Question Routing in Community Based QA: Incorporating Answer Quality and Answer Content [*]

Lei Fang Minlie Huang Xiaoyan Zhu

State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Sci. and Tech.
Department of Computer Sci. and Tech., Tsinghua University
fang-l10@mails.tsinghua.edu.cn, {aihuang, zxy-dcs}@tsinghua.edu.cn

ABSTRACT

Community based question and answering (cQA) services provide a convenient way for online users to share and exchange information and knowledge, which is highly valuable for information seeking. User interest and dedication act as the motivation to promote the interactive process of question and answering. In this paper, we aim to address a key issue about cQA systems: routing newly asked questions to appropriate users that may potentially provide answer with high quality. We incorporate answer quality and answer content to build a probabilistic question routing model. Our proposed model is capable of 1) differentiating and quantifying the authority of users for different topic or category; 2) routing questions to users with expertise. Experimental results based on a large collection of data from Wenwen demonstrate that our model is effective and has promising performance.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*information filtering*; H.3.5 [**Information Storage and Retrieval**]: On-line Information Services—*web-based services*; I.5.1 [**Pattern Recognition**]: Models—*statistical*

General Terms

Algorithms, Experimentation

Keywords

Community based Question Answering, Question Routing, Recommendation, Probabilistic Topic Modeling

[*]This work is supported by Chinese 973 project under No.2012CB316301, National Chinese Science Foundation projects with No.60803075 and No.60973104, the fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

1. INTRODUCTION

The booming Web 2.0 technology gives an enormous impetus to the prosperity of the online communities. To facilitate people share information more conveniently, some emerging community services such as Yahoo! Answers[1], Baidu Zhidao [2] and Soso Wenwen[3], etc, have become popular frameworks. They provide community-based question and answering services for users to exchange and share knowledge online. For instance, Soso Wenwen is a popular cQA service platform in China, integrated with a substantial amount of online instant messaging users (claimed that the active accounts amounted to 674.3 million while its peak concurrent users reached 137.2 million[4]). It is also reported that more than 17 millions of questions have been asked and answered on Wenwen by online users in less than four years, with tens of thousands of newly asked questions posted on it every day. The continuously flourish of cQA services gives rise to the fact that more and more research efforts have been dedicated to cQA systems.

In order to bridge the gap of knowledge sharing and exchanging between question asker and answerer, in this paper, we present a quality-aware question routing model considering answer content for question routing. The lack of question routing may prevent cQA services from providing prompt and high-quality answers. As for newly asked questions, appropriate question routing methods considering topic and user interest simultaneously will certainly encourage users' willingness and devotion to community, reducing the effort and time for askers to get desired answers.

However, a great number of questions have remained unsolved in cQA websites, many of which are closed even before answered due to the low participation rate[4]. From the statistics of our dataset, about 66% of the users only answered once. It is common that few users are willing to answer questions irrelevant to their interest. Actually, users browse the web community and answer question mainly follow their interest. Even in case that a number of users are voluntary to share their knowledge, chances of receiving a high quality answer from users with various levels of knowledge are slim. It should also be noticed that, authoritative actors in web communities are a small proportion compared to that of all active users for a specific topic or category [2]. Therefore, accurate user interest analysis and question

[1]http://answers.yahoo.com
[2]http://zhidao.baidu.com
[3]http://wenwen.soso.com
[4]http://www.tencent.com/en-us/at/abouttencent.shtml

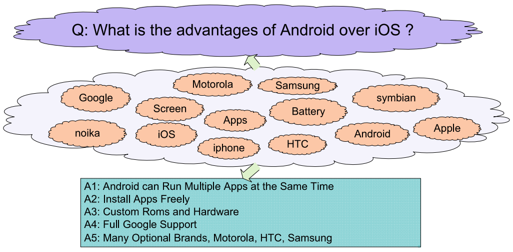


Figure 1: Relevance between Question and Answers

routing model is the key to improve the cQA services.

It is worthwhile to note that in cQA, each answer is marked as best or not, and as for best answers, some metadata like stars voted by the asker may signify the quality of answer. Meanwhile, every answer can be further evaluated and commented by online users, like support or against, interesting or not. Question routing will lack the ability to identify users with expertise and authority, if discarding answer quality.

For any specific user, the questions he/she posted usually indicate a blind spot of interest, which differs from those questions he/she answered. Consequently, modeling user authority and interest should mainly be based on the answered questions (question title and description). We claim that only the question (title, description) is not enough to model the user interest. For the example shown in Figure 1, the question is about the advantage of Android over iOS, and suppose there is a topic about phone, covering some related aspects such as iphone, Apps, Symbian and so on. The posted answers to this question are talking about the advantages of Android. In this scenario, the semantic consistence between question and answer is reflected by the topic. Thus, incorporating answer content will enrich the information that contained in the topic mined from a collection of questions. However, in comparison with title and description of questions, there exists much noise or spam in cQA answers, especially those with low quality. Filtering such noise and incorporating the extracted comprehensive and highly informative keywords from the answers will promote a better modeling of user interest.

