

# A Classification-based Approach to Question Routing in Community Question Answering

Tom Chao Zhou<sup>1</sup>, Michael R. Lyu<sup>1</sup>, Irwin King<sup>1,2</sup>

<sup>1</sup>Department of Computer Science and Engineering  
The Chinese University of Hong Kong  
Shatin, N.T., Hong Kong  
{czhou, lyu, king}@cse.cuhk.edu.hk

<sup>2</sup>AT&T Labs Research  
201 Mission Street  
San Francisco, CA, USA  
irwin@research.att.com

## ABSTRACT

Community-based Question and Answering (CQA) services have brought users to a new era of knowledge dissemination by allowing users to ask questions and to answer other users' questions. However, due to the fast increasing of posted questions and the lack of an effective way to find interesting questions, there is a serious gap between posted questions and potential answerers. This gap may degrade a CQA service's performance as well as reduce users' loyalty to the system. To bridge the gap, we present a new approach to *Question Routing*, which aims at routing questions to participants who are likely to provide answers. We consider the problem of question routing as a classification task, and develop a variety of local and global features which capture different aspects of questions, users, and their relations. Our experimental results obtained from an evaluation over the Yahoo! Answers dataset demonstrate high feasibility of question routing. We also perform a systematical comparison on how different types of features contribute to the final results and show that question-user relationship features play a key role in improving the overall performance.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - search process; H.3.5 [Online Information Services]: Web-based services; H.4.m [Information Systems and Applications]: Miscellaneous

## General Terms

Algorithms; Design; Experimentation; Measurement

## Keywords

Community Question Answering; Question Routing; Classification

## 1. INTRODUCTION

Community-based Question and Answering services (CQA) have emerged as an effective way for knowledge dissemination and information seeking. Examples of CQA services

include Yahoo! Answers<sup>1</sup>, Google Confucius [29] and Baidu Zhidao<sup>2</sup>, as well as more social-oriented newcomers such as Quora<sup>3</sup>. By allowing users to ask complex natural language questions and to answer other users' questions, users' information needs are met by explicit, self-contained answers instead of lists of Web pages or documents. Thus, CQA services have provided a viable alternative to general purpose Web search [3, 23].

CQA sites such as Yahoo! Answers and Baidu Zhidao have archived hundreds of millions of questions, whereas these sites continue to receive a large number of questions [1]. However, due to the fast growth of the number of posted questions in CQA services, users might not get their posted questions resolved in a short period. We randomly sampled 3,640 questions from one popular CQA service Yahoo! Answers, and kept track of the status of the questions. These 3,640 questions were sampled from 26 first-level categories of Yahoo! Answers, with 140 questions in each category. After one day, we observed that only 434 (11.95%) of questions got resolved, and 726 (19.95%) questions in total got resolved in two days. This finding shows that a large number of posted questions cannot get resolved in a short period. Similar problem was also found in previous research works [18, 32]. As a result, some users may not post new questions but rely on other means of finding information if they cannot get their questions resolved during a reasonable time period.

Due to the lack of an effective question routing mechanism, a user is easily overwhelmed by the large number of open questions, and cannot easily find questions he/she is interested in answering even if he/she is willing to contribute his/her knowledge. Thus, there is a serious gap between the existing open questions and potential answerers. To bridge the gap, we present a new approach to *Question Routing*, which aims at routing open questions to suitable CQA users who may answer these questions. Question routing boasts several benefits. From the seeker's perspective, it can reduce the time lag between the time a question is posted and the time it is answered, and it can potentially increase the asker's satisfaction to CQA services [24]. In return, the asker may be more willing to contribute knowledge to the CQA service in the future. From the answer provider's perspective, because he/she will receive questions he/she is interested in instead of a large number of unfiltered questions,

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.  
CQA'2012, a WWW'2012 workshop, April 17, 2012, Lyon, France  
ACM 978-1-4503-1230-1/12/04.

