of users' authoritativeness and the effectiveness of network-based methods depend on the condition that the interactions between users concern only one topic. Second, although the link structure can, to some extent, imply the correlation among users and questions, the user-user network is not directly related to a user's topical preference. To recommend experts for a given problem, they still need some approach to connect the question to users or their historical questions to make a personalized recommendation. The more recent expertise-aware network-based methods often hybridize with relevance models (e.g., language models and topic models) to overcome the above deficiencies.

Another possible issue with the hybridization of network-based methods with other methods is that authoritativeness is a vague concept, which already, to some extent, implies the interest and expertise of users. Therefore, the combination of techniques may cause the hybrid methods to consider multiple times of the same aspects. The rationale and the effect of such hybridization are yet to be examined.

5.5 Reliance on Sufficient Data

The cold start and data sparsity issues concern both the ground truth and the number of users' activity records. They turn out to be the common challenge for all the investigated expert recommendation methods. For the amount of ground truth, the rule of thumb is that the more straightforward methods tend to be less affected by the small ground truth size, as the more complicated methods usually require a larger train set. For example, the classification methods generally perform better under high-dimensional features given sufficient training data. When the training data is limited, these methods need to restrain the dimensionality to avoid over-fitting. In contrast, the voting-base techniques require no training, and thus unaffected by the size of training data.

The effect of the historical record number of users on the recommendation methods is closely related to the early detection of experts, i.e., promoting the participation rate of new experts or rising stars, i.e., the low profile users who have strong potential to contribute to the community later after short observations in CQA[69]. The lack of sufficient information in the user profile is, in fact, the primary obstacle towards identifying such early-career experts. A previous study[22] showed that only 15.67% of all users in Yahoo! Answers answered more than four questions. This observation indicates that all these approaches involve only a small portion of highly active users and therefore cannot recommend new questions to the rest of the community. If the recommendation method can also involve the good users with few activity records, these users can become motivated to take more intensive participation in a community or even develop into highly active experts.

Despite the significance of early expert detection issue, the markers that reflect the expertise of an ordinary user (e.g., the number of answers and the number of best answers) are not that strong for a newly joined user. Therefore, few researchers have studied the discovery of potential experts at an early stage of CQA[64-65,71]. Many existing methods bypass the cold start and data sparsity issues due to their reliance on sufficient data from CQA systems. For example, some approaches consider only those users who previously provided more than 5[40], 10[3] or even 20 answers[122]. Other methods take only users with a significant number of the best answers into consideration (e.g., more than 10[30] or 20 best answers[31]).

Among the existing methods, we identify two promising categories of methods that can potentially better detect experts early. One is semi-supervised learning methods (e.g., [68]), which regard users who provide above average-best-answers on a topic tag as topical experts. They apply a data-driven approach to predict whether a user will become an expert in the long term. The other is expertise propagation methods (e.g., [123]), which infer or consolidate the expertise of low-profile users by propagating the expertise of old users through their shared activities in a community.

5.6 Method Complexity

Generally, the more aspects considered, the better a method can potentially perform, and the more complicated the method could be. The expertise-aware techniques based on QLL usually combine the two aspects linearly using the simple-weight-sum method. The primary issue with the method is the difficulty in allocating the weights wisely among the two aspects. Usually, they need to resort to human experience or repeated trials in real applications to determine the optimal weights.

Though applicable to the expert recommendation problem, recommendation techniques face severe challenges besides the fundamental issues like the cold start problem. For example, considering multiple aspects of concerns could make a recommendation technique complex and challenging to optimize. More recently recommendation methods such as factorization machines may help resolve the problem but have not yet been applied to the expert recommendation in CQA.

Despite the ability to incorporate multiple aspects of concern, there is a lack of universal principle regarding which features to use for the classification methods. Consequently, the performance of classification methods largely depends on the features used and whether the technique and features fit the size of the labeled data.

6 Future Directions

After reviewing the state-of-the-art methods, we identify several challenging yet promising issues for the future research. We summarize them as realistic user modeling, recommending experts as a collaborative group, coping with dynamicity, utilization of external data, and comprehensive expert recommendation solutions. In the following, we review the limited related studies to the above challenges, highlight the significance of and new opportunities for addressing these challenges, and finally outlook the promising directions for future research. We hope this discussion could provide novel viewpoints to the existing studies and encourage more future contributions to this promising field of research.

6.1 Realistic User Modeling

Expert recommendation relies on effective user modeling. Intuitively, there exist three aspects of concerns that affect whether a user gives a high-quality answer to a question in a real Q&A scenario as follows.

Chance of a User Noticing the Question. Since a user may not have an opportunity to see a question, the user may not be an answerer to this question even though the user is an expert. The expert recommendation problem in CQA, however, is based on a different assumption from the real-world scenarios, i.e., what is the possibility that a user would answer a question and meanwhile provide a high-quality answer to the question if the user is invited to answer the question. Due to the above difference, when using the real-world labeled data to train the recommendation models, the recommendation methods should better take into account the possibility that a user may not have answered a question just because the user does not have the chance to notice the question. The likelihood that a user would see a question in real-world scenarios depends on various factors such as user availability (e.g., how often a user is online and available to answer questions), user behaviors (e.g., whether the user looks for new questions to answer actively), and other users' activities related to the question (e.g., how widespread the question is among users).

User's Willingness to Answer the Question. Even if a user has noticed a question, the user may choose not to answer it. A user's willingness to answer a question also depends on various factors such as how well the question fits the user's interest, user's self-confidence on the quality of answers, and user's expected gains from answering the question.

User's Expertise Level on the Question. Even if a user has noticed a question and is willing to answer it, the user may not have the adequate knowledge to give a high-quality answer. That is the first and foremost reason that we need an expert recommendation approach in CQA.

