Question Routing in Community Question Answering: Putting Category in Its Place

Baichuan Li
The Chinese University of Hong Kong, Hong Kong
bcli@cse.cuhk.edu.hk

Irwin King*
AT&T Labs Research, USA &
The Chinese University of Hong Kong, Hong Kong
irwin@research.att.com
king@cse.cuhk.edu.hk

Michael R. Lyu
The Chinese University of Hong Kong, Hong Kong
lyu@cse.cuhk.edu.hk

ABSTRACT
This paper investigates a ground-breaking incorporation of question category to Question Routing (QR) in Community Question Answering (CQA) services. The incorporation of question category was designed to estimate answerer expertise for routing questions to potential answerers. Two category-sensitive Language Models (LMs) were developed with large-scale real world data sets being experimented. Results demonstrated that higher accuracies of routing questions with lower computational costs were achieved, relative to traditional Query Likelihood LM (QLLM), state-of-the-art Cluster-Based LM (CBLM) and the mixture of Latent Dirichlet Allocation and QLLM (LDALM).

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—information filtering, selection process

General Terms
Algorithms, Experimentation, Performance

Keywords
question routing, community question answering, question category, category-sensitive language model

1. INTRODUCTION
Since the inception of forums for asking and answering questions, Community Question Answering (CQA) services have been providing users with web platforms to obtain useful information, for example, the development of Yahoo! Answers1 and Quora2. In recent years, the efficiency of CQA services, however, is challenged by a sharp increase of questions raised in the communities. Such increasing amount of questions have thus influenced access of answerers to their appropriate questions, with the process of question answering being hindered in CQA services [2]. To facilitate answerer access to proper questions, an approach of Question Routing (QR) has been initiated and developed in CQA services [2, 4, 6, 3, 7].

The concept of QR refers to routing newly posted questions to potential answerers; the appropriateness of potential answerers (expertise estimation, hereafter) is estimated based on archives of their previously answered questions. Volumes of studies have been conducted regarding expertise estimation, including Query Likelihood Language Model (QLLM) [5], Cluster-Based Language Model (CBLM) [7], mixture of Latent Dirichlet Allocation (LDA) and QLLM [4]. However, as for an answerer, a complete set of questions the answerer has answered is utilized in the models, although certain amount of answered questions might be irrelevant to questions to be routed. To solve this problem, question category will be utilized to sift out irrelevant questions of an answerer to enhance the efficiency of expertise estimation. In CQA services, askers have to choose a category for the question they asked. As shown in Fig. 1, each question is classified into a particular category. The categories of new questions would allow much latitude in screening irrelevant questions of an answerer to enhance the efficiency of expertise estimation. To date, few attempts have been made regarding category

*Irwin King is currently on leave from CUHK to be with AT&T Labs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2011 ACM 978-1-4503-0717-8/11/10 ...$10.00.

Figure 1: An example of question category in CQA services (captured from Yahoo! Answers on January 20, 2011)

1http://answers.yahoo.com
2http://www.quora.com
information in studies of QR. This study was thus designed to fill the gap.

The paper is organized as follows. Related work is first reviewed in Section 2. Category-sensitive LMs are developed in Section 3. Experimental setup as well as results are then reported in Section 4 and 5. In the end, a conclusion is drawn in Section 6.

2. RELATED WORK

Expertise estimation, as mentioned, has been of paramount importance to assess potential of answerers for solving questions to be routed [2, 4, 6, 3, 7]. In studies of expertise estimation, two families of models have been widely employed: Language Models [3, 7] and Topic Models [2, 6]. Cao et al. [1] leveraged question category to enhance estimations to be routed [2, 4, 6, 3, 7]. In studies of expertise estimation, two families of models have been widely employed:

3. QUESTION CATEGORY FOR ROUTING QUESTIONS

Let $C = \{c_1, c_2, c_3, ..., c_n\}$ represents all leaf categories, the basic category-sensitive LM (BCS-LM) is defined as follows:

$$E(u_i, q_r, c_j) \equiv P_{bcx}(u_i|q_r, c_j),$$

(1)

$$P_{bcx}(u_i|q_r, c_j) \propto P_{bcx}(q_r, c_j|u_i)P(u_i),$$

(2)

$$P_{bcx}(q_r, c_j|u_i) = P_{bcx}(q_r, c_j, u_i)P(c_j|u_i),$$

(3)

$$P_{bcx}(q_r|c_j, u_i) = P_{bcx}(q_r|c_j, u_i) = \prod_{u \in q_r} P(q_u|u),$$

(4)

and

$$P(w|u) = (1 - \lambda)P_{all}(w|u) + \lambda P_{all}(w|Coll),$$

(5)