<sup>1</sup><http://answers.yahoo.com>

<sup>2</sup><http://zhidao.baidu.com>

<sup>3</sup><http://www.quora.com>

the answerer would become more enthusiastic in providing answers. From the CQA system’s perspective, by linking open questions with suitable answerers, the CQA system could fully leverage users’ answering passion, leading to the improvement of the CQA system, as well as the boost of the user’s adhesiveness and loyalty to the system. In addition, because CQA services share a lot of properties with other social media systems [2], the knowledge unraveled can be ultimately applied to improving the performance of other social media applications.

In this paper, we consider the problem of question routing as a classification task, and develop a variety of local features which capture different aspects of questions, users, and their relationships. In addition, motivated by Cao et al. [7] that categorization information can be leveraged to improve question retrieval in CQA services, we further develop several global features to enhance the classification ability. More specifically, our contributions include:

- Proposing a classification-based approach to question routing in CQA services.
- Developing and evaluating a variety of local features, including question features, user history features, and question-user relationship features.
- Developing several global features, and integrating them with local features to further enhance the classification performance.
- Evaluating results over a real world dataset, indicating the derived features are essential for the classification task.

The rest of the paper is organized as follows. We introduce the question routing problem and present the features we have developed in Section 2. Experimental results are reported in Section 3. We discuss related work in Section 4, and conclude the paper in Section 5.

## 2. THE QUESTION ROUTING PROBLEM

### 2.1 Problem Definition

We consider the question routing problem as a classification task, and the definition is as follows:

**Question Routing Problem:** Given a question and a user in CQA, determine whether the user will contribute his/her knowledge to answer the question.

The reason we employ a classification approach is that we could find out what kinds of features are essential to question routing, and the knowledge is also valuable to other approaches. Classification approaches are commonly employed in analyzing CQA services [2, 12]. We believe that this definition captures two key entities of question routing problem, namely the question and the user. Therefore, our approach can be directly employed to route questions to suitable answerers.

### 2.2 Feature Investigation

Given a question and a user, we derive a variety of local features and global features. In addition, features investigated in this paper commonly exist in popular CQA services.

#### 2.2.1 Local Features

Given a question and a user, we first develop a variety of features organized around these two key entities: question features, user history features, and question-user relationship features. Features introduced here are referred to as local features, because only local information about question, user history and question-user relationships are needed. The complete list is reported in Table 1.

**Question:** This group includes features about the question, such as the length of the title (subject), the length of the detail, and the 5W1H question type (why, what, where, when, who, how).

**User History:** Because users’ history would have implications for users’ interests and behaviors in CQA services, this group contains features describing users’ history. We include users’ profile features such as member since, total points, etc. We also employ users’ question and answering behaviors such as number of answers provided, number of best answers provided, number of questions asked, etc. In addition, features indicating whether a user is enthusiastic in answering questions are included, such as the ratio of the number of answers provided and the number of questions asked.

**Question-User Relationship:** This group captures the relationship between a question and a user. We employ features adapted from the existing CQA service, such as whether the user is a top contributor in the category the question belongs to. We also include surrogate features that measure the extent the user is interested in the category the question belongs to, such as the ratio of the number of answered questions in the category the question belongs to and the number of answered questions. In addition, we adopt features describing the similarity of the question’s language model and the user’s language model, for example, we measure the KL-divergence between the current question’s title/detail and all the questions’ title/detail the user has provided answers.

#### 2.2.2 Global Features

Besides the local features developed above, we also design several global features to enhance the classification performance. These features are referred to as global features because they take into account the global information of the CQA service. In a CQA service, typically a question belongs to a certain category, and questions in the same category would discuss similar topics and are very likely to be semantically related [7, 18]. Thus, we believe that incorporating the global information would act as the smoothing effect and improve the robustness of the classification algorithm. The complete list is reported in Table 2.

**Question:** This group includes category-level features that could smooth each question such as average title length and average detail length. In addition, this group also contains the feature that could measure whether the question is representative in the category, such as the KL-divergence of the question title/detail with respect to other questions in the same category.

