Knowledge-Based Systems 211 (2021) 106535

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Topic sensitive hybrid expertise retrieval system in community question answering services

Dipankar Kundu [a,*], Rajat Kumar Pal [a], Deba Prasad Mandal [b]

[a] Department of Computer Science and Engineering, University of Calcutta, Kolkata 700106, India
[b] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India

A R T I C L E I N F O

Article history:
Received 2 November 2019
Received in revised form 27 July 2020
Accepted 13 October 2020
Available online 14 October 2020

Keywords:
Community question answering
Expertise retrieval
Social network analysis
Topic sensitive model

A B S T R A C T

Here, we propose a topic sensitive hybrid expertise retrieval system in community question answering services. We introduce three new expertise signatures: knowledge, reputation, and authority. These signatures consider the questions, and hence, their answerers from a topic sensitive perspective. We estimate the knowledge of an answerer on a new question based on the previously answered subset of questions with similar topic distributions to the new question. The reputation of an answerer, moreover, is derived from the qualities of previously answered questions by the answerer with similar distributions of topics. Furthermore, we propose a topic sensitive authority model. It considers some topic related information associated with questions and the relationships among their answerers. We compare the proposed method with 26 existing methods on 4 real-world datasets using 5 performance measures. It outperforms the comparing algorithms in 91.73% (477 out of 520) cases.

1. Introduction

Community question answering (CQA) services allow a user to post her questions and other users to answer those questions. Some of the popular CQA portals are Yahoo! Answers,[1] Answers.com,[2] Quara,[3] Stack Overflow,[4] and Baidu Zhidao.[5] In the recent past, CQA services have gained a lot of popularity for their usefulness. The existence of a large number of unanswered questions, however, is a major problem in CQA services. To address this problem, several approaches, including *question routing* [1,2], have been proposed. Question routing involves the routing of a new question to its experts. *Expertise retrieval* [3,4], an essential component of question routing, is the process of finding potential answerers (experts) for a given question. There are several approaches to expertise retrieval [1–16]. These approaches usually consider different expertise signatures at the category level. In our opinion, however, an answerer is usually an expert on a few topics (subcategories) within a subject (category). Therefore, considering signatures at category levels may not always represent the expertise of an answerer properly. Using this philosophy/assumption, here, we consider that each question belongs to a set of topics, and each answerer is an expert for a set of topics. Thus, instead of category level signatures, here, we hybridize multiple topic-sensitive expertise signatures for expertise retrieval.

In this paper, we introduce a topic sensitive hybrid expertise retrieval system (TSHER) in CQA services. Given a new question, TSHER considers three new expertise signatures: *knowledge*, *reputation*, and *authority*. We hybridize these three signatures and estimate final expertise score. We compare the proposed method with 26 existing approaches on 4 real-world datasets using 5 performance measures. TSHER is found to outperform the comparing algorithms in 91.73% (477 out of 520) cases. We discuss the exact contribution of this work in Section 3, after presenting the philosophies/assumptions of this work.

2. Related work

There are different categories of approaches for expertise retrieval in CQA services. Network-based [5–13] and topic-based [17–25] approaches are two prominent categories among them. Recently, there have been some attempts to hybridize these two approaches [26–32]. Note that, there are several works [1–3,14–16,33,34] that do not fall into these three categories. We do not discuss such works here to keep this paper concise. In CQA services, there are also some other associated tasks, such as, retrieval of alike questions [35], prediction of response time [36], and

* Corresponding author.
E-mail addresses: d.kundu7681@gmail.com (D. Kundu), pal.rajatk@gmail.com (R.K. Pal), dpmandal@isical.ac.in (D.P. Mandal).

1 http://answers.yahoo.com/.
2 https://www.answers.com/.
3 http://www.quora.com/.
4 http://stackoverflow.com/.
5 http://zhidao.baidu.com/.

https://doi.org/10.1016/j.knosys.2020.106535
0950-7051/

quality estimation of replies [37], which we do not discuss here due to the same reason. Moreover, other then CQA services there are some other approaches that provide an alternative way to users for satisfying their information needs. For instance, question answering (QA) services [38] accept questions in the form of natural languages and provide machine generated answers unlike our work. We also excluded these works.

2.1. Network based approaches

In these approaches, a user-to-user network is initially built by employing an inherent relationship among the users. This relationship is based on their activities in the community. Then, a link analysis technique is applied on the constructed network to calculate the authority of each user. A notable network-based approach is community expertise network (CEN) [5], where experts are chosen based on either "z-score" or "ExpertiseRank" algorithm [5]. The "z-score" measures the authority of an answerer based on in-degrees and out-degrees, and "ExperiseRank" is a ranking strategy based on PageRank algorithm [39]. Jurczyk and Agichtein [6,7] constructed a CEN and employed the hyperlink-induced topic search (HITS) [40] based algorithm to estimate the authority of a user. Zhu et al. [8,9] measured the relevances among categories, constructed an "extended category link graph", and estimated the authority of a user in the extended network. The "Personalized PageRank" method was proposed in [10]. It uses a personalization vector to assign preferences of individuals while computing their authorities. Aslay et al. [11] utilized the inherent competition among the co-answerers to select the best answerer. They proposed a competition based expertise network (CBEN). Then, they applied a link analysis technique to the CBEN for identifying authoritative users. Shahriari et al. [12] proposed a method that first detects a set of overlapping communities, and then, estimates the community-aware authority of a user. An expertise retrieval system proposed in [13] finds expert answerers via social referral chains considering a dynamic *k*-hop social network. First, the system allows users to exchange information to find known neighbors and to build social referral chains. Then, to find experts, it forwards each question in the social referral chains.

2.2. Topic based approaches

Topic models are powerful tools to construct knowledge based systems. Hence, it has been extensively used in CQA services [17–25] as well as in other applications [41,42]. During the development of a topic model, two assumptions are *usually* made. First, each document in the collection represents a probabilistic mixture of topics. Second, a topic itself is a probabilistic distribution of words. Expertise retrieval systems use extracted topics to associate a new question with an answerer to determine her expertise on the given question. A generalized topic model for expert finding in CQA services was introduced in [17], where the content of an answer is related to the content of the corresponding question. Then, they developed a Bayesian network for finding experts. In another prominent work, Qu et al. [18] employed a probabilistic latent semantic analysis based scheme for determining the interest of a user. They also proposed a question recommendation system. In a notable work on a topic model, a probabilistic dual role model [19] has been proposed. This model considers two roles of a user: the role of an asker and the role of an answerer. Later, Zongyang et al. [22] introduced another role based topic model, which considers three roles: the role of as an asker, the role of an answerer, and the role of a voter. Fatemeh et al. [20] investigated the performance of two topic models along with two word-based methods for expertise retrieval. In [21], Zhao et al. proposed a "topic-level expert learning" (TEL) model to find experts in CQA services. This model first generates some topic-specific information, and then, uses this information for finding experts on different topics. A multivariate beta mixture model to distinguish between authoritative and nonauthoritative users has been presented in [23]. This method first extracts features associated with each answerer using LDA as the associated topic model. For this, it also uses some statistics. Next, it models the estimated set of feature vectors using the beta mixture model. The components of the beta mixture model are used to identify authoritative users. In [24], the authors proposed an LDA-style user intimacy model for expertise retrieval, which learns the intimacy among users over topics using their social interactions in the community. In [25], a user correlation model has been proposed to adequately measure the level of the relationships among users in terms of their preferences of topics. It also considers the relevance of contents between each question and each of its answers. Note that, unlike our work, this work modified an existing topic model.

2.3. Hybrid approaches

In this category, we consider the approaches that hybridizes of network-based and topic model-based components. Such models are usually called topic sensitive network models. Zhou et al. [26] proposed a topic-sensitive algorithm for finding experts in CQA services. This algorithm first extracts the topic distributions of the users. Then, it applies an algorithm similar to PageRank [39] to estimate the authority of a user. Liu et al. [27] proposed a probabilistic "topic expertise model" (TEM), which uses both information from tags to help in learning topics and a hybridization of Gaussian mixtures to model the information regarding voting. Based on TEM, the authors in [27] proposed CQARank, an extension of the PageRank algorithm, to estimate the expertise scores of users on different topics. In [28], a "hybrid analysis model" (HAM) has been developed to solve the problem of expert finding in CQA services. It combines a topic analysis approach with a topic sensitive network model. ZhihuRank [29], a topic sensitive probabilistic model, has been developed to estimate the authority of a user on a new question. It considers a link analysis technique along with the similarities based on topics between users and questions. A personalized recommendation method for routing a new question to a group of experts has been proposed by Lin et al. in [30]. This method considered a topic model and improved the traditional HITS to better adapt the structural characteristics of CQA services. Wang et al. [31] proposed "TPLMRank" algorithm to measure the expertise score of a user to find experts in CQA services. This algorithm uses topic and professional level information by combining textual contents and link analysis techniques. In [32], researches incorporated the outputs from four different approaches that include a topic model and a network model.

