Joint Relevance and Answer Quality Learning for Question Routing in Community QA

Guangyou Zhou, Kang Liu, and Jun Zhao
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, China
{gyzhou, kliu, jzhao}@nlpr.ia.ac.cn

ABSTRACT

Community question answering (cQA) has become a popular service for users to ask and answer questions. In recent years, the efficiency of cQA service is hindered by a sharp increase of questions in the community. This paper is concerned with the problem of question routing. Question routing in cQA aims to route new questions to the eligible answerers who can give high quality answers. However, the traditional methods suffer from the following two problems: (1) word mismatch between the new questions and the users' answering history; (2) high variance in perceived answer quality.

To solve the above two problems, this paper proposes a novel joint learning method by taking both word mismatch and answer quality into a unified framework for question routing. We conduct experiments on large-scale real world data set from Yahoo! Answers. Experimental results show that our proposed method significantly outperforms the traditional query likelihood language model (QLLM) as well as state-of-the-art cluster-based language model (CBLM) and category-sensitive query likelihood language model (TCSLM).

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Storage and Retrieval

General Terms

Algorithms, Experimentation, Performance

Keywords

Question Routing, Translation Model, Language Model, Answer Quality

1. INTRODUCTION

Community question answering (cQA) provides an online service for users to share their knowledge in the form of questions and answers. cQA portals such as Yahoo! Answers[1] and Baidu Zhidao[2] have attracted increasing number of users and accumulated a large number of questions over the last few years. For example in Yahoo! Answers, it has more than 200 million users worldwide and around 15 million visits daily [12].

Although cQA service has brought significant benefits for users to seek information online, there are still several drawbacks in current systems. The most important problem is the efficiency of solving a new question. Previous study [12] has shown that more than 80% new questions cannot be resolved efficiently within 48 hours. On the other hand, with the rapidly increasing number of new questions, users who know well the answers to a particular domain are not easy to find their interested questions, which leads to the low participation rate (that is, most answers or knowledge in cQA comes from minority users) [4]. Besides, answer quality in cQA ranges from very high to low quality, sometimes abusive content or even spam [1]. Although cQA provides many mechanisms for community feedback ("thumbs up" and "thumbs down" votes), such community feedback requires some time to accumulate, and often remains sparse for obscure or unpopular topics.

To address the above problems, several approaches have been proposed in both industry and academic communities. In industry community, Horowitz and Kamvar [5] developed a social search engine, called Aardvark[3], which routed the question to the person in the user's extended social network most likely to be able to answer that question. Recently, a new question answering social network called Quora[4] has gained increasing popular. Users in Quora can follow topics and experts as well as following people in Twitter, and then answer the questions of the specified topics or route the new questions to experts.

In academic community, **question routing** has been conducted to tackle the above problems. The task of **question routing** is to route new questions to the eligible answerers who can give high quality answers [13, 20]. The traditional methods include the query likelihood language model (QLLM) [12], the cluster-based language model (CBLM) [20], and state-of-the-art category-sensitive language model (TCSLM) [13]. However, two problems of applying these methods to question routing are noted:

- **(1) Word mismatch:** As illustrated in Figure 1, we can see that the number of users who answered more than 4 questions is only 15.67% of the total number of users. The sparse data may lead to the word mismatch between the new questions and the answerer profiles.[5] Therefore, the traditional

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.

[1] http://answers.yahoo.com/
[2] http://zhidao.baidu.com/
[3] http://vark.com
[4] http://quora.com/
[5] An answerer profile usually consists of a small number of ques-

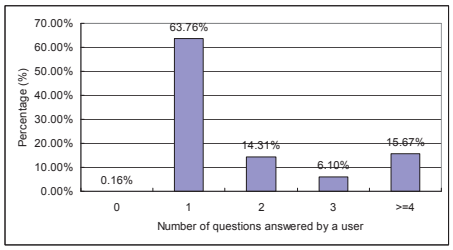

Figure 1: User participation rate in our obtained data set.

methods (e.g., QLLM, CBLM, TCSLM) fail to achieve satisfactory results.

- **(2) Answer quality variance:** Answer quality in cQA ranges from very high to low quality [1]. A user may answer a great number of questions which are relevant to the new questions, but we cannot conclude that the user must be an eligible answerer if his/her previous answers are of low quality.