**User History:** This group of features are used to capture uniqueness of a user. For example, features such as the

**Table 1: Local Features**

Feature	Description
<b>Question</b>	
Q: Title length	Length of the question title (subject).
Q: Detail length	Length of the question detail.
Q: 5W1H-type	5W1H of the question (why, what, where, when, who and how).
<b>User History</b>	
UH: Member since	How long the user has registered in the CQA service.
UH: Percentage of best answer	Percentage of answers selected as best answers among all answers the user has provided.
UH: Total points	Total points the user has gained.
UH: # of answers	# of answers the user has provided.
UH: # of best answers	# of best answers the user has provided.
UH: # of asked questions	# of questions asked by the user.
UH: # of resolved questions	# of questions get resolved of the user.
UH: # of stars received	# of stars the user received.
UH: Answer/question ratio	# of answers provided/# of questions asked.
UH: Best answer/question ratio	# of best answers provided/# of questions asked.
<b>Question-User Relationship (The category means category the question belongs to.)</b>	
QU: Top contributor	Whether the user is recognized as a top contributor in the category.
QU: Ratio of answered question in the category	# of answers provided in the category/# of answers provided.
QU: Ratio of best answered question in the category	# of best answers provided in the category/# of best answers.
QU: Ratio of asked question in the category	# of asked questions in the category/# of asked questions.
QU: Ratio of starred question in the category	# of starred questions in the category/# of starred questions.
QU: Question’s KL-divergence with the user’s answered questions’	Question title/detail’s KL-divergence value with the user’s answered questions’ title/detail.
QU: Question’s KL-divergence with the user’s background language model	Question title/detail’s KL-divergence value with the user’s answered, asked, and starred questions’ title/detail.

KL-divergence value of the user’s answered questions’ title and detail with respect to all users’ answered questions’ title/detail are included.

**Question-User Relationship:** We hypothesize that the more similar between the language model of a user’s answered questions’ title/detail and that of the questions in a category, the more probable a user would answer the questions from the category. Thus, we include the feature of KL-divergence value of a user’s answered questions’ title/detail with questions’ title/detail in the category the given question belongs to.

### 3. EXPERIMENTS

#### 3.1 Classification algorithm

Support vector machines (SVM) [11] are widely used in many classification tasks due to its robustness in the presence of noisy data. In this paper, SVM was employed as the classification method to investigate the utility of derived features. Each feature value was scaled to the same range. More specifically, libsvm [9] with linear kernel was engaged. The reason we employed the linear kernel is that we could rank features’ importance by sorting the absolute weight values of the SVM model, and the weight value of the  $j$ -th feature could be calculate according to Eq. (1),

$$w_j = \sum_{i=1}^n \alpha_i y_i x_i^j, \quad (1)$$

where  $n$  is the number of training samples,  $\alpha_i$  is the support vector,  $y_i$  is the label, and  $x_i^j$  is the value of  $j$ -th feature of observation  $x_i$ .

#### 3.2 Evaluation Metrics

The task in the paper is a two-class classification problem, but we focus on the positive class, which means given a question-user pair, the user would answer the question. For the question routing problem, it is more important to have higher certainty about the positive class so that questions could be routed to suitable answerers effectively. We measure the *Precision*, *Recall* and *F1* for the positive class, and the overall *Accuracy* for both classes.

**Precision:** the fraction of the predicted *positive question-user pair* that were indeed observed in the data.

**Recall:** the fraction of all answered question-user pair that were correctly predicted by the system.

**F1:** the geometric mean of Precision and Recall measures, computed as  $2PR/(P + R)$ .

**Accuracy:** the overall fraction of instances classified correctly into the proper class.

#### 3.3 Datasets

Our data was crawled from the popular CQA service Yahoo! Answers during year 2009. The initial board categories to start the crawl were “Cars & Transportation”, “Computers & Internet”, and “Consumer Electronics”. We recorded all the user ids we found until the total number of users was 50,000. We then crawled all these 50,000 users’ profile pages and found 42,517 users were valid. A user was considered as valid if the following two conditions were satisfied: (1) the user still exists in the system; (2) the user sets his/her “Answers” page to be public and the user has answered at least one question. Among the 42,517 valid users, we randomly sampled 3,500 users, and systematically crawled their “Answers” page, “Questions” page, and “Starred Questions” page. The statistics of the crawled 3,500 users is presented