Besides identifying different aspects, we need to find a reasonable way to combine them to recommend real experts more comprehensively given a new question. Unfortunately, the existing expert recommendation methods usually consider only the second, the last, or both the above aspects. For example, most language models and topic models focus on recommending users who are most likely to answer the given question. However, a high possibility of a user answering a question does not necessarily mean the user would be able to provide a high-quality answer. Many link analysis methods identify experts as the most authoritative users in a user-user network, where authoritativeness is a different concept from both relevance and user expertise. Therefore, the most authoritative users are neither guaranteed to be willing to answer a question nor being able to give a high-quality answer. Besides, some classification methods merely rely on the previous answer quality and some metadata features without considering users' topical distributions. For this reason, a recommended expert by such methods may not want to answer a question even if the user is a good answerer in general. Worse still, to the best of our knowledge, the first aspect has not been considered by any previous research efforts. A promising research direction is to incorporate expert recommendation method with models that could effectively predict user behaviors, just like the prediction of check-ins in a Point of Interest recommendation problem[124].

6.2 Coping with Dynamicity

Currently, the vast majority of researchers consider the expert recommendation problem in a static context, where they use a snapshot of users' previously asked or answered questions for the expert recommendation. However, the real-world question answering websites are dynamic, with new users joining and leaving, users' interest changing, users' roles transforming, users' mutual interactions evolving, and the content on the website never stopping updating[125]. Therefore, it is especially promising to develop methods that leverage the temporal information to make the expert recommendation methods adaptive in a dynamic context in a real-time fashion.

Currently, user availability is the most commonly considered dynamic aspect for the expert recommendation problem. Several studies used temporal features to estimate the availability of users for a given day[6,93,123] or for a specific time of the day[93,126]. For example, Sung *et al.*[123] used all replies of users to train a sigmoid function and Chang and Pal[93] built binary classifiers using all responses of users within a fixed time frame (previous day). Different from the above work, Yeniterzi and Callan[94] used only the particular question-related replies to estimate availability.

Besides user availability, we identify two promising directions to cope with the dynamicity in CQA.

User Interest Drifts. Similar to general recommendation systems, the expert recommendation problem for CQA also faces the user interest drift issue[127]. A straightforward solution is to include a decaying factor to suppress questions answered in remote history and focus on users' recent interest (reflected by their shifting answering behavior)[128]. Although the topic drift issue has been studied in multiple areas such as social networks and mobile crowdsourcing[129], it is almost an unexplored topic in the CQA context. More factors such as fluctuations in user behaviors and more sophisticated time series prediction methods could be employed to gain better results.

Dynamic User Expertise. Generally, users' expertise may not be consistent as well, as users' skills may improve over time and the quality of their answers may depend on various factors such as the users' status and other users' behaviors on a given question, not mentioning the impact from the evolution of the Q&A community. Pal *et al.*[130] analyzed the evaluation of experts over time and showed that estimating expertise using temporal data outperforms that using static snapshots of the data. In a very recent work[94], Yeniterzi and Callan incorporated temporal information to model dynamic user expertise and apply two models, namely exponential and hyperbolic discounting models, to discount the effect of older records in calculating z-scores. This method is still rather straightforward being equivalent to using a decaying factor. In particular, the z-scores are calculated for each time interval and then discounted according to its temporal distance from the question's hosting interval. In the field of answer quality prediction, Szpektor[131] used a more advanced set of temporal features calculated between the time at which a question and its replies are posted to improve prediction results. The similar method applies to the expert recommendation problem.

In summary, all the above dynamic aspects are suggesting an expert recommendation method suitable to self-evolve in an online fashion[132]. However, none of the above methods is designed to be online-friendly, and it could take tremendous time to retrain the new model when new information becomes available, which is unacceptable in practice as most Q&A systems in the real-world involve a massive amount of data overall. Therefore, a promising research direction is to introduce novel methods that are capable of incrementally learning about users and continuously adapting their recommendation behaviors over time effectively and efficiently. Besides the short of related work, we believe the dynamicity-related research for CQA is still at a preliminary stage, as most of the methods used are relatively simple and predict different aspects of consideration, such as user availability, user interest, and user expertise, separately. Moreover, they have not considered the possible correlations among these aspects. Therefore, another potential point of research is to predict different aspects of features simultaneous using a single, comprehensive model for better results. As an example, Tensor Factorization (TF)[133] may model the correlations among high-dimensional features better than Matrix Factorization (MF).

6.3 Recommending Experts as a Collaborative Group

Since a single user may not be able to well or fully address a given question, most methods would recommend either a ranked list or an unranked group of users instead of one user. The recommended user group is expected to address the question better than a single user, and their answers are supposed to be better than most or ideally all the other possible user groups that contain the same number of members when considered as a whole. An ideal expert group should satisfy the following conditions. First, the question must appeal to all group members so that they are likely to answer the question. Second, the group members should be compatible with one another so that the existence of one user in the group would not discourage another user to answer the question. Third, it is desirable for the group members to complement one another in the knowledge domains required to address the question, given that users may have different sets of skills and different levels of expertise on different skill aspects[134]. Another benefit of group recommendation is the potential to make the recommending technique adaptive to specific scenarios and reduce the generation of redundant information in Q&A communities. For example, by capturing a global insight into an expert group, a method can automatically adjust the number of experts to recommend. In this way, difficult questions may get more answers than easy ones.

To better answer a question, it is necessary to evaluate and select a competitive combination of potential answerers as a collaborative group. Unfortunately, there are rarely any studies on this topic, and the group recommendation methods for traditional recommender system assume known user groups before making a recommendation[135-137]. For example, Pal proposed an expert group recommendation for CQA[138], aiming to find the experts from predefined communities to provide an answer. Given a new question, the authors computed its similarity with the three aspects of features of each community, namely question features, user features, and community features. Then, they used the two k-NN algorithms over the similarity metrics to build a vector of community scores for each community. Finally, they used linear regression, Borda Count, and SVM ranking, as well as the combinations of the above three methods to train a binary classifier for distinguishing the desired from the non-desired communities for the given question. The issue with this method, as well as the traditional group recommendation methods, is that the users are evaluated separately and later put together to form a group. Consequently, the workers' answers may not collectively better address the question from the requester's perspective as the second and third conditions above may not be well met.

