Multiview Physician-Specific Attributes Fusion for Health Seeking

Liqiang Nie, *Member, IEEE*, Luming Zhang, *Member, IEEE*, Yan Yan, *Member, IEEE*,
Xiaojun Chang, *Member, IEEE*, Maofu Liu, *Member, IEEE*, and Ling Shao, *Senior Member, IEEE*

Abstract—Community-based health services have risen as important online resources for resolving users health concerns. Despite the value, the gap between what health seekers with specific health needs and what busy physicians with specific attitudes and expertise can offer is being widened. To bridge this gap, we present a question routing scheme that is able to connect health seekers to the right physicians. In this scheme, we first bridge the expertise matching gap via a probabilistic fusion of the physician-expertise distribution and the expertise-question distribution. The distributions are calculated by hypergraph-based learning and kernel density estimation. We then measure physicians attitudes toward answering general questions from the perspectives of activity, responsibility, reputation, and willingness. At last, we adaptively fuse the expertise modeling and attitude modeling by considering the personal needs of the health seekers. Extensive experiments have been conducted on a real-world dataset to validate our proposed scheme.

Index Terms—Adaptive fusion, attitude modeling, community-based health services (CHSs), expertise modeling, question routing.

I. INTRODUCTION

THE RISE of digital technologies advances many disciplines including healthcare. Nowadays, when patients struggle with health concerns, some of them start to explore the Internet on their health problems rather than directly visit physicians. In particular, 70% of Canadians turned to the Internet for health-related information in 2009 [1] and 72% of American Internet users searched for health information in 2012 [2]. These studies somehow signal the prevalence of online health seeking behaviors. Thereinto, 92% of the Canadian online health seekers [1] and 77% of the American online health seekers [2] gather their health solutions from the popular and general search engines, such as Google and Yahoo, instead of specialized health portals. However, the debate over accuracy and trustworthiness of health information returned by general search engines never stops [3]. For instance, the study in [4] warns health seekers that most of the sports-related medical information returned by general search engines is incorrect. Therefore, vertical health services that can offer health seekers reliable and personalized knowledge are highly desired.

On the other hand, recent years have witnessed the revolutions in online healthcare systems brought about by the community-based health services (CHSs). HealthTap[1] and HaoDF[2] are the typical examples. They not only allow health seekers to freely post health-oriented questions, but also encourage physicians to provide answers of high-quality. With these services, health seekers can either instantly search similar cases from the archived question answer (QA) repositories or obtain personalized health solutions from the real physicians. Physicians, in turn, can enhance their expertise via interaction with their peers and diverse health seekers. In such context, CHSs play a pivotal role in promoting the health knowledge dissemination and sharing.

In CHSs, the volume of health seekers and their posted questions usually increases at a faster pace than that of physicians. Considering HealthTap as an example, as of March 2013, it had served 7.5 million unique health seekers and received more than 10 million questions[3] monthly. However, it has only attracted less than 68 thousand physicians since its foundation in 2010. Physicians are thus frequently overwhelmed by numerous questions that are beyond their expertise or interests. Meanwhile, health seekers often suffer from a long time waiting for the professional solutions to their specific questions. The time varies from hours to days according to our observation. Such problem, in fact, can be well-addressed by the

L. Nie is with the School of Computer Science and Technology, Shandong University, Jinan 250101, China (e-mail: nieliqiang@gmail.com).

L. Zhang is with the Department of Electric Engineering and Information System, Hefei University of Technology, Hefei 230009, China (e-mail: zglumg@gmail.com).

Y. Yan is with the Department of Information Engineering and Computer Science, University of Trento, Trento 38123, Italy, and also with the Advanced Digital Sciences Center, Singapore 138632 (e-mail: yan@disi.unitn.it).

X. Chang is with the Centre for Quantum Computation and Intelligent Systems, University of Technology Sydney, Sydney, NSW 2007, Australia (e-mail: cxj273@gmail.com).

M. Liu is with the College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: liumaofu@wust.edu.cn).

L. Shao is with the Computer Vision and Artificial Intelligence Group, Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K., and also with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: ling.shao@northumbria.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

---

[1]https://www.healthtap.com/
[2]www.haodf.com
[3]http://techcrunch.com/2013/03/16/

"right" physician-patient match. As reported in this paper,[4] the right physician-patient match can produce better results. On the contrary, an ineffective physician-patient matching mechanism may lead to an extra cost of patients, interest loss of physicians, and hence quality deterioration in community content. Therefore, an effective approach to instantly routing questions to the right physicians is necessary.

Question routing in CHSs is, however, nontrivial due to the following reasons. First, medical concepts in CHSs expressed by the low-literacy laypersons and the highly-technical experts are, more often than not, variant. For example, "shortness of breath" and "breathless" may be used by different users to refer to the same concept "dyspnea." Conventional term-based representation approaches are not applicable to capture such variants or synonyms of medical concepts. Second, different from the general community-based question answering services (cQA), such as Yahoo! Answers, the roles of askers and answerers in CHSs are not exchangeable. Besides, the askers' profiles are inaccessible due to the privacy issues. Hence, it is hard to collect the profiles and past records of the health seekers and leverage them to train the routing models. Third, the qualities of answers are usually jointly affected by physicians' expertise and attitude. Most importantly, health seekers expect personalized health solutions regarding their specific question contexts. Question context refers to the health cues conveyed by the given question, probably including the health seekers' demographic information, physical and mental symptoms, diseases, and medical histories, through which health seekers expect to express their health conditions [5]. These factors together pose tough challenges for health question routing in CHSs.

It is worth mentioning that question routing approaches have been initiated and developed in various cQAs. Roughly speaking, they can be divided into two categories: 1) the global expert discovery [6] and 2) the topic-level expert discovery [7]. The philosophy behind the global expert discovery is to measure the experts' authorities using the links acquired from the past posts, regardless of the given question context [8], [9]. In contrast, topic-level expert discovery approaches solve the problem of question-aware expertise matching in a finer-granularity. Even though great success has been achieved by the existing methods, they will encounter the following challenges when applied to CHSs.

1) The first category of approaches do not take the question context into consideration, which is the key in healthcare. This is because the health conditions, such as symptoms and other demographic cues, are usually conveyed in the question contexts.

2) Those in the second category could find experts according to the similarities between the given question context and the profiles of the candidate experts, but none of them [7], [10]–[13], as far as we know, have encoded the experts' attitudes into their models.

Even worse, the profiles and historical behaviors of health seekers are unable to be leveraged, as they are invisible to the public.

To address the aforementioned problems, in this paper, we present, a question routing scheme for CHSs. This scheme is able to route the given question to the appropriate physicians by jointly considering the personal needs of the health seekers, physicians' expertise, and physicians' attitudes. As illustrated in Fig. 1, our scheme consists of three main components. The first one attempts to model the expertise matching between the given question and candidate physicians. This is accomplished via fusing hypergraph-based learning and kernel density estimation, KDE for short. The hypergraph-based learning calculates the medical expertise distribution through a fuzzy style partition. The constructed hypergraph is capable of characterizing the higher-order relations among physicians, and it is able to seamlessly integrate multifaceted heterogeneous cues, such as physicians' social connections, biographies, and historical experience. On the other hand, inspired by the soft voting strategy, we employ the KDE to estimate the expertise-question distribution. The second component assumes that physicians' attitudes significantly affect their behaviors and thus have key influences on the qualities of their provided answers. We model the expert attitudes from multiple angles, namely, activity, responsibility, reputation, and willingness. Regarding the third component, it adaptively weighs the effects between expertise modeling and attitude modeling according to the context of the given question. This is motivated by the observation that question contexts hold a bias toward expertise or attitude. It is worth emphasis that data are represented by topic-level features, which are robust toward data inconsistency and can capture the data semantics. Extensive experiments on a real-world dataset have validated our routing scheme and demonstrated its advantages over several state-of-the-art baselines.

The main contributions of this paper are threefold.

1) To bridge the expertise matching gap, we propose a probabilistic model that seamlessly integrates the physician-expertise distribution and expertise-question distribution. It is noteworthy that the former distribution was estimated by a hypergraph-based learning algorithm, which explores the higher-order relations among physicians and the heterogeneous information of physicians.