**Table 2: Global Features**

Feature	Description
<b>Question</b>	
Q: Average title length	Average title length in the category the question belongs to.
Q: Average detail length	Average detail length in the category the question belongs to.
Q: Question’s KL-divergence value with the category	Question’s title/detail’s KL-divergence value with this category’s questions’ title/detail.
<b>User History</b>	
UH: KL-divergence value of the user’s answered questions with all answered questions	KL-divergence of the user’s answered questions’ title/detail with all answered questions’ title/detail
UH: KL-divergence value of the user’s background language model with all users’ background language model	KL-divergence of the user’s answered, asked, and starred questions with all users’ answered, asked, and starred questions.
<b>Question-User Relationship</b>	
QU: KL-divergence value of the user’s answered questions with questions in the category the question belongs to.	KL-divergence of the user’s answered questions’ title/detail with questions in the category the question belongs to.

in Table 3, and in total there are 1,325,225 questions from 538 categories answered by these users and 88,852 questions asked by these users. From Table 3, we can find the average number of questions these users answered is much larger than the average number of questions they asked.

**Table 3: Statistics of users**

Name	Mean
# of answers provided	378
# of questions provided	25
% of best answers	12.34%
# of Total Points	1,660

After crawling the raw data, we prepared the positive instances and negative instances based on the crawled data. If a user answered a question, we considered the question-user pair as positive instance, and if a user asked a question, we considered the question-user pair as negative instance. The reason is that if a user asked a question, it might mean that he/she did not possess the knowledge about the question, and this indicates the user was unable to answer the question. We admit that the procedure of choosing negative instances is arguable, and we plan to investigate how to select negative instances more reasonably in the future.

In our experiments, we used the most frequent 1,000 unigram and bigram text features to represent questions’ and users’ language model, and we employed Porter Stemmer [27] to stem the words. We adopted the stop word list used by SMART system [6], but we removed the  $5W1H^4$  words from the stop word list.

### 3.4 The Effect of Local Features

We randomly selected 80% samples as training data, and 20% samples as testing data, and the average results of three rounds were reported. We used 10-folder cross-validation and employed a grid search on SVM parameter  $c$  over values of 0.0001, 0.001, 0.01, 0.1, 1, 100, 1000 to find the best parameter, and 1 is found to be optimal and consequently applied through out the following experiments.

We first evaluated the performance of groups of local features introduced in Section 2.2.1 individually. Table 4 demonstrates the results of precision, recall, F1 score and accuracy. It is important to note that we focus on positive class, and F1 should be considered as the most important metric here.

<sup>4</sup>who,what,where,when,why and how.

From Table 4, we can find the Question-User Relationship group of features achieves the best F1 score and Recall. The result is reasonable because Question-User Relationship features capture the user’s performance and interests in the category of the given question. In addition, Question-User Relationship features contain the KL-divergence value of the question and the user’s answered questions, and the KL-divergence value of the question and the user’s answered, asked and starred questions. Consequently, this group of features could capture the semantic relatedness of the given question and the user, and achieve the best F1 score. In addition, User History group of features achieve the highest precision, and this is because some users are quite active in the system. Given a question and a highly active user, the user would have very high probability to answer the question. According to the power law [1], however, these highly active users only account for a few percentage among all users. As a result, the recall rate for User History Group is relatively low compared with Question-User Relationship group, and the accuracy could be explained with similar reasons.

Besides the single group of features, we also investigate the performance of combined groups of features. Since Question-User Relationship group achieves the best F1, we combine the other two groups with it separately, and combine all the local features together. Table 5 presents the results. From Table 5, we find that the combination of User History and Question-User Relationship achieves better F1 than the combination of Question and Question-User Relationship. This is because User History has the best precision and Question-User Relationship has the best recall as shown in Table 4, and the combination could obtain the advantages of both. The combination of all the local features achieves the best F1 because the classification method could benefit from the most complete information.