Intuitively, the users who have frequently answered the similar question are likely to be compatible with one another[139]. A possible solution following this insight is to propose an expert recommendation scheme that aims at selecting the best subset (e.g., of a size of k) of collaborative users by learning their co-occurrence in the same thread and topical expertise simultaneously. The selection of a group of collaborative users could also borrow ideas from two closely related topics, namely optimal task decomposition[140] and user group evaluation. Task decomposition is the opposite approach of group formation, which aims to break the knowledge requirement into sub-requirements and find a user for every sub-requirement to compose the final group. User group evaluation aims to set better heuristics to promote better recommendation results and answers to the given question. On this aspect, Chang and Pal[93] hypothesized that valuable question-answer threads are those where several people collaborate. A natural application of this hypothesis in group expert recommendation is that, instead of aiming to maximize the requester's satisfaction on the given question, we could select the group of collaborative users in such a way that maximizes the long-term value to a broader audience of the thread.

6.4 Leveraging External Information

To address the cold start problem and the sparsity of data, especially for users with the low-level activity, it is crucial to leverage external data to facilitate the better expert recommendation in CQA. The types of non-CQA data may vary depending on the external services and platforms. Typically, they include the "about me" description, homepage, blogs, micro-blogs, or social networking sites. For example, many users make their GitHub homepages public in Stack Overflow, which could be an external source of information to estimate the users' expertise.

Until now, existing expert recommendation researches for CQA only uses external information in a narrow scope in a relatively simplistic manner. They mostly focus on users' social attributes including interrelationship in social networks (either within or outside the Q&A system)[141], and they either obtain user attributes through heuristic statistics outside of the model or combine users' social attributes with the original expert recommendation model by linear interpolation. The limited related work includes the followings. Srba *et al.*[119] used non-CQA sources of data in topic-model-based approaches as a supplement for Q&A activities in expertise estimation; Zhao *et al.*[96] considered both topical interest and the "following relations" between the users to build a user-to-user graph. The graph is then used in combination with past question-answering activities of users to evaluate and rank users. Instead of using social attributes as heuristics to estimate user expertise, both Liu and Jansen[142] and Luo *et al.*[143] directly used users' social characteristics as additional features in addition to user expertise for the expert recommendation in CQA; Zhao *et al.*[95] combined heuristics from two different social networks, i.e., social following and social friendship on Twitter and Quora for ranking answerers.

Besides the explicit links to external information sources for the users in a Q&A system, we identify two promising fields of research that could avail the detection and use of external information by an expert recommendation method. The first is account linkage techniques, which aim to identify and link to accounts of the same users on different systems such as websites. By automatically detecting the linked account of a user in CQA, the external information for this user could be efficiently extracted and utilized. The second is cross-domain learning, represented by transfer learning techniques, which aim to utilize the information in the related or similar domains to help learn user models for a targeted domain. Though great potentials in the Q&A domain, both techniques have not yet currently been introduced to CQA.

6.5 More Comprehensive Solutions

Despite hybrid methods have considered multiple aspects of concern comprehensively, the research in this area is still at a preliminary stage as many of those methods simple combine the calculation results on different aspects as a weighted sum. Considering this deficiency, it is beneficial to develop more comprehensive methods. To this end, we advocate several approaches beyond the existing techniques for the expert recommendation in CQA: factorization machines, ensemble learning, graph embedding, and deep learning models. We will briefly discuss them in the following.

Factorization machines (FM)[144] is a matrix factorization based machine learning models similar to linear regression models. It represents a generic approach that combines the generality of feature engineering with the superiority of factorization models in estimating interactions between the massive variables in a vast domain. FM model has the advantages of embedded variable interactions, reliable estimation of a linear number of parameters under high sparsity, and applicability to a variety of prediction tasks including regression, binary classification, and ranking. All these advantages make FM a better replacement of traditional recommendation methods such as matrix factorization and tensor factorization for the expert recommendation in CQA.

Ensemble learning[145] is a method that combines multiple learning algorithms to obtain better performance compared with each individual learning algorithm alone. It generally includes the parallel and sequential ensemble learning and applies to various problems such as classification, regression, feature selection, and anomaly detection. Therefore, it could be potentially used to recommend experts in CQA. Following the sequential or parallel ensemble paradigms, the candidate experts are either filtered by one learning algorithm after another or obtained by merging the recommended lists of experts by different algorithms. Gradient tree boosting[146], represented by XGBoost[147], has been another class of ensemble models that shines in many applications in recent years. It is also promising yet unexplored in the context of CQA.

Graph embedding[148] is an effective and efficient way to solve various graph analytics problems such as node classification, node recommendation, and link prediction while well overcoming the high computation and space cost issues of traditional graph analytics methods. It converts the graph data into a low dimensional space which maximally preserves the graph's properties and structural information. In fact, the graph is an ideal form to represent the complicated interactions among various aspects of users, questions, and answers in CQA, especially when considering multiple aspects of concern and numerous information sources. A benefit of applying the graph embedding approach is that various approximation algorithms and probabilistic solutions are readily available to address complex problems. Besides graph embedding, word embedding[149] is another promising direction that has hardly been explored for the expert recommendation problem.

Deep learning[150] has proven successful in many applications, and various deep learning models such as those based on autoencoders and neural autoregressive models have been applied to recommender systems[151]. Deep learning models have the advantage of utilizing multimodal heterogeneous features and thus have the potential of solving complex problems such as the expert recommendation problem on a large scale. Until now, convolutional neural network (CNN) has been the only deep learning model applied to recommending experts for a given question in CQA[152].

A closely related topic to the expert recommendation in CQA is question answering, which aims to find or generate answers to a given question automatically. This topic is more classic and also heavily researched in the Q&A research domain. Other related research topics include question retrieval, answer quality/probability prediction, and expert finding in broader contexts. In fact, though rare adoption for the expert recommendation problem, deep learning models have been widely applied for question answering in the domain of CQA. Therefore, it could be a good idea to borrow and adapt various sophisticated methods in these related domains to address the expert recommendation problem in CQA.