To the best of our knowledge, no existing studies have utilized the aforementioned issues (answer quality, answer content, and noise filtering) in cQA and therefore may be unable to perform accurate question routing. We propose a novel unified probabilistic model, called the Question Routing Model (abbreviated as QRM). The QRM could discover topics from questions and corresponding answer content and represent user authority with the learnt topics simultaneously for further user ranking and question routing. The major contributions of this paper are as follows: 1) differentiating and quantifying the authority of users with respect to topic or category level; 2) routing questions to users with expertise; 3) evaluating the quality of answers for posted questions. Experimental results based on a large collection of data from Wenwen demonstrate that our model is effective and has promising performance in real cQA services.

The rest of this paper is organized as follows: in Section 2, we briefly introduce some related work. Section 3 presents the problem statement, Section 4 and Section 5 discuss details of QRM for cQA services: question routing considering **A**nswer **Q**uality, namely QRM-AQ and question routing considering **A**nswer **C**ontent, denoted as QRM-AC. Section 6 illustrates the question routing model and Section 7 presents our experimental setup and results. We provide our conclusion and future work in the last section.

2. RELATED WORK

Much related work has been dedicated to cQA related research, previous research such as [6]focused on question retrieval. Recently, many researchers lay efforts on cQA services systems, such as question routing or answer quality evaluation, for instance,[11] systematically presents several key technologies for designing and developing a QA system which is meant to leverage the social network of QA users to produce high-quality content that may benefit Web search. [12] designs a new algorithm called "majority-based perceptron algorithm" which can avoid the influence of noisy instances by emphasizing its learning over data instances from the majority users to find category representative questions when user browsing questions by categories.

For question routing and community expert finding, [2]propose to automatically identify authoritative users in web community and model the authority scores of users as a mixture of gamma distributions rather than output a ranked list of users using graph-based ranking algorithm such as HITS or PageRank or one of its variants, AuthorRank [7] and ExpertiseRank [13]. [4] develop a generative model to simulate user behaviors in cQA, for both question asking and answering, and obtain topic analysis of questions/answers and users. Then, recommend answer providers for new questions according to discovered topics as well as term-level information of questions and users. [5] presents a social search engine for question routing which rely more on social relationship. [3] introduces a multi-channel recommender system approach for question routing in Yahoo! Answers by exploiting a wide variety of content and social signals in the community.

For our work reported here, we propose a generative model to route questions to users with expertise and willingness using a novel metric representing user interest and authority simultaneously. To our knowledge, our work is the first to introduce quality to traditional generative topic models.

3. PROBLEM FORMULATION

3.1 Problem Definition

Entities in cQA services consist of a finite set of users and questions, denoted by $U = \{u_1, u_2, \ldots, u_L\}$ and $Q = \{q_1, q_2, \ldots, q_D\}$ respectively, for each question q_i, A_i represents the associated set of answerer and corresponding answer pairs as

$$A_i = \{(w_a, u_a)|\ u_a \text{ provides answer } w_a \text{ for } q_i\}$$

other symbols that may be used in this paper is defined in table 1, where "#". means "the number of".

We propose to solve the following problem:

Question Routing : ***Given a question*** q_i ***in cQA, recommend a list of suitable users in*** U ***to provide answers.***

To resolve the addressed problem, we have to deal with several sub problems as:

Question and answer content modeling —Semantically, answers are consistent with questions. For this issue, noise

Table 1: Symbol Notations

Symbols	Description
U, Q	question set and user set
l	individual user in cQA
q	a question document
T	#.topics
N_d	#.words in question q
N_{ka}	#.extracted keywords in answer a
Ans_q	#.answers for q
V	#.terms in vocabulary q
z	topic
$w_q(w_a)$	word in question (answer)
v_q^l	contribution weight for l on q
ψ, ϕ	distribution of user over topics
φ	distribution of topic over terms
α, β	Dirichlet prior

in answers should be firstly filtered; the main challenge lies in how to incorporate the content of answers to promote user interesting modeling and question routing.

User interest and authority representation —Better representation of the user interest and authority may potentially improve the quality of answers provided by the recommended users. In our model, we propose a novel routing model, unifying user interest and expertise, to rank the user for a given question.

3.2 Basic Assumptions

Before introducing our model, we present several basic model assumptions as follows:

Assumption 1. For a specific user, interest can be modeled by the questions (question title, description, etc) he/she answered and the answers (answer content) he/she proposed.

Assumption 2. Higher quality answers can better characterize the expertise for users.

Based on Assumption 1, the answering history indicates the focus of user interest. Assumption 2 addresses that the answer quality information shall be taken into account when modeling user interest.