We also investigate what kinds of features are the most significant ones among all local features. We rank importance of these features by sorting the absolute weight values of the SVM model as introduced before, and the 10 most significant features for the classification of question routing problem are listed as follows:

**QU** Question title/detail’s KL-divergence value with the user’s answered questions’ title/detail.

**QU** Question title/detail’s KL-divergence value with the user’s answered, asked, and starred questions’ title/detail.

**Table 4: Single Group of Local Features**

	Precision	Recall	F1	Accuracy
Question	0.5314	0.3896	0.4496	0.5157
User History	<b>0.8278</b>	0.4682	0.5981	<b>0.6805</b>
Question-User Relationship	0.5824	<b>0.935</b>	<b>0.7178</b>	0.6267

**Table 5: Combined Groups of Local Features**

	Precision	Recall	F1	Accuracy
Question + Question-User Relationship	0.5974	<b>0.9134</b>	0.7223	0.6435
User History + Question-User Relationship	0.7362	0.8275	0.7792	0.7619
Question + User History + Question-User Relationship	<b>0.7418</b>	0.8253	<b>0.7814</b>	<b>0.7655</b>
Top 10 features in Local Features	0.6964	0.8095	0.7487	0.7241

**Q** Length of the question title (subject).

**UH** How long the user has registered in the CQA service.

**UH** Percentage of answers selected as best answers among all answers the user has provided.

**UH** # of best answers the user has provided.

**UH** # of questions get resolved of the user.

**UH** # of questions asked by the user.

**UH** Total points the user has gained.

**UH** # of answers the user has provided.

The first two most important features are in line with our intuition. Question title/detail’s KL-divergence value with the user’s answered questions’ title/detail captures the most accurate semantic relatedness between the given question and the knowledge of the user. Besides sharing similar property with the most important feature, the second most important feature considers the user’s interests as well by incorporating factors such as starred questions.

Another interesting finding is that the length of the question title is shown to be very important, and this agrees with the previous finding of other researchers that question title length could be considered as an important feature to measure the quality of the question [2]. In other words, a question of good quality is easier to get answered.

We also investigated the classification performance by employing only the top 10 features, and the results are shown in Table 5. We can find that the result of employing the top 10 features is already encouraging.

### 3.5 The Effect of Local and Global Features

We also evaluated the performances of employing global features and the combination of local and global features. The experimental setting was the same with the previous section, namely 80% samples were randomly selected as training data, and the remaining 20% samples are considered as testing data. The average results of three rounds are reported, as shown in Table 6.

From Table 6, we observe that utilizing local features only achieves the best precision. The reason is as follows: given a question-user pair, local features could capture both the question’s and the user’s individual characteristic, as well as their relationship accurately. Consequently, local features could obtain the best precision. However, compared with global features, local features suffer the weakness that the

certainty about the true positive likelihood is not high for those question-user pairs whose local features’ information is not sufficient, for example, the user has only answered a few questions, or the question’s title/detail are very short, resulting in a relatively low recall compared with global features.

By considering the global information, such as category-level language models, global features achieve the best recall. This result is coincide with our intuition that in a CQA service, questions belonging to the same category are focused on similar topics and share similar language models. The effect of global features is quite similar to that in the smoothing technique [26, 33].

From Table 6, we find that the combination of local features and global features promises to maintain the best elements of the two, and the best F1 score is consequently achieved. In addition, this indicates that local features and global features could complement each other perfectly. We rank importance of the features with the same approach of the previous section, and the 10 most important features are shown as follows:

**Local-QU** Question title/detail’s KL-divergence value with the user’s answered questions’ title/detail.

**Local-QU** Question title/detail’s KL-divergence value with the user’s answered, asked, and starred questions’ title/detail.

**Global-Q** Question’s title/detail’s KL-divergence value with this category’s questions’ title/detail.

**Local-UH** How long the user has registered in the CQA service.

**Global-Q** Average detail length in the category the question belongs to.

**Local-Q** Length of the question title (subject).

**Local-UH** Percentage of answers selected as best answers among all answers the user has provided.

**Global-UH** KL-divergence value of the user’s answered questions’ title/detail with questions’ title/detail in the category the question belongs to.