Now, we discuss the differences as well as novelties of TSHER compared to these related hybrid approaches. First, the works presented in [27,28,30] utilize different modified versions of the LDA model. Nonetheless, the work in [31] uses the LDA model along with a modified version of it. However, similar to the works in [26,29,32], in this work, we use the original LDA model as a topic model without any modification. Second, works presented in [26,28–31] extract topics form the entire user profile, whereas we extract the topics from each question. Third, the works in [26, 28–30] incorporate the topic interests of users and the topic-specific similarity between users in the network model. Again, the works in [27,31] incorporate the topic interests of users, the topic expertise of users, and the similarity between users on a specific topic in the network model. In contrast to these works, the network model of TSHER includes dynamic link weights, topic-specific preferences of answerers, the association between two answerers on a particular topic, and the topic-wise reliabilities of the answerers. Fourth, while all of the above hybrid methods construct a single network, in this work, we construct a network corresponding to each topic.

3. Objectives, philosophies, and contributions

3.1. Objectives of this work

Let there be a CQA service. We denote the set of all archived questions as $Q = \{q_1, q_2, \ldots, q_{|Q|}\}$ and the set of all answerers as $A = \{a_1, a_2, \ldots, a_{|A|}\}$. Let us assume that for each archived question $q \in Q, A_q \in A$ is the set of answerers who have answered q. For a new question $\hat{q}$, expertise retrieval is the process of suggesting answerers having relevant expertise on $\hat{q}$. Here, we propose a method of expertise retrieval, and hence, for a new question $\hat{q}$, the objective of this work is to estimate expertise $\mathcal{E}(a, \hat{q})$ for all $a \in A$. Then, for $\hat{q}$, the system decides a subset of answerers $A^*_{\hat{q}} \subseteq A$ with top expertise. The objective of this work is to propose an intuitive, convenient, and useful technique for expertise retrieval in CQA services. The proposed system should be enriched by the philosophies discussed next.

3.2. Philosophies of this work

To achieve the above objectives, in this work, we make use of the following philosophies/assumptions:

Knowledge on a Topic Most of the expertise retrieval systems assume that an answerer has the expertise for answering a question on a particular category if she has already contributed quality information on that category. Different systems, however, realize this philosophy in different ways. In our opinion, each question corresponds to different topics (subcategories) with different degrees. Let there be a question $\hat{q}$ and a set of the probability distribution of topics $P^{\mathcal{T}}(\hat{q}) = \{p(T_k|\hat{q}) : k \in \{1, 2, \ldots, K\}\}$ associated with it where $\mathcal{T} = \{T_1, T_2, \ldots, T_K\}$ is the set of all topics. Here, $p(T_k|\hat{q})$ is the probability of $\hat{q}$ to be associated with topic T_k and K is the number of topics that we want to be associated with $\hat{q}$. An answerer who has provided a high number of quality answers to the archived questions that are highly associated with $T \in \mathcal{T}$, would be a potential answerer to $\hat{q}$. In other words, a signature that corresponds to a topic (subcategory) level expertise is more effective compared to a signature that corresponds to a category level expertise. For example, it would be more useful to note "Riemann had expertise on *number theory*" and "Riemann had expertise on *differential geometry*" compared to note "Riemann had expertise on *mathematics*". Here, *number theory* and *differential geometry* are two topics (subcategories) associated with the category *mathematics*. We note here that two topics associated with a category may correspond to overlapping concepts. Therefore, instead of using a category wise network, it might be better to construct a set of topic wise networks for identifying experts.

Reputation of an Answerer Let us assume that an answerer a has answered to a set of questions Q_a. Let $\mathcal{T}_a$ be the set of all topics that are associated with the questions in Q_a. Then, a would have a significant expertise on a topic $T \in \mathcal{T}_a$, if a has provided quality answers to a considerable number of questions associated with T. We note here that quantity may also be considered as quality.

Standard of a Question We assume that if a question has obtained a large number of answers, then the question is either easy or has an inherent ambiguity. Therefore, the co-answerers should get a low credit for answering such a question. Note that, a similar assumption has been adopted in [4].

Preference of a Topic An answerer, who has attended a large number of questions associated with a particular topic, should be taken as having a high preference for that topic.

Credit for the Best Answers Consider three answerers a_1, a_2, and a_3, such that, for a given topic T, a_1 has a much higher expertise compared to a_2. Now, consider two cases. In the first case, for a given question, a_1 and a_3 have answered, and in the second case, for a question, a_2 and a_3 have answered. We also assume that in both the cases a_3 has provided the best answers. Then, in the first case, a_3 should obtain a higher credit compared to the second case.

Association of a Question with a Topic Consider two questions, q_1 and q_2, such that, q_1 has a high correspondence to topic T and q_2 has a less correspondence to topic T. Suppose an answerer a has answered both q_1 and q_2. When we assess the expertise of a on T, a should obtain a higher credit for answering q_1.

3.3. Contributions of this work

We contribute in the following ways to incorporate and validate the above assumptions:

1. We define a measure (called *knowledge*) to estimate the knowledge of an answerer on a given question. Here, a high knowledge indicates a high expertise. Note that, each question corresponds to multiple topics with different degrees. We compute the knowledge considering the set of topics, which the new question corresponds to.
2. We propose a strategy to estimate the reputation of an answerer on a set of topics, which the new question corresponds to. For this, we define a measure called the *reputation* of an answerer. Here, a high *reputation* indicates a (probable) high expertise of an answerer.
3. Given a new question, we propose a topic sensitive authority model for estimating the authority of an answerer based on the following considerations. First, instead of considering the whole network structure, we construct a network structure for each topic. Second, for a given network, we propose and employ a scheme to determine the weights of the links dynamically. Third, we introduce a score to estimate the *contribution* of each answerer on a particular topic. We utilize it to estimate the preferences of answerers on a particular topic. Fourth, we define a way to estimate the association between two answerers on a particular topic. Fifth, we propose a way to estimate the topic-wise reliability of an answerer. Finally, we incorporate these schemes to determine the authority of an answerer.
4. Finally, we propose a topic sensitive hybrid expertise retrieval (TSHER) model in CQA services by considering the knowledge, reputation, and authority of an answerer.
5. We compare the proposed method with 26 state-of-the-art algorithms on four real-world datasets using five measures to validate the proposed method.

In this article, for comparison purpose we extensively make use of some results, which we reported earlier in [4]. The work in [4] has the same objective as this work. It, however, is neither a network-based approach or a topic-based approach, nor hybridizes these

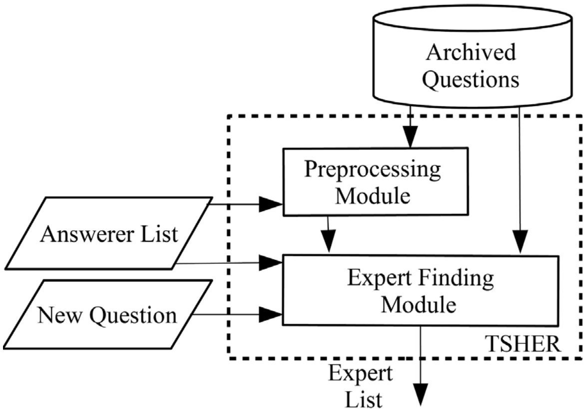

Fig. 1. Architecture of the proposed system.

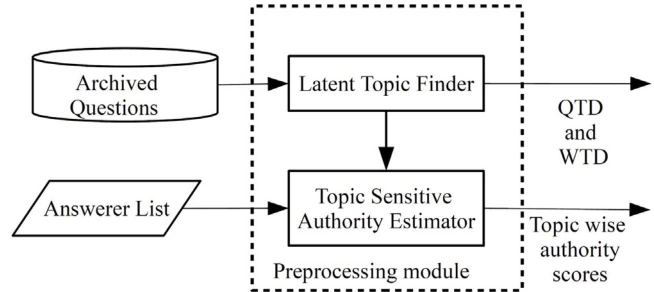

Fig. 2. Architecture of the preprocessing module.

two approaches. Therefore, we choose not to discuss it while discussing the literature. However, as we use the results reported in [4], now we explicitly mark the major differences of our previous work in [4] with this current work. First, in [4], we considered a question as a set of words, whereas, here we consider each question as a set of topics with different probabilities associated with them. Therefore, to find similarities between each of the archived questions and a new question, the previous work considers terms, whereas the current work considers topics. As the number of topics is usually much smaller than the number of terms, the previous work is usually much more computationally costly compared to the current work. Second, our previous work used query likelihood language (QLL) model that we do not use here. Third, in [4], we used a single network corresponding to the entire category. On the contrary, here, we use multiple networks, where each network corresponds to a topic. Finally, we note that although the current work addresses some drawbacks of our previous work, it is not an extension of that work.

4. The proposed system

The proposed system, as illustrated in Fig. 1, is composed of two modules: a preprocessing module and a expert finding module. The preprocessing module is independent of the new question, whereas, the expert finding module depends on the new question. As Fig. 1 illustrates, the proposed system requires a set of archived questions and a list of answerers, and provides a possible list of experts. Next, we describe these two modules. In the following description, we illustrate the architecture of different components visually using the diagrams. Though we discuss different components in a detailed fashion, the connectivity, and hence, the dependencies among these components need to be realized from the diagrams. We choose not to discuss them to avoid redundancy and to keep the description concise.

