

Improving Question Recommendation by Exploiting Information Need

Shuguang Li

Department of Computer Science
University of York, YO10 5DD, UK
sgli@cs.york.ac.uk

Suresh Manandhar

Department of Computer Science
University of York, YO10 5DD, UK
suresh@cs.york.ac.uk

Abstract

In this paper we address the problem of question recommendation from large archives of community question answering data by exploiting the users' information needs. Our experimental results indicate that questions based on the same or similar information need can provide excellent question recommendation. We show that translation model can be effectively utilized to predict the information need given only the user's query question. Experiments show that the proposed information need prediction approach can improve the performance of question recommendation.

1 Introduction

There has recently been a rapid growth in the number of community question answering (CQA) services such as Yahoo! Answers¹, Askville² and WikiAnswer³ where people answer questions posted by other users. These CQA services have built up very large archives of questions and their answers. They provide a valuable resource for question answering research. Table 1 is an example from Yahoo! Answers web site. In the CQA archives, the title part is the user's query question, and the user's information need is usually expressed as natural language statements mixed with questions expressing their interests in the question body part.

In order to avoid the lag time involved with waiting for a personal response and to enable high quali-

ty answers from the archives to be retrieved, we need to search CQA archives of previous questions that are closely associated with answers. If a question is found to be interesting to the user, then a previous answer can be provided with very little delay. Question search and question recommendation are proposed to facilitate finding highly relevant or potentially interesting questions. Given a user's question as the query, *question search* tries to return the most semantically similar questions from the question archives. As the complement of question search, we define *question recommendation* as recommending questions whose information need is the same or similar to the user's original question. For example, the question "What aspects of my computer do I need to upgrade ..." with the information need "... making a skate movie, my computer freezes, ..." and the question "What is the most cost effective way to expend memory space ..." with information need "... in need of more space for music and pictures ..." are both good recommendation questions for the user in Table 1. So the recommended questions are not necessarily identical or similar to the query question.

In this paper, we discuss methods for question recommendation based on using the similarity between information need in the archive. We also propose two models to predict the information need based on the query question even if there's no information need expressed in the body of the question. We show that with the proposed models it is possible to recommend questions that have the same or similar information need.

The remainder of the paper is structured as fol-

¹<http://answers.yahoo.com>

²<http://askville.amazon.com>

³<http://wiki.answers.com>

Q Title	If I want a faster computer should I buy more memory or s- torage space? ...
Q Body	I edit pictures and videos so I need them to work quickly. Any advice?
Answer	... If you are running out of s- pace on your hard drive, then ... to boost your computer speed usually requires more RAM ...

Table 1: Yahoo! Answers question example

lows. In section 2, we briefly describe the related work on question search and recommendation. Section 3 addresses in detail how we measure the similarity between short texts. Section 4 describes two models for information need prediction that we use for the experiment. Section 5 tests the performance of the proposed models for the task of question recommendation. Section 7 is the conclusion of this paper.

2 Related Work

2.1 Question Search

Burke et al. (1997) combined a lexical metric and a simple semantic knowledge-based (WordNet) similarity method to retrieve semantically similar questions from frequently asked question (FAQ) data. Jeon et al. (2005a) retrieved semantically similar questions from Korean CQA data by calculating the similarity between their answers. The assumption behind their research is that questions with very similar answers tend to be semantically similar. Jeon et al. (2005b) also discussed methods for grouping similar questions based on using the similarity between answers in the archive. These grouped question pairs were further used as training data to estimate probabilities for a translation-based question retrieval model. Wang et al. (2009) proposed a tree kernel framework to find similar questions in the CQA archive based on syntactic tree structures. Wang et al. (2010) mined lexical and syntactic features to detect question sentences in CQA data.

2.2 Question Recommendation

Wu et al. (2008) presented an incremental automatic question recommendation framework based on probabilistic latent semantic analysis. Question recommendation in their work considered both the users' interests and feedback. Duan et al. (2008) made use of a tree-cut model to represent questions as graphs of topic terms. Questions were recommended based on this topic graph. The recommended questions can provide different aspects around the topic of the query question.

The above question search and recommendation research provide different ways to retrieve questions from large archives of question answering data. However, none of them considers the similarity or diversity between questions by exploring their information needs.

3 Short Text Similarity Measures

In question retrieval systems accurate similarity measures between documents are crucial. Most traditional techniques for measuring the similarity between two documents mainly focus on comparing word co-occurrences. The methods employing this strategy for documents can usually achieve good results, because they may share more common words than short text snippets. However the state-of-the-art techniques usually fail to achieve desired results due to short questions and information need texts.