2) To comprehensively model physicians' attributes toward answering general questions, we heuristically define some metrics, including activity, responsibility, reputation, and willingness. We adaptively adjust the effects between expertise and attitudes by considering question contexts.

3) To the best of our knowledge, this is the first work that targets at routing questions to appropriate physicians in CHSs.

The remainder of this paper is organized as follows. Section II reviews the related work. In Section III, we briefly introduce our question routing scheme. Expertise modeling and attitude modeling are detailed in Sections IV and V, respectively. In Section VI, we investigate the relations between expertise and attitude. Section VII presents the experimental results and analysis, followed by our concluding remarks and future work in Section VIII.

---

[4]http://tinyurl.com/zbybmkq
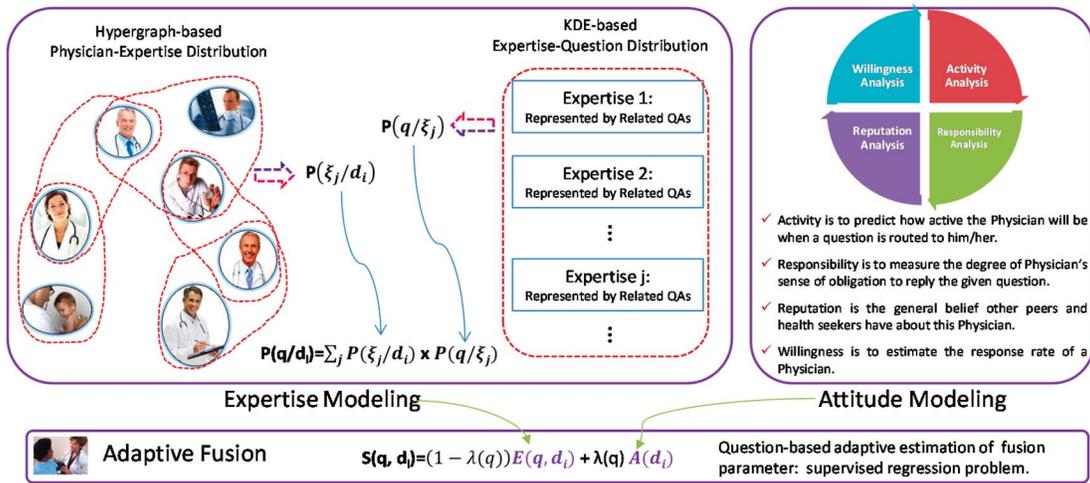
Fig. 1. Schematic of our proposed question routing scheme, consisting of three component. The first two components estimate the matching score between the given question and each physician from the perspective of expertise and attitude, respectively. The third component adaptively balances the impact between these two factors via a supervised regression model.

II. RELATED WORK

A. Question Routing

It is worth mentioning that, to bridge the matching gap between questions and answers, several research efforts have been dedicated to the question routing approaches in the past decades, with well theoretical underpinnings and great practical success. These efforts can be roughly divided into two categories: 1) global expert discovery approaches [14] and 2) topic-level expert discovery approaches [10].

The idea behind the approaches of global experts discovery is to measure question-independent experts' authorities using links acquired from posts and replies at specialized forums [8]. Most of the existing efforts represent the environments as graphs where nodes represent all users and edges represent the interactions between them. Such representations benefit several applications of "link analysis" techniques and graph-based ranking algorithms such as hyperlink-induced topic search (HITS) [15] and PageRank [16]. The output of these algorithms is a ranked list of experts based on their expertise on subjects of interest. Based on the list, the top K experts are considered as the most reliable ones. For example, to measure the prestige of an author in the ACM/IEEE research community, Liu *et al.* [17] employed a graph to represent the co-authorship network, where nodes and edges stand for authors and their collaboration relationships, respectively. Meanwhile, they explored various edge representations, spanning from nonweighted undirected graph, nonweighted directed graph, to weighted directed graph. Another example was introduced in [18], which aims to identify experts from cQAs. Similarly, the data were represented as a graph where nodes correspond to users and edges represent the interactions between askers and answerers. In addition, systems developed in [6] and [19] achieved promising performance for expertise identification from email and the authors concluded that graph-based algorithms seem more suitable for the purpose of expert identification. Bouguessa *et al.* [9] noticed that a major problem of this research line: we have to determine how many users can be chosen as authoritative experts from a ranked list. Toward this

end, they proposed to model the expertise score of each user as a mixture of Gamma distributions. This method enables automatic discrimination between experts and nonexperts.

On the other hand, topic-level expert discovery approaches consider the question-aware expertise matching in a finer-granularity as compared to those in the first category. These approaches usually construct expert profiles first by aggregating historical information. For a given question, all the candidate experts are returned in a ranking list according to the similarity between the given question and the expert profiles. For example, the work in [10] seeks to recommend the right expert with two steps. Its first step attempts to discover latent topics in the content of QA pairs as well as latent interests of users to build user profiles. It then recommends experts for newly posted questions according to the discovered topics as well as term-level information of questions and experts. Another example was introduced in [11]. This paper presents three language model-based approaches to represent the expertise of users relying on their previous question answering activities, and it then reranks the user expertise measured in terms of probability by utilizing the structural relations among users in a forum system. Experimental results on the real-world data show that this approach can effectively find promising experts for new questions. Meanwhile, the routing system developed in [7] treats the question routing problem as a classification task, and extracts a rich set of local features and global features which are able to capture different aspects of the questions, the users, and the relationships between them. Extensive experimental results indicate high feasibility of this approach. Beyond expertise estimation, Li *et al.* [12], [13] also incorporated the availabilities of experts and category-sensitive language models to boost the routing performance.

In summary, the approaches of the first category [6], [8], [9], [11], [15]–[17], [19], only take experts' information into consideration and do not leverage the question contexts to estimate the matching score between questions and physicians. They are thus not applicable to the health domain, since precise expert-question matching is the key in the health sectors.

In addition, none of the aforementioned approaches pay attention to the experts' attitudes, which in fact play pivotal roles in the task of question routing.

B. Hypergraph-Based Learning

Given a collection of data samples, it is a common practice to represent the similarity relations between its elements using either a weighted or an unweighted simple graph, whereby vertices are samples and edge weights indicate the similarity between two samples [20]. Some machine learning methods for unsupervised and semi-supervised learning can then be formulated in terms of operations on this simple graph. However, the learning methods based on simple graph consider only the pairwise relationship between two samples, and they ignore the relationship in a higher order. For example, from a graph, we can easily find two close samples according to the pairwise similarities, but it is not easy to predict whether there are three or more close samples. Essentially, modeling the high-order relationship among samples will significantly improve classification performance. Hypergraph addresses this problem, which is a generalization of a pairwise simple graph, where an edge can connect any number of vertices [21]. The expressive power of the hypergraph models places a special emphasis on the relationship among three or more samples, which has made hypergraphs better models of choice in a lot of problems. This is in sharp contrast with the conventional simple graph representation of samples where only pairwise connectivity between samples is captured.

Recently, there has been a lot of interest in learning with hypergraph [22], [23], which has been proven to be a successful tool to represent and model concepts and structures in various areas of computer science. For example, Agarwal *et al.* [24] first introduced the hypergraph idea to the field of computer vision, and solved it by transferring the hypergraph to the simple graph using "clique average." Zhou *et al.* [25] developed a general framework which applies to classification, clustering, and embedding on hypergraph data, and has been adopted widely to solve the unsupervised video object segmentation problem in [26]. The work in [27] cast the image matching problem to a hypergraph-based convex optimization problem. Tian *et al.* [28] introduced a hypergraph-based learning algorithm to classify arrayCGH data with spatial priors modeled as correlations among variables for cancer classification and biomarker identification. Huang *et al.* [29] formulated the task of image clustering as a problem of hypergraph partition, where image and its nearest neighbors form two kinds of hyperedges based on the descriptors of shape or appearance. The work in [30] leverages a hypergraph to depict the attribute relations in the data, and successfully transformed the attribute prediction problem as a regularized hypergraph cut problem. Besides, several hypergraph-based ranking models have been proposed and validated, such as the hypergraph-based image retrieval in [31] and hypergraph-based question annotation in [32].