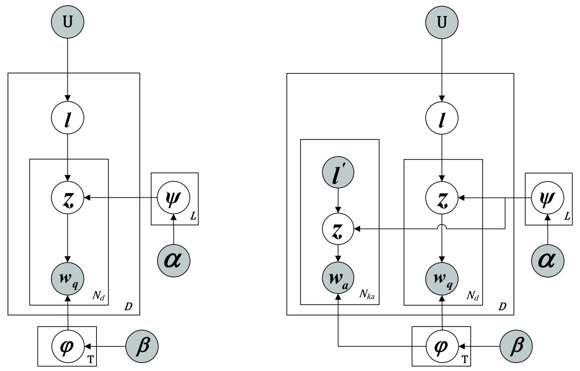
Quality Differentiation.

In order to differentiate the quality for each answerer, we regard the process of answering question in cQA as online user's contribution to the community. For a specific question, non-best answer providers share equal contribution while best answerers contribute more. We define that the total contribution weight of a question q is equal to its number of answerers as

$$weight(q) = Ans_q \tag{1}$$

To further quantify the contribution v_q^l for user l on question q, we introduce quality factor, or can be called as contribution weight factor λ. Firstly, high quality answerers are highlighted as contribution weight $\lambda(\lambda \geqslant 1)$ and others as 1. Then, the total weight of contribution is normalized to $weight(q)$. After that, we quantify the contribution weight v_q^l for user l as

$$v_q^l = \begin{cases} \frac{\lambda Ans_q}{Ans_q+(\lambda-1)} & \text{if l provides the best answer,} \\ \frac{Ans_q}{Ans_q+(\lambda-1)} & \text{otherwise} \end{cases} \tag{2}$$



(a) QRM-AQ (b) QRM-AC

Figure 2: Graphical representation of QRM

v_q^l can be modulated by changing the value of λ. Such definition for v_q^l is in accordance with commonsense that users with expertise provide more valuable knowledge in cQA and therefore deserve more contribution weight.

4. QRM-AQ, MODELING ANSWER QUALITY IN CQA

In this section, we firstly introduce the generative process of question and answering in cQA, and then present the details of incorporating answer quality.

4.1 Generative Model, QRM-AQ

Following the existing work on probabilistic topic models [9, 10] to model user interest, we assume that question is a mixture of topics, where each topic is a multinomial distribution over terms; each corresponding answerer is represented by a mixture of topics. This assumption helps to give a sound description for the generative process of questions in cQA. We use the mixture weight of multinomial distribution to characterize the answerer's contribution over topics.

The graphical representation of QRM-AQ model is shown in Figure 2(a), following the Author-Topic[9], we define the corresponding generative process of a collection of questions Q which consists of D questions covering L unique users and V unique terms with several steps: firstly, each topic z is a multinomial distribution over terms drawn independently from the Dirichlet prior β , denoted by φ_z, and each user l is associated with a multinomial distribution over topics, denoted by ψ_l choosen from Dirichlet prior α. Then for each question $q \in \{q_1, q_2, \ldots, q_D\}$, and for each word $w_{q,n}$ in question q, draw a user $l_{q,n}$ and a topic $z_{q,n}$ according to the mixture weight of question over corresponding users and the user distribution over topics.

The generative process of QRM-AQ is similar to Author-Topic[9] of modeling documents and authors, it should be noticed that these users are not the authors of the question, actually, they make contributions to topics that contained in questions, we employ multinomial distribution to characterize each user's contribution over topic.

4.2 Parameter Estimation

We employ Gibbs Sampling algorithm to obtain parameter estimates under Dirichlet priors:

$$p(z_i = z, l_i = l | w_i = w, l_{\setminus i}, z_{\setminus i}, w_{\setminus i}, \alpha, \beta) \propto \frac{n^z_{l,\setminus i} + \alpha}{n_{l,\setminus i} + T\alpha} \cdot \frac{n^w_{z,\setminus i} + \beta}{n_{z,\setminus i} + V\beta} \quad (3)$$

where $n^z_{l,\setminus i}$ represents the number of times topic z is assigned to answerer l, and similarly, $n^w_{z,\setminus i}$ is the number of times term w is assigned to topic z, $n_{l,\setminus i}$ is the number of all the topic assignments for l and $n_{z,\setminus i}$ represents the number of all the term assignments for topic z. Subscript $\setminus i$ indicates not including the i-th word for all the counts.

After Gibbs Sampler reaches burn-in, we can harvest several samples and count the topic and user assignments for the estimates of parameters:

$$\varphi_{z,w} \propto n^w_z + \beta \quad (4)$$

$$\psi_{l,z} \propto n^z_l + \alpha \quad (5)$$

4.3 Modeling Answer Quality

Traditional topic models such as Latent Dirichlet Allocation (LDA)[1] and Author-Topic[9] do not take the "weight" of each word or entity into consideration, and in cQA, "weight" of each word or topic can be propagated from the contribution weight of its assigned user, indicating the quality of content. From the perspective of askers, they would appreciate their questions answered by authoritative users. We claim that user authority is the combination of answer quantity and quality. Even if the popularity or answer count of user can be well captured on the large dataset by traditional topic models[1, 9], quality of answering lack appropriate representation.