To this end, we propose a joint relevance and answer quality learning method for question routing. First, we develop a general probabilistic model by taking both word mismatch and answer quality into a unified framework. We derive two scoring functions based on the framework: user interest score and answer quality score. Second, we estimate the user interest score by using the improved translation model, which models the exact matched words and the translated semantically related words using different approaches. Then, we propose to estimate the answer quality score by taking into account the expertise of answerers and the non-textual features of answers. Finally, we propose a refined strategy to better rank the eligible answerers for a new question. To the best of our knowledge, little work has addressed both the word mismatch and the answer quality variance problems into a unified framework in studies of question routing, which remains an under-explored research area. This paper is thus designed to fill the gap.

The rest of this paper is organized as follows. Section 2 presents our proposed joint relevance and answer quality learning method for question routing. Experimental results are presented in Section 3. Finally, we conclude with ideas for future work.

2. OUR APPROACH

2.1 A General Probabilistic Framework

In this paper, question routing is unified by means of a probability model. Specifically, we will rank the users according to the probabilities that a candidate user is "interest" and "answer quality" to a new question, and the key challenge is to compute these probabilities.

Formally, let $\mathcal{U} = \{u_1, u_2, \cdots, u_{|\mathcal{U}|}\}$ is the set of users in the community. Let I be a binary random variable to denote interest (1 for interest and 0 for non-interest). Let A be another binary random variable to denote answer quality (1 for high quality and 0 for low quality). Given a new question q and a candidate answerer $u \in \mathcal{U}$, we are interested in estimating the conditional probability of a user u being an eligible answerer:

$$\begin{aligned} & P(I=1, A=1|u,q) \\ = \; & P(I=1|u,q)P(A=1|u,q,I=1) \end{aligned} \tag{1}$$

tions. Also, the questions and answers themselves are always very short. Therefore, the insufficient word co-occurrence may lead to the word mismatch.

Here, we assume that A and I are independent with each other since the user's interest and the user's answer quality do not have directly relationship. Therefore, we have:

$$\begin{aligned} & P(I=1, A=1|u,q) \\ = \; & \underbrace{P(I=1|u,q)}_{\text{interest score}} \; \underbrace{P(A=1|u,q)}_{\text{answer quality score}} \end{aligned} \tag{2}$$

Based on the Bayes' Theorem and some assumption, we have

$$\begin{aligned} & \log P(I=1, A=1|u,q) \equiv \Big\{ \log P(q|u, I=1) \\ & + \log P(q|u, A=1) + \log \frac{P(u|I=1, A=1)}{P(u|I=0, A=0)} \Big\} \end{aligned} \tag{3}$$

Therefore, the question routing framework $P(I=1, A=1|u,q)$ is decomposed into three components, an interest score $P(q|u, I=1)$, an answer quality score $P(q|u, A=1)$, and a user prior ratio $\frac{P(u|I=1,A=1)}{P(u|I=0,A=0)}$. Following the literature [13, 20], we model the user prior ration as the number of answers u provided divided by the total number of answers. That is,

$$\frac{P(u|I=1, A=1)}{P(u|I=0, A=0)} = \frac{N_{ans}(u)}{N_{total}} \tag{4}$$

where $N_{ans}(u)$ denotes the number of answers provided by u, N_{total} denotes the total number of previously answered answers.

2.2 Interest Score Estimation as a Statistical Translation Model

In this paper, we borrow the idea of statistical machine translation and give a thorough analysis how to model the interest score estimation and learn the translation probabilities.

2.2.1 Model Formulation

Formally, let $D(u) = \{(u, q_1, a(u, q_1)), (u, q_2, a(u, q_2)), \cdots, \}$ denote the answerer profile of user u, which contains all questions previously answered by u. $a(u, q_j)$ denotes an answer of question q_j. For a new question q, u's interest on q is defined as follows:

$$P(q|u, I=1) = P(q|D(u), I=1) \tag{5}$$

$$P(q|D(u), I=1) = P_{TR}(q|D(u)) \tag{6}$$

Let $q = t_1 \cdots t_{|q|}$ and $D(u) = w_1 \cdots w_{|D(u)|}$. The statistical machine translation model [7, 17] assumes that both q and $D(u)$ are bag of words. However, the previous work [7, 17] cannot differentiate the importance between the exact matched terms and the translated semantically related words in ranking the relevancy of user profile to a new question q. Therefore, we define a improved model $P_s(t|D(u))$ used for "seen" words that occur in the user profile $D(u)$ (i.e., $\#(t, D(u)) > 0$), and a model $P_u(t|D(u))$ is used for "unseen" words that do not occur in the user profile (i.e., $\#(t, D(u)) = 0$). The improved translation model (ITR) can be rewritten as follows:

$$\begin{aligned} & \log P_{ITR}(q|D(u)) = \sum_{t \in q} \log P(t|D(u)) \\ = \; & \sum_{t \in q}^{\#(t,D(u))>0} \log P_s(t|D(u)) + \sum_{t \in q}^{\#(t,D(u))=0} \log P_u(t|D(u)) \\ = \; & \sum_{t \in q}^{\#(t,D(u))>0} \log \frac{P_s(t|D(u))}{P_u(t|D(u))} + \sum_{t \in q} \log P_u(t|D(u)) \end{aligned} \tag{7}$$