where $c_j$ is $q_r$’s category, $P(c_j|u_i)$ denotes the probability of answering questions in $c_j$ for $u_i$, and $q_{u_j}$ represents the question texts of all previously answered questions in $c_j$ for $u_i$. It is noted that BCS-LM is based on the same-leaf-category assumption, with potential answerers under similar leaf categories being omitted. As shown in Fig. 2, CQA portals like Yahoo! Answers set refined category hierarchy. Under one main category, there exist similar leaf categories. For example, the leaf categories of “Programming & Design” and “Software” in Fig. 2. Answerers with expertise in “Programming & Design” may also be an expert on questions asked in “Software”. To supply such omissions, we have come up with a transferred category-sensitive QLLM (TCS-LM) as follows:

$$P_{txt}(q_r, c_j|u_i) = \frac{\beta P_{bcx}(q_r, c_j|u_i) + \sum_{c_k \in Trans(c_j)} T(c_k \rightarrow c_j) P_{bcx}(q_r, c_k|u_i)}{\beta + \sum_{c_k \in Trans(c_j)} T(c_k \rightarrow c_j)}$$

(6)

where $\beta$ adjusts the weight between the original leaf category and other similar leaf categories, the lower $\beta$, more weights are given to similar categories. $Trans(c_j)$ denotes the set of categories which are transferable from (similar to) $c_j$ and $T(c_k \rightarrow c_j)$ represents the probability of transferring from category $c_k$ to $c_j$.

We define $c_k \in Trans(c_j)$ if $T(c_k \rightarrow c_j) \geq \delta$. (7)

where $\delta$ is a threshold between 0 and 1.

We use an answerer-based approach to estimate the transferring probability between two categories, which assumes that if there are many same answerers posting answers in two categories, these two categories should be similar with each other. To be specific, we construct a category-answerer matrix $E$ from resolved questions, each row of $E$ represents one (leaf) category and each column represents one answerer. In addition, the value of $e_{ij}$ denotes the number of answers $u_i$ provided in category $c_j$. Let $e_j$ and $e_k$ denote two row vectors of $c_j$ and $c_k$, we apply the cosine similarity to estimate the transferring probability ($T_{ans}()$) between two categories:

$$T_{ans}(c_j \rightarrow c_k) = T_{ans}(c_k \rightarrow c_j) = \frac{e_j \cdot e_k}{|e_j||e_k|}. \quad (8)$$

4. EXPERIMENTAL SETUP

4.1 Data Collection

The data comprised over 400,000 resolved questions (June to October 2010) from Computers & Internet and Entertainment & Music categories of Yahoo! Answers through provided API. The two categories included 20 and 25 leaf categories respectively. Table 1 reports the statistics of datasets. As for all selected questions, the information regarding affiliated category, texts and answerer IDs was available. Those questions were further classified into Set A (Test dataset, 50,377 questions, and 243,167 answerers) and Set B (Archive data, remaining questions: 174,639 answers and 49,476 answerers). In addition, answerers in Set A were used as ground truth.

| Table 1: Description of the Yahoo! Answers data set (after stop words removing and stemming) |
|---------------------------------|-----------------|-----------------|-----------------|
| Number of questions             | 443,072         | 50,377          |
| Number of answerers             | 1,510,534       | 240,277         |
| Average number of answers for one question | 3.49             | 4.87             |
| Maximum number of answers for one question | 50               | 30.08            |
| Mean first reply duration (in minutes) | 197.32           |                  |
| Average question length in words (both subject and content) | 43.87             |
| Average answer length in words  | 30.08            |                  |
| Number of askers                | 240,277         | 270,043         |
| Number of answers               | 171,726         | 68,551          |
| Number of both askers and answerers | 201,492         |                  |

4.2 Methods Compared

Cluster-based language model (CBLM) [7] and mixture of LDA and QLLM (LDALM) [4] were selected to be compared.

http://developer.yahoo.com/answers/

The leaf category Polos & Surveys was excluded since this leaf category was used to elicit public opinion. The dataset was thus composed of 44 leaf categories.
5. EXPERIMENTAL RESULTS

5.1 Category-Sensitive LMs

Table 2 reports $\text{Prec}@K$ for all algorithms with different $K$s from 1 to 100, and Table 3 presents the MRR and MAP of all methods. Table 4 gives the time-efficiency of each method in QR based on MQRT.

5.1.1 Higher Accuracies

From Table 2 we observe that, for various of $K$s, both BCS-LM and TCS-LM outperform QLLM significantly on $\text{Prec}@K$. For instance, when routing questions to the top 1 answerers, on average QLLM gives less than 8 successful routings per 100; BCS-LM and TCS-LM make more than 11 and 12 successful routings, which improve QLLM by 40.13% and 54.34% respectively. For other $K$s, category sensitive LMs also perform better than QLLM.