In this paper, we employ hypergraph to characterize the higher-order relations among physicians, and adopt it to seamlessly integrate multifaceted heterogeneous cues, such

Fig. 2. Illustration of a QA example selected from HealthTap. Even if it has multiple answers, we do not list all of them due to the limited space.

as physicians' social connections, biographies, and historical experience.

C. Multiview Information Fusion

In this paper, we fuse physicians' multiview information, consisting of expertise and attitude, to estimate the "matching score" between the physician candidate and the given health-oriented question. Frankly speaking, multiview learning methods have been well-studied, such as the multiview learning on object classification [33], [34], multiple social network learning [35]–[38]. They consider the relationships among distinct views. But it is hard to predefine or discover the relations between expertise and attitude of physicians. Consequently, the conventional multiview learning approaches are not applicable to this paper.

III. QUESTION ROUTING SCHEME

To formulate our problem, we declare some notations in advance. In particular, we use bold capital letters (e.g., $\mathbf{X}$) and bold lowercase letters (e.g., $\mathbf{x}$) to denote matrices and vectors, respectively. We employ nonbold letters (e.g., x) to represent scalars, and Greek letters (e.g., λ) as parameters. If not clarified, all vectors are in column form.

Suppose that we are given a collection of n physicians in the target CHS, and it is denoted as $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$. Physicians are represented by their profiles,[5] comprising of their biographies and QA pairs they involved before. A biography contains a physician's education, publications, awards, endorsements, and other information. Regarding the QA pair, it is composed of a question, answers, tags associated to answers, and agree votes to answers. Fig. 2 shows one typical QA pair example. We view QA pairs involved by a physician as his/her accumulated experience. $\mathbf{d_i}$ is the vectorized feature representation of physician d_i.

---

[5]A profile example of a physician is accessible here: www.healthtap.com/experts/10003568-dr-james-w-ferguson.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

NIE *et al.*: MULTIVIEW PHYSICIAN-SPECIFIC ATTRIBUTES FUSION FOR HEALTH SEEKING

5

For a given question q in natural language, we aim to select a small set of matched physician candidates from $\mathcal{D}$ and route question q to them. The matched score is estimated as

$$S(d_i, q) = (1 - \lambda(q)) \cdot E(d_i, q) + \lambda(q) \cdot A(d_i) \qquad (1)$$

where $E(d_i, q)$ is an expertise model that captures how well the physician d_i can potentially answer the given question q from a professional perspective. $A(d_i)$ is an attitude model that models the physician's choice of action and the way of behaving toward general questions. Attitudes are not directly observable but can be inferred from the historical behaviors. In addition, $\lambda(q)$ is an adaptive parameter to balance the effects between expertise and attitude. It is question-aware and somehow reflects the personalized healthcare.

IV. EXPERTISE MODELING

In latent Dirichlet allocation (LDA) [39], each document is viewed as a distribution of various topics. For example, a document is related to topic 1 with probability 0.7, and topic 2 with probability 0.3. Meanwhile, LDA posits that each topic is a distribution over words. For example, a cat-related topic has probabilities of generating the words of milk, meow, and kitten, while a dog-related topic likewise has high probabilities of generating the words of puppy, bark, and bone. In practice, only the words in each document are observable. The topic mixture of each document and the topic for each word in each document are latent variables that need to be inferred from the observations.

Inspired by the principle of LDA, we assume that each physician can be characterized by a particular set of expertise. For example, a physician has 70% of confidence to solve questions of breast cancer, and has 30% of confidence in hernia-related concerns. In addition, we also assume that each expertise can be interpreted by a group of semantically similar questions that can be professionally handled by this expertise. Take the expertise on breast cancer as an example. It is reasonable to use questions like "Can men without breasts get breast cancer?" to represent the semantics of skills on breast cancer. Analogous to topics and words in LDA, expertise is an unobservable variable and questions are observable. Inspired by this, we formulate the expertise matching score between each physician and the given question as

$$E(d_i, q) = p(q|d_i) = \sum_j p(\mathcal{E}_j|d_i) \times p(q|\mathcal{E}_j) \qquad (2)$$

where $p(\mathcal{E}_j|d_i)$ estimates medical expertise distribution, i.e., how skillful the given physician d_i is with respect to expertise $\mathcal{E}_j$; and $p(q|\mathcal{E}_j)$ calculates the expertise-question distribution, i.e., to what degree question q can be professionally solved by expertise $\mathcal{E}_j$.

A. Medical Expertise Distribution

We observed that a physician in CHSs usually holds multiple skills with different confidences.[6] For example, a physician

[6]According to our statistics on the collected data, each physician in HealthTap on average has 3.4 professional skills.

may be proficient in obstetrics, good at pediatrics, and familiar with breast cancer, simultaneously. To obtain the quantified score $p(\mathcal{E}_j|d_i)$, we have to solve two problems: 1) estimating how many fields of expertise are hidden in the physician set $\mathcal{D}$ and 2) inferring the physician confidence at each specific expertise. We deem them as a "soft" physician partition task on $\mathcal{D}$. That is, a physician may fall into multiple fields of expertise with different confidence scores $p(\mathcal{E}_j|d_i)$.

Several clustering techniques can be utilized to calculate $p(\mathcal{E}_j|d_i)$, such as k-means [40], LDA, and graph-based clustering [41]. These approaches, however, suffer from the following two intrinsic limitations. First, they, more often than not, hypothesize pairwise relations among the entities of the interest. In CHSs, relations among physicians are much more sophisticated and complex. For example, one question may be replied by more than two physicians, and they might partially share one common expertise. Naively squeezing the grouping relations into the pairwise ones may lead to a information loss. Second, even though great success has been achieved for homogeneous relations, they are unable to handle heterogeneous relations. Generally speaking, the relations among physicians in CHSs can be reflected from multifaceted aspects, such as similar experiences and co-answering. In the light of this, we build a hypergraph and perform a probabilistic partition based on this hypergraph. It enables each hyperedge to connect more than two physicians and hence is capable of summarizing the local grouping information. In addition, different hyperedges can represent heterogeneous relations.

A hypergraph $G(\mathcal{D}, \mathcal{U}, \mathbf{W})$ is composed of a vertex set $\mathcal{D}$, a hyperedge set $\mathcal{U}$, and a diagonal matrix of hyperedge weights $\mathbf{W}$ [42]. Here, $\mathcal{U}$ is a family of hyperedges e that connect arbitrary subsets of $\mathcal{D}$, and each hyperedge e is assigned a weight $W(e)$. In this paper, physicians are treated as vertices. Three types of hyperedges are constructed. Regarding the first type of hyperedges, we treat each physician as a centroid and use a hyperedge to circle around his/her k-nearest neighbors based on their profile similarities. Whereby, a physician's profile is constructed via concatenating his/her biography and historical QA pairs. Here k is empirically set to 25. For each physician, the second and third types, respectively, utilize one hyperedge to group those physicians who have voted his/her answers before, or have co-answered some questions before.

A probabilistic hypergraph G can be represented by an incidence matrix $\mathbf{H}$ with $|\mathcal{D}| \times |\mathcal{U}|$ entries

$$h(d_i, e_j) = \begin{cases} p(d_i, e_j) & \text{if } d_i \text{ is linked by } e_j \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where $p(d_i, e_j)$ describes the probability that physician d_i falls into one community grouped by the hyperedge e_j. Let us denote d_j as the centroid of e_j, and define $p(d_i, e_j)$ as

$$p(d_i, e_j) = \begin{cases} K(d_i, d_j) & \text{Biography/Experience} \\ 1 & \text{Co-answer/Voting-based.} \end{cases} \qquad (4)$$

$K(\cdot, \cdot)$ is the Gaussian similarity function [20], defined as

$$K(d_i, d_j) = exp\left(-\frac{||\mathbf{d_i} - \mathbf{d_j}||^2}{\sigma^2}\right) \qquad (5)$$

where the radius parameter, σ, is simply set as the median of the Euclidean distances among all physician pairs.