The improved model $P_s(t|D(u))$ can be computed like Jelinek-Mercer smoothed maximum likelihood estimation:

$$P_s(t|D(u)) = (1-\lambda)P_{ml}(t|D(u)) + \lambda P_{ml}(t|C) \quad (8)$$

The improved model $P_u(t|D(u))$ can be computed with Jelinek-Mercer smoothed translation model:

$$P_u(t|D(u)) = (1-\lambda)\Big[\sum_{w\in D(u)} P(t|w)P_{ml}(w|D(u))\Big] + \lambda P_{ml}(t|C) \quad (9)$$

where $P_{ml}(w|D(u))$ is the unigram probability of word w in $D(u)$, and $P(t|w)$ denotes the word-to-word translation probability. $\lambda \in [0,1]$ is the Jelinek-Mercer smoothing parameter [18]. $C = \sum_{u\in\mathcal{U}} D(u)$ is question-answer collection.

Previous work [7, 17] treat the exact matched words and the translated semantically related words equally, which may lead to non-optimal ranking performance because it is possible that a user profile that matches a new question word exactly ($P(t|t)$) gets less score contribution than a user profile that "matches" a new question word through translation ($P(t|w)$). On the contrary, the improved models defined in equation (8) and equation (9) treat the exact matched words and the translated semantically related words using different approaches. That is to say, the improved models only translate the unmatched words. It is thus reasonable to expect that using such improved models is likely to improve the performance of interest estimation, as we will show in our experiments. To learn the translation probabilities, we use the same way of [19].

2.3 Answer Quality Score Estimation

In subsection 3.2, we propose a improved translation model for answerers' interest estimation. This model assumes that a user has high interest on question q if he/she has previously answered some similar questions, but it does not consider the quality of the previous answers. In cQA, a user may answer a great number of questions which are relevant to q, but we cannot draw the conclusion that the user must be an eligible answerer if the previous answers are of low quality. We propose to estimate $P(q|u, A = 1)$ from previous answers' qualities of user u_i. Similar to [12], we assume that the user's answer quality on the new question q is the weighted average answers' qualities of similar questions he/she has previously answered:

$$P(q|u, A=1) = \frac{\sum_{u\leftarrow q_i} Q(a(u,q_i)) \cdot sim(q_i, q)}{\sum_{u\leftarrow q_i} sim(q_i, q)} \quad (10)$$

where $u \leftarrow q_i$ denotes question q_i answered by u with answer $a(u, q_i)$. $Q(a(u, q_i)) \in [0, 1]$ is the answer quality score of u's previous answer $a(u, q_i)$. $sim(q_i, q)$ is the relevance between q_i and q. To overcome the problem of word mismatch between the two questions, equation (7) can be used to measure the relevance $sim(q_i, q)$, except that we change u's profile $D(u)$ into a single question q_i. Due to its symmetry, we thus define a symmetric metric to better capture the relevance:

$$sim(q_i, q) = \frac{1}{2}(P_{ITR}(q_i|q) + P_{ITR}(q|q_i)) \quad (11)$$

Now we turn to calculate the answer quality score $Q(a(u, q_i))$ of u's previous answer $a(u, q_i)$. In this paper, $Q(a(u, q_i))$ can be derived from the asking expertise (Ask) and answering expertise (Ans) of its answerer u:

$$Q(a(u, q_i)) = \gamma Ask(u, q_i) + (1-\gamma)Ans(u, q_i) \quad (12)$$

where u answers q_i with answer $a(u, q_i)$, γ controls the relative importance of asking expertise. Based on the question answering relationships in cQA, the asking expertise and the answering expertise can be calculated using the HITS model [10] on the whole question-answer collections as proposed by Jurczyk and Agichtein [9].