4.1. Preprocessing module

Fig. 2 illustrates the preprocessing module. It is composed of two segments: a latent topic finder and a topic sensitive authority estimator.

4.1.1. Latent topic finder

In this segment, we find the latent topics of each archived question. In order to capture the latent topics of each archived question, we employ latent Dirichlet allocation (LDA) model [43]. The LDA model has been widely used for mining the latent topics from documents. In the present investigation, we treat each archived question as a document and apply the LDA model to find the distribution of topics for each of the archived questions. The LDA model provides two matrices: word–topic distribution (WTD) and question–topic distribution (QTD). The dimensions of these two matrices are $|\mathcal{V}| \times K$ and $|Q| \times K$, respectively. Here, $\mathcal{V}$ is the set of lexicon on the set of archived questions Q, and K is the number of topics. The distribution of topics associated with a question $q \in Q$ is denoted as $P^{\mathcal{T}}(q) = \{p(T_1|q), p(T_2|q), \dots, p(T_K|q)\}$, where $\mathcal{T} = \{T_1, T_2, \dots, T_K\}$ is the set of all topics. Note that, the size of a question is usually short and it normally corresponds to a limited number of topics. Therefore, for an archived question, a small set of topics with significant probability values are enough to reflect its topic distribution. The same assumption can be found in [44]. With this assumption, here, we consider the topics with the top (highest) K' probability values for all archived questions. In other words, we consider $P^{\mathcal{T}}_{K'}(q) = \{p(T_k|q) : p(T_k|q) \geq \mathtt{max}[P^{\mathcal{T}}(q), K']\}$, where $\mathtt{max}[P^{\mathcal{T}}(q), K']$ denotes the kth largest element of $P^{\mathcal{T}}(q)$, and $K' \in \{1, 2, \dots, K\}$. We divide the set of archived questions Q into a K number of overlapped sub-collections depending on the topic distribution of each archived question, i.e., $Q = Q^{T_1} \cup Q^{T_2} \cup \dots \cup Q^{T_K}$. Here, Q^{T_i}, $i \in \{1, 2, \dots, K\}$, is the set of questions associated with the topic T_i. Note that, an archived question may correspond to more than one sub-collections, i.e., $\forall_{i \neq j} Q^{T_i} \cap Q^{T_j} = \emptyset$ need not to hold true. Throughout this paper, $(\cdot)^T$ indicates that $(\cdot)$ corresponds to the topic T.

4.1.2. Topic sensitive authority estimator

This segment constructs a network corresponding to a topic, and then, estimates the authority of each answerer using a link analysis technique. Here, we adopt the philosophies behind CBEN [11] to construct topic-wise networks. Moreover, we introduce an authority analysis approach to measure the authority of each answerer on a topic. Next, we discuss how we construct the topic-wise network and calculate the weights of the links. Then we discuss how we find the topic-wise preferences of an answerer. After that, we introduce the association between two answerers corresponding to each topic. The reliability of each answerer on a particular topic is presented next. Finally, we introduce an iterative approach to estimate the topic-sensitive authority of each answerer.

Topic-wise network. For each topic $T \in \mathcal{T}$ and the associated set of questions $Q^T \subseteq Q$, we construct a network/graph $G^T = (A^T, E^T)$. Here, $A^T \subseteq A$ is the set of vertices and E^T is the set of directed edges. For this, we follow CBEN [11] that incorporates the competition among the answerers to be selected as the best answerer. Here, we establish a directed edge $e^T_{ij} \in E^T$ from a^T_i to a^T_j if $\exists q \in Q^T$ such that, $a^T_i, a^T_j \in A_q$ and a^T_j is its best answerer. Thus, we have a set of graphs $\{G^{T_1}, G^{T_2}, \dots, G^{T_K}\}$.

We identify a question with a high number of co-answerers as easy. For such a question, therefore, the best answerers should be assigned a low credit. To implement this concept, we assign a weight $w(q)$ to an archived question q based on the number of co-answerers of q as follows:

$$w(q) = \begin{cases} 1 & \text{for } |A_q| \leq 2, \\ \dfrac{1}{|A_q| - 1} & \text{for } |A_q| \geq 2. \end{cases} \tag{1}$$

Now, we assign an initial (un-normalized) weight v_{ij}^T to the edge e_{ij}^T as follows:

$$v_{ij}^T = \sum_{q \in Q^T} \delta_{ij}(q) w(q), \tag{2}$$

where

$$\delta_{ij}(q) = \begin{cases} 1 & \text{if } a_i^T \text{ and } a_j^T \text{ are co-answerers of } q \\ & \text{and } a_j^T \text{ is the best answerer,} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Here, we assume $v_{ii}^T = 0, \ \forall_i$, to avoid self-connection. The final weight of the edge e_{ij}^T is then defined by normalizing the corresponding weight as:

$$f_{ij}^T = \begin{cases} \dfrac{v_{ij}^T}{\sum_{m=1}^{|A^T|} v_{im}^T} & \text{if } \sum_{m=1}^{|A^T|} v_{im}^T \neq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where, $f_{ij}^T \neq f_{ji}^T$, and $\sum_j f_{ij}^T = 1, \ \forall_i$.

Topic-wise preferences of answerers. For a specific topic T, we assign a topic-wise preference to each answerer. Here, the answerers who are highly interested in answering questions on T are assigned higher preference. We utilize this preference while calculating the authority of an answerer. To incorporate this philosophy, first, we introduce a score, named as *contribution*, that measures the effort of an answerer to answer questions on a particular topic. The *contribution*, of answerer a_j on T (i.e., $\mathcal{C}(a_j^T)$) as follows:

$$\mathcal{C}(a_j^T) = \sum_{q \in Q^T \cap Q_{a_j}} p(T|q), \tag{5}$$

where Q_{a_j} is the set of archive questions answered by a_j, and $p(T|q)$ is the probability value obtained from the LDA model for question q on T. A high value of $\mathcal{C}(a_j^T)$ indicates a high contribution of a_j on T. Now, an answerer with a high contribution on T has a preference on T. To measure the preference of a_j on T, we define a preference of a_j (i.e., $\texttt{pref}[a_j^T]$) as follows:

$$\texttt{pref}[a_j^T] = \frac{\mathcal{C}(a_j^T)}{\sum_{a \in A_T} \mathcal{C}(a^T)}. \tag{6}$$

Note that, a user who is more interested in the topic T should be assigned a higher probability [45]. Here, we introduce topic-wise preferences of answerers, i.e., $\texttt{pref}[\cdot]$ to address this issue.

Topic-wise association between answerers. We introduce an association between co-answerers using the distribution of topics among the archived questions. For this, we associate another weight u_{ij}^T with the link e_{ij}^T in G^T. It takes into account the association of a_i^T and a_j^T on topic T and is defined as

$$u_{ij}^T = \sum_{q \in Q^T} \delta_{ij}(q) p(T|q), \tag{7}$$

where $\delta_{ij}(q)$ is as defined in (3).

Topic-wise reliability of answerers. Next, we introduce a concept of topic-wise reliability. It quantifies the reliability of an answerer in terms of producing valuable information on a given topic. We formulate the topic-wise reliability of a_j^T on T (i.e., $\mathcal{R}(a_j^T)$) as

$$\mathcal{R}(a_j^T) = \sum_{q \in Q^T \cap Q_{a_j}} p(T|q) \rho(q, a_j), \tag{8}$$

where

$$\rho(q, a_j) = \begin{cases} 1 & \text{if } a_j^T \text{ is the best answer of } q, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Now, we normalize the topic-wise reliability of a_j^T on T as

$$\texttt{reli}[a_j^T] = \frac{\mathcal{R}(a_j^T) + \mathcal{R}_{max}^T}{2 \times \mathcal{R}_{max}^T}. \tag{10}$$

Here, $\mathcal{R}_{max}^T$ is the highest reliability score received by any answerer corresponding to T. Note that $\texttt{reli}[\cdot]$ is normalized in the range [0.5, 1]. We do this to ensure that the minimum reliability value is 0.5.

Authority of an answerer. Here, we introduce a simple iterative approach to determine the authority of each answerer considering their activity on a particular topic. Moreover, this approach takes into account all the concepts that we discussed above. We define the authority of a_j^T on T (i.e., $\texttt{AT}[a_j^T]$) as

$$\begin{aligned} \texttt{AT}[a_j^T]_{r+1} = d \sum_{i=1}^{|A^T|} \texttt{AT}[a_i^T]_r \times f_{ij}^T \times u_{ij}^T \times \texttt{reli}[a_i^T] \\ + (1 - d) \times \texttt{pref}[a_j^T], \end{aligned} \tag{11}$$

where $d \in [0, 1]$ is a damping factor. Here, f_{ij}^T, u_{ij}^T, $\texttt{reli}[a_i^T]$, and $\texttt{pref}[a_j^T]$ are as defined in (4), (7), (10), and (6), respectively. Moreover, $\texttt{AT}[a_j^T]_r$ or $\texttt{AT}[a_j^T]_{r+1}$ indicate the authority of a_j on T in the rth and the $(r+1)^{th}$ step, respectively. We iterate $\texttt{AT}[a_j^T]$ while $|\texttt{AT}[a_j^T]_{r+1} - \texttt{AT}[a_j^T]_r| > \tau$ and $r < r_{max}$. Here, τ and r_{max} are two user defended parameters, which denotes the threshold for convergence and the number of iterations for convergence, respectively.