Based on $\mathbf{H}$, the vertex degree of $d_i \in \mathcal{D}$ is estimated as

$$d(d_i) = \sum_{e_j \in \mathcal{U}} W(e_j) h(d_i, e_j). \tag{6}$$

The weight for each hyperedge is computed as

$$W(e_j) = \sum_{d_i \in e_j} h(d_i, e_j) \tag{7}$$

where $d_i \in e_j$ means that vertex d_i involves in hyperedge e_j. If the hyperedge e_j has a higher inner group similarity, it will be assigned a larger weight [21], which indicates its compact and dense connections within its group. For a hyperedge $e_j \in \mathcal{U}$, its degree, $\delta(e_j)$, is defined as the number of physicians connected by this hyperedge.

Several methods thus far have been proposed for hypergraph partition, and achieved great success [43]. In this paper, we adopt the framework in [29] to partition the hypergraph due to its efficiency. Beyond that, our constructed hypergraph is a probabilistic graph. Inspired by the normalized cost function of a simple graph [44], [45], we define the regularizer $\Omega(f)$ on the given hypergraph as

$$\frac{1}{2} \sum_{e \in \mathcal{E}} \sum_{u,v \in e} \frac{w(e) h(u, e) h(v, e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}} \right)^2 \tag{8}$$

where vector $\mathbf{f} \in \mathbb{R}^D$ contains the relevance probabilities between each physician and a specific latent expertise category that we want to learn. By defining $\Theta = \mathbf{D}_v^{-(1/2)} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_v^{-(1/2)}$, we can further derive that

$$\sum_{e \in \mathcal{E}} \sum_{u,v \in e} \frac{w(e) h(u, e) h(v, e)}{\delta(e)} \left(\frac{f^2(u)}{d(u)} - \frac{f(u)f(v)}{\sqrt{d(u)d(v)}} \right)$$
$$= \sum_{u \in \mathcal{V}} f^2(u) \sum_{e \in \mathcal{E}} \frac{w(e) h(u, e)}{d(u)} \sum_{v \in \mathcal{V}} \frac{h(v, e)}{\delta(e)}$$
$$- \sum_{e \in \mathcal{E}} \sum_{u,v \in e} \frac{f(u) h(u, e) w(e) h(v, e) f(v)}{\sqrt{d(u)d(v)}\delta e}$$
$$= \mathbf{f}^T (\mathbf{I} - \Theta) \mathbf{f} \tag{9}$$

where $\mathbf{I}$ is an identity matrix. Let $\Delta = \mathbf{I} - \Theta$, which is a positive semi-definite matrix, the so-called hypergraph Laplacian [25], then $\Omega(\mathbf{f})$ can be rewritten as

$$\arg \min_{\mathbf{f} \in \Re^D} \Omega(\mathbf{f}) = \mathbf{f}^T \Delta \mathbf{f}. \tag{10}$$

By minimizing $\Omega(\mathbf{f})$, we can ensure that the relevance probability function is continuous and smooth in the semantic space. This implies that the scores of physicians, who have similar expertise, should be close. The theoretical solution of the above optimization problem is the eigenvector associated with the smallest nonzero eigenvalue of Δ. To make a multiway clustering of vertices in the given hypergraph, we take T eigenvectors corresponding to the T smallest nonzero eigenvalues of Δ as our solutions $\{\mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(T)}\}$. We set $\mathbf{f}_i^{(j)} = p(\mathcal{E}_j | d_i)$ with $(1 \leq j \leq T)$. To adaptively determine the threshold T, we first sort all the nonzero eigenvalues in an ascending order, and then select the cutoff at the point that achieves the largest

increasing gap. Thereinto, T denotes the number of expertise hidden in a given physician collection $\mathcal{D}$.

B. Expertise-Question Distribution

So far, we have obtained $\mathbf{f}_i^{(j)}$, which refers to the confidence of physician d_i with expertise j. We thus can generate a physician ranking list for each expertise. To semantically represent the expertise $\mathcal{E}_j$, we first merge all the QA pairs of the top 50 physicians from the jth expertise, and denote such a collection as $\mathcal{QA}_j$. We observe that physicians with the same expertise frequently use common medical concepts to convey the same skill. We hence select ten concepts with high frequencies from $\mathcal{QA}_j$. Afterwards, we use these concepts as queries to search over $\mathcal{QA}_j$ to retrieve the top 100 relevant QA pairs (Apache Lucene indexing and search). These selected 100 QA pairs are ultimately used to represent the expertise $\mathcal{E}_j$, and are denoted as $\mathcal{X}_j$.

We employ the KDE approach [46] to estimating the relations between a question and each given expertise. KDE has been widely applied in various applications, such as image ranking [20]. It is formally stated as

$$p(q|\mathcal{E}_j) = \frac{1}{|\mathcal{X}_j|} \sum_{q_i \in \mathcal{X}_j} K(q_i, q) \tag{11}$$

where q_i is the ith QA pair from $\mathcal{X}_j$. The above equation can be intuitively interpreted as follows: $\mathcal{X}_j$ and each of its QA pair can be viewed as a family and a family member, respectively. Then the closeness of an unknown question to this family is estimated by averaging the soft voting from all family members. As yet, we have derived the expertise matching score $E(q, d_i)$ between a given question and each physician.

V. ATTITUDE MODELING

Attitude is defined as a relatively enduring organization of beliefs, feelings, and behavioral tendencies toward socially significant objects, groups, events, or symbols.[7] Social psychologists have justified that attitudes and actual behaviors are usually consistent [47]. In view of this, we assume that physicians behave in accordance with their attitudes. Psychologists frequently describe attitudes with the ABC model[8] in terms of three components: 1) affective component (A); 2) behavioral component (B); and 3) cognitive component (C). This mode is, however, difficult to be quantized. Based on the available information from CHSs, in this paper, we comprehensively model physicians' attitudes from four complementary perspectives: 1) activity; 2) responsibility; 3) reputation; and 4) willingness. They are estimated based on physicians' historical data.

Activity estimates how active or fast the physicians' responses will be once questions are routed to them. It is defined as

$$\text{Activity}(d_i) = \frac{\#\text{first}(d_i) + \epsilon}{\#\text{answered}(d_i) + 1} \tag{12}$$

where $\#\text{first}(d_i)$ and $\#\text{answered}(d_i)$, respectively, denote the number of questions that physician d_i has been the first

[7]http://www.simplypsychology.org/attitudes.html
[8]http://www.gerardkeegan.co.uk/glossary/gloss_t.htm

to answer and participated to answer before. The larger Activity(d_i) is, the shorter time physician d_i may use to reply the given question d_i. In addition, ϵ is a smoothing factor which is set to 10^{-3} in this paper.

Responsibility measures the degree of physicians' obligation to satisfactorily solve the given questions. It characterizes the answer qualities. If a provided answer was selected as the best answer by peer votes, physician d_i is deemed as having been responsible for that answer. Responsibility of d_i can be formulated as

$$\text{Responsibility}(d_i) = \frac{\#\text{bestAnswer}(d_i) + \epsilon}{\#\text{answered}(d_i) + 1} \quad (13)$$

where $\#\text{bestAnswer}(d_i)$ denotes how many answers provided by d_i have been selected as the best answers.

Activity and responsibility metrics, however, have one limitation. Specifically, a physician, who only answered one question, and whose answer might be the first or the best answer, would achieve a higher attitude value. This does not make sense. Willingness is devised specifically to compensate for such a problem. It estimates the willingness of a given physician to take up one routed question from the perspective of response rate

$$\text{Willingness}(d_i) = \frac{\#\text{answered}(d_i) + \epsilon}{\#\text{maxAnswers}} \quad (14)$$

where $\#\text{maxAnswers}$ is the maximum number of answers that have been provided by someone in $\mathcal{D}$.

A physician's reputation is the general belief or opinion that their peers and patients have. In CHSs, a physician' online reputation is usually recognized via the votes from others. We utilize the sigmoid function with output ranging between 0 and 1 to denote the reputation

$$\text{Reputation}(d_i) = \frac{1}{1 + e^{-(\rho \times \#\text{Peer} + \#\text{Seeker})}} \quad (15)$$

where $\#\text{Peer}$ and $\#\text{Seeker}$, respectively, denote the number of total votes from physicians and health seekers for the given physician. ρ is a parameter emphasizing endorsements from peers, since they are more professional and reliable. It is set to the ratio between the average number of votes from physicians and health seekers in our whole data collection. In particular, it is set as 1.06 in this paper.