In order to measure the similarity between short texts, we make use of three kinds of text similarity measures: TFIDF based, Knowledge based and Latent Dirichlet Allocation (LDA) based similarity measures in this paper. We will compare their performance for the task of question recommendation in the experiment section.

3.1 TFIDF

Baeza-Yates and Ribeiro-Neto (1999) provides a TFIDF method to calculate the similarity between two texts. Each document is represented by a term vector using TFIDF score. The similarity between two text D_i and D_j is the cosine similarity in the vector space model:

$$\cos(D_i, D_j) = \frac{D_i^T D_j}{\|D_i\| \|D_j\|}$$

This method is used in most information retrieval systems as it is both efficient and effective. However if the query text contains only one or two words this method will be biased to shorter answer texts (Jeon et al., 2005a). We also found that in CQA data short contents in the question body cannot provide any information about the users’ information needs. Based on the above two reasons, in the test data sets we do not include the questions whose information need parts contain only a few noninformative words

3.2 Knowledge-based Measure

Mihalcea et al. (2006) proposed several knowledge-based methods for measuring the semantic level similarity of texts to solve the lexical chasm problem between short texts. These knowledge-based similarity measures were derived from word semantic similarity by making use of WordNet. The evaluation on a paraphrase recognition task showed that knowledge-based measures outperform the simpler lexical level approach.

We follow the definition in (Mihalcea et al., 2006) to derive a text-to-text similarity metric mcs for two given texts D_i and D_j :

$$mcs(D_i, D_j) = \frac{\sum_{w \in D_i} \max Sim(w, D_j) * idf(w)}{\sum_{w \in D_i} idf(w)} + \frac{\sum_{w \in D_j} \max Sim(w, D_i) * idf(w)}{\sum_{w \in D_j} idf(w)}$$

For each word w in D_i , $\max Sim(w, D_j)$ computes the maximum semantic similarity between w and any word in D_j . In this paper we choose *lin* (Lin, 1998) and *jcn* (Jiang and Conrath, 1997) to compute the word-to-word semantic similarity.

We only choose nouns and verbs for calculating mcs . Additionally, when w is a noun we restrict the words in document D_i (and D_j) to just nouns. Similarly, when w is a verb, we restrict the words in document D_i (and D_j) to just verbs.

3.3 Probabilistic Topic Model

Celikyilmaz et al. (2010) presented probabilistic topic model based methods to measure the similarity between question and candidate answers. The candidate answers were ranked based on the hidden

topics discovered by Latent Dirichlet Allocation (LDA) methods.

In contrast to the TFIDF method which measures “common words”, short texts are not compared to each other directly in probabilistic topic models. Instead, the texts are compared using some “third-party” topics that relate to them. A passage D in the retrieved documents (document collection) is represented as a mixture of fixed topics, with topic z getting weight $\theta_z^{(D)}$ in passage D and each topic is a distribution over a finite vocabulary of words, with word w having a probability $\phi_w^{(z)}$ in topic z . Gibbs Sampling can be used to estimate the corresponding expected posterior probabilities $P(z|D) = \hat{\theta}_z^{(D)}$ and $P(w|z) = \hat{\phi}_w^{(z)}$ (Griffiths and Steyvers, 2004).

In this paper we use two LDA based similarity measures in (Celikyilmaz et al., 2010) to measure the similarity between short information need texts. The first LDA similarity method uses KL divergence to measure the similarity between two documents under each given topic:

$$sim_{LDA1}(D_i, D_j) = \frac{1}{K} \sum_{k=1}^K 10^{W(D_i^{(z=k)}, D_j^{(z=k)})}$$

$$W(D_i^{(z=k)}, D_j^{(z=k)}) = -KL(D_i^{(z=k)} \parallel \frac{D_i^{(z=k)} + D_j^{(z=k)}}{2}) - KL(D_j^{(z=k)} \parallel \frac{D_i^{(z=k)} + D_j^{(z=k)}}{2})$$

$W(D_i^{(z=k)}, D_j^{(z=k)})$ calculates the similarity between two documents under topic $z = k$ using KL divergence measure. $D_i^{(z=k)}$ is the probability distribution of words in document D_i given a fixed topic z .

The second LDA similarity measure from (Griffiths and Steyvers, 2004) treats each document as a probability distribution of topics:

$$sim_{LDA2}(D_i, D_j) = 10^{W(\hat{\theta}^{(D_i)}, \hat{\theta}^{(D_j)})}$$

where $\hat{\theta}^{(D_i)}$ is document D_i ’s probability distribution of topics as defined earlier.