4.2. Expert finding module

We visually illustrate the architecture of the expert finding module in Fig. 3. As Fig. 3 shows, the expert finding module has there segments. They are topic finder, question similarity estimator, and online expertise estimator. Next, we describe these three segments.

4.2.1. Topic finder

For a new question $\hat{q}$, this segment uses the WTD of the LDA model [43] trained in the latent topic finder segment (discussed in Section 4.1.1) to infer the distribution of its topics. The distribution of topics associated with a word/term $t \in \mathcal{V}$ is denoted as $P^{\mathcal{T}}(t)$. Using the distribution of topics $P^{\mathcal{T}}(t)$ = $\{p(T_1|t), p(T_2|t), \ldots, p(T_K|t)\}$ associated with the word/term t, we determine the distribution of topics associated with new question $\hat{q}$ as $P^{\mathcal{T}}(\hat{q})$ = $\{p(T_1|\hat{q}), p(T_2|\hat{q}), \ldots, p(T_K|\hat{q})\}$, where

$$p(T|\hat{q}) = \frac{\phi(T, \hat{q})}{\sum_{\bar{T} \in \mathcal{T}} \phi(\bar{T}, \hat{q})} \tag{12}$$

and

$$\phi(T|\hat{q}) = \frac{1}{|\hat{q} \cap \mathcal{V}|} \sum_{t \in \hat{q} \cap \mathcal{V}} p(T|t). \tag{13}$$

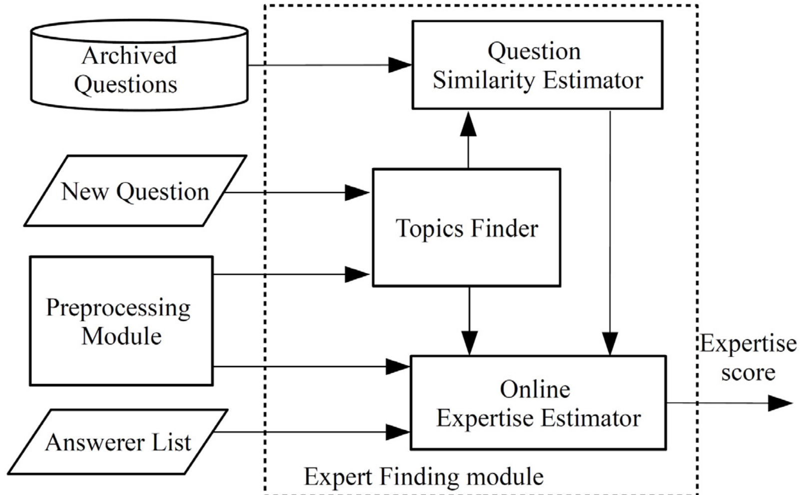

Fig. 3. Architecture of the expert finding module.

Here, $p(T|t)$ is the probability of t to be associated with T. Now, we select the top K' number of topics for $\hat{q}$. That is, we consider $P^{\mathcal{T}}_{K'}(\hat{q}) = \{p(T_k|\hat{q}) : p(T_k|\hat{q}) \geq \mathtt{max}[P^{\mathcal{T}}(\hat{q}), K']\}$, where $\mathtt{max}[P^{\mathcal{T}}(\hat{q}), K']$ denotes the kth largest item of $P^{\mathcal{T}}(\hat{q})$, and $K' \in \{1, 2, \ldots, K\}$.

4.2.2. Question similarity estimator

This segment finds the similarity $\mathtt{sim}[\hat{q}, q]$ between a new question $\hat{q}$ and each of the archived questions $q \in Q$. Using normalized Kullback–Leibler (KL) divergence [46] as

$$\mathtt{sim}[\hat{q}, q] = \frac{1}{2}\{D_{KL}(q \parallel \hat{q}) + D_{KL}(\hat{q} \parallel q)\} \tag{14}$$

where

$$D_{KL}(q \parallel \hat{q}) = \sum_{T \in P^{\mathcal{T}}_{K'}(q)} p(T|q) \log\left(\frac{p(T|q)}{p(T|\hat{q})}\right). \tag{15}$$

4.2.3. Online expertise estimator

Here, the term "online" implies that it works in the expertise estimation (testing) phase, and does not indicate an incremental learning. Fig. 4 presents the architecture of this segment. As Fig. 4 illustrates, this segment consists of five blocks, where each block produces a separate score. The final block, called expertise calculator, produces the final expertise for ranking the answerers by given a score. We next discuss these five blocks.

Knowledge estimator. The purpose of this block is to estimate the knowledge of an answerer corresponding to a new question. Here, the knowledge is realized with respect to the question topics. For this, we aggregate the similarity scores of the new question with each of the archived questions that are answered by the answerer. We call it the knowledge of the answerer a_j to answer the new question $\hat{q}$ (i.e., $\mathtt{know}[a_j, \hat{q}]$) as

$$\mathtt{know}[a_j, \hat{q}] = \sum_{q \in Q_{a_j}} \mathtt{sim}[\hat{q}, q]. \tag{16}$$

Reputation estimator. We estimate the reputation of each answerer on a new question considering the topic of the new question. This is done by aggregating the similarity score of the new question with the set of archived questions that have been answered by the answerer and have been select as the best answerer. We define the reputation of an answerer a_j for a new question $\hat{q}$ (i.e., $\mathtt{repu}[a_j, \hat{q}]$) as

$$\mathtt{repu}[a_j, \hat{q}] = \sum_{q \in Q_{a_j}} \mathtt{sim}[\hat{q}, q] \times \rho(q, a_j), \tag{17}$$

where $\rho(q, a_j)$ is as defined in (9).

Trustworthiness estimator. We combine the knowledge and the reputation to find the trustworthiness of each answerer using the reciprocal rank fusion (RRF) technique [47]. The trustworthiness of a_j on $\hat{q}$ (i.e., $\mathtt{trust}[a_j, \hat{q}]$) is estimated as

$$\mathtt{trust}[a_j, \hat{q}] = \frac{\gamma}{\mathtt{rank}[\mathtt{know}[a_j, \hat{q}]]} + \frac{(1-\gamma)}{\mathtt{rank}[\mathtt{repu}[a_j, \hat{q}]]}, \tag{18}$$

where $\mathtt{rank}[\cdot]$ denotes the rank of $[\cdot]$ in the corresponding scenario. Here, $\gamma \in [0, 1]$ is a factor for balance between the importance of knowledge and reputation.

Authority estimator. Corresponding to a question we estimate the authority of each answerer. For this, we utilize the topic-sensitive authority (see Section 4.1.2). We define the authority of a_j on $\hat{q}$ (i.e., $\mathtt{auth}[a_j, \hat{q}]$) as

$$\mathtt{auth}[a_j, \hat{q}] = \sum_{T \in P^{\mathcal{T}}(\hat{q})} p(T|\hat{q}) \times \mathtt{AT}[a_j^T]. \tag{19}$$

where $\mathtt{AT}(a_j^T)$ is estimated using (11).

Expertise calculator. Finally, we combine the trustworthiness and the authority of an answerer to determine her expertise of a_j for $\hat{q}$ (i.e., $\mathcal{E}(a_j, \hat{q})$ as

$$\mathcal{E}(a_j, \hat{q}) = \max\left(\frac{1}{\mathtt{rank}[\mathtt{trust}[a_j, \hat{q}]]}, \frac{1}{\mathtt{rank}[\mathtt{auth}[a_j, \hat{q}]]}\right). \tag{20}$$

Based on $\mathcal{E}(a_j, \hat{q})$, a ranked list for $\hat{q}$ is generated as the output of the proposed system.

5. Experiments, results, and discussions

5.1. Datasets and measures

We have used four datasets: Movie, Music, Celebrity, and History, which have been generated and used by us in [4]. More information about these datasets are available in Table 1 of [4]. Here, we follow the same training and testing sets and preprocessing steps as presented in [4]. We consider five evaluation measures in this work: mean reciprocal rank (MRR) [48], accuracy [49], precision at the top N (P@N) [50], recall at the top N (R@N) [50], and matching set count at the top N (MSC@N) [50].