The multiplication results of the aforementioned factors reflect physicians' attitudes to general questions. In summary, our proposed physicians' attitude modeling is question independent. It predicts physicians' behaviors toward answering general questions.

VI. ADAPTIVE FUSION

According to our observation, different questions hold different biases in favor of either expertise or attitude. Table I displays some question examples. From this table, it can be seen that the first three questions ask for some factoid problems or basic knowledge in the health domain. It is obvious that junior physicians can well answer such questions with high confidences. Meanwhile, the quality in terms of intuition and comprehension is of vital importance. This kind of quality is strongly dependent on physicians' attitudes rather

TABLE I
ILLUSTRATION OF THE QUESTION BIAS TOWARD EITHER EXPERTISE OR ATTITUDE. WE DID NOT LIST THE CORRESPONDING ANSWERS DUE TO THE LIMITED SPACE

Biased towards	Question Examples
Attitude	*What are the causes of breast cancer?*
Attitude	*Is there any foods or vitamins that increase resistance against infections (cough and cold) in babies?*
Attitude	*Is it true that if you smoke while you are pregnant, the child will grow up to be a smoker too?*
Expertise	*I just found a lump in my breast that is very tender. I'm pregnant, so my breasts are chang -ing. Could it be breast cancer ?*
Expertise	*I drink a lot of water and don't take vitamins, but my urine is neon yellow. I had a bad kidney infection, could that be related to it?*
Expertise	*Do symptoms of birth control pills make you feel like you are pregnant?*

than expertise. On the other hand, the last three questions seek for reasoning results or very professional knowledge based on the given manifested symptoms and signals. In such scenarios, expertise is the key to the answer quality.

The parameter λ in (1) is an adaptive function of the given question context, which balances the importance between expertise and attitude. When $\lambda = 1$, all the newly posted questions will be routed to the physicians with optimal attitudes and their expertise is completely overlooked. On the contrary, if λ tends to zero, physicians' attitudes will not be considered.

We regard the estimation problem of λ as a supervised regression task [48]. The objective is to predict a weight for each newly asked question. For each question in the training data, we first obtain its optimal value of λ, which is derived by sweeping $[0, 1]$ with a small and fixed step size. These optimal values are utilized as the ground truth to train the regression models. Various regression models are examined, spanning from linear regression [49], isotonic regression [50], to pace regression [51].

VII. EXPERIMENTS

In this section, we first discuss the experimental settings, including the data collection and feature extraction. Following that, we detail the objective and subjective evaluation metrics. We then validate the whole scheme and each of its components.

A. Data Collection and Feature Extraction

To avoid the data sparseness, we manually selected 100 active physicians as seeds from HealthTap, and iteratively extended the seeds by crawling/adding other physicians who have co-answering or voting relations with the physicians in the seed set. We ultimately collected a set of 3123 physician profiles from HealthTap. Each profile contains a physician's biography and a set of involved QA pairs. Table II shows the statistics of our data. It can be seen that the average number of votes is high, which tells that physicians in CHSs are very active to establish and maintain their social relations. Meanwhile, we can derive that each physician on

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8   IEEE TRANSACTIONS ON CYBERNETICS

TABLE II
STATISTICS OF OUR DATA COLLECTION. THE SYMBOLS OF # AND #^{a} DENOTE THE
NUMBER AND THE AVERAGE NUMBER OF SOMETHING, RESPECTIVELY

Physician #	Question #	Answer #	Tag #^{a} Each Answer	Vote #^{a} from Other Peers	Vote #^{a} from Health Seekers
$3,123$	$142,025$	$235,162$	7.06	82.37	77.65
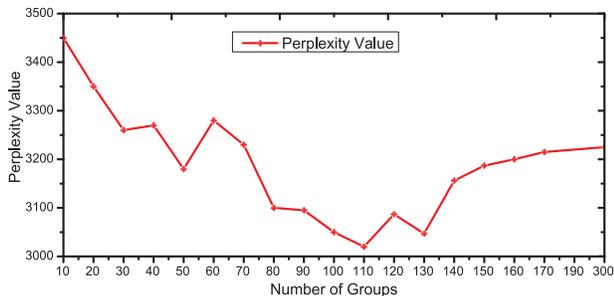
Fig. 3.   Curve of perplexity values with respect to the number of topics. As can be seen, we obtain the optimal perplexity value when the number of topics is at 110.

average answered more than 75 questions, which implies that physicians are enthusiastic to contribute their knowledge and support the health seekers online.

We employed the LDA-based topic-level features for physician profile representation. In particular, each latent topic was deemed as one feature. The number of topics was tuned according to the widely-adopted perplexity metric [39]. Regarding perplexity, a lower value usually indicates a better LDA model. We divided the physician profiles into two subsets: 1) 80% was used to train the LDA models with various numbers of latent topics and 2) 20% was used for evaluation in terms of perplexity. In addition, the LDA model and perplexity metric were implemented with the help of the Stanford topic modeling toolbox.[9] As shown in Fig. 3, when the number of latent topics arrives at 110, the perplexity curve reaches the trough. Each physician was hence represented as a 110 dimensional semantic feature vector.

As compared to the representation of physician profiles, topic-level features are unable to be precisely extracted from each single question or QA pair, since it is relatively short and lack of contexts. In addition, users with diverse backgrounds do not necessarily share the same vocabulary in CHSs [52]. Oftentimes, the same medical subjects may be colloquially expressed with distinct medical concepts. For example, "birth control" and "family planning" are commonly used by individuals to refer to the same medical terminology "contraception." In a sense, the traditional context representations such as n-grams are unable to capture the variation of medical concepts and may lead to an explosion of feature dimension (i.e., the curse of feature dimensionality). To alleviate such problems, we employed the MetaMap tool [53] to detect medical attributes that are noun phrases in the health domain, and then normalized them to standardized terminologies in the SNOMED CT Metathesaurus.[10] The work in [52] detailed

this procedure. The semantic types of these terminologies span from symptom, treatment, medication, body parts, to demographics. In this paper, we utilized these normalized medical attributes to represent the question or QA content. In the end, we obtained 5036 dimensional bag-of-terminologies. Before feeding such features into our model, we performed principal component analysis (PCA)[11] and reduced the feature dimension to 800, which will significantly boost the learning efficiency.

For the subsequent subjective evaluations, we invited three volunteers with basic medical backgrounds. The volunteers were trained with short tutorials and a set of typical examples before their labeling procedure. A majority voting scheme among the three volunteers was adopted to partially alleviate the subjectivity problem. For cases where there were two volunteers having an understanding problem regarding the same instance, a discussion was carried out among the volunteers to obtain the final decision.

B. Evaluation Metrics

It is well known that, for the expert recommendation task, precision is usually more important than recall. We thus adopted two metrics that are able to capture precisions from different aspects: objective and subjective evaluations.

Regarding the objective evaluation, we employed the average $H@K$ [54]. For a given question, $H@K$ is defined as 1, if there exists at least one physician, who really replied the given question before, being ranked in one of the top $K\%$ positions; otherwise $H@K$ is defined as 0. The metric benefits from the real-world data and does not need the construction of ground truth. In this paper, for all evaluations in terms of average $H@K$, we randomly selected 1000 questions from our dataset as the test set and reported the average $H@K$ over the testing samples. However, this metric may suffer from this scenario: physician d_i did reply the given question q, meanwhile physician d_j has the ability to solve q but he/she did not answer it due to some unknown reasons. Even though our scheme has a large probability to rank physician d_j in a higher position than d_i, $H@K$ overlooks that. Therefore, the only $H@K$ metric cannot evaluate our routing scheme comprehensively and fairly.

