Translation Language Model Enhancement for Community Question Retrieval Using User Adoption Answer

Ming Chen[1], Lin Li[1(✉)], and Qing Xie[1]

School of Computer Science and Technology, Wuhan University of Technology, Wuhan, China
{erlangera,cathylilin,felixxq}@whut.edu.cn

Abstract. Community Question Answering (CQA) services on Web provide an important alternative for knowledge acquisition. As an essential component of CQA services, question retrieval can help users save much time by finding relevant questions. However, there is a "gap" between queried questions and candidate questions, which is called lexical chasm or word mismatch problem. In this paper, we improve traditional Topic inference based Translation Language Model (T^2LM) by using the topic information of queries. Moreover, we make use of user information, specifically the number of user adoption answers, for further enhancing our proposed model. In our model, the translation model and the topic model "bridge" the word gap by linking different words. Besides, user information that has no direct relation with semantics is used to help us "bypass" the gap. By combining both of them we obtain a considerable improvement for the performance of question retrieval. Experimental results on a real Chinese CQA data set show that our proposed model improves the retrieval performance over T^2LM baseline by 7.5% in terms of Mean Average Precision (MAP).

Keywords: Question retrieval · Translation model · Topic model · User information

1 Introduction

Community Question Answering (CQA) services have become a popular alternative for online information access, such as Yahoo! Answers[1], Zhihu[2] and Sogouwenwen[3]. Meanwhile, a huge number of User Generated Content (UGC) has been accumulated in the form of Question and Answer (QA) pairs. Their large amount of access historical data makes it possible that users can find the answers of their questions from answered questions. As we know, question retrieval in CQA services returns several relevant questions with possible answers directly.

[1] http://answers.yahoo.com
[2] http://www.zhihu.com
[3] http://wenwen.sogou.com

By that way, users do not need to wait for answers from human, which helps users save a lot of time. Therefore, the retrieval of relevant issues and their corresponding answers become an important task for CQA services. Here we define question retrieval as a task where new questions are used as queries to find relevant questions that have already been provided answers. For simplicity and consistency, we use the term "query" to denote a new question raised by a user and "question" to denote the answered question in the CQA archive [16].

Meanwhile, question retrieval can also be considered as a solution of traditional Question Answering problem, by transforming the focus of the Question Answering task from answer extraction, answer matching and answer ranking to searching relevant questions with good ready answers [14]. One of the major challenges for question retrieval is the lexical gap, i.e., the word mismatch between queried questions and candidate questions. For example, "Where can I listen to rock for free online?" and "I need a music sharing website." probably have the same meaning but in different word forms. In addition, the limited length of questions causes the sparsity of word features [17]. Therefore, retrieval models based on word frequency and document frequency statistics are no longer suitable for question retrieval tasks. Jeon et al. [2] proposed language model based question retrieval model. In their model, questions are ranked by their similarities to a query which depends on the exact match of words. However, the model cannot solve the word mismatch problem between the query and their relevant QA pairs in an archive.

To conquer the gap, on the one side, researchers are constantly trying to develop more enhanced models that can bridge the chasm by linking different words. Since the relationship between different words can be modeled through word-to-word translation probabilities, translation-based approaches have obtained some good results [3–7]. To control the noises in translation model, some researchers introduced potential topic information in translation-based model, namely, the topic inference based translation model [11] that is the state-of-the-art translation-based language model. On the other side, researchers are incessantly looking for new ways that can help people bypass the gap. Omari et al. [18] proposed an approach that ranks answers based on their novelty. Other work mainly resorts to query expansion. For example, category Information [19], Key Concept [16, 22], Dependency Relation [20] and Topic Information [21] were used to expand queries.

In this paper, we focus on the improvement of translation-based language model by metadata information. We take two major actions to solve the word mismatch problem in question retrieval. Firstly, we improve topic inference based translation model by introducing the topic information of queries. Our improved approach controls the translation noises by leveraging the topic information and balances the impact of each topic by using the topic information of query as weights. Secondly, we add user adoption answer number information into the improved model. Questions from users who have more adoption answers will be ranked higher than others. We do that not only for inspiring users to answer questions, but also because of our assumption that questions from users who have better answer behavior may have higher quality. Specifically detailed analysis will be introduced in section 3.3. By combining both, we further improve the performance of question retrieval in CQA.

The rest of this paper is organized as follows: Section 2 introduces the related work about question retrieval. Section 3 describes our improved retrieval model and the strategy of combining the model with user adoption answer number. Experiments and result analysis are reported in Section 4. Finally, conclusions and future work are discussed in Section 5.