Inspired by that user interest can be characterized by $\psi_{l,z}$ as a mixture of weight over topics, we propose to represent the user quality as certain distribution over topics, denoted by $\phi_{l,z}$. To statistically represent the quality for cQA answerers, we propose three strategies based on Gibbs Sampling algorithm assuming that the "weight" of each word or topic can be propagated from the corresponding user contribution weight.

4.3.1 Post Sampling

Actually, the quality of answer provided by different users varies over topics even if they share same interest, high quality answer providers contribute more on topics than normal answerers.

Based on the harvested samples of terms along with topic and user assignment, we regard the contribution weight on corresponding question of the assigned user as the topic weight, for the calculation of $\phi_{l,z}$, we only need to count the assignment of topic and user, and substitute each assignment of count "1" with v^l_q representing the contribution for user l over topics, therefore, we have

$$\phi_{l,z} \propto m^z_l + \alpha \quad (6)$$

where m^z_l is the total contribution for user l on topic z by summing up the weight of terms assigned to user l and topic z for all the answered questions, and the weight of the term is propagated from the contribution weight for l on corresponding question q, the value is equal to v^l_q, as defined in Equation (2).

Intuitively, the value of $\phi_{l,z}$ is greater than $\psi_{l,z}$ if the answers provided by user l on topic z is usually marked as best answers, otherwise less. We define this method of incorporating answer quality as Post Sampling since it is the re-process of the collected samples.

4.3.2 Quality Sampling

As authority is the combination of user involvement in web community with the answer quality, if the quality factor is introduced during the sampling procedure, parameters obtained from the collected samples will simultaneously represent the user interest and the answer quality. We regard such strategy as Quality Sampling because we sample each user and topic considering the quality factor. The sampling equation has the following extended forms depending on whether considering the quality for topics distribution over terms:

Sampling Topic without Quality.

For this method, we assume that the previous user interest distribution over topics $\psi_{l,z}$ is the same as quality distribution over topics $\phi_{l,z}$, unifying the user interest and expertise. Quality is introduced to the user-topic layer while the topic distribution over terms remains unaffected from answer quality. Therefore, the aforementioned sampling Equation (3) can be rewritten as

$$p(z_i = z, l_i = l | w_i = w, l_{\setminus i}, z_{\setminus i}, w_{\setminus i}, \alpha, \beta) \propto \frac{m^z_{l,\setminus i} + \alpha}{m_{l,\setminus i} + T\alpha} \cdot \frac{n^w_{z,\setminus i} + \beta}{n_{z,\setminus i} + V\beta} \quad (7)$$

For the estimation of parameters, we employ Equation (4) and

$$\phi_{l,z} = \psi_{l,z} \propto m^z_l + \alpha$$

As is mentioned before, m^z_l is the total contribution for user l on topic z by summing up the weight all the terms assigned to user l and topic z. The Equation (7) show that topics obtained from this model are free from quality and quality information is only contained in $\phi_{l,z}$.

Sampling Topic with Quality.

Compared with sampling topic without quality, another strategy is introducing the quality factor into the topic-term layer as well as user-topic layer. That is, for each specific user, quality distribution highlights on topics that he is expert at, and for each topic, distribution over terms emphasize more on those high quality ones. Therefore, the sampling equation has the form as:

$$p(z_i = z, l_i = l | w_i = w, l_{\setminus i}, z_{\setminus i}, w_{\setminus i}, \alpha, \beta) \propto \frac{m^z_{l,\setminus i} + \alpha}{m_{l,\setminus i} + T\alpha} \cdot \frac{m^w_{z,\setminus i} + \beta}{m_{z,\setminus i} + V\beta} \quad (8)$$

To obtain the parameter estimates, we have

$$\varphi_{z,w} \propto m^w_z + \beta$$
$$\phi_{l,z} = \psi_{l,z} \propto m^z_l + \alpha$$

As an analogy drawn from m^z_l, m^w_z associated with each topic is the total sum of the term weight propagated from the corresponding assigned user contribution weight. Therefore, semantics conveyed by these topics also contain answer quality information, and terms assigned to authoritative have a high share in the mixture weight.

5. QRM-AC, MODELING ANSWER CONTENT IN CQA

In this section, we introduce the generative model for question routing incorporating answer content, QRM-AC. We firstly introduce the selection of answer content to filter the noise, after that we present details about modeling the answer content.

5.1 Answer Content Filtering

Answer content is semantically relevant with title or description of question. Answers and questions are linked by the topic. Inevitably, there exists noise in answers from community users with various levels of knowledge. As a result, high semantic content of answers shall be extracted into the model. TextRank[8] is an extension of traditional PageRank model, applying graph-based ranking algorithm for unsupervised keyword and sentence extraction, such method is suitable and can be easily applied to extract high semantic keywords in answers. In our approach, we build a new document D_i for question q_i by concatenate all the corresponding answer content as

$$D_i = \sum_{(w_a, l_a) \in A_i} w_a$$

where $\sum$ is the string concatenation operation for all the answers.