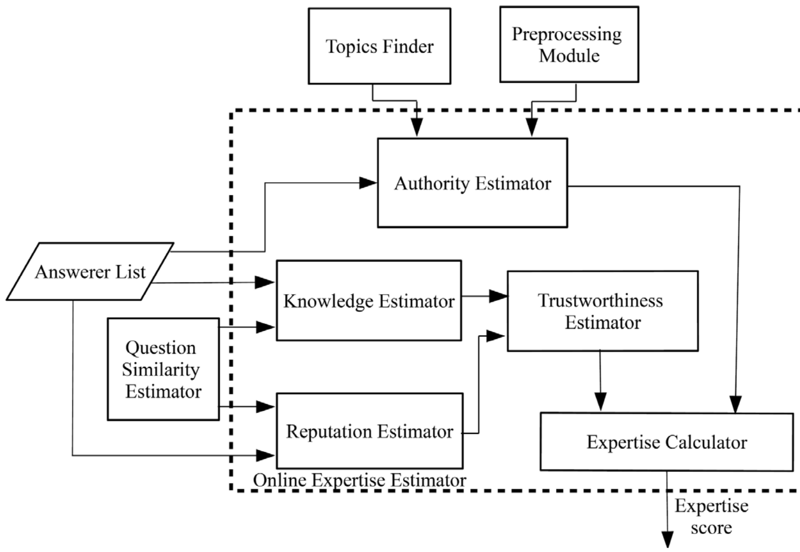

Fig. 4. Architecture of the online expertise estimator segment.

5.2. Comparison with other methods

5.2.1. Comparing methods

We compare the proposed method with following 26 state-of-the-art methods: "(1) document model (DM) [51], (2) CBEN-PageRank [11], (3) CBEN-HITS [11], (4) Adaptive HITS [52], (5) cluster-based language model (CBDM) [3], (6) community-aware PageRank based on speaker listener label propagation algorithm (CA-PR-SLPA) [12], (7) community-aware HITS based on speaker listener label propagation algorithm (CA-HITS-SLPA) [12], (8) ExpertiseRank [53], (9) ExpertRank [54], (10) HITS [7], (11) latent link analysis (LLA) approach [55], (12) NEWHITS without topic model [56], (13) PageRank [39], (14) personalized PageRank (P-PageRank) [10], (15) query likelihood (QL) model [3], (16) TF-IDF [20], (17) TopicRank-based document priors (TBDP) [57], (18) Z-score [53]" [4], (19) hybrid expertise retrieval system using PageRank (HER-PR) [4], (20) hybrid expertise retrieval system using HITS (HER-HITS) [4]. Here, we borrow the results of these 20 methods from [4]. Moreover, we execute the following six methods for comparison: (1) topical PageRank (TPR) [45], (2) a topic-sensitive random surfer model abbreviated as TSPR [26], (3) topic-sensitive expert ranking (TSER) [28], (4) ZhihuRank [29], (5) QRec [32], and (6) active expert finding (AEF) [34].

5.2.2. Experimental protocol

The parameter settings of the 20 algorithms for which we borrow results from [4] are available in Sections 4.1.2 and 4.1.3 of [4]. To execute TPR [45], TSPR [26] and TSER [28], we set the damping factor as 0.85 following [28,39]. However, following [29], we set the damping factor as 0.75 for ZhihuRank. Similarly, following [26], to implement topic models (LDA) for TPR, TSPR, TSER, and ZhihuRank, we set the number of topics as 15. In the LDA model, we use Dirichlet priors $\alpha = 50/K$ (K is number of topics) and $\beta = 0.05$; and 200 iterations following [26]. Now, for implementing QRec [32] we set $c = 0.90$ the probability of jumping back to the seed node and to implement topic model (LDA) we set number of topic as $K = 50$, Dirichlet priors $\alpha = 50/K$ and $\beta = 0.05$; and 200 iterations. Finally, for implementing AEF, we set balancing parameter $\lambda = 0.5$ and time window $\delta = 2$ as in [34]. Moreover, we only consider the texts of the questions for implementing the topic models. Note that, in [4], all of the 20 methods were repeated 10 times whenever a randomness was caused by the LDA model, and then, the average results were reported. To make this comparison fair, here, we execute TPR, TSPR, TSER, ZhihuRank, QRec, and the proposed algorithm 10 times and report the average results. This is because each of these six methods incorporates randomness due to the use of LDA model. We, however, do not repeat AEF as it has no associated randomness.

TSHER needs nine parameters as described in Table 1 along with their chosen values. Four parameters, α, β, max_iter, and d, have been chosen following [26,26,26], and [39], respectively. The remaining five parameters, K, K', τ, r_{max}, and γ have been chosen based on our initial ad-hoc experiments. Two out of these five parameters, τ and r_{max}, needs to be small and large enough, respectively, for our method to have satisfactory performance. With a decrease in τ and with an increase in r_{max}, therefore, the performance of TSHER is supposed to either increase or remain unaltered. Moreover, we choose $\gamma = 0.5$ to have equal importance on both the ranks determined with knowledge and reputation while estimating the trustworthiness. To keep this work concise, we do not investigate how the choices of K and K' impact the performance of TSHER. However, as the performance of TSHER is found to be satisfactory with the chosen nine parameter values, they can be considered to be a good choice. Note that, a judicious selection of these values, specifically depending on the dataset, may result in a better performance of TSHER.

In a very few cases, it may happen that none of the terms of a new question $\hat{q}$ is present in the corpus vocabulary $\mathcal{V}$. Then, the proposed method cannot determine the distribution of $\hat{q}$. To address this issue, we use ExpertiseRank [5] algorithm to rank answerers in such extreme cases.

5.2.3. Comparison outcome

The borrowed results of the 20 comparing algorithms are available in Tables 4, 5, 6 and 7 of [4]. The results of the six state-of-the-art algorithms that we execute here are provided in Table 2. The results of the proposed method are provided in Table 3. The integer values provided within the parentheses corresponding to each entry in Table 3 indicates the number of comparing cases for which our method has outperformed the comparing one. An entry of $x/130$ in the last column of Table 3 indicates that in x out of 130 (= 26 methods × 5 measures) cases, the proposed method has outperformed the comparing algorithm on the corresponding dataset. Similarly, an entry of $x/104$ in the last row of Table 3 indicates that in x out of 104 (= 26 methods × 4 datasets) cases, the proposed method has outperformed the comparing algorithm in terms of the corresponding measure.

Table 1
Parameter settings of the proposed TSHER system.

Description	Notation	Value	Way of choosing the value
The number of topics [43]	K	50	Based on ad-hoc experiments
Dirichlet prior on the per-document topic distributions in the LDA model [43]	α	$50/K$	Following [26]
Dirichlet prior on the per-topic word distribution in the LDA model [43]	β	0.05	Following [26]
The number of iterations for the LDA model	max_iter	200	Following [26]
Topic control parameters	K'	5	Based on ad-hoc experiments
Damping factor in (11)	d	0.85	Following [39]
Threshold of convergence in (11)	τ	0.001	Based on ad-hoc experiments
The number of iterations for convergence in (11)	r_{max}	1000	Based on ad-hoc experiments
Controlling parameter for fusion in (18)	γ	0.5	Based on ad-hoc experiments

Table 2
The results of TPR, TSPR, TSER, ZhihuRank, QRec, and AEF.

Dataset	Methods	MRR	P@30	R@30	Accuracy	MSC@30
Movie	TPR	0.0915	0.0049	0.2036	0.8506	0.4057
	TSPR	0.0920	0.0047	0.1987	0.8470	0.3990
	TSER	0.0903	0.0047	0.1970	0.8526	0.3965
	ZhihuRank	0.1053	0.0050	0.2097	0.8197	0.4132
	QRec	0.0912	0.0046	0.1500	0.8541	0.2913
	AEF	0.0281	0.0019	0.0755	0.8872	0.1663
Music	TPR	0.0740	0.0051	0.1712	0.8678	0.3770
	TSPR	0.0686	0.0045	0.1613	0.8631	0.3551
	TSER	0.0751	0.0049	0.1678	0.8695	0.3711
	ZhihuRank	0.0797	0.0053	0.1795	0.8284	0.3820
	QRec	0.0941	0.0063	0.1883	0.8736	0.3893
	AEF	0.0456	0.0031	0.1058	0.8825	0.2277
Celebrity	TPR	0.1047	0.0045	0.2302	0.8134	0.4048
	TSPR	0.1053	0.0043	0.2243	0.8086	0.3939
	TSER	0.1058	0.0043	0.2239	0.8161	0.3927
	ZhihuRank	0.1072	0.0045	0.2226	0.7641	0.3964
	QRec	0.0523	0.0023	0.0775	0.7632	0.1876
	AEF	0.0378	0.0021	0.0619	0.8534	0.1636
History	TPR	0.1525	0.0320	0.3174	0.9523	0.6523
	TSPR	0.1569	0.0324	0.3220	0.9490	0.6568
	TSER	0.1548	0.0316	0.3158	0.9525	0.6488
	ZhihuRank	0.1692	0.0326	0.3248	0.9193	0.6575
	QRec	0.0513	0.0123	0.1217	0.8627	0.2832
	AEF	0.0091	0.0027	0.0210	0.9469	0.0709

Finally, an entry of 477/520 in the cell corresponding to the last row and the last column denotes that in 477 out of 520 (= 26 methods × 5 measures × 4 datasets), i.e., 91.73% cases the proposed method has outperformed the comparing algorithm. The proposed method has performed the best on Celebrity dataset (has performed the best in 122 out of 130, i.e., 93.85% cases) and has performed the worst on History dataset (has performed the best in 115 out of 130, i.e., 88.46% cases). Again, the proposed method has performed the best when measured in terms of MRR (has performed the best in 102 out of 104, i.e., 98.08% cases) and has performed the worst when measured in terms of accuracy (has performed the best in 85 out of 104, i.e., 81.73% cases).