2 Related Work

Ponte and Croft [1] proposed language model based information retrieval model firstly. Then it is widely used in all relevant areas of information retrieval. Jeon et al. [2] took the lead in applying the language model to question retrieval. They exploited Unigram language model to model QA pairs in CQA services and it was applied to find similar questions. However, the above approaches cannot bridge the lexical chasm well. In other words, these methods cannot find relevant questions that have different words from queries.

2.1 Translation-based Model

To compensate for the lack of traditional retrieval methods, researchers introduced statistical machine translation model into information retrieval model. They used word-to-word translation probabilities to model the relationship between different words.

Berger et al. [3] introduced statistical techniques to bridging the lexical gap in FAQ retrieval. They studied similarity calculation method in question retrieval from the lexical level towards the semantic level firstly. Riezler et al. [4] availed of monolingual translation based retrieval model for answer retrieval. They utilized sentence level paraphrasing approach to capture similarities between questions and answers. Xue et al. [5] presented a question retrieval model that combined a translation-based language model for the question part with a query likelihood method for the answer part. Since they used QA pairs as training data of translation model, the improvements are limited. Bernhanrd and Gurevych [6] relied high quality parallel corpora that has QA pairs collected from the WikiAnswer website, the definitions and glosses of the same term in different lexical semantic resources, and then used this corpora to train a translation model. Finally they obtained a better result. Zhou et al. [7] availed of phrase level translation model to capture similarities between query and question and get a better result than word level.

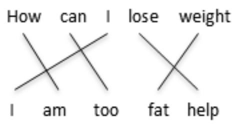

Fig. 1. An aligned example of IBM Model 1

However, the current translation models in question retrieval only rely on statistics co-occurrence information in parallel corpus to capture word-to-word translation probabilities, which makes the word-to-word translation information have lots of noises. Especially in the present circumstances, monolingual parallel corpus with high quality is few and researchers usually use QA pairs or question title and question description pairs as training data. Figure 1 show an aligned example of IBM Model 1. From Figure 1 we can see that there is no correlation between word pairs except "weight" and "fat". That makes the word-to-word translation probabilities produced by translation model are biased and affects the performance of retrieval.

2.2 Topic-based Model

Topic modelling based approaches, such as PLSA [12] and LDA [13], provide an elegant mathematical tool to analyze shallow semantics. Naturally, these techniques have attracted question retrieval researchers attention for a long time.

Wei and Croft [8] proposed an LDA-based document model within the language modeling framework, and evaluated it on several TREC collections. Cai et al. [9] combined the semantic similarity based latent topics with the translation-based language model to improve the retrieval performance. Ji et al. [10] presented Question-Answer Topic Model to model the question-answer relationships, with the assumption that a question and its paired answer share the same topic distribution. However, questions and answers are different in many aspects. They do not share the same topic distribution in many cases. Caused by this, the disadvantage of their model is obvious. Zhang et al. [11] proposed a model that controls translation noise by leveraging the topic information. They focused on similarity of topic distribution between word in query and question. They utilized word distribution information under topic to improve accuracy of word-to-question similarity and further obtained better performances. But their model did not consider the topic information of query that is also valuable for question retrieval.

In this paper, we utilize the topic distribution information of queries to improve the performance of retrieval on the basis of Zhang's model [11]. We add the topic information of queries as weights into the process of word-to-question similarity to balance the impact of each topic.

2.3 Metadata Enhanced Model

Taking into account the social properties of CQA services, QA pairs are accompanied by metadata in real CQA archives usually. The metadata include information of categories, information of askers, information of responders, and information of other users' feedback and so on. It is natural to think that this kind of metadata information is valuable to improve the performance of question retrieval. The category information attracts the attention of researchers immediately because of their relationship with semantics. Cai et al. [9] presented topic model involved category information to discover the latent topics in the content of questions. Cao et al. [14] proposed the category smoothing based and question classification based methods to enhance the performances of existing models. Certainly the methods depend on metadata information that is not always feasible for researchers.

In this paper, we focus on exploiting user information, specifically askers' adoption answer number, to improve the performance of question retrieval. Because questions from user who have better answer behavior could have higher value and they are more acceptable to user.

From the above work in question retrieval, using translation model to capture the similarity of words and using the metadata to improve the performance are two state-of-the-art techniques for question retrieval. In this paper, we take the advantages of the previous approaches and further integrate topic information and user information into a unified probabilistic ranking model to tackle the lexical gap in question retrieval.

3 User Adoption Answer Number for Translation Language Enhancement Model

In this section, we give a brief introduction about the Topic inference based Translation Language Model (T^2LM) [11]. Then we present two our main contributions. Firstly, by using the topic information of queries as weights to balance the impact of each topic, we improve the T^2LM. Secondly, by introducing the adoption answer information of a question asker into the improved model, we further optimize the ranking list.