7 Conclusions

In this survey, we focused on the expert recommendation problem, one of the most significant issues in community question answering (CQA), and reviewed the main techniques and state-of-the-art efforts on addressing the problem. We summarized and compared the existing methods in various aspects, including the datasets, input and output, evaluation metric, the covered aspects of concern, robustness over data distributions, and complexity, followed by discussing the advantages and shortcomings of these methods and pointing out the open issues and promising future research directions. We hope this survey can help readers gain a quick and comprehensive understanding of the state-of-the-art research in the expert recommendation in CQA and inspire more future research in this area.

References

[1] Srba I, Bielikova M. A comprehensive survey and classification of approaches for community question answering. *ACM Transactions on the Web*, 2016, 10(3): Article No. 18.

[2] Liu Q, Agichtein E, Dror G, Maarek Y, Szpektor I. When web search fails, searchers become askers: Understanding the transition. In *Proc. the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2012, pp.801-810.

[3] Guo J, Xu S, Bao S, Yu Y. Tapping on the potential of Q&A community by recommending answer providers. In *Proc. the 17th ACM International Conference on Information and Knowledge Management*, Oct. 2008, pp.921-930.

[4] Su Q, Pavlov D, Chow J H, Baker W C. Internet-scale collection of human-reviewed data. In *Proc. the 16th International Conference on World Wide Web*, May 2007, pp.231-240.

[5] Agichtein E, Castillo C, Donato D, Gionis A, Mishne G. Finding high-quality content in social media. In *Proc. the International Conference on Web Search and Data Mining*, May 2008, pp.183-194.

[6] Li B, King I. Routing questions to appropriate answerers in community question answering services. In *Proc. the 19th ACM International Conference on Information and Knowledge Management*, Oct. 2010, pp.1585-1588.

[7] Fisher D, Smith M, Welser H T. You are who you talk to: Detecting roles in Usenet newsgroups. In *Proc. the 39th Annual Hawaii International Conference on System Sciences, Volume 3*, Jan. 2006.

[8] Viégas F B, Smith M. Newsgroup crowds and AuthorLines: Visualizing the activity of individuals in conversational cyberspaces. In *Proc. the 37th Annual Hawaii International Conference on System Sciences*, Jan. 2004.

[9] Welser H T, Gleave E, Fisher D, Smith M. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 2007, 8(2): 1-32.

[10] Anderson A, Huttenlocher D, Kleinberg J, Leskovec J. Discovering value from community activity on focused question answering sites: A case study of Stack Overflow. In *Proc. the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2012, pp.850-858.

[11] Adamic L A, Zhang J, Bakshy E, Ackerman M S. Knowledge sharing and Yahoo Answers: Everyone knows something. In *Proc. the 17th International Conference on World Wide Web*, Apr. 2008, pp.665-674.

[12] Movshovitz-Attias D, Movshovitz-Attias Y, Steenkiste P, Faloutsos C. Analysis of the reputation system and user contributions on a question answering website: StackOverflow. In *Proc. the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2013, pp.886-893.

[13] Yimam-Seid D, Kobsa A. Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 2003, 13(1): 1-24.

[14] Zhou Y, Cong G, Cui B, Jensen C S, Yao J. Routing questions to the right users in online communities. In *Proc. the 25th IEEE International Conference on Data Engineering*, Apr. 2009, pp.700-711.

[15] Qu M, Qiu G, He X, Zhang C, Wu H, Bu J, Chen C. Probabilistic question recommendation for question answering communities. In *Proc. the 18th International Conference on World Wide Web*, Apr. 2009, pp.1229-1230.

[16] Horowitz D, Kamvar S D. The anatomy of a large-scale social search engine. In *Proc. the 19th International Conference on World Wide Web*, Apr. 2010, pp.431-440.

[17] Bayati S. Security expert recommender in software engineering. In *Proc. the 38th International Conference on Software Engineering Companion*, May 2016, pp.719-721.

[18] Rjab A B, Kharoune M, Miklos Z, Martin A. Characterization of experts in crowdsourcing platforms. In *Proc. the 4th International Conference on Belief Functions: Theory and Applications*, Sept. 2016, pp.97-104.

[19] Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval (1st edition). Addison Wesley, 1999.

[20] Balog K, Azzopardi L, de Rijke M. Formal models for expert finding in enterprise corpora. In *Proc. the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2006, pp.43-50.

[21] Miller D R, Leek T, Schwartz R M. A hidden Markov model information retrieval system. In *Proc. the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 1999, pp.214-221.

[22] Zhou G, Liu K, Zhao J. Joint relevance and answer quality learning for question routing in community QA. In *Proc. the 21st ACM International Conference on Information and Knowledge Management*, Oct. 2012, pp.1492-1496.

[23] Liu X, Croft W B, Koll M. Finding experts in community-based question-answering services. In *Proc. the 14th ACM International Conference on Information and Knowledge Management*, Oct. 2005, pp.315-316.

[24] Lavrenko V, Croft W B. Relevance based language models. In *Proc. the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sept. 2001, pp.120-127.

[25] Liu X, Croft W B. Cluster-based retrieval using language models. In *Proc. the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2004, pp.186-193.

[26] Petkova D, Croft W B. Hierarchical language models for expert finding in enterprise corpora. *International Journal on Artificial Intelligence Tools*, 2008, 17(1): 5-18.

[27] Li B, King I, Lyu M R. Question routing in community question answering: Putting category in its place. In *Proc. the 20th ACM International Conference on Information and Knowledge Management*, Oct. 2011, pp. 2041-2044.

[28] Zheng X, Hu Z, Xu A, Chen D, Liu K, Li B. Algorithm for recommending answer providers in community-based question answering. *Journal of Information Science*, 2012, 38(1): 3-14.

[29] Zhai C, Lafferty J. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems*, 2004, 22(2): 179-214.

[30] Liu M, Liu Y, Yang Q. Predicting best answerers for new questions in community question answering. In *Proc. the 11th International Conference on Web-Age Information Management*, Jul. 2010, pp.127-138.