However, *Jurczyk and Agichtein [9] assumes that each user has the same level of expertise for different topics of questions*. While in real applications, we assume that a user should have different levels of expertise when he/she answers different topics of questions. Our claim sounds reasonable since users usually have diverse background knowledge. So the asking expertise and answering expertise by taking the question topics into account at the $(t+1)^{th}$ iteration are computed based on the answering expertise and asking expertise at the i^{th} iteration as follow:

$$Ask^{(t+1)}(u, q_i) = \sum_{u\rightarrow q_j} Rel(c_j, c_i) \cdot \sum_{u_k \leftarrow a(u_k, q_j)} Ans^{(t)}(u_k, q_i) \quad (13)$$

$$Ans^{(t+1)}(u, q_i) = \sum_{\substack{u\leftarrow a(u,q_j) \\ u_k \rightarrow q_j}} Rel(c_j, c_i) \cdot Ask^{(t)}(u_k, q_i) \quad (14)$$

where $u \rightarrow q_j$ represents u asking question q_j, and $u \leftarrow a(u, q_j)$ represents u answering q_j with the answer $a(u, q_j)$. $Rel(c_j, c_i)$ reflects the topic relevance between q_j and q_i, where c_j is leaf category of question q_j, and c_i is leaf category of question q_i.[6]

The difference between our method (equation (13) and equation (14)) and previous work [15] is that we leverage the leaf category to represent the question topic instead of representing each question with a vector space model. In this way, our proposed method can substantially alleviate the data sparseness problem and thus make a more accurate estimation, as we will shown in the experiments.

Jeon et al. [8] argued that answers' goodness was also an important factor for answer quality prediction. Therefore, we incorporate answers' goodness score w_{kj} derived from the non-textual features of $a(u_k, q_j)$ into equations (13) and (14):

$$Ask^{(t+1)}(u, q_i) = \sum_{u\rightarrow q_j} Rel(c_j, c_i) \cdot \sum_{u_k \leftarrow a(u_k, q_j)} w_{kj} \cdot Ans^{(t)}(u, q_i) \quad (15)$$

$$Ans^{(t+1)}(u, q_i) = \sum_{\substack{u\leftarrow a(u,q_j) \\ u_k \rightarrow q_j}} w_{kj} \cdot Rel(c_j, c_i) \cdot Ask^{(t)}(u, q_i) \quad (16)$$

2.3.1 Question Topic Similarity Estimation

To estimate the question topic similarity $Rel(c_j, c_i)$ between two categories, answerer-based and content-based methods described in Li et al. [13] can be employed. However, we observe that some leaf categories consist of only a small number of questions, which may lead to the data sparseness. In this paper, we use the widely studied topic model $--$ Latent Dirichlet Allocation (LDA) [3] to identify the latent topic information from the large scale question-answer collection. To identify the topics that each leaf category is about using LDA, we aggregate all questions under the same leaf category into a big document. Thus, each document essentially corresponds to a leaf category. After utilizing LDA, each leaf category c can be represented as a Z-dimension vector topic distribution $P(z|c)$, where Z is the topic number. Thus, the task of

[6] In cQA, questions are equipped with hierarchical and systematic categories. When an asker posts a question, he/she is required to choose a leaf category that the question belongs to. Therefore, it is reasonable to use the leaf categories to represent the question topics.

Table 1: The features for estimating the parameters.

Features	Descriptions
Best answer ratio	The ratio of best answers given by the answerer
Answer length	Number of words of the answer
†Overlap	Words shared between the question and answer
Number of answers	Number of answers received for this question
†Number of comments	Number of comments added by other users
Total thumbs up	Total number of thumbs up for the answer
Total thumbs down	Total number of thumbs down for the answer
Number of categories	Number of categories that the answerer is declared
Stars	Number of stars given by the asker to the answer
Total points	Total points that the answerer has received

topic similarity is converted to calculate the distance between two leaf category vectors. Here, we propose to use normalized Kullback Leibler (KL) divergence [11]. The KL-divergence from c_j to c_i is computed by $KL\left(c_j||c_i\right) = \sum_z \ P\left(z|c_j\right)\log\frac{P(z|c_j)}{P(z|c_i)}$. Then we calculate the similarity between leaf categories c_j to c_i using Jensen Shannon divergence, which shows the superior performance than others. Thus, we have

$$Rel(c_j, c_i) = \frac{1}{2}\Big\{KL(c_j||c_i) + KL(c_i||c_j)\Big\} \qquad (17)$$

2.3.2 Answers' Goodness Estimation

We use logistic regression to measure each answer's goodness. Jeon et al. [8] proposed to predict the goodness of answers using 13 features. Here, we use 8 of 13 features used in Jeon et al. [8] and two additional features marked by † sign. The other five features are not used because they are either not available or not provided in Yahoo! Answers (e.g., click, copy and print counts). The features we used for estimating the parameters are listed in Table 1.