3.1 Topic Inference Based Translation Language Model

In the T^2LM, given a query $query$ and a QA pair (q, a) consisted of a question q and an answer a, a ranking score $P(query|(q, a))$ is computed as follows:

$$P(query|(q,a)) = \prod_{w \in query} \left(\frac{|(q,a)|}{|(q,a)| + \lambda} P_{t^2lm}(w|(q,a)) + \frac{\lambda}{|(q,a)| + \lambda} P_{ml}(w|C) \right) \tag{1}$$

$$\begin{aligned} P_{t^2lm}(w|(q,a)) = \mu_1 P_{ml}(w|q) + \mu_2 \left(\sum_{t \in q} p(w|t) P_{ml}(t|q) \right) \\ + \mu_3 \left(\sum_{t \in q} P_{ml}(t|q) \sum_{i=1}^{K} p(w|z_i) p(t|z_i) \right) + \mu_4 P_{ml}(w|a) \end{aligned} \tag{2}$$

$$P_{ml}(w|q) = \frac{tf_{w,q}}{|q|}, P_{ml}(w|a) = \frac{tf_{w,a}}{|a|}, P_{ml}(w|C) = \frac{tf_{w,C}}{|C|} \tag{3}$$

The explanations of terms in Equation 1 to 3 are showed in Table 1. w is a word in query $query$, and t is a word in question q. $|q|$, $|a|$, $|C|$ have similar meanings to $|(q, a)|$; $tf_{w,a}$ and $tf_{w,C}$ have similar meanings to $tf_{w,q}$; $P_{ml}(w|a)$

Table 1. Explanation of Terms

Term	Explanation
C	the background collection
λ	the smoothing parameter
$\|(q,a)\|$	the word lengths of (q,a)
$tf_{w,q}$	the frequency of term w in q
$P_{ml}(w\|q)$	the maximum likelihood estimate of word w in q
$p(w\|t)$	the probability that t is the translation of word w
$p(w\|z_i)$	the distribution probabilities of word w under topic z_i
K	the topic number

and $P_{ml}(w|C)$ have similar meanings to $P_{ml}(w|q)$; $p(t|z_i)$ has a similar meaning to $p(w|z_i)$; and μ_1, μ_2, μ_3 and μ_4 balance the impact of each component and $\mu_1 + \mu_2 + \mu_3 + \mu_4 = 1$.

3.2 Incorporating the Topic Information of Query

From the Equation 2 we can see that the significance of each topic is equal in T^{2}LM. And the word topic distribution probabilities from static corpus are used statically for a dynamic query. Although the diverse topic information of various queries is beneficial for question retrieval generally, the information is ignored. In this paper, we propose an approach to exploit this topic information. In our approach, the topic information of queries is used as weights of each topic to improve the process of capture word-to-question similarity. More formally, given a query *query* and a topic z_i, $P(query|z_i)$ denoting the weight of topic z_i for query *query* is computed as follows:

$$P(query|z_i) = \frac{\prod_{w \in query} p(w|z_i)}{\sum_{j=1}^{K} \prod_{w \in query} p(w|z_j)} \tag{4}$$

Here w is a word in query *query*; $p(w|z_i)$ and $p(w|z_j)$ are the distribution probability of word w under topic z_i and z_j; and K is topic number. The denominator in Equation 4 may be zero in some cases. To solve the problem, we make a compromise that we set $P(query|z_i) = 1/K$ for each topic z_i if the problem happens. Then the $P(q|z_i)$ is used as weights of each topic in the process of capturing word-to-question similarity. The specific method is showed as follows:

$$\begin{aligned} P_{t^2lm^*}(w|(q,a)) = {} & \mu_1 P_{ml}(w|q) + \mu_2 \left(\sum_{t \in q} P(w|t) P_{ml}(t|q) \right) \\ & + \mu_3 \left(\sum_{t \in q} P_{ml}(t|q) \cdot K \cdot \sum_{i=1}^{K} P(query|z_i) \cdot p(w|z_i) \cdot p(t|z_i) \right) \\ & + \mu_4 P_{ml}(w|a) \end{aligned} \tag{5}$$

Here w is a word in query *query*. By multiplying K we control the range of topic part unchanged so that scope of μ_3 is the same as before. We denote the improved model as T^2LM^* in this paper.

3.3 Incorporating the User Information

Thinking about PageRank, a technique for Web search, we see that the importance of web pages relied on web structure building by links have nothing to do with the web content. This technique optimizes the ranking lists of search results significantly. Inspiring by that, we consider a way to measure the importance of questions. In CQA archive, there are no links between QA pairs. What come into our mind is that users' feedback information and user answer number information. For feedback information, there is a problem that users' feedback object are answers usually, users can thumb up or down an answer. It has no direct relation with questions. However, user answer number information can be contacted with questions by question askers. In all kinds of user information, user adoption answer number is most representative for user answer behavior. Therefore, we develop a technique to measure the importance of question by its asker adoption answer number.