[31] Riahi F, Zolaktaf Z, Shafiei M, Milios E. Finding expert users in community question answering. In *Proc. the 21st International Conference on World Wide Web*, Apr. 2012, pp.791-798.

[32] Hofmann T. Probabilistic latent semantic indexing. In *Proc. the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 1999, pp.50-57.

[33] Deerwester S, Dumais S T, Furnas G W, Landauer T K, Harshman R. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 1990, 41(6): 391-407.

[34] Wu H, Wang Y, Cheng X. Incremental probabilistic latent semantic analysis for automatic question recommendation. In *Proc. the ACM Conference on Recommender Systems*, Oct. 2008, pp.99-106.

[35] Xu F, Ji Z, Wang B. Dual role model for question recommendation in community question answering. In *Proc. the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2012, pp.771-780.

[36] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, 3: 993-1022.

[37] Du L, Buntine W, Jin H. A segmented topic model based on the two-parameter Poisson-Dirichlet process. *Machine Learning*, 2010, 81(1): 5-19.

[38] Pitman J, Yor M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 1997, 25(2): 855-900.

[39] Sahu T P, Nagwani N K, Verma S. TagLDA based user persona model to identify topical experts for newly posted questions in community question answering sites. *International Journal of Applied Engineering Research*, 2016, 11(10): 7072-7078.

[40] Tian Y, Kochhar P S, Lim E P, Zhu F, Lo D. Predicting best answerers for new questions: An approach leveraging topic modeling and collaborative voting. In *Proc. International Workshops at the International Conference on Social Informatics*, Nov. 2013, pp.55-68.

[41] Zhang J, Ackerman M S, Adamic L. Expertise networks in online communities: Structure and algorithms. In *Proc. the 16th International Conference on World Wide Web*, May 2007, pp.221-230.

[42] Jeon J, Croft W B, Lee J H, Park S. A framework to predict the quality of answers with non-textual features. In *Proc. the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2006, pp.228-235.

[43] Borodin A, Roberts G O, Rosenthal J S, Tsaparas P. Link analysis ranking: Algorithms, theory, and experiments. *ACM Transactions on Internet Technology*, 2005, 5(1): 231-297.

[44] Kleinberg J M. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 1999, 46(5): 604-632.

[45] Jurczyk P, Agichtein E. Discovering authorities in question answer communities by using link analysis. In *Proc. the 16th ACM Conference on Information and Knowledge Management*, Nov. 2007, pp.919-922.

[46] Jurczyk P, Agichtein E. Hits on question answer portals: Exploration of link analysis for author ranking. In *Proc. the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2007, pp.845-846.

[47] Haveliwala T H. Topic-sensitive PageRank. In *Proc. the 11th International Conference on World Wide Web*, May 2002, pp.517-526.

[48] Choetkiertikul M, Avery D, Dam H K, Tran T, Ghose A. Who will answer my question on Stack Overflow? In *Proc. the 24th Australasian Software Engineering Conference*, Sept. 2015, pp.155-164.

[49] Lempel R, Moran S. SALSA: The stochastic approach for link-structure analysis. *ACM Transactions on Information Systems*, 2001, 19(2): 131-160.

[50] Cheng T, Yan X, Chang K C C. EntityRank: Searching entities directly and holistically. In *Proc. the 33rd International Conference on Very Large Data Bases*, Sept. 2007, pp.387-398.

[51] Weng J, Lim E P, Jiang J, He Q. TwitterRank: Finding topic-sensitive influential twitterers. In *Proc. the 3rd ACM International Conference on Web Search and Data Mining*, Feb. 2010, pp.261-270.

[52] Liu X, Bollen J, Nelson M L, de Sompel H. Co-authorship networks in the digital library research community. *Information Processing & Management*, 2005, 41(6): 1462-1480.

[53] Shahriari M, Parekodi S, Klamma R. Community-aware ranking algorithms for expert identification in question-answer forums. In *Proc. the 15th International Conference on Knowledge Technologies and Data-Driven Business*, Oct. 2015, Article No. 8.

[54] Shahriari M, Krott S, Klamma R. Disassortative degree mixing and information diffusion for overlapping community detection in social networks (DMID). In *Proc. the 24th International Conference on World Wide Web*, May 2015, pp.1369-1374.

[55] Bouguessa M, Dumoulin B, Wang S. Identifying authoritative actors in question-answering forums: The case of Yahoo! Answers. In *Proc. the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2008, pp.866-874.

[56] Fagin R, Lotem A, Naor M. Optimal aggregation algorithms for middleware. *Journal of Computer and System Sciences*, 2003, 66(4): 614-656.

[57] Zhu H, Cao H, Xiong H, Chen E, Tian J. Towards expert finding by leveraging relevant categories in authority ranking. In *Proc. the 20th ACM International Conference on Information and Knowledge Management*, Oct. 2011, pp.2221-2224.

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

[59] Liu J, Song Y I, Lin C Y. Competition-based user expertise score estimation. In *Proc. the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2011, pp.425-434.

[60] Lai L C, Kao H Y. Question routing by modeling user expertise and activity in CQA services. In *Proc. the 26th Annual Conference of the Japanese Society for Artificial Intelligence*, Jun. 2012.

[61] Lin Y, Shen H. SmartQ: A question and answer system for supplying high-quality and trustworthy answers. *IEEE Transactions on Big Data*. doi:10.1109/TBDATA.2017.2735442.

[62] Liu D R, Chen Y H, Kao W C, Wang H W. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites. *Information Processing & Management*, 2013, 49(1): 312-329.

[63] Liu Z, Li K, Qu D. Knowledge graph based question routing for community question answering. In *Proc. the 24th International Conference on Neural Information Processing*, Nov. 2017, pp.721-730.

[64] Pal A, Konstan J A. Expert identification in community question answering: Exploring question selection bias. In *Proc. the 19th ACM International Conference on Information and Knowledge Management*, Oct. 2010, pp.1505-1508.

[65] Pal A, Farzan R, Konstan J A, Kraut R E. Early detection of potential experts in question answering communities. In *Proc. the 19th International Conference on User Modeling, Adaptation, and Personalization*, Jul. 2011, pp.231-242.