As a complement to $H@K$, we adopted one subjective metric $S@K$. It measures the probability of finding a matched physician among the top K recommended physicians. To be more specific, for each given question, $S@K$ is assigned as 1 if a matched physician was ranked in one of the top K positions and 0 otherwise. Distinguished from the objective evaluation, the ground truth here was manually constructed by three invited volunteers. In particular, they were presented with a

---

[9]http://nlp.stanford.edu/downloads/tmt/tmt-0.4/

[10]http://www.ihtsdo.org/snomed-ct/

[11]http://sebastianraschka.com/Articles/2014_pca_step_by_step.html

TABLE III
INTERVOLUNTEER AGREEMENT EVALUATION IN TERMS OF THE KAPPA METRIC. TOP 20 PHYSICIANS FOR
EACH TESTING QUESTION WERE MANUALLY LABELED BY OUR INVITED VOLUNTEERS

Number of Cases	Number of Categories	Number of Volunteers	Overall Agreement	Fixed-marginal Kappa	Free-marginal Kappa
50	2	3	97.33%	76.36%	94.67%

ranked list of physicians for each question generated by our scheme. They were then asked to go through each physician's biography and historical behaviors. Meanwhile, they were encouraged to use the Internet to facilitate specific terminology understanding. They were then suggested to label each physician in the ranking list as "matched," if they believed that the physician is able to and would like to answer that question based on their understanding; otherwise, they labeled it as "not matched." In this paper, for all evaluations in terms of $S@K$, we randomly selected 50 questions from our dataset to serve as the test set. Here we evaluated the intervolunteer agreement with the Kappa method [55]. The Kappa metric is a chance corrected statistic to quantitatively measure the degree of intervolunteer agreement. A Kappa result ranges from 0 to 1. The higher the value of Kappa is, the stronger the agreement will be. A Kappa value more than 0.7 typically indicates that the agreement is strong. In this paper, we employed the online Kappa calculator tool.[12] Table III shows the analytical results. It can be seen from Table III that there are sufficient intervolunteer agreements for our labeling task.

C. Component-Wise Analysis: Expertise Modeling

To justify the effectiveness of our proposed hypergraph-based expertise modeling, we conducted experiments to compare it against the following state-of-the-art techniques.

1) *k-means:* The medical expertise relations were estimated by the fuzzy-k-means [56]. The traditional k-means approach [40] produces a crisp result: for a given sample x, it either is or is not a member of a particular cluster. However, in the health domain, each physician is probably an expert [56] across multiple specific fields. We thus employed the fuzzy-k-means procedure that allows each sample to have a degree of membership in each cluster.
2) *LDA [39]:* The medical expertise relations were estimated by the LDA approach. Similar to the topic modeling, here each physician and expertise are, respectively, viewed as a document and a hidden topic.
3) *Ours:* It is our proposed expertise modeling with probabilistic hypergraph learning.

Tables IV and V, respectively, illustrate the comparison results in terms of $H@K$ and $S@K$. They are the results without considering the physicians' attitudes and personal care, since this section aims to evaluate the expertise modeling only. Jointly analyzing these two tables, it can be seen that our proposed hypergraph-based expertise modeling consistently outperforms others in $H@K$ and $S@K$ across different depths. The possible reason is that our approach is able to capture the high-order

TABLE IV
PERFORMANCE COMPARISON WITHOUT PHYSICIANS' ATTITUDES
IN TERMS OF $H@K$ (COMPONENT-WISE EVALUATION ON
EXPERTISE MODELING)

Approaches	H@5	H@10	H@15	H@20
K-means	0.078	0.146	0.226	0.284
LDA	0.073	0.147	0.207	0.279
Ours	**0.091**	**0.158**	**0.242**	**0.298**

TABLE V
PERFORMANCE COMPARISON WITHOUT PHYSICIANS' ATTITUDES
IN TERMS OF $S@K$ (COMPONENT-WISE EVALUATION ON
EXPERTISE MODELING)

Approaches	S@5	S@10	S@15	S@20
K-means	0.60	0.84	0.90	0.94
LDA	0.64	0.80	0.92	0.96
Ours	**0.74**	**0.86**	**0.94**	**0.98**

relations among physicians and is able to seamlessly integrate heterogeneous information. From Table IV, we can see our approach achieves approximately 30% in terms of $H@20$. That means 30% of physicians, who really answered the questions previously, are ranked in the top 20 positions. On the other hand, from Table V, we can see that our approach yields the precision in terms of $S@10$ as high as 0.9. It reveals that nine out of top ten recommended physicians on average were judged as matched physicians by our volunteers.

D. Component-Wise Analysis: Attitude Modeling

To examine the effects of physicians' attitudes and fairly compare the performance, we still use k-means, LDA, and Ours. The results are illustrated in Tables VI and VII, respectively. From Tables IV and VI, and Tables V and VII, it can be seen that the approaches considering physicians' attitudes perform stably better than those without considering physicians' attitudes. This verifies that physicians' attitudes really affect the answering quality and hence play pivotal roles in the problem of question routing.

It is worth emphasizing that the parameter λ in this section is not question-aware, but is an optimal and fixed parameter over all questions, since we aim to validate the effectiveness of attitude modeling. In particular, we randomly selected a set of 1000 questions from our dataset, which serve as training samples to learn the fixed and optimal λ. For each λ within [0, 1] with a fixed step size of 0.05, we calculated the average $H@5$ over all these questions. The λ corresponding to the optimal average $H@5$ was selected to fit in our models. Theoretically, for different metrics, we should learn different corresponding parameters. Here we learned the optimal parameter only based on $H@5$ and applied that to both $H@K$ and $P@K$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON CYBERNETICS

TABLE VI
PERFORMANCE COMPARISON WHEN CONSIDERING PHYSICIANS'
ATTITUDES IN TERMS OF $H@K$. THE SYMBOLS OF ^{f} AND ^{a},
RESPECTIVELY, DENOTE FIXED AND ADAPTIVE λ

Approaches	H@5	H@10	H@15	H@20
K-means^f	0.082	0.147	0.229	0.287
K-means^a	0.088	0.152	0.244	0.311
LDA^f	0.080	0.145	0.220	0.300
LDA^a	0.086	0.161	0.238	0.315
Ours^f	0.094	0.166	0.255	0.297
Ours^a	**0.103**	**0.179**	**0.280**	**0.347**

TABLE VII
PERFORMANCE COMPARISON WHEN CONSIDERING PHYSICIANS'
ATTITUDES IN TERMS OF $S@K$. THE SYMBOLS OF ^{f} AND ^{a},
RESPECTIVELY, DENOTE FIXED AND ADAPTIVE λ

Approaches	S@5	S@10	S@15	S@20
K-means^f	0.70	0.86	0.92	0.94
K-means^a	0.72	0.90	0.96	0.98
LDA^f	0.68	0.82	0.90	0.96
LDA^a	0.72	0.88	0.92	1.00
Ours^f	0.78	0.90	0.96	1.00
Ours^a	**0.86**	**0.94**	**1.00**	**1.00**

This is because $H@K$ does not require labor and time consuming ground truth labeling.

E. Component-Wise Analysis: Adaptive Fusion

To validate the importance of question-aware adaptive fusion implemented via the supervised regression models, we randomly selected 1000 questions from our dataset. These selected questions were divided into two subsets, 80% for training, and 20% for testing. For each question, we searched its optimal routing performance by sweeping λ within $[0, 1]$ at a fixed step size of 0.05. Different from the fixed parameter learning, where all the questions globally share the same parameter, here each question adaptively holds its own parameter. Four regression models were evaluated based on mean absolute error

$$\text{mAE} = \frac{\sum_{i=1}^{n} |p_i - a_i|}{n} \tag{16}$$

where a_i and p_i denote actual value and predicted value, respectively. Typically, a lower mean absolute error corresponds to a better regression model.

The adaptive parameter learning performance of the four models is summarized in Table VIII. It can be seen that the pace regression model achieves the best performance. For a newly posted question, we use the pace regression model to predict its adaptive parameter and use the predicted value to fit our model for question routing. It reflects personalized healthcare, since the question content somehow conveys the personal health information about the health seekers.

As shown in Tables VI and VII, three approaches with adaptive λs to balance the effects between expertise and attitude show superiority over those with fixed λs. This further demonstrates the correctness of our observation: question context holds a bias toward either physician expertise or attitude. Meanwhile, it underscores the importance of personalized question routing.