Our experiment data set contains metadata of user information including the answer number of each user and their adoption rate. We can get adoption answer number of each question's asker by letting answer number multiply adoption rate. We suspect that the data are related to the retrieval results. So we did some experiments about our conjecture. We selected 100 questions from the data set as queries, and utilized Lucene, TLM and T^2LM to get top 10 candidate questions. Then each question in the result was labelled with "relevant" or "irrelevant" by persons. Table 2 and Figure 2 show the statistical results of a survey on retrieval result.

Table 2. Distributions of question askers' adoption answer number (Sum denotes the summary of the three models; Background indicates the results of all 578926 questions in data set).

number	Lucene		TLM		T^2LM		Sum		Background
	relevant	irrelevant	relevant	irrelevant	relevant	irrelevant	relevant	irrelevant	
0	0.6307	0.6994	0.6410	0.6311	0.6642	0.6192	0.6453	0.6499	0.6614
1-5	0.1477	0.1600	0.1667	0.2017	0.1418	0.2002	0.1521	0.1873	0.1722
6-10	0.0682	0.0303	0.0513	0.0415	0.0448	0.0359	0.0547	0.0359	0.0412
11-15	0.0284	0.0218	0.0192	0.0237	0.0224	0.0336	0.0233	0.0264	0.0228
16-20	0.0284	0.0145	0.0256	0.0261	0.0299	0.0208	0.0280	0.0205	0.0183
21-30	0.0114	0.0218	0.0128	0.0178	0.0224	0.0220	0.0155	0.0205	0.0191
31-40	0.0227	0.0073	0.0192	0.0095	0.0149	0.0081	0.0190	0.0083	0.0117
41-50	0.0057	0.0048	0.0064	0.0071	0.0000	0.0069	0.0040	0.0063	0.0072
51-100	0.0114	0.0145	0.0256	0.0166	0.0224	0.0255	0.0198	0.0189	0.0174
101-200	0.0114	0.0109	0.0128	0.0083	0.0075	0.0081	0.0105	0.0091	0.0101
201-	0.0341	0.0145	0.0192	0.0166	0.0299	0.0197	0.0277	0.0169	0.0187

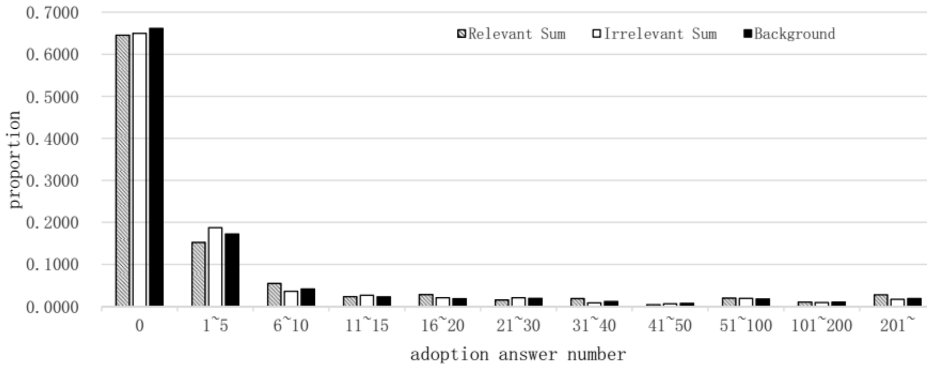

Fig. 2. Distribution of question askers' adoption answer number

Table 3. Average adoption answer number of question asker (Average* indicates that the averages are captured after removing outliers; Sum denotes the summary of the three models; Background denotes the result came from entire data set)

	Lucene		TLM		T^2LM		Sum		Background
	relevant	irrelevant	relevant	irrelevant	relevant	irrelevant	relevant	irrelevant	
Average	17.381	12.644	14.160	58.676	31.657	53.317	21.066	41.546	39.025
Average*	10.382	3.926	6.183	4.738	7.786	6.162	8.117	4.942	5.035

From Table 2 and Figure 2, we can find that most askers have no adoption answer. And relevant question and irrelevant question have similar interval distribution on asker's accepted answer number. To find the difference between them, we calculate their average. Unexpectedly, we get that the average number of irrelevant question is bigger than relevant. That is contrary to our conjecture. However, by studying the retrieval result carefully we find the reason is that there are some outliers in our data. The outliers often have more than 20,000 adoption answers which make a huge impact on the results. To eliminate the effects of outliers, we decided to clear it. In our plan, the top 2% users in adoption answer number are removed. After removing outliers, we get Average* which show us askers of relevant question have more accepted answer than irrelevant, on average. The result is showed in Table 3.