The features listed in Table 1 are non-monotonic features. Following [8], we convert these non-monotonic features into monotonic features using Kernel Density Estimation [6].

3. EXPERIMENTS

3.1 Data Set and Evaluation Metrics

We use the getByCategory function from the Yahoo! Answers API[7] to obtain the questions for the evaluation. More specifically, the data set consists of 359,152 resolved questions crawled from March 20, 2011 to September 18, 2011 under the top-level category at Yahoo! Answers, namely "Cars & Transportation". Under this category, there are 44 leaf categories. In this study, for all the resolved questions, the information of each question includes: (1) Texts of question and the associated answers, with stop words being excluded[8] and the words being stemmed.[9] (2) Answerers' IDs of each question. (3) Users' rating information (e.g., thumbs up, thumbs down, the best answers and so on.). (4) Time information about the question posted and answered.

To demonstrate the effectiveness of our approach, we split all questions into two disjoint sets:

Test Set: questions posted after August 10, 2011, used as questions to be routed.

Archive Set: the remaining data set.

Finally, test set is made up of 32,405 questions, 91,399 answers and 17,315 answerers. Archive set consists of 326,747 questions, 926,904 answers and 170,906 answerers. Similar to the previous work [12, 13], the ground truth for each question in test set is the answerers who actually answer it.[10]

Recall that we use a logistic regression to measure each answer's goodness. To train the model, sufficient labeled instances are needed. Following the literature [12], we use the community and the askers' choices to avoid manually labeling. For each question in archive set, the answer is labeled as "good" only if the following two conditions are met: (1) it is selected as the best answer; (2) it receives more than 50% of thumbs up for all answers of the question. Also, one answer is labeled as "bad" if it receives more than 50% of thumbs down for all answers of the question. Finally, 26,872 "good" instances and 31,093 "bad" instances used as training data to estimate the parameters of the logistic regression model.

Evaluation Metrics: we adopt Mean Reciprocal Rank (MRR) and Precision at N (P@N) as evaluation metrics for question routing, as they are widely used in evaluating the performance of question routing [13, 20].

Table 2: Performance of our proposed methods with traditional methods using two measures MRR and P@10.

#	Methods	MRR	P@10
1	QLLM	0.139	0.185
2	LK2010	0.151	0.192
3	TCSLM	0.160	0.229
4	LDALM	0.155	0.218
5	CBLM	0.146	0.205
6	**ITR-CSAQ**	**0.250**	**0.371**

3.2 Parameter Selection

The experiments use two parameters. The first is the smoothing parameter λ; and the second γ, controls the relative importance of asking expertise and answering expertise in equation (12)Following the literature, we set $\lambda = 0.2$ [18].

For parameter γ, we conduct experiments on a small development set of 440 questions (10 questions from each leaf category) to determine the best value among 0.1, 0.2, $\cdots$, 0.9 in terms of MRR. This set is also extracted from Yahoo! Answers at the top-level category of "Cars & Transportation", and it is not included in the test set. Finally, we set $\gamma = 0.2$ empirically as these settings yield the best performance.

3.3 Experimental Results

3.3.1 Comparing Joint Learning Model with Traditional Methods

We first look into how well our proposed joint relevance and answer quality learning model performs as compared with the traditional methods for question routing. Table 2 presents the results

[7]http://developer.yahoo.com/answers/

[8]http://truereader.com/manuals/onix/stopwords1.html

[9]http://tartarus.org/martin/PorterStemmer/

[10]In evaluation phase, we also remove the questions in test set whose answerers all do not appear in archive set.

Table 3: Performance of the improved translation model with the query query likelihood language model using two measures MRR and P@10.

#	Methods	MRR	P@10
1	QLLM	0.139	0.185
2	**ITR**	**0.224 (+0.085)**	**0.326 (+0.141)**

for different traditional methods and our proposed joint learning method according to two measures MRR and P@10.