TABLE VIII
PERFORMANCE COMPARISON BETWEEN VARIOUS REGRESSION
MODELS IN TERMS OF MEAN ABSOLUTE ERROR

Methods	Pace Regression	Linear Regression	Isotonic Regression	Simple Linear Regression
mAE	**0.0274**	0.0279	0.0277	0.0297

F. Overall Scheme Evaluation

As aforementioned, most of the global expert discovery approaches rely on graph-based analysis. Representative algorithms include HITS [8] and PageRank [57]. The former one assumes that there exist two types of graph nodes: 1) hubs which group edges to authoritative nodes and 2) authorities which are sources of information on a given topic. HITS is applicable to the traditional general cQA services, since the askers' IDs are trackable and some askers might also be the answerers, where askers and answerers are, respectively, regarded as "hubs" and "authorities." On the other hand, PageRank-like algorithms not only consider how many other people one helped, but also whom he/she helped. The philosophy behind is that if B is able to answer A's question, and C is able to answer B's question, the rank of C's expertise should be boosted not just because C is able to answer a question, but because C can answer a question of B who had some expertise. In a sense, it propagates expertise scores through the QA network. However, in CHSs, the health seekers' profiles are invisible due to privacy reasons, and they are prohibited to provide answers since they are patients instead of physicians. Therefore, it is hard to apply global experts discovery approaches to CHSs.

We thereby compare our approach against the following state-of-the-art topic-level expert discovery approaches.

1) *Classification-Based:* Zhou *et al.* [7] applied the classification techniques to solve the question routing problem. To be more specific, given a question and an expert, they explored whether the expert has the capability to contribute his/her knowledge to the question with global and local features.
2) *Profile-Based:* The work in [11] first created profiles for experts based on historical contents. It then proposed a probabilistic model to estimate matching relation between each expert profile and the given question.
3) *Expertise-Availability:* This approach in [12] attempted to route newly posted questions to experts who are most likely to reply within a short period. The authors modeled this problem as a typical trend analysis task and utilized an autoregressive model for forecasting.
4) *Ours-Without-Attitude:* Our expertise modeling only.
5) *Our Approach^f:* Our proposed scheme but the fusion parameter is optimally fixed for all the questions.
6) *Our Approach^a:* Our proposed scheme with adaptive fusion by considering the given question context.

The experimental results of several question routing approaches are comparatively reported in Table IX. From this table, the following observations can be made.

1) The approaches considering physicians' attitudes perform stably better than those without considering

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

NIE *et al.*: MULTIVIEW PHYSICIAN-SPECIFIC ATTRIBUTES FUSION FOR HEALTH SEEKING

11

TABLE IX
PERFORMANCE COMPARISON AMONG DIFFERENT QUESTION ROUTING APPROACHES IN TERMS OF $H@K$ AND $S@K$.
THE SYMBOLS OF ^{f} AND ^{a}, RESPECTIVELY, DENOTE FIXED AND ADAPTIVE λ

Approaches	H@5	H@10	H@15	H@20	S@5	S@10	S@15	S@20
Classification-based	0.076	0.149	0.245	0.311	0.70	0.86	0.92	1.00
Profile-based	0.087	0.153	0.238	0.304	0.72	0.82	0.94	0.98
Expertise-Availability	0.094	0.134	0.230	0.298	0.78	0.84	0.90	0.96
Ours-Without-Attitude	0.091	0.158	0.242	0.298	0.74	0.86	0.94	0.98
Our Approach^f	0.094	0.166	0.255	0.297	0.78	0.90	0.96	1.00
Our Approach^a	**0.103**	**0.179**	**0.280**	**0.347**	**0.86**	**0.94**	**1.00**	**1.00**

physicians' attitudes. This verifies that physicians' attitudes really affect the answering quality and hence the question routing performance.

2) Our approach with an adaptive λ to balance the effects between expertise and attitude shows superiority over that with a fixed λ. This further justifies the correctness of our observation: question context holds a bias toward physician expertise or attitude.

3) Our proposed approach with adaptive fusion consistently achieves better performance as compared to the first three baselines.

One possible reason is that none of them explicitly take the physicians' attitude into consideration. The other reason is that our approach is able to capture the high-order relations among physicians and is able to seamlessly integrate heterogeneous information cues via the hypergraph-based structure, while the first three baselines are unable to characterize such sophisticated relations in CHSs.

From Table IX, we can see our approach obtains 0.347 at $H@20$. This means around 35% of physicians, who really answered the questions previously, are placed in the top 20% positions by our scheme. On the other hand, we can see that our approach yields a precision as high as 100% in terms of $S@15$. This implies that at least one physician in the top 15 ones recommended by our routing scheme was judged as the matched expert by our volunteers, on average.

We also performed a pairwise significance test. It was conducted over average $H@K$ ($1 \leq K \leq 20$) between Our approach^a and each of the baselines. We noticed that all the p values are much smaller than 0.05, which indicates that the improvements are statistically significant.

G. Discussion

Regarding the computational complexity of our model, it is in the scale of $(E^3 + V^3 + dV^2)$, where d denotes the feature dimension, N and E, respectively, represent the number of physicians and relations among them. In this paper, the routing process can be completed in less than 5 s if we do not take the feature extraction part into account on a system with (3.4 GHz and 16 GB memory). It is worth noting that, our work is extendable to large-scale datasets by preclustering, whereby the routing will be performed on a small scale only.

On the other hand, physicians are human beings and hence their interests, expertise and attitudes naturally evolve over time. In this paper, we did not consider such evolution due to two reasons. First, the changes are usually with a slow progression. This paper was conducted on the data posted in recent two years. It is thus reasonable to assume that the changes can be negligible. Second, it is expensive and nearly impractical to label the evolution progress. The evaluation thus faces a tough challenge.

VIII. CONCLUSION

This paper has presented a question routing scheme for CHSs, which is capable of bridging the matching gap between physicians and health seekers. It comprises of three components. The first component models the matching relations between a given question and each physician from the perspective of expertise. The second one analyzes physicians' attitudes from multifaceted aspects. The last one adaptively balances the effects between expertise matching and physicians' attitudes by considering the given question context. Extensive experiments on a real-world dataset have shown the effectiveness of our proposed scheme. In addition, we have released our dataset to facilitate other researchers to repeat our experiments and validate their own ideas.

In the future, we are planing to study the physicians' attitude bias toward specific questions. Besides, we will construct large-scale data to validate the scalability our scheme, and apply it to other domains to justify its generalization.

REFERENCES

[1] M. Law, "Online drug information in Canada," Pharmaceut. Advertising Advisory Board, Pickering, ON, Canada, Tech. Rep., 2012. [Online]. Available: http://www.paab.ca/

[2] S. Fox and M. Duggan, "*Health Online 2013*," Pew Research Center, Survey, Washington, DC, USA, 2013.

[3] R. W. White and E. Horvitz, "Cyberchondria: Studies of the escalation of medical concerns in Web search," *ACM Trans. Inf. Syst.*, vol. 27, no. 4, 2009, Art. no. 23.

[4] A. A. of Orthopaedic Surgeons, "Much Internet-based sports medicine information incorrect," *J. Bone Joint Surg.*, 2010.

[5] L. Nie *et al.*, "Disease inference from health-related questions via sparse deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 8, pp. 2107–2119, Aug. 2015.

[6] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom, "Expertise identification using email communications," in *Proc. Int. Conf. Inf. Knowl. Manag.*, New Orleans, LA, USA, 2003, pp. 528–531.

[7] T. C. Zhou, M. R. Lyu, and I. King, "A classification-based approach to question routing in community question answering," in *Proc. Int. Conf. Companion World Wide Web*, Lyon, France, 2012, pp. 783–790.

[8] P. Jurczyk and E. Agichtein, "Discovering authorities in question answer communities by using link analysis," in *Proc. Int. ACM Conf. Inf. Knowl. Manag.*, Lisbon, Portugal, 2007, pp. 919–922.

[9] M. Bouguessa, B. Dumoulin, and S. Wang, "Identifying authoritative actors in question-answering forums: The case of Yahoo! Answers," in *Proc. ACM Int. Conf. Knowl. Discov. Data Min.*, Las Vegas, NV, USA, 2008, pp. 866–874.

[10] J. Guo, S. Xu, S. Bao, and Y. Yu, "Tapping on the potential of q&a community by recommending answer providers," in *Proc. ACM Conf. Inf. Knowl. Manag.*, Napa County, CA, USA, 2008, pp. 921–930.