By combining Table 2 and Table 3, it is appropriate to consider if question from asker have more adoption answers, it may have higher correlation probability. The reason may be that those users who have answered more questions correctly will give higher quality questions. And those questions have higher likelihood of being accepted. So the data of question asker's adoption answer may be beneficial to question retrieval. Based on this we further enhance our proposed T^2LM* by adding the information. More formally, given a QA pair (q, a), the $S_{ui}((q, a))$ that denotes user information part score in our model is computed as follows:

$$S_{ui}((q,a)) = \begin{cases} s(u_j) & (s(u_j) \le 20) \\ 20 & (s(u_j) > 20) \end{cases}, \ where\ s(u_j) = \sqrt{A_{u_j} \cdot R_{u_j}} \tag{6}$$

Here user u_j is asker of q; A_{u_j} and R_{u_j} are answer number and the adoption rate of u_j. In order to eliminate the impact of abnormal data and enhance the applicability of the model, instead of clearing the outliers directly, we set a limit for the score. The score will not be more than 20, for the reason that the maximum of adoption answer number near 400 after removing top 2%. After normalizing, we get $P_{ui}((q,a))$. Then we introduce it into our model. More formally, given a query $query$ and a QA pair (q,a), the ranking score $P(query|(q,a))$ is computed as follows:

$$P(query|(q,a)) = (1-\theta)P_{ui}((q,a)) + \\ \theta \prod_{w \in query} \left(\frac{|(q,a)|}{|(q,a)|+\lambda} P_{t^2lm^*}(w|(q,a)) + \frac{\lambda}{|(q,a)|+\lambda} P_{ml}(w|C) \right) \tag{7}$$

Here θ controls the impact of user information component. We call our final improved model as User information for Topic inference based Translation Language Model (UT^{2}LM).

4 Experiments

In this section, experiments are conducted on a real CQA archive to demonstrate the effectiveness of our proposed two question retrieval models.

4.1 Experimental Setup

The data set in our experiment comes from NDBC CUP2016, specifically comes from Sogouwenwen that is one of most popular CQA services in China. It is a Chinese data set with two parts. One is QA pairs and the other is user information. For QA pairs, each QA pair consists of three fields: "question" and "answer", as well as metadata "asker ID". For user information, each of user information consists of three fields: "ID", "answer number" and "adoption rate". The data set contains 1,729,263 QA pairs, which have 578608 single questions, and 2,047,669 user information.

For data preprocessing, we segment words and remove the stop words for each QA pair in the data set firstly. Then we select 300 QA pairs from the data set randomly. After removing QA pairs that question lengths are too short and questions are same, we get 253 questions as queries. Among them, 210 questions which are selected by randomly are used as test set to test and other 43 questions are used as development set to adjust the parameters.

We use the GIZA++ toolkit [15] for learning the IBM Translation Model 1 to get the word-to-word translation probabilities. We pool the QA pairs and the answer-question pairs together as the input to this toolkit [5]. We get a word-to-word translation probability list after training. For topic model part, we think of questions as documents and utilize LDA [13] model to model question set. All 578608 single questions are used as data of Gibbs sampling to get the word topic distribution probabilities.

For a practical Question Retrieval System, search results should be presented to the user in a hierarchical way where a list of questions is first presented, and after the user selects a specific question the corresponding answers are presented [5]. Question list is more concise and efficient than QA pair list. Thus, in our experiments, relevance judgments are based on questions. It is easy to transform a QA pair rank into a question rank by taking the highest rank among a group of QA pairs for the same question. In this paper, all the experiments are conducted by getting QA pair ranks firstly and then transforming it into question ranks.

We use three question retrieval models, the Language Model (LM) [2], the Translation-based Language Model (TLM) [5] and the Topic Inference-based Translation language Model (T^{2}LM) [11], as baseline methods. We conduct experiments to demonstrate the effect of our proposed two models in Section 3, T^{2}LM* and UT^{2}LM. For each method, the top 10 retrieval results are kept.

As a common evaluation process used by other researchers in QA field, we recruit three students who their mother tongue is Chinese to label the relevance of the candidate questions regarding to queries. Given a query and its candidate questions list, a student is asked to label the questions in list with "relevant" or "irrelevant". If a candidate question is considered semantically similar to the query, the student will label it as "relevant"; otherwise, the student will label it as "irrelevant". As a result, each candidate question gets three labels and the majority of the labels are taken as the final decision for a query candidate pair.

