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! Answers[1] and Quora[2]. 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 sifted out irrelevant questions in profile of *an answerer* for 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

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

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

[1]http://answers.yahoo.com
[2]http://www.quora.com

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

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 question retrieval in CQA and the experimental results ensured this approach's effectiveness on various of retrieval models. To our knowledge, no previous work, however, estimates answerer expertise using question category for QR.

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:

$$\begin{aligned} E(u_i, q_r, c_j) &\equiv P_{bcs}(u_i|q_r, c_j), &(1)\\ P_{bcs}(u_i|q_r, c_j) &\propto P_{bcs}(q_r, c_j|u_i)P(u_i), &(2)\\ P_{bcs}(q_r, c_j|u_i) &= P_{bcs}(q_r|c_j, u_i)P(c_j|u_i), &(3)\\ P_{bcs}(q_r|c_j, u_i) &= P_{bcs}(q_r|c_j, q_{u_i}) = \prod_{\omega \in q_r} P(\omega|q_{u_{ij}}), &(4)\end{aligned}$$

and

$$P(\omega|q_{u_{ij}}) = (1-\lambda)P_{ml}(\omega|q_{u_{ij}}) + \lambda P_{ml}(\omega|Coll), \quad (5)$$

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 $q_{u_{ij}}$ 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:

$$P_{tcs}(q_r, c_j|u_i) = \frac{\beta P_{bcs}(q_r, c_j|u_i) + \sum_{c_k \in Tran(c_j)} T(c_k \to c_j)P_{bcs}(q_r, c_k|u_i)}{\beta + \sum_{c_k \in Tran(c_j)} T(c_k \to c_j)}, \quad (6)$$

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

We define

$$c_k \in Tran(c_j)\ if\ T(c_k \to c_j) \geq \delta, \quad (7)$$

where δ 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_{ji} denotes the number of answers u_i provided in category c_j. Let $\mathbf{e}_j$ and $\mathbf{e}_k$ denote two row vectors of c_j and c_k, we apply the cosine similarity to estimate the transferring probability ($T_{ans}(\cdot)$) between two categories:

$$T_{ans}(c_j \to c_k) = T_{ans}(c_k \to c_j) = \frac{\mathbf{e}_j \cdot \mathbf{e}_k}{|\mathbf{e}_j||\mathbf{e}_k|}. \quad (8)$$

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[3]. The two categories included 20 and 25 leaf categories respectively[4]. 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 data, questions posted after 6 May: 382,695 questions, 1,335,892 answers and 243,167 answerers) and Set B (Archive data, remaining questions: 50,377 questions, 174,639 answers and 49,466 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	433,072
Number of answers	1,510,531
Average number of answers for one question	3.49
Maximum number of answers for one question	50
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
Number of answerers	270,043
Number of both askers and answerers	68,551
Number of askers only	171,726
Number of answerers only	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

[3]http://developer.yahoo.com/answers/
[4]The leaf category *Polls & Surveys* was excluded since this leaf category was used to elicit public opinion. The dataset was thus composed of 44 leaf categories.

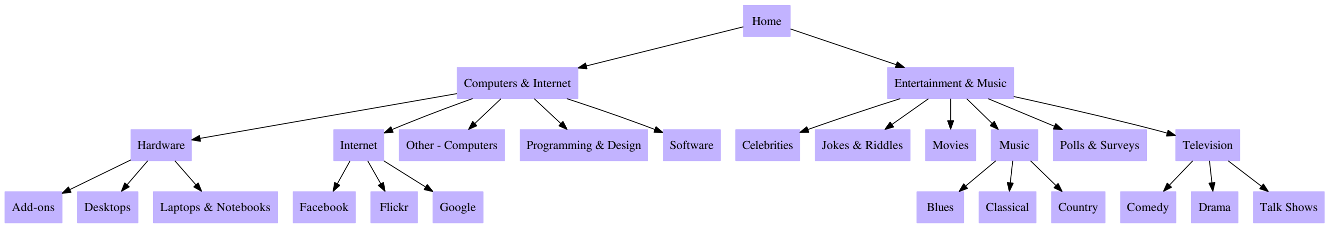

Figure 2: Part of category hierarchy in Yahoo! Answers

with category-sensitive LMs for expertise estimation based on the following two considerations:

1. In CBLM [7], similar questions under same topic are clustered and answerer expertise is estimated through calculating answerer's contribution to each cluster and the similarity between the routed question and each cluster. In CQA portals, each leaf category could be treated as a cluster and thus CBLM could be employed. Therefore, we applied CBLM to explore whether such "category-sensitive" setting was comparable with our category-sensitive LMs.

2. Experimental results in [4] showed that utilizing latent topics boosted the performance of QLLM for expertise estimation. Therefore, we intended to compare the effectiveness of latent topics and explicit categories as they both consider semantic in expertise estimation.

In addition, the original QLLM was included in comparisons as the baseline method. We used the tool GibbsLDA++[5] to estimate the posterior probabilities of LDA (say, θ of each answerer and ϕ of each topic). The default setting was adopted and the number of latent topics was set as 100 empirically.

4.3 Evaluation Metrics

We adopted *Precision at* K, *Mean Average Precision* and *Mean Reciprocal Rank* as evaluation metrics for the ranking lists generated by various LMs in expertise estimation. Furthermore, we employed the mean QR time (MQRT) which calculates the average time spent on routing one question (including expertise estimation and ranking) as the metric of time efficiency for all methods.