To extract high informative keywords in answers, D_i is first tokenized and annotated with part of speech tags, then we employ TextRank to obtain the high informative parts of answers. Details of the parameters settings will be presented in the experiment section.

5.2 Generative Model, QRM-AC

For the answer content, we propose to link the extracted keywords of answers to the topic space that reflects the semantics of question. As the answer content is generated by the corresponding answerer, for each extracted keyword in answers, we assign it directly to the already known answerer, and sample a topic z from the topic space indicating the semantic relevance between questions and answers.

Graphical representation of proposed QRM-AC is shown in Figure 2(b). For the right part(generative process for questions), it is the same as QRM-AQ, we add an additional step after QRM-AQ to describe the generative process for answer content: for each word $w_{a,n}$ in answer content of question q, draw a topic $z_{a,n}$ according to the mixture weight of topics corresponding to user l', where $w_{a,n}$ occurs in the answer provided by l' for q.

5.3 Parameter Estimation

For parameter estimation, we have Equation (3) for words in questions. For answers, since w_a is generated by l', the probability of assigning w_a to other answerer is 0, that is, $p(z_i = z, l_i \neq l | w_i = w) = 0$. Thus, we employ an empirical inference as

$$p(z_i = z, l_i = l | w_i = w, l_{\setminus i}, z_{\setminus i}, w_{\setminus i}, \alpha, \beta) \propto \frac{n^z_{l',\setminus i} + \alpha}{n_{l',\setminus i} + T\alpha} \cdot \frac{n^w_{z,\setminus i} + \beta}{n_{z,\setminus i} + V\beta} \quad (9)$$

Similarly, parameter estimates can be obtained by Equation (4) and (5).

6. QUESTION ROUTING USING QRM

In this section, we utilize the parameters estimated from collected samples to build a question routing model, and then employ the routing model to further evaluate the answer quality.

6.1 Quality Distribution over Words

Each user in cQA can be represented by a multinomial distribution over topics, while each topic is also a multinomial distribution over terms. By this generative process shown in Figure 2, we obtain user's distribution over terms as

$$p(w|l) = \sum_z p(w|z)p(z|l) \quad (10)$$

since

$$p(w|z) \propto \varphi_{z,w}$$
$$p(z|l) \propto \psi_{l,z}$$

We rewrite Equation (10) as

$$p(w|l) = \sum_z \varphi_{z,w}\psi_{l,z} \quad (11)$$

For a new arrival question q_n, we propose to rank a list of users by user interest and answer quality. In order to incorporate the quality information contained in parameter $\phi_{l,z}$ to recommend suitable candidates, we substitute $\psi_{l,z}$ with $\phi_{l,z}$ and rewrite the Equation (11) as

$$p(w|l) = \sum_z \varphi_{z,w}\phi_{l,z} \quad (12)$$

We still denote quality distribution over words as $p(w|l)$ for convenience to measure user interest and authority for user ranking.

6.2 Routing Models

To predict a list of user to answer questions, two factors have to be taken into consideration: the relevance between user and the given q_n, the willingness of user to provide answer for q_n. The former can be written as $p(l|q_n)$ while the latter $p(q_n|l)$. Based on the bag of words assumption, we have

$$p(q_n|l) = \prod_{w \in q_n} p(w|l)$$

Therefore, we define the relevance score $ReleScore_{q_n}(l)$ for user l with question q_n as

$$ReleScore_{q_n}(l) = p(l|q_n) = \frac{p(q_n|l)p(l)}{p(q_n)} \propto p(q_n|l)p(l)$$

However, from the user perspective, high relevance between user l and question q_n does not indicate that l has the willingness to provide answers for q_n. We measure user's willingness as

$$WillScore_{q_n}(l) = p(q_n|l)$$

The willingness score measures the likelihood for l to answer q_n. In that cQA is a kind of user generated content, user dedication contributes to the prosperity of cQA, which mainly relies on the desire or willingness of individual user, we assume that ranking users according to the willingness tallies with real data better than relevance, which will be testified in the experiment section.

Table 2: Statistics about Raw Dataset

#.questions	14,280,752
#.questions with best answer	7,716,682(54.0%)
average #.anwers per question	3.33
#.users	14,582,092
#.users answered	9,245,912(63.4%)
#.users only answered	6,561,621(45.0%)
#.users without best answer	5,152,475(55.7% of users that answered questions)

7. EXPERIMENTS

In this section, we present details about the experiment setup, evaluating and discussing the result by QRM.