**Local-UH** # of best answers the user has provided.

**Local-UH** # of questions get resolved of the user.

**Table 6: Local Features and Global Features**

	Precision	Recall	F1	Accuracy
Local Features	<b>0.7418</b>	0.8253	0.7814	0.7655
Global Features	0.5779	<b>0.8713</b>	0.6949	0.6109
Local + Global Features	0.7279	0.8499	<b>0.7842</b>	<b>0.7689</b>

We can find that both global features and local features play a crucial role in the classification task. Measured by KL-divergence, the semantic relatedness of a question and a user’s previously answered questions is the most important feature.

KL-divergence value of question’s title/detail with that of this category’s questions’ title/detail is also considered as a very important feature. This could be explained as follows: If a question is quite typical in the category, it would have higher chance to be answered by users, and this could also partially explain the reason why CQA services all have well-structured categories. In this way, users could post their questions to a suitable category and increase the chance of the question being answered.

We can also find that, the KL-divergence value of the user’s answered questions’ title/detail with questions’ title/detail in the category the given question belongs to, is also very important feature. The reason is that normally a user’s interests and knowledge would focus on a certain number of categories. This finding explains why the popular CQA sites such as Yahoo! Answers and Baidu Zhidao would recognize a few number of users as top contributors in each category, and some other popular CQA sites such as Answers.com would provide the category-level Really Simple Syndication (RSS) feed to enable users to get updates about questions on certain categories.

### 3.6 Consistency Analysis

To understand whether the proposed classification approach is consistent, we varied the size of training data and testing data, and report the average of three experiments. Table 7 presents the results of employing only local features, as well as the combination of local features and global features. 20% means we randomly selected 20% data as training, and left the remaining as testing data. From Table 7, we can find the combination of local and global features consistently outperform the one with only local features, particularly, on the most important metric F1. In addition, with the growth of the size of the training data, the classification method’s performance on the most important metric F1 increases monotonically. It is also observed in Table 7 that when varying the size of training data, the performance of employing only local features is also satisfactory. Considering the types of local features introduced in Table 1, we can see the computational cost for obtaining the local features is not very high, and with the development of the fast classification methods [10], the findings in this paper could be quite valuable at the practical level.

## 4. RELATED WORK

To facilitate answerer access to proper questions, approaches of question routing have been initiated and developed in CQA services. Zhou et al. [37] proposed expertise-based question routing in online forums. Li and King [18] proposed a language model based framework by combining expertise

estimation and availability estimation. Li et al. [19] proposed category-sensitive language models. However, their approaches focused on content analysis, and did not investigate what kind of features are important to question routing. Horowitz and Kamvar [13] developed a social search engine Aardvark, which routes the question to persons in the asker’s extended social network such as Facebook. Richardson and White [28] studied several prediction problems in a synchronous social Q&A system. Compared with social Q&A systems, our work focus on the community in CQA services rather than the social circles of the asker.

Community-based Question and Answering (CQA) has become an active area of research. Adamic et al. [1] studied Yahoo! Answers’ knowledge sharing activity, and they found that interactions in some categories resemble expertise sharing forums, while others incorporate discussion. Lou et al. [25] identified three types of motivations to share knowledge in CQA. Li et al. [20] studied how to identify questions on twitter. Bian et al. [3] proposed a general ranking framework for factual information retrieval from CQA. In order to find high quality content, Eugene et al. [2] took the first step to employ a supervised classification approach to analyze what kind of features would affect the quality of questions and answers, and there is subsequent work proposed by Bian et al. [4] in which the authors developed a semi-supervised coupled mutual reinforcement framework to find high-quality answers, questions and users from CQA sites. Liu et al. [22] proposed a competition-based user expertise score estimation approach. Wei et al. [30] studied how to integrate CQA archives with different taxonomies. Our work is related to above work because these efforts are all intended to improve users’ satisfaction to the CQA service. While previous approaches focused on how to retrieve high quality content on the existing CQA site, the proposed approach in this paper would help the CQA site to generate more high quality content by routing questions to suitable users to answer.