In order to evaluate the performance of different models, we employ Mean Average Precision (MAP), Mean Reciprocal Rank (MRR) and Precision at 1 (P@1) as evaluation measures [22]. These measures are widely used in the literature for question retrieval in CQA.

By experiments on development set, we adjust the parameters of the baseline systems and our systems to the best. Some results are showed in Figure 3, Figure 4 and Figure 5. From Figure 3 we can see that MAP reduces with μ_2 growing overall. Figure 5 shows that with the growth of θ MAP grows first and then decreases and maximizes at θ=0.66. Finally for T^{2}LM* we set $\mu_1 = 0.5,\ \mu_2 = 0.1,\ \mu_3 = 0.3,\ \mu_4 = 0.1,\ K = 70$, and for UT^{2}LM we set $\mu_1 = 0.5,\ \mu_2 = 0.1,\ \mu_3 = 0.3,\ \mu_4 = 0.1,\ \theta = 0.66,\ K = 70$. We set smoothing parameter $\lambda = 1$ according to the previous work [5].

4.2 Experimental Results

Table 4 shows the compare results of our models and baseline models.We can see that among the three baseline models, T^{2}LM performs the best and LM performs the worst; TLM between LM and T^{2}LM. TLM and T^{2}LM has better performance since they are able to retrieval question that do not share common words with the query, but are semantically similar to the query. T^{2}LM introduces the topic information to control the translation noise in translation model, it has the best performance. Comparing results of LM and TLM, we find the improvement is very small. We speculate that this is related to our training data of translation model. We use QA pair to train IBM Model 1. That makes word-to-word translation probabilities inaccurate. Showing in the T^{2}LM*, the

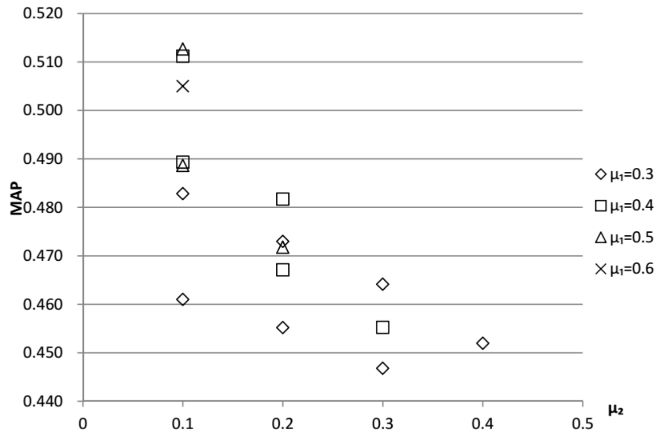

Fig. 3. The influence of μ_1, μ_2 on the performance of T^{2}LM*

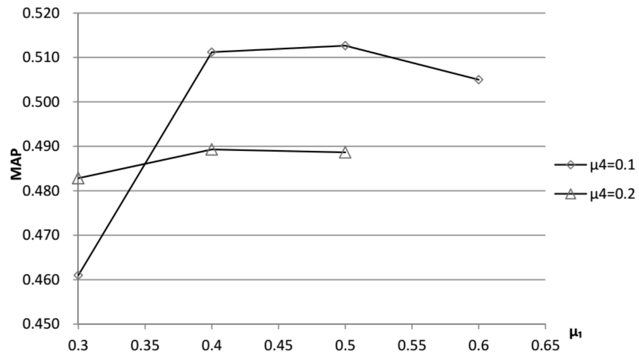

Fig. 4. The influence of μ_1 on the performance of T^{2}LM* when $\mu_2 = 0.1$

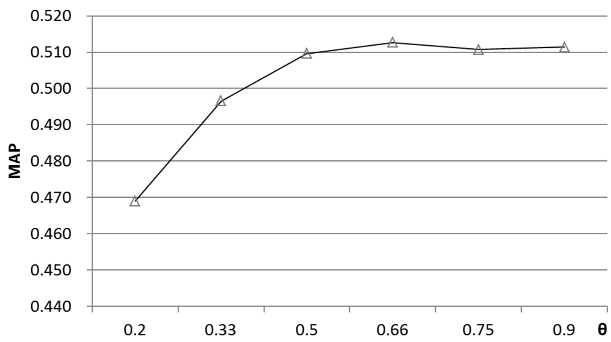

Fig. 5. The influence of θ on the performance of UT^{2}LM

μ_3 is very small. That means translation model part has very small weight. The results are consistent with the reported in previous work [11].