7.1 Dataset

To demonstrate the performance of QRM, we use the data from Soso Wenwen offered by Tencent, Inc.[5] These questions are posted between Sep 20, 2010 and Nov 20, 2010 on Soso Wenwen. Table 2 shows some statistics about the raw dataset. During this interval, 14,582,092 users were involved in the community interactivity, including posting and answering questions. However, there are about 63.4% users that have answered questions; the remaining only posted their questions on the community. Meanwhile, about 55.7% of users with answering history are never marked as the best answer providers and only 54.0% of the total questions come with a best answer, which indicates that chance of receiving high quality answers from community users is slim.

To alleviate unnecessary sparsity of the raw dataset and better demonstrate the performance of the QRM, we filter inactive and non-authoritative users with number of answers less than 20 or have never provide best answers, as for questions, we select those ones with at least one best answer and no less than 5 tokens. After that, we get 2,044,507 questions, among these questions, we randomly reserve 10% for future evaluation and the remaining as training set. Finally, we train our model on 1,840,437 questions covering 323,088 users and a vocabulary of 66,924 terms, the size of the reserved dataset is 204,070, with the average number of answers for each question is 3.32.

After some text pre-processing steps such as word segmentation, part-of-speech tagging and stop-words removal, we train QRM on the dataset. Since the parameters selection is not our focus, they are empirically set as $T = 400$, $\alpha = 50/T$ and $\beta = 0.01$.

For the evaluation of our model, we utilize the metric Precision@K(abbreviated as P@K) and Best Answer Coverage@K (abbreviated as C@K) to test the performance of QRM. P@K is the fraction of the top K recommended candidates that answered the question indeed. For example, P@10 is 0.1 means that only one of the top 10 recommended users answered the question in real situation. C@K is the percentage of the questions that its best answerer is in top K recommended users, and C@10 is 10% means that for 10% of the total questions, the corresponding best answer provider is in the top 10 ranked users.

7.2 Routing Model Selection

To illustrate that cQA relies more on user dedication, we employ the P@K to evaluate the user recommendation performance by *ReleScore* and *WillSocre*; the contribution weight factor λ is set to 1 using QRM-AQ, that is, discarding the difference between best answers and non-best answers.

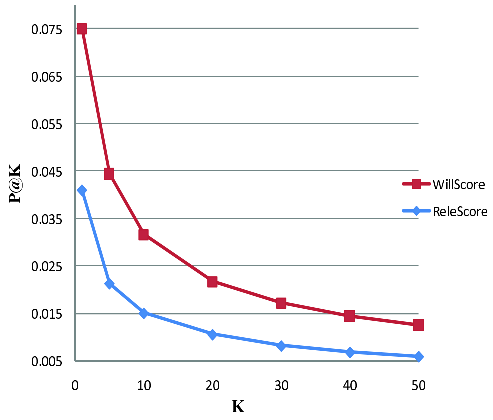


Figure 3: Performance of Routing Models

Figure 3 shows the comparison of ranking users by willingness to answer and relevance to the given question. Higher value means high accuracy in prediction with real data. It explicitly shows that ranking according the willingness outperforms relevance score, which is consistent with our assumption. Therefore, we perform question routing by calculating the willingness score for each user in the following demonstration of our experiment.

7.3 QRM-AQ, Incorporating Answer Quality

High performance for P@K is not equivalent to that the recommended candidates are capable to answer the questions with high quality. On this account, we exam the performance of P@K only for best answers providers, i.e., the fraction of the top K recommended candidates that provided best answers indeed. We evaluate the performance for QRM-AQ of incorporating answer quality.

7.3.1 *Post Sampling*

Post sampling is a straightforward method to recommend user with quality. We evaluate the performance of post sampling in consideration of different quality factor λ. Figure 4 presents the relationship between P@K and λ. It should be noticed that there is little chance of matching the predicted candidates with user that really answered question as best answer since there are a great many of users. A slight improvement will result in a great advancement in real situation. The results demonstrate that Post Sampling can efficiently improve the performance of question routing.

7.3.2 *Quality Sampling*

For Quality Sampling, Figure 5 presents the comparison of choosing different quality factor λ. Compared with the result shown in Figure 3, about half of our recommended users answer the best answer using the metric P@K, which demonstrates that QRM-AQ is capable to recommend authoritative users to provide relatively high quality answers.

Strategies Comparison.