[66] Zhou T C, Lyu M R, King I. A classification-based approach to question routing in community question answering. In *Proc. the 21st International Conference on World Wide Web*, Apr. 2012, pp.783-790.

[67] Ji Z, Wang B. Learning to rank for question routing in community question answering. In *Proc. the 22nd ACM International Conference on Information & Knowledge Management*, Oct. 2013, pp.2363-2368.

[68] Dijk D, Tsagkias M, de Rijke M. Early detection of topical expertise in community question answering. In *Proc. the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2015, pp.995-998.

[69] Le L T, Shah C. Retrieving rising stars in focused community question-answering. In *Proc. the 8th Asian Conference on Intelligent Information and Database Systems*, Mar. 2016, pp.25-36.

[70] Dror G, Koren Y, Maarek Y, Szpektor I. I want to answer, who has a question? Yahoo! Answers recommender system. In *Proc. the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2011, pp.1109-1117.

[71] Pal A, Harper F M, Konstan J A. Exploring question selection bias to identify experts and potential experts in community question answering. *ACM Transactions on Information Systems*, 2012, 30(2): Article No. 10.

[72] Burel G, Mulholland P, He Y, Alani H. Predicting answering behaviour in online question answering communities. In *Proc. the 26th ACM Conference on Hypertext & Social Media*, Sept. 2015, pp.201-210.

[73] Burges C J, Ragno R, Le Q V. Learning to rank with nonsmooth cost functions. In *Proc. the Neural Information Processing Systems Conference*, Dec. 2006, pp.193-200.

[74] Cao Z, Qin T, Liu T Y, Tsai M F, Li H. Learning to rank: From pairwise approach to listwise approach. In *Proc. the 24th International Conference on Machine Learning*, Jun. 2007, pp.129-136.

[75] Cheng X, Zhu S, Chen G, Su S. Exploiting user feedback for expert finding in community question answering. In *Proc. the IEEE International Conference on Data Mining Workshop*, Nov. 2015, pp.295-302.

[76] Wang N, Abel M H, Barthés J P, Negre E. An answerer recommender system exploiting collaboration in CQA services. In *Proc. the 20th IEEE International Conference on Computer Supported Cooperative Work in Design*, May 2016, pp.198-203.

[77] Dom B, Paranjpe D. A Bayesian technique for estimating the credibility of question answerers. In *Proc. SIAM International Conference on Data Mining*, Apr. 2008, pp.399-409.

[78] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*, 2009, 42(8): 30-37.

[79] Cho J H, Li Y, Girju R, Zhai C. Recommending forum posts to designated experts. In *Proc. IEEE International Conference on Big Data*, Oct. 2015, pp.659-666.

[80] Singh A P, Gordon G J. Relational learning via collective matrix factorization. In *Proc. the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2008, pp.650-658.

[81] Yang B, Manandhar S. Tag-based expert recommendation in community question answering. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2014, pp.960-963.

[82] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization. In *Proc. the 20th International Conference on Neural Information Processing Systems*, Dec. 2008, pp.1257-1264.

[83] Zhou G, Liu K, Zhao J. Topical authority identification in community question answering. In *Proc. Chinese Conference on Pattern Recognition*, Sept. 2012, pp.622-629.

[84] Liu X, Ye S, Li X, Luo Y, Rao Y. ZhihuRrank: A topic-sensitive expert finding algorithm in community question answering websites. In *Proc. the 14th International Conference on Web-Based Learning*, Nov. 2015, pp.165-173.

[85] Yang J, Peng S, Wang L, Wu B. Finding experts in community question answering based on topic-sensitive link analysis. In *Proc. the 1st IEEE International Conference on Data Science in Cyberspace*, Jun. 2016, pp.54-60.

[86] Rao Y, Xie H, Liu X, Li Q, Wang F L, Wong T L. User authority ranking models for community question answering. *Journal of Intelligent & Fuzzy Systems*, 2016, 31(5): 2533-2542.

[87] Zhao T, Bian N, Li C, Li M. Topic-level expert modeling in community question answering. In *Proc. the SIAM International Conference on Data Mining*, May 2013, pp.776-784.

[88] Zhou G, Lai S, Liu K, Zhao J. Topic-sensitive probabilistic model for expert finding in question answer communities. In *Proc. the 21st ACM International Conference on Information and Knowledge Management*, Oct. 2012, pp.1662-1666.

[89] Yang L, Qiu M, Gottipati S, Zhu F, Jiang J, Sun H, Chen Z. CQArank: Jointly model topics and expertise in community question answering. In *Proc. the 22nd ACM International Conference on Information & Knowledge Management*, Oct. 2013, pp.99-108.

[90] Yan Z, Zhou J. A new approach to answerer recommendation in community question answering services. In *Proc. the 34th European Conference on Information Retrieval*, Apr. 2012, pp.121-132.

[91] Yin H, Hu Z, Zhou X, Wang H, Zheng K, Nguyen Q V H, Sadiq S. Discovering interpretable geo-social communities for user behavior prediction. In *Proc. the 32nd IEEE International Conference on Data Engineering*, May 2016, pp.942-953.

[92] Suryanto M A, Lim E P, Sun A, Chiang R H. Quality-aware collaborative question answering: Methods and evaluation. In *Proc. the 2nd ACM International Conference on Web Search and Data Mining*, Feb. 2009, pp.142-151.

[93] Chang S, Pal A. Routing questions for collaborative answering in community question answering. In *Proc. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2013, pp.494-501.

[94] Yeniterzi R, Callan J. Moving from static to dynamic modeling of expertise for question routing in CQA sites. In *Proc. the 9th International AAAI Conference on Web and Social Media*, May 2015, pp.702-705.

[95] Zhao Z, Zhang L, He X, Ng W. Expert finding for question answering via graph regularized matrix completion. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(4): 993-1004.

[96] Zhao Z, Wei F, Zhou M, Ng W. Cold-start expert finding in community question answering via graph regularization. In *Proc. the 20th International Conference on Database Systems for Advanced Applications*, Apr. 2015, pp.21-38.