ZhihuRank performs the best in 5 out of 20 (= 4 datasets × 5 measures) cases among 26 comparing methods that we consider in this article. ZhihuRank stands out to be the best comparing algorithm among these comparing methods. We, consequently, emphasize on comparing the performance of our method with ZhihuRank. Out of 20 cases, TSHER performs better than, equal to, and worse than ZhihuRank for 13, 1, and 6 cases, respectively.

5.3. How different philosophies affect the proposed authority model

Now, we answer two important questions. First, there are several philosophies used in the proposed authority model. Does each of the philosophies help to enhance the performance? Second, there could be many ways to incorporate these philosophies. Does the proposed method take into account these philosophies in an efficient way? To answer this two questions, we investigate how different philosophies affect the proposed authority model. For this, we consider a baseline model and five improvements over the baseline model. We consider each of these models as individual expertise retrieval systems and use the same computational protocols to generate the results. Note that, here, each improved model incorporates all the philosophies accumulated by its preceding models. Next, we discuss these models, and then, we discuss the experimental outcomes. The parameter values used for this experiments are the same as our previous experiment (provided in Table 1).

5.3.1. Authority models

Baseline. It follows CBEN [11] to construct the network considering the entire category. Then, it applies PageRank [39] as a link analysis technique to find the authority of an answerer.

Topic-wise authority (TA). It follows CBEN [11] to construct a network corresponding to each topic in the category. Then, it applies the following link analysis technique on each of the networks. It estimates the topic-wise authority ($\mathtt{auth}_{\mathtt{TA}}[a_j, \hat{q}]$) of an answerer as follows:

$$\mathtt{auth}_{\mathtt{TA}}[a_j, \hat{q}] = \sum_{T \in P^{\mathcal{T}}_{K'}(\hat{q})} p(T|\hat{q})\mathcal{P}(a^T_j). \tag{21}$$

Here, $\mathcal{P}(a^T_j)$ is updated iteratively as follows:

$$\mathcal{P}(a^T_j)_{r+1} = d\sum_{i=1}^{|A^T|} \mathcal{P}(a^T_i)_r \hat{f}_{ij} + (1-d)\frac{1}{|A^T|}, \tag{22}$$

where $d \in [0, 1]$ is a damping factor and r denotes the iteration index. Moreover, $\mathcal{P}(a^T_j)$ is similar to the PageRank algorithm [39].

Furthermore, the normalized weight $\hat{f}_{ij}$ calculated as

$$\hat{f}_{ij} = \begin{cases} \dfrac{\sum_{q \in Q^T} \delta_{ij}(q)}{\sum_{m=1}^{|A^T|}\sum_{q\in Q^T}\delta_{im}(q)} & \text{if } \displaystyle\sum_{m=1}^{|A^T|}\sum_{q\in Q^T}\delta_{im}(q) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

where $\delta_{ij}(q)$ is as defined in (3).

Weighted topic-wise authority (WTA). It follows CBEN [11] to construct a network corresponding to each topic in the category. Then, it applies the following link analysis technique on the networks. It estimates the weighted topic-wise authority ($\mathtt{auth}_{\mathtt{WTA}}[a_j, \hat{q}]$) of an answerer considering the dynamic link weights as follows:

$$\mathtt{auth}_{\mathtt{WTA}}[a_j, \hat{q}] = \sum_{T \in P^{\mathcal{T}}(\hat{q})} p(T|\hat{q})\hat{\mathcal{P}}(a^T_j), \tag{24}$$

where $\hat{\mathcal{P}}(a^T_j)$ is updated iteratively as follows:

$$\hat{\mathcal{P}}(a^T_j)_{r+1} = d\sum_{i=1}^{|A^T|} \hat{\mathcal{P}}(a^T_i)_r f_{ij} + (1-d)\frac{1}{|A^T|}. \tag{25}$$

Here, f_{ij} is define as in (4).

Table 3
The results of the proposed TSHER system.

Dataset	MRR	P@30	R@30	Accuracy	MSC@30	# Best
Movie	0.1018 (25)	0.0050 (25)	0.2020 (24)	0.8933 (21)	0.4029 (24)	119/130
Music	0.0870 (25)	0.0057 (25)	0.1716 (23)	0.8925 (23)	0.3874 (25)	121/130
Celebrity	0.1391 (26)	0.0049 (26)	0.2192 (22)	0.8737 (22)	0.4142 (26)	122/130
History	0.2148 (26)	0.0331 (24)	0.3219 (23)	0.9755 (19)	0.6579 (23)	115/130
# Best	102/104	100/104	92/104	85/104	98/104	477/520

Topic preference-based authority (TPA). It uses CBEN [11] to construct a network corresponding to each topic in the category. Then, it applies a link analysis technique on the networks. It estimates the topic preference-based authority ($\mathtt{auth}_{\mathtt{TPA}}[a_j, \hat{q}]$) of an answerer considering her preferences as follows:

$$\mathtt{auth}_{\mathtt{TPA}}[a_j, \hat{q}] = \sum_{T \in P^{\mathcal{T}}(\hat{q})} p(T|\hat{q})\tilde{\mathcal{P}}(a_j^T), \tag{26}$$

where $\tilde{\mathcal{P}}(a_j^T)$ is updated iteratively as follows:

$$\tilde{\mathcal{P}}(a_j^T)_{r+1} = d\sum_{i=1}^{|A^T|} \tilde{\mathcal{P}}(a_j^T)_r f_{ij} + (1-d) \times \mathtt{pref}^{\mathrm{T}}[a_j], \tag{27}$$

Here, f_{ij} is define as in (4) and $\mathtt{pref}^{\mathrm{T}}[a_j]$ is as defined in (6).

Topic association-based authority (TAA). It uses CBEN [11] to construct a network corresponding to each topic in the category. Then, it applies a link analysis approach on each of the networks. It estimates the topic association-based authority ($\mathtt{auth}_{\mathtt{TAA}}[a_j, \hat{q}]$) of an answerer considering the association between two answerers as follows:

$$\mathtt{auth}_{\mathtt{TAA}}[a_j, \hat{q}] = \sum_{T \in P^{\mathcal{T}}(\hat{q})} p(T|\hat{q})\bar{\mathcal{P}}(a_j^T), \tag{28}$$

where $\bar{\mathcal{P}}(a_j^T)$ is updated iteratively as follows:

$$\bar{\mathcal{P}}(a_j^T)_{r+1} = d\sum_{i=1}^{|A^T|} \bar{\mathcal{P}}(a_j^T)_r f_{ij} u_{ij}^T + (1-d) \times \mathtt{pref}[a_j^T]. \tag{29}$$

Here, f_{ij}, $\mathtt{pref}^{\mathrm{T}}[a_j]$, and u_{ij}^T are as defined in (4), (6), and (7), respectively.

Reliability-aware topic sensitive authority (RTSA). It uses assumptions of CBEN [11] to construct the network corresponding to each topic in the category. Then, it applies a link analysis technique as defined in (11) on each of the networks. It estimates the authority of an answerer as in (19).

5.3.2. Experimental outcomes

Table 4 presents the results of the five authority models (TA, WTA, TPA, TAA, and RTSA) when they are considered independent expertise retrieval systems. In each cell of Table 4, we also provide the percentage improvement compared to the baseline system in the corresponding scenario. A negative percentage, therefore, indicates a deterioration in the performance. Table 4 contains results on four datasets using five measures, and hence, we have 100 ($=4$ datasets $\times$ 5 measures $\times$ 5 models) cases. Additionally, we also perform t-tests (two-tailed) at statistical significance level $\alpha = 0.001$ between each case and the corresponding baseline one. A $*$ symbol corresponding to a cell of Table 4 indicates that the corresponding p-value is less than the chosen level of significance.

From Table 4, we find improvements in 88 cases and statistically significant improvements in 80 cases. On the other hand, among the 12 cases of deterioration, in 7 cases the decline is significant. When we consider MRR, in all of the 20 ($=4$ datasets $\times$ 5 models) cases the performance is improved compared to the baseline system, and in 19 cases the improvements are statistically significant. The worst performance is found when we use accuracy to measure the same — in 10 out of 20 cases, the performance is improved. In all of these 10 cases, the improvements are statistically significant. However, in 8 out of 10 cases, the deteriorations are statistically significant.

Now, we examine the results from a dataset oriented perspective. The best performance is obtained for the Music dataset — in 24 out of 25 ($=$ 5 measures $\times$ 5 models) cases the performance is significantly better than the baseline system. The only case (Table 4, dataset: Music, measure: Accuracy, method: WTA), when the performance has deteriorated, is not statistically significant. Again, the performance is the worst on the History dataset — in 20 out of 25 cases the performance is statistically better compared to the baseline system. Note that, on History dataset, when measured in terms of accuracy in each of the 5 cases, the performance is deteriorated.