Our work is related to but distinct from link analysis and expert finding. Link-based algorithms PageRank [5] and HITS [17] were successfully applied in social media to find experts [15, 16, 34]. However, above approaches only considered from the user’s perspective, if a particular question is presented, above approaches could not utilize the specific characteristic of the current question to determine whether a user would answer the current question. Thus, previous approaches could not tackle the question routing problem. The proposed approach in this paper is different because by considering question feature, user history feature, and question-user relationship feature, given a question-user pair, the proposed approach could leverage these features to make a classification and decide whether to route a question to a user to answer.

Our work is also related to question search and question recommendation. Given a queried question, question search is to find the questions that are semantically equiv-

**Table 7: Precision, Recall, and F1 of Local Features, and Local+Global Features for varying amount of training data and testing data.**

		Precision	Recall	F1	Accuracy
20%	Local	0.8063	0.4506	0.5781	0.6712
	Local+Global	0.809	0.4832	0.6051	0.6846
40%	Local	0.8218	0.4559	0.5865	0.6815
	Local+Global	0.8155	0.5029	0.6222	0.6974
60%	Local	0.841	0.5111	0.6358	0.7055
	Local+Global	0.8414	0.513	0.6374	0.7064

alent to the queried questions. Jeon et al. [14] first studied question search problem in CQA services, and they employed translation model to tackle the problem. Xue et al. [31] proposed translation-based language model to find good answers for a user’s question. Zhou et al. [35] proposed phrase-based translation model for question retrieval. Zhou et al. [36] proposed topic-enhanced translation-based language model. Momtazi and Klakow proposed trained trigger language model to address the word mismatch problem. Recommending questions using the MDL-based tree cut model was proposed by Cao et al. [8]. Li and Manandhar [21] exploited information need for question recommendation. Although these approaches provide an alternative way to the asker to find out whether there are similar questions existing before the asker posted the question, they could not help the asker to solve the problem if the asker does not find similar questions. By adopting language model features from above methods into the classification framework, the proposed method provides a viable way to help users solve their questions effectively.

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, realizing that a large percentage of questions do not get resolved during a short period and the potential answerers do not have an effective way to find questions he/she is capable to answer, we present a new approach to question routing. We consider the question routing problem as a classification task, and derive a variety of local features and global features which capture different aspects of the question, the user and the relationship between them. We also analyze the contributions from different sources. Thorough experimental analysis indicates high feasibility of our approach. Our work opens a promising direction towards users’ knowledge modeling, personalized question routing, and can potentially lead to improvements to CQA services. In the future, we plan to employ a semi-supervised approach to leverage the large amount of unlabeled data to further improve the performance of the classification model. We also plan to investigate whether incorporating the probabilistic aspect into the classification model to combine different types of features would improve the performance.

## 6. ACKNOWLEDGEMENT

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 413210 and CUHK 415311) and two grants from Google Inc. (one for Focused Grant Project "Mobile 2014" and one for Google Research Awards).

## 7. REFERENCES

- [1] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th International Conference on World Wide Web*, pages 665–674, 2008.
- [2] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *Proceedings of the International Conference on Web Search and Web Data Mining*, pages 183–194, 2008.
- [3] J. Bian, Y. Liu, E. Agichtein, and H. Zha. Finding the right facts in the crowd: Factoid question answering over social media. In *Proceeding of the 17th International Conference on World Wide Web*, pages 467–476, 2008.
- [4] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *Proceedings of the 18th International Conference on World Wide Web*, pages 51–60, 2009.
- [5] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceeding of the 7th International Conference on World Wide Web*, pages 107–117, 1998.
- [6] C. Buckley, A. Singhal, M. Mitra, and G. Salton. New retrieval approaches using smart: Trec 4. In *Proceedings of the 4th Text Retrieval Conference (TREC-4)*, pages 25–48, 1995.
- [7] X. Cao, G. Cong, B. Cui, C. S. Jensen, and C. Zhang. The use of categorization information in language models for question retrieval. In *Proceedings of the 18th ACM International Conference on Information and Knowledge Management*, pages 265–274, 2009.
- [8] Y. Cao, H. Duan, C. Y. Lin, Y. Yu, and H. W. Hon. Recommending questions using the mdl-based tree cut model. In *Proceedings of the 17th International Conference on World Wide Web*, pages 81–90, 2008.
- [9] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [10] H.-P. Graf, E. Cosatto, L. Bottou, I. Dourdanovic, and V. Vapnik. Parallel support vector machines: The cascade svm. In *Advances in Neural Information Processing Systems (NIPS)*, 2005.
- [11] M. Hearst, S. Dumais, E. Osman, J. Platt, and B. Scholkopf. Support Vector Machines. *IEEE Intelligent systems*, 13(4):18–28, 1998.
- [12] L. Hong and B. D. Davison. A classification-based approach to question answering in discussion boards.