[97] Nam K K, Ackerman M S, Adamic L A. Questions in, knowledge in? A study of Naver's question answering community. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, Apr. 2009, pp.779-788.

[98] Li X L, Foo C S, Tew K L, Ng S K. Searching for rising stars in bibliography networks. In *Proc. the 14th International Conference on Database Systems for Advanced Applications*, Apr. 2009, pp.288-292.

[99] Daud A, Abbasi R, Muhammad F. Finding rising stars in social networks. In *Proc. the 18th International Conference on Database Systems for Advanced Applications*, Apr. 2013, pp.13-24.

[100] Deng H, King I, Lyu M R. Formal models for expert finding on DBLP bibliography data. In *Proc. the 8th IEEE International Conference on Data Mining*, Dec. 2008, pp.163-172.

[101] Hashemi S H, Neshati M, Beigy H. Expertise retrieval in bibliographic network: A topic dominance learning approach. In *Proc. the 22nd ACM International Conference on Information & Knowledge Management*, Oct. 2013, pp.1117-1126.

[102] Mimno D, McCallum A. Expertise modeling for matching papers with reviewers. In *Proc. the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2007, pp.500-509.

[103] Bagdouri M. Cross-platform question routing for better question answering. In *Proc. the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2015, pp.1053-1053.

[104] Richardson M, White R W. Supporting synchronous social Q&A throughout the question lifecycle. In *Proc. the 20th International Conference on World Wide Web*, Mar. 2011, pp.755-764.

[105] Java A, Kolari P, Finin T, Oates T. Modeling the spread of influence on the blogosphere. In *Proc. the 15th International World Wide Web Conference*, May 2006, pp.22-26.

[106] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In *Proc. the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug 2003, pp.137-146.

[107] Pal A, Counts S. Identifying topical authorities in microblogs. In *Proc. the 4th ACM International Conference on Web Search and Data Mining*, Feb. 2011, pp.45-54.

[108] Campbell C S, Maglio P P, Cozzi A, Dom B. Expertise identification using email communications. In *Proc. the 12th International Conference on Information and Knowledge Management*, Nov. 2003, pp.528-531.

[109] Dom B, Eiron I, Cozzi A, Zhang Y. Graph-based ranking algorithms for e-mail expertise analysis. In *Proc. the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, Jun. 2003, pp.42-48.

[110] Shetty J, Adibi J. Discovering important nodes through graph entropy the case of Enron email database. In *Proc. the 3rd International Workshop on Link Discovery*, Aug. 2005, pp.74-81.

[111] Mockus A, Herbsleb J D. Expertise browser: A quantitative approach to identifying expertise. In *Proc. the 24th International Conference on Software Engineering*, May 2002, pp.503-512.

[112] Wei W, Lee J, King I. Measuring credibility of users in an e-learning environment. In *Proc. the 16th International Conference on World Wide Web*, May 2007, pp.1279-1280.

[113] Fu Y, Xiang R, Liu Y, Zhang M, Ma S. A CDD-based formal model for expert finding. In *Proc. the 16th ACM Conference on Information and Knowledge Management*, Nov. 2007, pp.881-884.

[114] Balog K, Bogers T, Azzopardi L, de Rijke M, van den Bosch A. Broad expertise retrieval in sparse data environments. In *Proc. the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2007, pp.551-558.

[115] Pasca M A, Harabagiu S M. High performance question/answering. In *Proc. the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sept. 2001, pp.366-374.

[116] Fang H, Zhai C. Probabilistic models for expert finding. In *Proc. the 29th European Conference on Information Retrieval*, Apr. 2007, pp.418-430.

[117] Macdonald C, Ounis I. Voting for candidates: Adapting data fusion techniques for an expert search task. In *Proc. the 15th ACM International Conference on Information and Knowledge Management*, Nov. 2006, pp.387-396.

[118] Liu Y, Bian J, Agichtein E. Predicting information seeker satisfaction in community question answering. In *Proc. the 31st International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2008, pp.483-490.

[119] Srba I, Grznar M, Bielikova M. Utilizing non-QA data to improve questions routing for users with low QA activity in CQA. In *Proc. the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Aug. 2015, pp.129-136.

[120] Fagin R, Kumar R, Sivakumar D. Comparing top k lists. *SIAM Journal on Discrete Mathematics*, 2003, 17(1): 134-160.

[121] Herlocker J L, Konstan J A, Terveen L G, Riedl J T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 2004, 22(1): 5-53.

[122] Fang L, Huang M, Zhu X. Question routing in community based QA: Incorporating answer quality and answer content. In *Proc. the ACM SIGKDD Workshop on Mining Data Semantics*, Aug. 2012, Article No. 5.

[123] Sung J, Lee J G, Lee U. Booming up the long tails: Discovering potentially contributive users in community-based question answering services. In *Proc. the 7th International AAAI Conference on Weblogs and Social Media*, Jul. 2013, pp.602-610.

[124] Yin H, Zhou X, Shao Y, Wang H, Sadiq S. Joint modeling of user check-in behaviors for point-of-interest recommendation. In *Proc. the 24th ACM International Conference on Information and Knowledge Management*, Oct. 2015, pp.1631-1640.

[125] Yin H, Cui B. Spatio-Temporal Recommendation in Social Media (1st edition). Springer Singapore, 2016.

[126] Liu Q, Agichtein E. Modeling answerer behavior in collaborative question answering systems. In *Proc. the 33rd European Conference on Information Retrieval*, Apr. 2011, pp.67-79.

[127] Yin H, Zhou X, Cui B, Wang H, Zheng K, Nguyen Q V H. Adapting to user interest drift for POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(10): 2566-2581.

[128] Szpektor I, Maarek Y, Pelleg D. When relevance is not enough: Promoting diversity and freshness in personalized question recommendation. In *Proc. the 22nd International Conference on World Wide Web*, May 2013, pp.1249-1260.

[129] Tong Y, Cao C C, Chen L. TCS: Efficient topic discovery over crowd-oriented service data. In *Proc. the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2014, pp.861-870.