Next, we consider the five models together. Thus, we have 20 ($=4$ datasets $\times$ 5 measures) cases. Let, $\mathtt{perf}[\cdot]$ denote the performance of $[\cdot]$. Then, in 12 cases we observe non-decreasing performance for these five model, i.e., $\mathtt{perf}[\mathrm{TA}] \leq \mathtt{perf}[\mathrm{WTA}] \leq \mathtt{perf}[\mathrm{TPA}] \leq \mathtt{perf}[\mathrm{TAA}] \leq \mathtt{perf}[\mathrm{RTTA}]$ holds. Therefore, from this experiments we can loosely say that with inclusions of the different philosophies/assumptions, the performance of the system usually increases. Moreover, this experiment illustrates that the philosophies are incorporated in the proposed method in an efficient way. However, there could be other ways to incorporate these philosophies/assumptions in a more effective fashion.

Now we compare the performance of the proposed system, i.e., TSHER with the final authority model, i.e., RTSA. The results of TSHER and RTSA are provided in Tables 3 and 4, respectively. After careful observation of the performance of TSHER and RTSA, we realize that the proposed method has performed better in 12 out of 20 ($=4$ datasets $\times$ 5 measures), i.e., 60% cases. Moreover, the proposed method has performed the best when measured in terms of MRR and accuracy and has performed worst when measured in terms of P@N. These results indicate that the integration of trustworthiness with RTSA improve the performance of the system.

6. Conclusion

Here, we propose a topic sensitive hybrid expertise retrieval system, which we abbreviate as TSHER. Corresponding to a new question, TSHER integrates the knowledge, reputation, and authority of an answerer to determine her expertise. TSHER combines these scores at the topic level. From the perspective of design, TSHER has two modules: an preprocessing module and an expert finding module. We decompose these two modules further and discuss their components with minute details. Different components of TSHER take into account different philosophies. We also discuss these philosophies in a detailed fashion. We compare TSHER with 26 state-of-the-art methods on 4 real-world datasets using 5 measures. TSHER is found to perform the best in 477 out of 520 cases, i.e., 91.73% cases.

This work has some shortfalls. First, we have used two intuitive schemes for fusion of the knowledge, reputation, and

Table 4
Results of authority models.

Dataset	Methods	MRR	P@30	R@30	Accuracy	MSC@30
Movie	TA	0.0861 (+045.79%)*	0.0042 (−003.30%)	0.1584 (−011.20%)*	0.8382 (−004.01%)*	0.3407 (−007.23%)*
	WTA	0.0873 (+047.97%)*	0.0046 (+005.52%)	0.1922 (+007.76%)*	0.8526 (−002.36%)	0.3845 (+004.68%)
	TPA	0.0925 (+056.78%)*	0.0049 (+012.52%)*	0.2022 (+013.34%)*	0.8914 (+002.09%)*	0.4026 (+009.61%)*
	TAA	0.0986 (+067.00%)*	**0.0052** (+017.88%)*	0.2066 (+015.82%)*	**0.8926** (+002.22%)*	0.4120 (+012.18%)*
	RTSA	**0.0995** (+068.49%)*	**0.0052** (+018.91%)*	**0.2080** (+016.59%)*	**0.8926** (+002.22%)*	**0.4144** (+012.84%)*
Music	TA	0.0519 (+066.25%)*	0.0052 (+051.01%)*	0.1604 (+079.92%)*	0.8844 (+003.75%)*	0.3692 (+050.93%)*
	WTA	0.0648 (+107.52%)*	0.0052 (+052.68%)*	0.1520 (+070.53%)*	0.8381 (−001.69%)	0.3565 (+045.75%)*
	TPA	0.0606 (+093.96%)*	0.0052 (+050.32%)*	0.1417 (+058.99%)*	0.8900 (+004.40%)*	0.3419 (+039.78%)*
	TAA	0.0757 (+142.32%)*	**0.0059** (+070.73%)*	0.1686 (+089.15%)*	0.8924 (+004.68%)*	0.3829 (+056.51%)*
	RTSA	**0.0795** (+154.58%)*	0.0058 (+069.31%)*	**0.1697** (+090.39%)*	**0.8925** (+004.69%)*	**0.3856** (+057.61%)*
Celebrity	TA	0.1041 (+008.93%)	0.0041 (+007.20%)	0.1749 (+004.39%)	0.8139 (−004.28%)*	0.3578 (+002.86%)
	WTA	0.1136 (+018.90%)*	0.0043 (+010.53%)*	0.1826 (+008.99%)	0.8147 (−004.19%)*	0.3684 (+005.89%)
	TPA	0.1333 (+039.48%)*	0.0046 (+019.72%)*	0.1957 (+016.85%)*	0.8714 (+002.48%)*	0.3885 (+011.67%)*
	TAA	**0.1350** (+041.27%)*	0.0049 (+027.64%)*	0.2092 (+024.90%)*	**0.8730** (+002.67%)*	0.4070 (+017.00%)*
	RTSA	0.1346 (+040.82%)*	**0.0050** (+029.21%)*	**0.2126** (+026.93%)*	**0.8730** (+002.67%)*	**0.4122** (+018.50%)*
History	TA	0.1850 (+002.05%)*	0.0300 (+006.57%)*	0.2898 (+005.76%)*	0.9598 (−001.93%)*	0.6176 (+003.62%)*
	WTA	0.1866 (+002.93%)*	0.0302 (+006.92%)*	0.2966 (+008.26%)*	0.9602 (−001.88%)*	0.6238 (+004.66%)*
	TPA	**0.1884** (+003.88%)*	0.0310 (+010.03%)*	0.3047 (+011.22%)*	0.9747 (−000.38%)*	0.6342 (+006.41%)*
	TAA	0.1869 (+003.07%)*	0.0333 (+017.97%)*	0.3227 (+017.76%)*	0.9753 (−000.34%)	0.6578 (+010.38%)*
	RTSA	0.1869 (+003.08%)*	**0.0340** (+020.43%)*	**0.3285** (+019.89%)*	0.9752 (−000.36%)*	**0.6655** (+011.66%)*

*p-value $< .001$.

authority of an answerer, there may be other ways to do the same, which may lead to a superior design of the system. However, exploring the design of such systems is not included in the scope of the work. Although, our experimental results have empirically validated the effectiveness of the proposed design. Second, there are nine parameters in TSHER (see Table 1). We have not inspected the parameter dependencies/sensitivities. The choice of these parameters either have been made depending on our initial ad-hoc experiments or following other works (e.g., parameter d from [39]). We have found satisfactory results with this parameter settings. Therefore, in our opinion, the chosen set of parameter values are sufficiently good for a sound performance of TSHER. A judiciously chosen set of parameter values, however, may increase the performance of the system. Third, TSHER considers the text of each question as a single document but the sizes of the questions are usually short. In some cases, it may cause failure in detection of topics using the LDA model.

In the future, we plan to investigate different fusion schemes to take into account the knowledge, reputation, and authority of an answerer. We, moreover, plan to investigate the parameter dependencies of TSHER. Furthermore, we intend to inspect on schemes for query expansion to address the issue caused by the short length of questions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Li, I. King, Routing questions to appropriate answerers in community question answering services, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM, 2010, pp. 1585–1588.

[2] B. Li, I. King, M.R. Lyu, Question routing in community question answering: putting category in its place, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, ACM, 2011, pp. 2041–2044.

[3] X. Liu, W. Croft, M. Koll, Finding experts in community-based question-answering services, in: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM, 2005, pp. 315–316.

[4] D. Kundu, D.P. Mandal, Formulation of a hybrid expertise retrieval system in community question answering services, Appl. Intell. 49 (2) (2019) 463–477.

[5] J. Zhang, M.S. Ackerman, L. Adamic, Expertise networks in online communities: Structure and algorithms, in: Proceedings of the 16th International Conference on World Wide Web, in: WWW '07, ACM, New York, NY, USA, 2007, pp. 221–230.

[6] P. Jurczyk, E. Agichtein, Discovering authorities in question answer communities by using link analysis, in: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, ACM, 2007, pp. 919–922.

[7] P. Jurczyk, E. Agichtein, Hits on question answer portals: exploration of link analysis for author ranking, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2007, pp. 845–846.

[8] H. Zhu, H. Cao, H. Xiong, E. Chen, J. Tian, Towards expert finding by leveraging relevant categories in authority ranking, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, in: CIKM '11, ACM, New York, NY, USA, 2011, pp. 2221–2224.

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

[10] D. Schall, F. Skopik, An analysis of the structure and dynamics of large-scale Q/A communities, in: J. Eder, M. Bielikova, A.M. Tjoa (Eds.), Advances in Databases and Information Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 285–301.

[11] c. Aslay, N. O'Hare, L.M. Aiello, A. Jaimes, Competition-based networks for expert finding, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, in: SIGIR '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 1033–1036.

[12] M. Shahriari, S. Parekodi, R. Klamma, Community-aware ranking algorithms for expert identification in question-answer forums, in: Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business, in: i-KNOW '15, ACM, New York, NY, USA, 2015, pp. 8:1–8:8.