The MRR of BCS-LM and TCS-LM increase that of QLLM by 29.66% and 34.59%. From the definition of MRR, each new question will be answered by at least one answerer in the top 5 answerers using BCS-LM or TCS-LM. However, with QLLM on average we have to route the question to almost top 7 answerers to get an answer.

As to MAP, BCS-LM and TCS-LM improve QLLM by 33.08% and 37.29% respectively and it shows that category sensitive LMs give more accurate rankings on the whole.

5.1.2 Lower Costs

Now let’s turn to the time costs of QLLM and category-sensitive LMs. Table 4 gives the average time of routing a question for each model. We find that BCS-LM saves 47.16% of time while TCS-LM costs 13.80% less time than QLLM, which demonstrates that category-sensitive LMs are more efficient than QLLM in expertise estimation and thus make QR faster. The lower costs of BCS-LM lie in only relevant profiles are utilized in expertise estimation, which reduces the computational cost. TCS-LM spends more time than BCS-LM because of employing profiles in relevant categories for expertise estimation. Although TCS-LM is more time-consuming than BCS-LM, it is possible to reduce the time through parallel computing since the expertise estimation with different categories’ profiles is independent with each other.

5.1.3 BCS-LM vs. TCS-LM

Looking at Table 2, we find that similar categories improve accuracies of expertise estimation when $K$ is small. In particular, the $\text{Prec}@1$ of TCS-LM is 10.14% higher than those of BCS-LM. In addition, the $\text{Prec}@10$ of TCS-LM is 2.04% more accurate than that of BCS-LM. Although when $K$ becomes large (say, high than 40), TCS-LM improves fewer or even a little worse than BCS-LM, the former is still a better choice as a QR system has to route a question to minimum number of potential answerers in practice. The MRR and MAP of TCS-LM are also better than those of BCS-LM from Table 3. TCS-LM utilizes similar categories’ profiles...
Table 2: Various methods’ Prec@K in QR versus various Ks (best results in bold)

<table>
<thead>
<tr>
<th>K</th>
<th>QLLM</th>
<th>BCS-LM</th>
<th>TCS-LM</th>
<th>LDALM</th>
<th>CBLM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0795</td>
<td>0.1114</td>
<td>0.1227</td>
<td>0.0989</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>0.1659</td>
<td>0.2364</td>
<td>0.2340</td>
<td>0.1950</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>0.2091</td>
<td>0.2727</td>
<td>0.2706</td>
<td>0.2455</td>
<td>0.0000</td>
</tr>
<tr>
<td>10</td>
<td>0.2705</td>
<td>0.3866</td>
<td>0.3455</td>
<td>0.3102</td>
<td>0.0000</td>
</tr>
<tr>
<td>20</td>
<td>0.3386</td>
<td>0.3999</td>
<td>0.3932</td>
<td>0.3110</td>
<td>0.0091</td>
</tr>
<tr>
<td>40</td>
<td>0.4136</td>
<td>0.4523</td>
<td>0.4591</td>
<td>0.3492</td>
<td>0.0273</td>
</tr>
<tr>
<td>60</td>
<td>0.4477</td>
<td>0.4818</td>
<td>0.4795</td>
<td>0.4649</td>
<td>0.0545</td>
</tr>
<tr>
<td>80</td>
<td>0.4727</td>
<td>0.4955</td>
<td>0.4909</td>
<td>0.4867</td>
<td>0.0727</td>
</tr>
<tr>
<td>100</td>
<td>0.4999</td>
<td>0.5159</td>
<td>0.5114</td>
<td>0.4979</td>
<td>0.0795</td>
</tr>
</tbody>
</table>

6. CONCLUSION

This paper reported here is an investigation of applying question category to QR in CQA services. The question category was adopted to the development of category-sensitive LMs for estimating answerer expertise. Experiments on large-scale real world data revealed that category-sensitive LMs obtained more accuracies of expertise estimation, relative to QLLM and state-of-the-art algorithms including CBLM and LDALM. Results of experiments have proven that higher accuracies with lower costs are achieved due to the inclusion of question category in routing questions, which have therefore provided empirical evidence to validate the incorporation of question category in QR for CQA services. In future work, effects of question category on the content quality of answers and questions in CQA services can be further detected.

ACKNOWLEDGEMENT

This work is supported by two grants from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK 413210 and Project No. CUHK 415410) and a grant supported by a research funding from Google Focused Grant Project “Mobile 2014”.

7. REFERENCES