4 Information Need Prediction using Statistical Machine Translation Model

There are two reasons that we need to predict information need. It is often the case that the query question does not have a question body part. So we need a model to predict the information need part based on the query question in order to recommend questions based on the similarity of their information needs. Another reason is that information need prediction plays a crucial part not only in Question Answering but also in information retrieval (Liu et al., 2008). In this paper we propose an information need prediction method based on a statistical machine translation model.

4.1 Statistical Machine Translation Model

$(\mathbf{f}^{(s)}, \mathbf{e}^{(s)})$, $s = 1, \dots, S$ is a parallel corpus. In a sentence pair (\mathbf{f}, \mathbf{e}) , source language String, $\mathbf{f} = f_1 f_2 \dots f_J$ has J words, and $\mathbf{e} = e_1 e_2 \dots e_I$ has I words. And alignment $\mathbf{a} = a_1 a_2 \dots a_J$ represents the mapping information from source language words to target words.

Statistical machine translation models estimate $Pr(\mathbf{f}|\mathbf{e})$, the translation probability from source language string \mathbf{e} to target language string \mathbf{f} (Och et al., 2003):

$$Pr(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} Pr(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

EM-algorithm is usually used to train the alignment models to estimate lexicon parameters $p(f|e)$.

In E-step, the counts for one sentence pair (\mathbf{f}, \mathbf{e}) are:

$$c(f|e; \mathbf{f}, \mathbf{e}) = \sum_{\mathbf{a}} Pr(\mathbf{a}|\mathbf{f}, \mathbf{e}) \sum_{i,j} \delta(f, f_j) \delta(e, e_{a_j})$$

$$Pr(\mathbf{a}|\mathbf{f}, \mathbf{e}) = Pr(\mathbf{f}, \mathbf{a}|\mathbf{e}) / Pr(\mathbf{a}|\mathbf{e})$$

In the M-step, lexicon parameters become:

$$p(f|e) \propto \sum_s c(f|e; \mathbf{f}^{(s)}, \mathbf{e}^{(s)})$$

Different alignment models such as IBM-1 to IBM-5 (Brown et al., 1993) and HMM model (Och and Ney, 2000) provide different decompositions of

$Pr(\mathbf{f}, \mathbf{a}|\mathbf{e})$. For different alignment models different approaches were proposed to estimate the corresponding alignments and parameters. The details can be found in (Och et al., 2003; Brown et al., 1993).

4.2 Information Need Prediction

After estimating the statistical translation probabilities, we treat the information need prediction as the process of ranking words by $p(w|Q)$, the probability of generating word w from question Q :

$$P(w|Q) = \lambda \sum_{t \in Q} P_{tr}(w|t) P(t|Q) + (1 - \lambda) P(w|C)$$

The word-to-word translation probability $P_{tr}(w|t)$ is the probability of word w is translated from a word t in question Q using the translation model. The above formula uses linear interpolation smoothing of the document model with the background language model $P(t|C)$. λ is the smoothing parameter. $P(t|Q)$ and $P(t|C)$ are estimated using the maximum likelihood estimator.

One important consideration is that statistical machine translation models first estimate $Pr(\mathbf{f}|\mathbf{e})$ and then calculate $Pr(\mathbf{e}|\mathbf{f})$ using Bayes' theorem to minimize ordering errors (Brown et al., 1993):

$$Pr(\mathbf{e}|\mathbf{f}) = \frac{Pr(\mathbf{f}|\mathbf{e}) Pr(\mathbf{e})}{Pr(\mathbf{f})}$$

But in this paper, we skip this step as we found out the order of words in information need part is not an important factor. In our collected CQA archive, question title and information need pairs can be considered as a type of parallel corpus, which is used for estimating word-to-word translation probabilities. More specifically, we estimated the IBM-4 model by *GIZA++*⁴ with the question part as the source language and information need part as the target language.