[13] I. Ali, R.Y. Chang, C. Hsu, SOQAS: Distributively finding high-quality answerers in dynamic social networks, IEEE Access 6 (2018) 55074–55089.

[14] D.-R. Liu, Y.-H. Chen, W.-C. Kao, H.-W. Wang, Integrating expert profile, reputation and link analysis for expert finding in question-answering websites, Inf. Process. Manage. 49 (1) (2013) 312–329.

[15] Z. Zhao, L. Zhang, X. He, W. Ng, Expert finding for question answering via graph regularized matrix completion, IEEE Trans. Knowl. Data Eng. 27 (4) (2015) 993–1004.

[16] D.P. Mandal, D. Kundu, S. Maiti, Finding experts in community question answering services: A theme based query likelihood language approach, in: 2015 International Conference on Advances in Computer Engineering and Applications, 2015, pp. 423–427.

[17] J. Guo, S. Xu, S. Bao, Y. Yu, Tapping on the potential of q&a community by recommending answer providers, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, in: CIKM '08, ACM, New York, NY, USA, 2008, pp. 921–930.

[18] M. Qu, G. Qiu, X. He, C. Zhang, H. Wu, J. Bu, C. Chen, Probabilistic question recommendation for question answering communities, in: Proceedings of the 18th International Conference on World Wide Web, in: WWW '09, ACM, New York, NY, USA, 2009, pp. 1229–1230.
[19] F. Xu, Z. Ji, B. Wang, Dual role model for question recommendation in community question answering, in: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, in: SIGIR '12, ACM, New York, NY, USA, 2012, pp. 771–780.
[20] F. Riahi, Z. Zolaktaf, M. Shafiei, E. Milios, Finding expert users in community question answering, in: Proceedings of the 21st International Conference on World Wide Web, in: WWW '12 Companion, ACM, New York, NY, USA, 2012, pp. 791–798.
[21] T. Zhao, N. Bian, C. Li, M. Li, Topic-level expert modeling in community question answering, in: Proceedings of the 2013 SIAM International Conference on Data Mining, SIAM, 2013, pp. 776–784.
[22] Z. Ma, A. Sun, Q. Yuan, G. Cong, A tri-role topic model for domain-specific question answering, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, in: AAAI'15, AAAI Press, 2015, pp. 224–230.
[23] T.P. Sahu, N.K. Nagwani, S. Verma, Multivariate beta mixture model for automatic identification of topical authoritative users in community question answering sites, IEEE Access 4 (2016) 5343–5355.
[24] C. Fu, User intimacy model for question recommendation in community question answering, Knowl.-Based Syst. 188 (2020) 104844.
[25] C. Fu, User correlation model for question recommendation in community question answering, Appl. Intell. 50 (2) (2020) 634–645.
[26] G. Zhou, S. Lai, K. Liu, J. Zhao, Topic-sensitive probabilistic model for expert finding in question answer communities, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, in: CIKM '12, ACM, New York, NY, USA, 2012, pp. 1662–1666.
[27] L. Yang, M. Qiu, S. Gottipati, F. Zhu, J. Jiang, H. Sun, Z. Chen, Cqarank: jointly model topics and expertise in community question answering, in: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, ACM, 2013, pp. 99–108.
[28] H. Li, S. Jin, S. L.I., A hybrid model for experts finding in community question answering, in: 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2015, pp. 176–185.
[29] X. Liu, S. Ye, X. Li, Y. Luo, Y. Rao, Zhihurank: A topic-sensitive expert finding algorithm in community question answering websites, in: F.W. Li, R. Klamma, M. Laanpere, J. Zhang, B.F. Manjón, R.W. Lau (Eds.), Advances in Web-Based Learning – ICWL 2015, Springer International Publishing, Cham, 2015, pp. 165–173.
[30] L. Wang, B. Wu, J. Yang, S. Peng, Personalized recommendation for new questions in community question answering, in: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, in: ASONAM '16, IEEE Press, Piscataway, NJ, USA, 2016, pp. 901–908.
[31] S. Wang, D. Jiang, L. Su, Z. Fan, X. Liu, Expert finding in CQA based on topic professional level model, in: Y. Tan, Y. Shi, Q. Tang (Eds.), Data Mining and Big Data, Springer International Publishing, Cham, 2018, pp. 459–465.
[32] L.T. Le, C. Shah, Retrieving people: Identifying potential answerers in community question-answering, J. Assoc. Inf. Sci. Technol. 69 (10) (2018) 1246–1258.
[33] D. Kundu, R.K. Pal, D.P. Mandal, Preference enhanced hybrid expertise retrieval system in community question answering services, Decis. Support Syst. 129 (2020) 113164.
[34] D. Kundu, R.K. Pal, D.P. Mandal, Finding active experts for question routing in community question answering services, in: Pattern Recognition and Machine Intelligence, Springer International Publishing, Cham, 2019, pp. 320–327.
[35] O. Chergui, A. Begdouri, D. Groux-Leclet, Integrating a Bayesian semantic similarity approach into CBR for knowledge reuse in community question answering, Knowl.-Based Syst. (2019) 104919.
[36] YueLiu, A. Tang, FeiCai, P. Ren, Z. Sun, Multi-feature based question–answerer model matching for predicting response time in CQA, Knowl.-Based Syst. 182 (2019) 104794.
[37] H. Fu, S. Oh, Quality assessment of answers with user-identified criteria and data-driven features in social Q&A, Inf. Process. Manage. 56 (1) (2019) 14–28.
[38] M. Esposito, E. Damiano, A. Minutolo, G.D. Pietro, H. Fujita, Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering, Inform. Sci. 514 (2020) 88–105.
[39] L. Page, S. Brin, R. Motwani, T. Winograd, The Pagerank Citation Ranking: bringing Order to the Web, Tech. rep., Stanford InfoLab, 1999.
[40] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.
[41] Q. Huang, C. Huang, J. Huang, H. Fujita, Adaptive resource prefetching with spatial–temporal and topic information for educational cloud storage systems, Knowl.-Based Syst. 181 (2019) 104791.
[42] X. Chen, Y. Yuan, M.A. Orgun, L. Lu, A topic-sensitive trust evaluation approach for users in online communities, Knowl.-Based Syst. (2020) 105546.
[43] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (Jan) (2003) 993–1022.
[44] Z. Yan, J. Zhou, Optimal answerer ranking for new questions in community question answering, Inf. Process. Manage. 51 (1) (2015) 163–178.
[45] G. Zhou, K. Liu, J. Zhao, Topical authority identification in community question answering, in: C.-L. Liu, C. Zhang, L. Wang (Eds.), Pattern Recognition, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 622–629.
[46] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[47] M. Zhang, R. Song, C. Lin, S. Ma, Z. Jiang, Y. Jin, Y. Liu, L. Zhao, S. Ma, Expansion-based technologies in finding relevant and new information: Thu trec 2002: Novelty track experiments, NIST Special Publ. SP 251 (2003) 586–590.
[48] E.M. Voorhees, Overview of the TREC 2001 question answering track, in: Proceedings of the Tenth Text REtrieval Conference, TREC 2001, Gaithersburg, Maryland, USA, November 13-16, 2001, p. 2001.
[49] Z. Zhao, Q. Yang, D. Cai, X. He, Y. Zhuang, Expert finding for community-based question answering via ranking metric network learning, in: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, in: IJCAI'16, AAAI Press, 2016, pp. 3000–3006.
[50] S. Chang, A. Pal, Routing questions for collaborative answering in community question answering, in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ACM, 2013, pp. 494–501.
[51] K. Balog, L. Azzopardi, M. De Rijke, Formal models for expert finding in enterprise corpora, in: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2006, pp. 43–50.
[52] R. Yeniterzi, Effective and efficient approaches to retrieving and using expertise in social media, SIGIR Forum 49 (2) (2016) 152–153.
[53] J. Zhang, M.S. Ackerman, L. Adamic, Expertise networks in online communities: structure and algorithms, in: Proceedings of the 16th International Conference on World Wide Web, ACM, 2007, pp. 221–230.
[54] G.A. Wang, J. Jiao, A.S. Abrahams, W. Fan, Z. Zhang, Expertrank: A topic-aware expert finding algorithm for online knowledge communities, Decis. Support Syst. 54 (3) (2013) 1442–1451.
[55] Y. Lu, X. Quan, X. Ni, W. Liu, Y. Xu, Latent link analysis for expert finding in user-interactive question answering services, in: Semantics, Knowledge and Grid, 2009. SKG 2009. Fifth International Conference on, IEEE, 2009, pp. 54–59.
[56] L. Wang, B. Wu, J. Yang, S. Peng, Personalized recommendation for new questions in community question answering, in: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2016, pp. 901–908.
[57] J. Liu, B. Jia, H. Xu, B. Liu, D. Gao, B. Li, A topicrank based document priors model for expert finding, in: M. Fei, S. Ma, X. Li, X. Sun, L. Jia, Z. Su (Eds.), Advanced Computational Methods in Life System Modeling and Simulation, Springer Singapore, Singapore, 2017, pp. 334–341.