We also make a comparison between the proposed three

[5]http://www.tencent.com/en-us/index.shtml

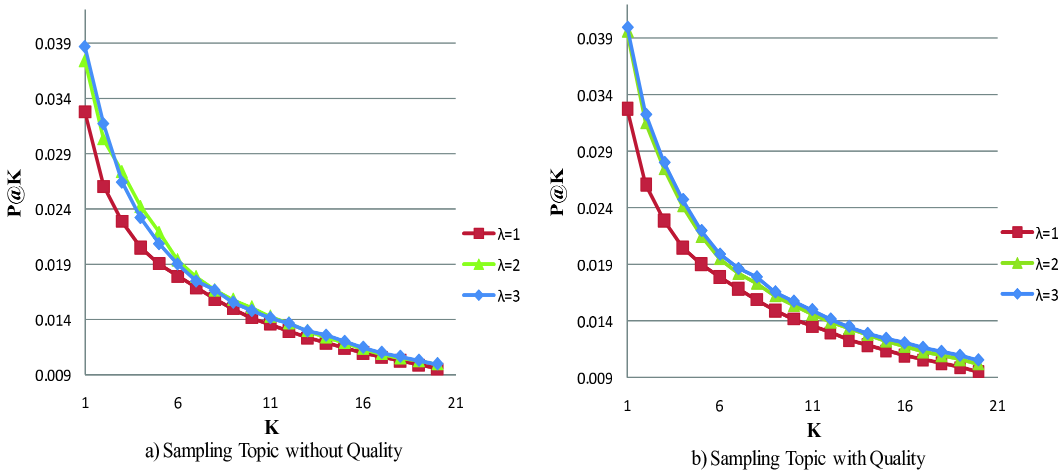


Figure 5: Quality Sampling

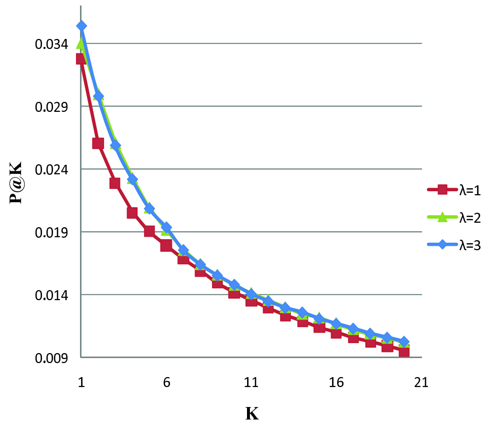


Figure 4: Post Sampling

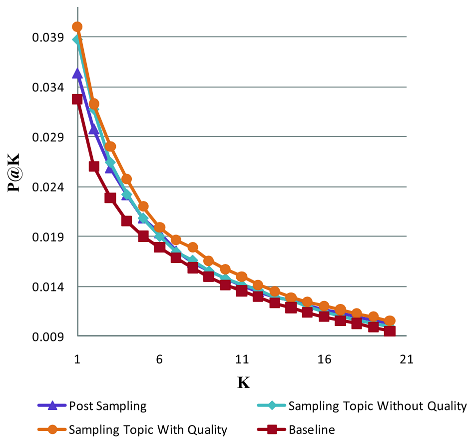


Figure 6: Strategies Comparison

strategies(Post Sampling, Quality Sampling without topic, and Quality Sampling with topic), the above results show that on average, setting $\lambda = 3$ reaches the best performance. Thus, we set λ to 3 for the comparison, and $\lambda = 1$ is chosen as the baseline. Figure 6 shows the comparison of the three proposed methods to incorporate answer quality. The method of quality sampling that sample the topic with quality reach the best performance. It is mainly due to the reason that incorporating quality to both user-topic layer and topic-term layer interprets real data better.

7.4 QRM-AC, Incorporating Answer Content

We utilize TextRank to extract keywords from answers. For each question in training set, we select nouns and adjectives in all the corresponding answers as vertices to build a graph, two vertices are connected if their corresponding lexical units co-occur within a window of 2, as reported in [8] to reach a relatively high performance. Then, we employ TextRank to bring order to all the selected nouns and adjectives. After that, we only re-rank nouns and adjectives for each answer according to the output order of TextRank, regardless of other words. As a result of noise in answers, we introduce the top ranked words only in best answers into the model during the training procedure.

Another intuitive problem that remains unsolved is how many extracted keywords shall be drawn into the model. To determine this, we build QRM-AC on four dataset: data with **a**nswer **c**ontent at varying degree by selecting top n keywords (AC-n,n=0,1,2,3). AC-0 means that no answer content is incorporated.

Figure 7 illustrates the performance comparison of incorporating answer content at different level. The result shows that with increasing answer content introduced, the performance reach the peak when incorporating top 2 words, which demonstrates that the semantic consistence between answers and questions contributes to the raise of performance in question routing; then it slowly goes down mainly due to the reason that though the points of answer and question are consistent, but the description from asker differs with answerer or there may exist noise in answers.