[11] Y. Zhou, G. Cong, B. Cui, C. S. Jensen, and J. Yao, "Routing questions to the right users in online communities," in *Proc. IEEE Int. Conf. Data Eng.*, Shanghai, China, 2009, pp. 700–711.

[12] B. Li and I. King, "Routing questions to appropriate answerers in community question answering services," in *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, Toronto, ON, Canada, 2010, pp. 1585–1588.

[13] B. Li, I. King, and M. R. Lyu, "Question routing in community question answering: Putting category in its place," in *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, Glasgow, U.K., 2011, pp. 2041–2044.

[14] W. Li, C. Zhang, and S. Hu, "G-finder: Routing programming questions closer to the experts," in *Proc. ACM Int. Conf. Object Orient. Program. Syst. Lang. Appl.*, Amsterdam, The Netherlands, 2010, pp. 62–73.

[15] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[16] G. Zhou, S. Lai, K. Liu, and J. Zhao, "Topic-sensitive probabilistic model for expert finding in question answer communities," in *Proc. ACM Int. Conf. Inf. Knowl. Manag.*, 2012, pp. 1662–1666.

[17] X. Liu, J. Bollen, M. L. Nelson, and H. V. de Sompel, "Co-authorship networks in the digital library research community," *Inf. Process. Manag.*, vol. 41, no. 6, pp. 1462–1480, 2005.

[18] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: Structure and algorithms," in *Proc. Int. Conf. World Wide Web*, Banff, AB, Canada, 2007, pp. 221–230.

[19] B. Dom, I. Eiron, A. Cozzi, and Y. Zhang, "Graph-based ranking algorithms for e-mail expertise analysis," in *Proc. ACM SIGMOD Workshop Res. Issues Data Min. Knowl. Discov.*, San Diego, CA, USA, 2003, pp. 42–48.

[20] L. Nie, S. Yan, M. Wang, R. Hong, and T.-S. Chua, "Harvesting visual concepts for image search with complex queries," in *Proc. ACM Int. Conf. Multimedia*, Nara, Japan, 2012, pp. 59–68.

[21] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. Int. Conf. Mach. Learn.*, Pittsburgh, PA, USA, 2006, pp. 17–24.

[22] Q. Liu, Y. Huang, and D. N. Metaxas, "Hypergraph with sampling for image retrieval," *Pattern Recognit.*, vol. 44, nos. 10–11, pp. 2255–2262, 2011.

[23] S. Huang, A. M. Elgammal, and D. Yang, "On the effect of hyperedge weights on hypergraph learning," *arXiv:1410.6736*, 2014.

[24] S. Agarwal *et al.*, "Beyond pairwise clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, pp. 838–845.

[25] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2006, pp. 1601–1608.

[26] Y. Huang, Q. Liu, and D. Metaxas, "Video object segmentation by hypergraph cut," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 1738–1745.

[27] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, 2008, pp. 1–8.

[28] Z. Tian, T. Hwang, and R. Kuang, "A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge," *Bioinformatics*, vol. 25, no. 21, pp. 2831–2838, 2009.

[29] Y. Huang, Q. Liu, F. Lv, Y. Gong, and D. N. Metaxas, "Unsupervised image categorization by hypergraph partition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1266–1273, Jun. 2011.

[30] S. Huang, M. Elhoseiny, A. Elgammal, and D. Yang, "Learning hypergraph-regularized attribute predictors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 409–417.

[31] Y. Huang, Q. Liu, S. Zhang, and D. N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, 2010, pp. 3376–3383.

[32] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua, "Learning to recommend descriptive tags for questions in social forums," *ACM Trans. Inf. Syst.*, vol. 32, no. 1, 2014, Art. no. 5.

[33] L. Zhang, M. Song, X. Liu, J. Bu, and C. Chen, "Fast multi-view segment graph kernel for object classification," *Signal Process.*, vol. 93, no. 6, pp. 1597–1607, 2013.

[34] Y. Xia, L. Zhang, W. Xu, Z. Shan, and Y. Liu, "Recognizing multi-view objects with occlusions using a deep architecture," *Inf. Sci.*, vol. 320, pp. 333–345, 2015.

[35] L. Nie, X. Song, and T. Chua, *Learning From Multiple Social Networks*. San Rafael, CA, USA: Morgan Claypool, 2016.

[36] X. Song, Z.-Y. Ming, L. Nie, Y.-L. Zhao, and T.-S. Chua, "Volunteerism tendency prediction via harvesting multiple social networks," *ACM Trans. Inf. Syst.*, vol. 34, no. 2, 2016, Art. no. 10.

[37] X. Song, L. Nie, L. Zhang, M. Akbari, and T.-S. Chua, "Multiple social network learning and its application in volunteerism tendency prediction," in *Proc. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, Santiago, Chile, 2015, pp. 213–222.

[38] X. Song, L. Nie, L. Zhang, M. Liu, and T.-S. Chua, "Interest inference via structure-constrained multi-source multi-task learning," in *Proc. Int. Joint Conf. Artif. Intell.*, Buenos Aires, Argentina, 2015, pp. 2371–2377.

[39] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2001, pp. 601–608.

[40] T. Kanungo *et al.*, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[41] P. Qian, F.-L. Chung, S. Wang, and Z. Deng, "Fast graph-based relaxed clustering for large data sets using minimal enclosing ball," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 3, pp. 672–687, Jun. 2012.

[42] L. Zhu, J. Shen, H. Jin, R. Zheng, and L. Xie, "Content-based visual landmark search via multimodal hypergraph learning," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2756–2769, Dec. 2015.

[43] H. Lee-Kwang and C. H. Cho, "Hierarchical reduction and partition of hypergraph," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 26, no. 2, pp. 340–344, Apr. 1996.

[44] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2004, pp. 321–328.

[45] S. Li, H. Lu, and X. Shao, "Human body segmentation via data-driven graph cut," *IEEE Trans. Cybern.*, vol. 44, no. 11, pp. 2099–2108, Nov. 2014.

[46] S. Wang, J. Wang, and F.-L. Chung, "Kernel density estimation, kernel methods, and fast learning in large data sets," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 1–20, Jan. 2014.

[47] G. A. Guagnano, P. C. Stern, and T. Dietz, "Influences on attitude-behavior relationships a natural experiment with curbside recycling," *J. Environ. Behav.*, vol. 27, no. 5, pp. 699–718, 1995.

[48] N. Limsopatham, C. Macdonald, and I. Ounis, "Learning to combine representations for medical records search," in *Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2013, pp. 833–836.

[49] W. Ploberger, W. Krämer, and K. Kontrus, "A new test for structural stability in the linear regression model," *J. Econometrics*, vol. 40, no. 2, pp. 307–318, 1989.

[50] R. Luss, S. Rosset, and M. Shahar, "Decomposing isotonic regression for efficiently solving large problems," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2010, pp. 1513–1521.

[51] M. Shahnawaz and K. Saxena, "A comparative study of various regression model for data farming," *Int. J. Wisdom Based Comput.*, vol. 2, no. 1, pp. 29–34, 2012.

[52] L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, and T.-S. Chua, "Bridging the vocabulary gap between health seekers and healthcare knowledge," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 2, pp. 396–409, Feb. 2015.

[53] A. R. Aronson and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *J. Amer. Med. Informat. Assoc.*, vol. 17, no. 3, pp. 229–236, 2010.

[54] H. Xuan, Y. Yang, and C. Peng, "An expert finding model based on topic clustering and link analysis in CQA website," *J. Netw. Inf. Security*, vol. 4, no. 2, pp. 165–176, 2013.

[55] M. J. Warrens, "Inequalities between multi-rater kappas," *Adv. Data Anal. Classif.*, vol. 4, no. 4, pp. 271–286, 2010.

[56] C.-T. Chang, J. Z. C. Lai, and M.-D. Jeng, "A fuzzy k-means clustering algorithm using cluster center displacement," *J. Inf. Sci. Eng.*, vol. 27, no. 3, pp. 995–1009, 2011.

[57] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the Web," Stanford Univ., Stanford, CA, USA, Tech. Rep., 1998. [Online]. Available: http://ilpubs.stanford.edu:8090/422/

Authors' photographs and biographies not available at the time of publication.