Table 4. Question retrieval results (Bold indicates our method and the corresponding results. T^{2}LM* and UT^{2}LM have a statistically significant improvement over the baseline using the t-test, p-value < 0.05)

	MAP	% of MAP improvements over				P@1	MRR
		LM	TLM	T^{2}LM	T^{2}LM*		
LM	0.3583	N/A	N/A	N/A	N/A	0.2516	0.1958
TLM	0.3746	+4.5	N/A	N/A	N/A	0.2846	0.2372
T^{2}LM	0.4361	+21.7	+16.4	N/A	N/A	0.3208	0.2641
T^{2}LM*	**0.4592**	**+28.1**	**+21.3**	**+5.2**	**N/A**	**0.3385**	**0.3195**
U^{2}TLM	**0.4687**	**+30.8**	**+25.1**	**+7.5**	**+2.1**	**0.3401**	**0.3244**

The T^{2}LM* performs better than the T^{2}LM, and the UT^{2}LM performs better than the T^{2}LM*. The underlying reasons are that the T^{2}LM* utilizes the topic information of query as weight to improve the topic component on the basis of T^{2}LM. UT^{2}LM introduce the user information into the T^{2}LM*. Because of the reason that we describe in section 3.3, the performance of question retrieval is improved further. Comparing with T^{2}LM*, the improvement of UT^{2}LM is small too. The reason is that main function of user information is optimization the ranking list, not affects result strongly. By observing data, we find the difference between Rank 1 and Rank 5 usually is between 1 to 10 in T^{2}LM*, here we take the logarithmic technology in the calculation process to prevent underflow. And the optimal value of θ is relatively small 0.5 that illustrates this problem from the side.

The topic number K in the LDA model also has an impact on retrieval result, so we conduct an experiment on development set to find the relationship between them. Figure 6 shows the result of the experiment. We can see that

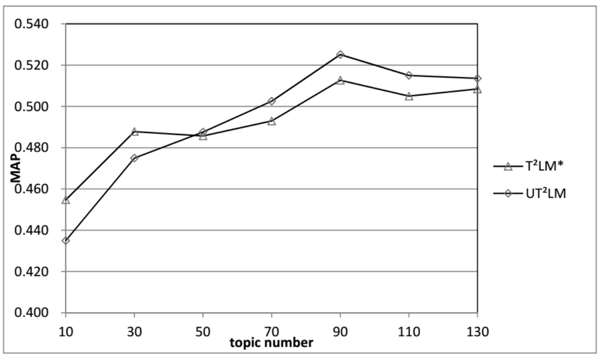

Fig. 6. The influence of topic number on the performance of T^{2}LM* and UT^{2}LM

MAP increases as the growth of K when K does not exceed 70. However, when exceeding 70, MAP no longer grows or even lower. Possible reasons are that distinction between topics reduces with topic number growing when topic number is too large. That causes the overlap between topics. After studying the result of LDA model, we find that some words repeat in different topic top words list when topic number exceeds 70. That confirms our previous conjecture.

Figure 7 shows a retrieval example of T^2LM, T^2LM* and UT^2LM. Here we briefly explain the content of results in English. Query is about children's English learning interest. Three Rank 1 of three models are same which is about helping children learn English. Rank 2 and 3 of T^2LM are about children's English learning institutions. Rank 2 of T^2LM* and Rank 3 of UT^2LM are same and both are about helping children learn English. Rank 3 of T^2LM* and Rank 2 of UT^2LM are same and they are about cultivating children's interest in learning English which are the most relevant result with the query. From Figure 7 we can see that T^2LM* performs better than T^2LM. The results all have a relationship with English, but the results of T^2LM* and UT^2LM focus the topic on the methods and interests in learning English. And we can see that UT^2LM improve ranking list by user information.

Query: 我想让孩子学英语，提高一下英语口语水平，但孩子不想学怎么办？

Rank	T^2LM	T^2LM*	UT^2LM
1	**想提高孩子的英语口语，该如何培养呢？**	**想提高孩子的英语口语，该如何培养呢？**	**想提高孩子的英语口语，该如何培养呢？**
2	暑假了 让孩子去那学点英语呢	**想帮助孩子学好英语口语，该怎么做呢？**	**我是想提高小孩对英语的兴趣，掌握学习英语的方法。**
3	昆明英语口语培训学校提高英语能力 孩子想上个英语补习班去那儿很有名？	**我是想提高小孩对英语的兴趣，掌握学习英语的方法。**	**想帮助孩子学好英语口语，该怎么做呢？**

Fig. 7. An example of T^2LM, T^2LM*, UT^2LM (Bold indicates relevant; Underline indicates the most relevant)

5 Conclusion and Future Work

Question retrieval is an important component in Community Question Answering (CQA) services. In this paper, we propose a novel approach by first using topic information of query to improve the quality of the word-to-question similarity estimates. In addition, we exploit user adoption answer number associated with questions in CQA archives for further improving the performance of question retrieval. Experiments on a real CQA data set from Sogouwenwen demonstrate the effectiveness of the proposed retrieval model.