7.5 Evaluation of Best Answer Coverage

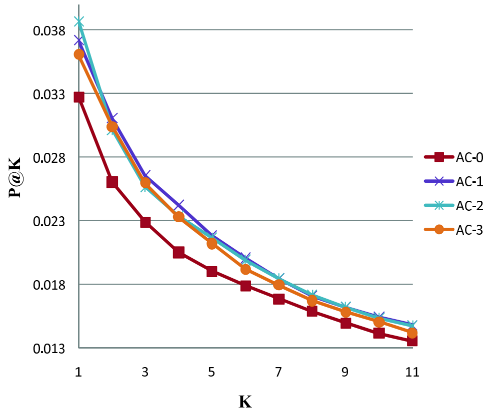


Figure 7: Incorporating Answer Content

Table 3: Percentage of best answer coverage

Model	C@K(%)					
	1	20	100	200	300	400
Baseline	3.27	18.29	32.78	40.20	44.43	47.52
PS-2	3.41	19.57	33.67	40.95	45.32	48.37
PS-3	**3.53**	**19.72**	**33.69**	**41.06**	**45.54**	**48.50**
QS-2	3.74	19.00	**33.79**	**40.65**	**44.78**	**47.67**
QS-3	**3.87**	**19.09**	33.38	40.56	44.58	47.39
QST-2	3.96	19.50	34.36	**41.62**	**45.86**	**48.67**
QST-3	**4.00**	**20.11**	**34.46**	41.47	45.54	48.45
AC-1	3.72	**20.17**	34.81	42.03	47.01	50.29
AC-2	**3.87**	19.80	**35.20**	42.83	**47.46**	**50.73**
AC-3	3.61	19.48	34.96	**43.09**	47.31	50.59

Now we test the performance of best answer coverage for QRM-AQ and QRM-AC using C@K. That is, given a collection of questions with multiple answers, calculate the percentage of questions that the corresponding best answer provider is in the top K rank.

Table 3 shows the results of best answer coverage. Note that PS, QS, QST represent Post Sampling, Quality Sampling for topic without quality and Quality Sampling for topic with quality respectively, QRM-AQ with $\lambda = 1$ is chosen as baseline. It shows that on average, question routing incorporating answer quality and answer content outperforms the baseline. For QRM-AQ, Post Sampling with quality factor 3 outperforms that with quality factor 2 when K ranges from 1 to 400. Quality Sampling is a little different, for QS and QST, when K is at a relative small value ($k \leqslant 20$ for QS, $k \leqslant 100$ for QST), the performance of Quality Sampling improves as λ increases. This is very crucial in real situation, cQA systems shall route questions to a minimum user set in order not to frequently bother the online users, and recommend user at a low cost. The results also illustrate that for QRM-AC, incorporating appropriate amount of answer content (top 2 keywords) helps improve the performance.

8. CONCLUSION AND FUTURE WORK

By building the model QRM-AQ, this paper is the first to introduce quality to entities incorporated in traditional topic model. Experimental results show that our methods have promising performance. Meanwhile, the QRM-AQ can be easily extended to other areas, for instance, in scientific area, the contribution weight of each author can be scaled by the sequence of author order in paper, the contribution or interest over topic can be better captured by QRM-AQ.

Noticing that answer content conveys some informative semantics, we propose the QRM-AC to model answer content. In that there exists much noise in answers, it is inappropriate to incorporate all the answer content. We firstly extract keywords that are highly informative, and then we train QRM-AC on a large collection of data with different amount of answer content. Performance on both question routing and best answer coverage demonstrate that QRM-AC has promising results.

For the future work, we plan to incorporate the mined user interest and authority from QRM with other information, such as the user online time, click history or other user behavior jointly for question routing in real products.

9. REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.

[2] M. Bouguessa, B. Dumoulin, and S. Wang. Identifying authoritative actors in question-answering forums - the case of yahoo! answers. In *Proceeding of KDD*, 2008.

[3] G. Dror, Y. Koren, Y. Maarek, and I. Szpektor. I want to answer, who has a question? yahoo! answers recommender system. In *Proceedings of KDD*, 2011.

[4] J. Guo, S. X. S. Bao, and Y. Yu. Tapping on the potential of q&a community by recommending answer providers. In *Proceedings of CIKM*, 2008.

[5] D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of WWW*, 2010.

[6] Q. Liu, E. Agichtein, G. Dror, E. Gabrilovichz, Y. Maarek, D. Pelleg, and I. Szpektor. Predicting web searcher satisfaction with existing community-based answers. In *Proceedings of SIGIR*, 2011.

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

[8] R. Mihalcea and P. Tarau. Textrank: Bringing order into texts. In *Proceedings of EMNLP*, 2004.

[9] M. Rosen-Zvi, T. Griffths, M. Steyvers, and P. Smyth. The author-topic model for authors and documents. In *Proceedings of UAI*, 2004.

[10] P. Serdyukov and D. Hiemstra. Modeling documents as mixtures of persons for expert finding. In *Proceedings of ECIR*, 2008.

[11] X. Si, E. Y. Chang, Z. Gyongyi, and M. Sun. Confucius and its intelligent disciples: Integrating social with search. In *Proceedings of the VLDB Endowment*, 2010.

[12] K. Sun, Y. Cao, X. Song, Y.-I. Song, X. Wang, and C.-Y. Lin. Learning to recommend questions based on user ratings. In *Proceedings of CIKM*, 2009.

[13] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: Structure and algorithms. In *Proceedings of WWW*, 2007.