[130] Pal A, Chang S, Konstan J A. Evolution of experts in question answering communities. In *Proc. the 6th International AAAI Conference on Weblogs and Social Media*, Jun. 2012, pp.274-281.

[131] Cai Y, Chakravarthy S. Answer quality prediction in Q/A social networks by leveraging temporal features. *International Journal of Next-Generation Computing*, 2013, 4(1): 127.

[132] Tong Y, She J, Ding B, Wang L, Chen L. Online mobile micro-task allocation in spatial crowdsourcing. In *Proc. the 32nd IEEE International Conference on Data Engineering*, May 2016, pp.49-60.

[133] Yin H, Chen H, Sun X, Wang H, Wang Y, Nguyen Q V H. SPTF: A scalable probabilistic tensor factorization model for semantic-aware behavior prediction. In *Proc. IEEE International Conference on Data Mining*, Nov. 2017, pp.585-594.

[134] Yin H, Cui B, Huang Y. Finding a wise group of experts in social networks. In *Proc. the 7th International Conference on Advanced Data Mining and Applications*, Nov. 2011, pp.381-394.

[135] O'Connor M, Cosley D, Konstan J A, Riedl J. PolyLens: A recommender system for groups of users. In *Proc. the 7th European Conference on Computer-Supported Cooperative Work*, Sept. 2001, pp.199-218.

[136] Ye M, Liu X, Lee W C. Exploring social influence for recommendation: A generative model approach. In *Proc. the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2012, pp.671-680.

[137] Gorla J, Lathia N, Robertson S, Wang J. Probabilistic group recommendation via information matching. In *Proc. the 22nd International Conference on World Wide Web*, May 2013, pp.495-504.

[138] Pal A. Metrics and algorithms for routing questions to user communities. *ACM Transactions on Information Systems*, 2015, 33(3): Article No. 14.

[139] Feng W, Zhu Q, Zhuang J, Yu S. An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. *Cluster Computing*. https://doi.org/10.1007/s10586-017-1576-y, June 2018.

[140] Tong Y, Chen L, Zhou Z, Jagadish H V, Shou L, Lv W. SLADE: A smart large-scale task decomposer in crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering*. doi:10.1109/TKDE.2018.2797962.

[141] Atkinson J, Maurelia A. Redundancy-based trust in question-answering systems. *Computer*, 2017, 50(1): 58-65.

[142] Liu Z, Jansen B J. Predicting potential responders in social Q&A based on non-QA features. In *Proc. ACM CHI Conference on Human Factors in Computing Systems*, Apr. 2014, pp.2131-2136.

[143] Luo L, Wang F, Zhou M, Pan Y, Chen H. Who have got answers?: Growing the pool of answerers in a smart enterprise social QA system. In *Proc. the 19th International Conference on Intelligent User Interfaces*, Feb. 2014, p.716.

[144] Rendle S. Factorization machines. In *Proc. IEEE International Conference on Data Mining*, Dec. 2010, pp.9951000.

[145] Zhou Z H. Ensemble Methods: Foundations and Algorithms (1st edition). CRC Press, 2012.

[146] Friedman J H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 2001, 29(5): 1189-1232.

[147] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In *Proc. the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp.785-794.

[148] Yan S, Xu D, Zhang B, Zhang H J, Yang Q, Lin S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(1): 40-51.

[149] Huang C, Yao L, Wang X *et al.* Expert as a service: Software expert recommendation via knowledge domain embedding in Stack Overflow. In *Proc. the 24th IEEE Internation Conference on Web Services*, June 2017, pp.317-324.

[150] Yin H, Wang W, Wang H, Chen L, Zhou X. Spatial-aware hierarchical collaborative deep learning for POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(11): 2537-2551.

[151] Lin S, Hong W, Wang D, Li T. A survey on expert finding techniques. *Journal of Intelligent Information Systems*, 2017, 49(2): 255-279.

[152] Zheng C, Zhai S, Zhang Z. A deep learning approach for expert identification in question answering communities. arXiv:1711.05350, 2017. https://arxiv.org/abs/1711.05350, Jun. 2018.

Xianzhi Wang is a lecturer with School of Software, University of Technology Sydney, Sydney. He received his B.E. degree from Xi'an Jiaotong University, Xi'an, M.E. and Ph.D. degrees from Harbin Institute of Technology, Harbin, all in computer science, in 2007, 2009, and 2014 respectively. His research interests include Internet of Things, data management, machine learning, and services computing. He received ARC Discovery Early Career Researcher Award (DECRA) in 2017 and IBM Ph.D. Fellowship Award in 2013.

Chaoran Huang is currently a Ph.D. candidate at School of Computer Science and Engineering, University of New South Wales, Sydney. He received his B.E. degree from Tianjin Polytechnic University, Tianjin, in 2014. His research interests include data mining, Internet of Things, and service-oriented computing.

Lina Yao is a lecturer at School of Computer Science and Engineering, University of New South Wales, Sydney. She received her Ph.D. degree in computer science from University of Adelaide, Adelaide, in 2014. Her research interests include data mining and machine learning applications with the focus on Internet of Things, recommender systems, human activity recognition, and Brain Computer Interface. She is the recipient of ARC Discovery Early Career Researcher Award (DECRA) and Inaugural Vice Chancellor's Women's Research Excellence Award in 2015.

Boualem Benatallah is a scientia professor and research group leader at University of New South Wales, Sydney. He received his Ph.D. degree in computer science from Grenoble University, Grenoble. His research interests include Web services composition, quality control in crowdsourcing services, crowdsourcing for vulnerability discovery, data curation, cognitive services engineering, and cloud services orchestration.

Manqing Dong is currently a Ph.D. candidate at School of Computer Science and Engineering, University of New South Wales, Sydney. She received her B.E. degree from Jilin University, Jilin, and her M.Sc. degree in the City University of Hong Kong, Hong Kong. Her research interests include anomaly detection, data mining, deep learning, statistical learning, and probabilistic graphical models. She is a student member of ACM and IEEE.