5 Experiments and Results

5.1 Text Preprocessing

The questions posted on community QA sites often contain spelling or grammar errors. These errors in-

⁴<http://fjoch.com/GIZA++.html>

Methods	Test_c			Test_t		
	MRR	Precision@5	Precision@10	MRR	Precision@5	Precision@10
TFIDF	84.2%	67.1%	61.9%	92.8%	74.8%	63.3%
Knowledge1	82.2%	65.0%	65.6%	78.1%	67.0%	69.6%
Knowledge2	76.7%	54.9%	59.3%	61.6%	53.3%	58.2%
LDA1	92.5%	68.8%	64.7%	91.8%	75.4%	69.8%
LDA2	61.5%	55.3%	60.2%	52.1%	57.4%	54.5%

Table 2: Question recommendation results without information need prediction

Methods	Test_c			Test_t		
	MRR	Precision@5	Precision@10	MRR	Precision@5	Precision@10
TFIDF	86.2%	70.8%	64.3%	95.1%	77.8%	69.3%
Knowledge1	82.2%	65.0%	66.6%	76.7%	68.0%	68.7%
Knowledge2	76.7%	54.9%	60.2%	61.6%	53.3%	58.2%
LDA1	95.8%	72.4%	68.2%	96.2%	79.5%	69.2%
LDA2	61.5%	55.3%	58.9%	68.1%	58.3%	53.9%

Table 3: Question recommendation results with information need predicted by translation model

fluence the calculation of similarity and the performance of information retrieval (Zhao et al., 2007; Bunescu and Huang, 2010). In this paper, we use an open source software *afterthedeathline*⁵ to automatically correct the spelling errors in the question and information need texts first. We also made use of Web 1T 5-gram⁶ to implement an N-Gram based method (Cheng et al., 2008) to further filter out the false positive corrections and re-rank correction suggestions (Mudge, 2010). The texts are tagged by Brill’s Part-of-Speech Tagger⁷ as the rule-based tagger is more robust than the state-of-art statistical taggers for raw web contents. This tagging information is only used for WordNet similarity calculation. Stop word removal and lemmatization are applied to the all the raw texts before feeding into machine translation model training, the LDA model estimating and similarity calculation.

5.2 Construction of Training and Testing Sets

We made use of the questions crawled from Yahoo! Answers for the estimating models and evaluation. More specifically, we obtained 2 million questions under two categories at Yahoo! Answers: ‘travel’

(1 million), and ‘computers&internet’ (1 million). Depending on whether the best answers have been chosen by the asker, questions from Yahoo! answers can be divided into ‘resolved’ and ‘unresolved’ categories. From each of the above two categories, we randomly selected 200 resolved questions to construct two testing data sets: ‘Test_t’ (‘travel’), and ‘Test_c’ (‘computers&internet’). In order to measure the information need similarity in our experiment we selected only those questions whose information needs part contained at least 3 informative words after stop word removal. The rest of the questions ‘Train_t’ and ‘Train_c’ under the two categories are left for estimating the LDA topic models and the translation models. We will show how we obtain these models later.

5.3 Experimental Setup

For each question (query question) in ‘Test_t’ or ‘Test_c’, we used the words in the question title part as the main search query and the other words in the information need part as search query expansion to retrieve candidate recommended questions from Yahoo! Answers website. We obtained an average of 154 resolved questions under ‘travel’ or ‘computers&internet’ category, and three assessors were involved in the manual judgments.

Given a question returned by a recommendation

⁵<http://afterthedeathline.com>

⁶<http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>

⁷<http://www.umiacs.umd.edu/~jimmylin/resources.html>

method, two assessors are asked to label it with ‘good’ or ‘bad’. The third assessor will judge the conflicts. The assessors are also asked to read the information need and answer parts. If a recommended question is considered to express the same or similar information need, the assessor will label it ‘good’; otherwise, the assessor will label it as ‘bad’.

Three measures for evaluating the recommendation performance are utilized. They are Mean Reciprocal Rank (MRR), top five prediction accuracy (precision@5) and top ten prediction accuracies (precision@10) (Voorhees and Tice, 2004; Cao et al., 2008). In MRR the reciprocal rank of a query question is the multiplicative inverse of the rank of the first ‘good’ recommended question. The top five prediction accuracy for a query question is the number of ‘good’ recommended questions out of the top five ranked questions and the top ten accuracy is calculated out of the top ten ranked questions.

5.4 Similarity Measure

The first experiment conducted question recommendation based on their information need parts. Different text similarity methods described in section 3 were used to measure the similarity between the information need texts. In TFIDF similarity measure (TFIDF), the idf values for each word were computed from frequency counts over the entire *Aquaint* corpus⁸. For calculating the word-to-word knowledge-based similarity, a WordNet::Similarity Java implementation⁹ of the similarity measures *lin* (Knowledge2) and *jcn* (Knowledge1) is used in this paper. For calculating topic model based similarity, we estimated two LDA models from ‘Train.t’ and ‘Train.c’ using *GibbsLDA++*¹⁰. We treated each question including the question title and the information need part as a single document of a sequence of words. These documents were preprocessed before being fed into LDA model. 1800 iterations for Gibbs sampling 200 topics parameters were set for each LDA model estimation.