In Table 2, row 1 is the query likelihood language model (QLLM). Row 2 is the answer quality smoothed query likelihood language model with user activity (LK2010) proposed by Li and King [12]. Row 3 is the category-sensitive QLLM (TCSLM) [13], which investigates a ground-breaking incorporation of question category to filter the irrelevant answerers. Row 4 is the mixture of LDA and QLLM (LDALM)[11] proposed by Liu et al. [14], and row 5 is the cluster-based language model (CBLM) [20]. The last one is our proposed joint learning model with the improved translation model and category-sensitive answer quality method (ITR-CSAQ).

From this table, we can see that the proposed ITR-CSAQ significantly outperforms all previous methods for question routing (row 1, row 3, row 4, and row 5 vs. row 6, all these comparisons are statistically significant at $p < 0.05$ by using t-test).

3.3.2 Comparing Improved Translation Model with Query Likelihood Language Model

We now look into how well the improved translation model (ITR) performs as compared with the query query likelihood language model (QLLM). Table 3 shows the results for QLLM and ITR methods according to two measures MRR and P@10. The comparison show that ITR significantly outperforms QLLM. Significant test using t-test show the difference between these two methods are statistically significant (row 1 vs. row 2).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel joint learning method by taking both word mismatch and answer quality into a unified framework for question routing. Experimental results conducted on cQA data set demonstrate that our proposed method significantly outperforms the traditional methods.

There are some ways in which this research could be continued. First, user reputation should be considered, so it is necessary to combine our proposed approach with the semi-supervised coupled mutual reinforcement framework [2] for question routing. Second, we will try to investigate the use of external sources of social relations between users to enhance the performance of question routing, such as the method of Horowitz and Kamvar [5]. Third, inspired by Wang and McCallum [16], we could take the temporal information of questions and answers into consideration. After this extension, we may find the interest change of users over time, and then route the questions by estimating the current interest of answerers.

5. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 61070106), the National Basic Research Program of China (No. 2012CB316300), Tsinghua National Laboratory for Information Science and Technology (TNList), Cross-discipline Foundation and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (No. 5026035403). We thank the anonymous reviewers for their insightful comments.

[11]We use GibbsLDA++ to estimate the posterior probability of LDA. The default parameter setting is used and the number of topics is set to 100 because these settings give the best performance.

6. REFERENCES

[1] E. Agichtein, C. Castillo, and D. Donato. Finding high-quality content in social media. In *WSDM*.

[2] 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 *WWW*.

[3] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] J. Guo, S. Xu, S. Bao, and Y. Yu. Tapping on the potential of q&a community by recommendation answer providers. In *CIKM*, pages 921–930, 2008.

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

[6] J. Hwang, S. Lay, and A. Lippman. Nonparametric multivariate density estimation: a comparative study. *IEEE Transactions of Signal Processing*, 42(10):2795–2810, 1994.

[7] J. Jeon, W. Croft, and J. Lee. Finding similar questions in large question and answer archives. In *CIKM*, 2005.

[8] J. Jeon, W. Croft, J. Lee, and S. Park. A framework to predict the quality of answers with non-textual features. In *SIGIR*, 2006.

[9] P. Jurczyk and E. Agichtein. Discovering authorities in question answer communities by using link analysis. In *CIKM*, pages 919–922, 2007.

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

[11] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[12] B. Li and I. King. Routing questions to approriate answerers in community question answering services. In *CIKM*, pages 1585–1588, 2010.

[13] B. Li, I. King, and M. Lyu. Question routing in community question answering: putting category in its place. In *CIKM*, 2011.

[14] M. Liu, Y. Liu, and Q. Yang. Predicting best answerers for new questions in community question answering. In *WAIM*, pages 127–138, 2010.

[15] M. Suryanto, E. Lim, A. Sun, and R. Chiang. Quality-aware collobarative question answering: methods and evaluation. In *WSDM*, pages 142–151, 2009.

[16] X. Wang and A. McCallum. Topic over time: a non-markov continuous-time model of topical trends. In *KDD*, pages 424–433, 2006.

[17] X. Xue, J. Jeon, and W. Croft. Retrieval models for question and answer archives. In *SIGIR*, pages 475–482, 2008.

[18] C. Zhai and J. Lafferty. A study of smooth methods for language models applied to ad hoc information retrieval. In *SIGIR*, pages 334–342, 2001.

[19] G. Zhou, L. Cai, J. Zhao, and K. Liu. Phrase-based translation model for question retrieval in community question answer archives. In *ACL*, pages 653–662, 2011.

[20] Y. Zhou, G. Cong, B. Cui, C. Jensen, and J. Yao. Routing questions to the right users in online communities. In *ICDE*, pages 700–711, 2009.