- In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development on Information Retrieval*, 2009.
- [13] D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the 19th International Conference on World Wide Web*, 2010.
- [14] J. Jeon, W. B. Croft, and J. H. Lee. Finding similar questions in large question and answer archives. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, pages 84–90, 2005.
- [15] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the 16th ACM International Conference on Information and Knowledge Management*, pages 919–922, 2007.
- [16] P. Jurczyk and E. Agichtein. Hits on question answer portals: exploration of link analysis for author ranking. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 845–846, 2007.
- [17] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.
- [18] B. Li and I. King. Routing questions to appropriate answerers in community question answering services. In *Proceeding of the ACM 19th conference on Information and Knowledge Management*, pages 1585–1588, 2010.
- [19] B. Li, I. King, and M. R. Lyu. Question routing in community question answering: Putting category in its place. In *Proceeding of the ACM 20th conference on Information and Knowledge Management*, pages 2041–2044, 2011.
- [20] B. Li, X. Si, M. R. Lyu, I. King, and E. Y. Chang. Question identification on twitter. In *Proceeding of the ACM 20th conference on Information and Knowledge Management*, 2011.
- [21] S. Li and S. Manandhar. Improving question recommendation by exploiting information need. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 1425–1434, 2011.
- [22] J. Liu, Y. Song, and C. Lin. Competition-based user expertise score estimation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development on Information Retrieval*, pages 425–434, 2011.
- [23] Q. Liu, E. Agichtein, G. Dror, E. Gabrilovich, Y. Maarek, D. Pelleg, and I. Szpektor. Predicting web searcher satisfaction with existing community-based answers. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development on Information Retrieval*, pages 415–424, 2011.
- [24] Y. Liu, J. Bian, and E. Agichtein. Predicting information seeker satisfaction in community question answering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 483–490, 2008.
- [25] J. LOU, K. Lim, Y. Fang, and Z. Peng. Drivers of knowledge contribution quality and quantity in online question and answering communities. In *Proceedings of the 15th Pacific Conference on Information Systems*, 2011.
- [26] C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, 2008.
- [27] M. Porter. An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
- [28] M. Richardson and R. White. Supporting synchronous social q&a throughout the question lifecycle. In *Proceedings of the 20th International Conference On World Wide Web*, pages 755–764, 2011.
- [29] X. Si, E. Chang, Z. Gyöngyi, and M. Sun. Confucius and its intelligent disciples: Integrating social with search. volume 3, pages 1505–1516, 2010.
- [30] W. Wei, G. Cong, X. Li, S. Ng, and G. Li. Integrating community question and answer archives. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.
- [31] X. Xue, J. Jeon, and W. B. Croft. Retrieval models for question and answer archives. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 475–482, 2008.
- [32] L. Yang, S. Bao, Q. Lin, X. Wu, D. Han, Z. Su, and Y. Yu. Analyzing and predicting not-answered questions in community-based question answering services. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.
- [33] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2):179–214, April 2004.
- [34] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceeding of the 16th International Conference on World Wide Web*, pages 221–230, 2007.
- [35] G. Zhou, L. Cai, J. Zhao, and K. Liu. Phrase-based translation model for question retrieval in community question answer archives. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 653–662, 2011.
- [36] T. C. Zhou, C.-Y. Lin, I. King, M. R. Lyu, Y.-I. Song, and Y. Cao. Learning to suggest questions in online forums. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.
- [37] Y. Zhou, G. Cong, B. Cui, C. Jensen, and J. Yao. Routing questions to the right users in online communities. In *Proceedings of the 25th IEEE International Conference on Data Engineering*, pages 700–711, 2009.