The results in table 2 show that TFIDF and LDA1 methods perform better for recommending questions than the others. After further analysis of the questions recommended by both methods, we discov-

Q1:	If I want a faster computer should I buy more memory or storage space?
InfoN	If I want a faster computer should I buy more memory or storage space? What's the difference? I edit pictures and videos so I need them to work quickly. ...
RQ1	Would buying 1gb memory upgrade make my computer faster?
InfoN	I have an inspiron B130. It has 512mb memory now. I would add another 1gb into 2nd slot ...
RQ2	whats the difference between memory and hard drive space on a computer and why is.....?
InfoN	see I am starting edit videos on my computer but i am running out of space. why is so expensive to buy memory but not external drives? ...
Q2:	Where should my family go for spring break?
InfoN	... family wants to go somewhere for a couple days during spring break ... prefers a warmer climate and we live in IL, so it shouldn't be SUPER far away. ... a family road trip. ...
RQ1	Whats a cheap travel destination for spring break?
InfoN	I live in houston texas and i'm trying to find i inexpensive place to go for spring break with my family.My parents don't want to spend a lot of money due to the economy crisis, ... a fun road trip...
RQ2	Alright you creative deal-seekers, I need some help in planning a spring break trip for my family
InfoN	Spring break starts March 13th and goes until the 21st ... Someplace WARM!!! Family-oriented hotel/resort ... North American Continent (Mexico, America, Jamaica, Bahamas, etc.) Cost= Around \$5,000 ...

Table 4: Question recommendation results by LDA measuring the similarity between information needs

⁸<http://ldc.upenn.edu/Catalog/docs/LDC2002T31>

⁹<http://cogs.susx.ac.uk/users/drh21/>

¹⁰<http://gibbslda.sourceforge.net>

ered that the ordering of the recommended questions from TFIDF and LDA1 are quite different. TFIDF similarity method prefers texts with more common words, while the LDA1 method can find the relation between the non-common words between short texts based on a series of third-party topics. The LDA1 method outperforms the TFIDF method in two ways: (1) the top recommended questions' information needs share less common words with the query question's; (2) the top recommended questions span wider topics. The questions highly recommended by LDA1 can suggest more useful topics to the user.

Knowledge-based methods are also shown to perform worse than TFIDF and LDA1. We found that some words were mis-tagged so that they were not included in the word-to-word similarity calculation. Another reason for the worse performance is that the words out of the WordNet dictionary were also not included in the similarity calculation.

The Mean Reciprocal Rank score for TFIDF and LDA1 are more than 80%. That is to say, we are able to recommend questions to the users by measuring their information needs. The first two recommended questions for Q1 and Q2 using LDA1 method are shown in table 4. InfoN is the information need part associated with each question.

In the preprocessing step, some words were successfully corrected such as “*What should I do this saturday? ... and staying in a hotell ...*” and “*my faimly is traveling to florda ...*”. However, there are still a small number of texts such as “*How come my Gforce visualization doesn't work?*” and “*Do i need an Id to travel from new york to maimi?*” failed to be corrected. So in the future, a better method is expected to correct these failure cases.

5.5 Information Need Prediction

There are some retrieved questions whose information need parts are empty or become empty or almost empty (one or two words left) after the preprocessing step. The average number of such retrieved questions for each query question is 10 in our experiment. The similarity ranking scores of these questions are quite low or zero in the previous experiment. In this experiment, we will apply information need prediction to the questions whose information needs are missing in order to find out whether we improve the recommendation task.

The question and information need pairs in both ‘Train_t’ and ‘Train_c’ training sets were used to train two IBM-4 translation models by *GIZA++* toolkit. These pairs were also preprocessed before training. And the pairs whose information need part become empty after preprocessing were disregarded.

During the experiment, we found that some of the generated words in the information need parts are themselves. This is caused by the self translation problem in translation model: the highest translation score for a word is usually given to itself if the target and source languages are the same (Xue et al., 2008). This has always been a tough question: not using self-translated words can reduce retrieval performance as the information need parts need the terms to represent the semantic meanings; using self-translated words does not take advantage of the translation approach. To tackle this problem, we control the number of the words predicted by the translation model to be exactly twice the number of words in the corresponding preprocessed question.