We set $\beta = 3.5$ for TCS-LM in the experiments empirically as this setting yields the best performance. As it is time-consuming to test all questions in Set A (Test set), we sampled 440 questions randomly from Set A (10 questions from each leaf category) as testing data. All algorithms ran in a PC with two 2.4GHz CPUs and 3G main memory.

5. EXPERIMENTAL RESULTS

5.1 Category-Sensitive LMs

Table 2 reports $Prec@K$ for all algorithms with different Ks 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]http://gibbslda.sourceforge.net/

5.1.1 Higher Accuracies

From Table 2 we observe that, for various of Ks, both BCS-LM and TCS-LM outperform QLLM significantly on $Prec@K$. 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 Ks, 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.

To sum up, the above results have assured the effectiveness of utilizing category information in expertise estimation.

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 time-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 $Prec@1$ of TCS-LM is 10.14% higher than those of BCS-LM. In addition, the $Prec@10$ 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 one 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: Different methods' $Prec@K$ in QR versus various Ks (best results in bold)

K	QLLM	BCS-LM	TCS-LM	LDALM	CBLM
1	0.0795	0.1114 (↑40.13%)	**0.1227** (↑54.34%)	0.0989 (↑24.40%)	0.0000
3	0.1659	**0.2364** (↑42.50%)	0.2340 (↑41.05%)	0.1950 (↑17.54%)	0.0000
5	0.2091	**0.2727** (↑30.42%)	0.2705 (↑29.36%)	0.2455 (↑17.41%)	0.0000
10	0.2705	0.3386 (↑25.18%)	**0.3455** (↑27.73%)	0.3102 (↑14.68%)	0.0000
20	0.3386	0.3909 (↑15.45%)	**0.3932** (↑16.13%)	0.3710 (↑9.57%)	0.0091
40	0.4136	0.4523 (↑9.36%)	**0.4591** (↑11.00%)	0.4392 (↑6.19%)	0.0273
60	0.4477	**0.4818** (↑7.62%)	0.4795 (↑7.10%)	0.4649 (↑3.84%)	0.0545
80	0.4727	**0.4955** (↑4.82%)	0.4909 (↑3.85%)	0.4867 (↑2.96%)	0.0727
100	0.4909	**0.5159** (↑5.09%)	0.5114 (↑4.18%)	0.4979 (↑1.43%)	0.0795

Table 3: MRR and MAP of various models (best results in bold)

Method	MRR	MAP
QLLM	0.1460	0.1070
BCS-QLLM	0.1893 (↑29.66%)	0.1424 (↑33.08%)
TCS-QLLM	**0.1965** (↑34.59%)	**0.1469** (↑37.29%)
LDALM	0.1695 (↑16.10%)	0.1281 (↑19.72%)
CBLM	0.0031	0.0024

Table 4: Various methods' MQRT in QR (in seconds)

QLLM	BCS-QLLM	TCS-QLLM	LDALM	CBLM
10.4271	5.5098	8.9884	16.7689	4.2488

and assign weights to these profiles according to the degree of similarities. Therefore, they give more precise expertise estimation and thus improve QR's performance.

5.1.4 Category Sensitive LMs vs. CBLM vs. LDALM

Across these four methods, CBLM performs the worst. The probable reason is that a great amount of answerers only answered in one cluster (leaf category), as such their contributions to this cluster are 1. Under this circumstance, these answerers' expertise is actually measured by those clusters' "expertise", which will cause many answerers to own the same expertise and thus make the ranking meaningless. LDALM increases $Prec@K$ of QLLM, which shows the impact of utilizing latent topics, but explicit question category provides more help than latent topics as category-sensitive LMs outperform LDALM at various Ks. MRR and MAP of these four methods report the similar results and detail will not be provided here.

When turning to MQRT, we find that CBLM works the best, followed by BCS-LM and TCS-LM, while LDALM costs much more time in inference. CBLM estimates answerer expertise through combining answerer's contribution to each cluster (which is pre-computed) and the probability of generating the routed question from each cluster (which is efficient to calculate), thus it makes the fastest estimation. However, the estimation made by CBLM is most inaccurate, as stated above. On the whole, category-sensitive LMs are time-efficient among the four methods.

In summary, category-sensitive LMs give more accurate expertise estimation than CBLM and LDALM and at the same time keep high time-efficiency.

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

[1] X. Cao, G. Cong, B. Cui, C. S. Jensen, and C. Zhang. The use of categorization information in language models for question retrieval. In *Proc. of CIKM*, 2009.

[2] J. Guo, S. Xu, S. Bao, and Y. Yu. Tapping on the potential of Q&A community by recommending answer providers. In *Proc. of CIKM*, 2008.

[3] B. Li and I. King. Routing questions to appropriate answerers in community question answering services. In *Proc. of CIKM*, 2010.

[4] M. Liu, Y. Liu, and Q. Yang. Predicting best answerers for new questions in community question answering. In *Proc. of WAIM*, 2010.

[5] X. Liu, W. B. Croft, and M. Koll. Finding experts in community-based question-answering services. In *Proc. of CIKM*, 2005.

[6] M. Qu, G. Qiu, X. He, C. Zhang, H. Wu, J. Bu, and C. Chen. Probabilistic question recommendation for question answering communities. In *Proc. of WWW*, 2009.

[7] Y. Zhou, G. Cong, B. Cui, C. S. Jensen, and J. Yao. Routing questions to the right users in online communities. In *Proc. of ICDE*, 2009.