Although the topic model can be used as a potential semantic extension to enhance the performance of question retrieval in CQA services, there is a problem of ambiguity between topics. It is interesting to find an approach to solve the problem. Meanwhile, thinking of keyword extraction techniques, it could be very meaningful to capture weight of each topic by keywords of query instead of all words in query. Besides, it is promising to explore other metadata in CQA archive for improving retrieval performance.

Acknowledgement. This research project is supported by the National Natural Science Foundation of China (Grant No:61602353,61303029), National Social Science Foundation of China (Grant No:15BGL048), Hubei Province Science and Technology Support Project(2015BAA072), 863 Program (2015AA015403).

References

1. Ponte, J.M., Croft, W.B.: A language modeling approach to information retrieval. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 275-281 (1998)
2. Jeon, J., Croft, W.B., Lee, J.H.: Finding similar questions in large question and answer archives. In: CIKM, pp. 84-90 (2005)
3. Berger A.L., Caruana R., Cohn D., Freitag D., Mittal V.O.: Bridging the lexical chasm: statistical approaches to answer-finding. In: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 192-199 (2000)
4. Riezler S., Vasserman A., Tsochantaridis I., Mittal V.O., Liu Y.: Statistical machine translation for query expansion in answer retrieval. In: ACL, pp. 464-471 (2007)
5. Xue X., Jeon J., Croft W.B.: Retrieval models for question and answer archives. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 475-482 (2008)
6. Bernhard D., Gurevych I.: Combining lexical semantic resources with question & answer archives for translation-based answer finding. In: ACL, pp. 728-736 (2009)
7. Zhou, G., Cai, L., Zhao, J., Liu, K.: Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives. In: ACL, pp.653-662 (2011)
8. Xing W., Croft W.B.: LDA-based document models for ad-hoc retrieval. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 178-185 (2006)
9. Cai, L., Zhou, G., Liu, K., Zhao, J.: Learning the Latent Topics for Question Retrieval in Community QA. In: IJCNLP, pp. 273-281 (2011)
10. Ji, Z., Xu, F., Wang, B., He, B.: Question-answer topic model for question retrieval in community question answering. In: CIKM, pp. 2471-2474 (2012)
11. Zhang, W.N., Yu, Z., Liu, T.: A topic inference based translation model for question retrieval in community-based question answering services. Chinese Journal of Computers, vol. 38, no. 2, pp. 313-321 (2015)
12. Hofmann T.: Unsupervised learning by probabilistic latent semantic analysis. In: Machine Learning, vol. 45, pp. 177-196 (2001)
13. Blei, M.D., Ng, Y.A., Jordan, I.M.: Latent Dirichlet Allocation. In: Journal of Machine Learning Research, 3, pp. 993-1022 (2003)
14. Cao X., Cong G., Cui B., Jensen C.S., Yuan Q.: Approaches to exploring category information for question retrieval in community question-answer archives. In: Acm Transactions on Information Systems, vol. 30, no. 2, pp. 1-38 (2012)

15. Och, F.J., Ney, H.: Improved statistical alignment models. In: ACL, pp. 440-447 (2000)
16. Zhang, W.N., Ming, Z.Y., Zhang, Y., Liu, T.: Capturing the semantics of key phrases using multiple languages for question retrieval. In: IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 4, pp. 888-900 (2016)
17. Chen, L., Jose, J.M., Yu, H., Yuan, F., Zhang, D.: A semantic graph based topic model for question retrieval in community question answering. In: the Ninth ACM International Conference on Web Search and Data Mining, pp. 287-296 (2016)
18. Omari, A., Carmel, D., Rokhlenko, O., Szpektor, I.: Novelty based Ranking of Human Answers for Community Questions. In: International ACM SIGIR Conference on Research and Development in Information Retrieval pp.215-224 (2016)
19. Yuan, Q., Cong, G., Sun, A., Lin, C.Y., Thalmann, N.M.: Category hierarchy maintenance: a data-driven approach. In: the 35th international ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 791-800 (2012)
20. Zhang, W., Ming, Z., Zhang, Y., Nie, L., Liu, T., Chua, T.S.: The Use of Dependency Relation Graph to Enhance the Term Weighting in Question Retrieval. In: COLING, pp. 3105-3120 (2012).
21. Dijk, D.V., Tsagkias, M., Rijke, M.D.: Early Detection of Topical Expertise in Community Question Answering. In: The International ACM SIGIR Conference, pp.995-998 (2015).
22. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval. Cambridge university press, pp. 139-159 (2008)