The predicted information need words for the retrieved questions are shown in Table 5. In Q1, the information need behind question “*recommend website for custom built computer parts*” may imply that the users need to know some information about building computer parts such as “*ram*” and “*motherboard*” for a different purpose such as “*gaming*”. While in Q2, the user may want to compare computers in different brands such as “*dell*” and “*mac*” or consider the “*price*” factor for “*purchasing a laptop for a college student*”.

We also did a small scale comparison between the generated information needs against the real questions whose information need parts are not empty. Q3 and Q4 in Table 5 are two examples. The original information need for Q3 is “*looking for beautiful beaches and other things to do such as museums, zoos, shopping, and great seafood*” in CQA. The generated content for Q3 contains words in wider topics such as ‘*wedding*’, ‘*surf*’ and the price information (‘*cheap*’). This reflects that there are some other users asking similar questions with the same or other interests.

From the results in Table 3, we can see that the performance of most similarity methods were improved by making use of information need predic-

tion. Different similarity measures received different degrees of improvement. LDA1 obtained the highest improvement followed by the TFIDF based method. These two approaches are more sensitive to the contents generated by a translation model.

However we found out that in some cases the LDA1 model failed to give higher scores to good recommendation questions. For example, Q5, Q6, and Q7 in table 5 were retrieved as recommendation candidates for the query question in Table 1. All of the three questions were good recommendation candidates, but only Q6 ranked fifth while Q5 and Q7 were out of the top 30 by LDA1 method. Moreover, in a small number of cases bad recommendation questions received higher scores and jeopardized the performance. For example, for query question “*How can you add subtitles to videos?*” with information need “... add subtitles to a music video ... got off youtube ...download for this ...”, a retrieved question “*How would i add a music file to a video clip. ...*” was highly recommended by TFIDF approach as predicted information need contained ‘youtube’, ‘video’, ‘music’, ‘download’,

The MRR score received an improvement from 92.5% to 95.8% in the ‘Test.c’ and from 91.8% to 96.2% in ‘Test.t’. This means that the top one question recommended by our methods can be quite well catering to the users’ information needs. The top five precision and the top ten precision scores using TFIDF and LDA1 methods also received different degrees of improvement. Thus, we can improve the performance of question recommendation by predicting information needs.

6 Conclusions

In this paper we addressed the problem of recommending questions from large archives of community question answering data based on users’ information needs. We also utilized a translation model and a LDA topic model to predict the information need only given the user’s query question. Different information need similarity measures were compared to prove that it is possible to satisfy user’s information need by recommending questions from large archives of community QA. The Latent Dirichlet allocation based approach was proved to perform better on measuring the similarity between short

Q1:	Please recommend A good website for Custom Built Computer parts?
InfoN	custom, site, ram, recommend, price, motherboard, gaming, ...
Q2:	What is the best laptop for a college student?
InfoN	know, brand, laptop, college, buy, price, dell, mac, ...
Q3:	What is the best Florida beach for a honeymoon?
InfoN	Florida, beach, honeymoon, wedding, surf, cheap, fun, ...
Q4:	Are there any good clubs in Manchester
InfoN	club, bar, Manchester, music, age, fun, drink, dance, ...
Q5:	If i buy a video card for my computer will that make it faster?
InfoN	nvidia, video, ati, youtube, card, buy, window, slow, computer, graphics, geforce, faster, ...
Q6:	If I buy a bigger hard drive for my laptop, will it make my computer run faster or just increase the memory?
InfoN	laptop, ram, run, buy, bigger, memory, computer, increase, gb, hard, drive, faster, ...
Q7:	Is there a way I can make my computer work faster rather than just increasing the ram or hardware space?
InfoN	space, speed, ram, hardware, main, gig, slow, computer, increase, work, gb, faster, ...

Table 5: Information need prediction examples using IBM-4 translation model

texts in the semantic level than traditional methods. Experiments showed that the proposed translation based language model for question information need prediction further enhanced the performance of question recommendation methods.

References

- Ricardo A. Baeza-Yates and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, Robert L. Mercer. 1993. *The mathematics of statistical machine translation: parameter estimation*. Computational Linguistics, v.19 n.2, June 1993.
- Razvan Bunescu and Yunfeng Huang. 2010. *Learning the Relative Usefulness of Questions in Community QA*. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Cambridge, MA.
- Robin D. Burke and Kristian J. Hammond and Vladimir A. Kulyukin and Steven L. Lytinen and Noriko Tomuro and Scott Schoenberg. 1997. *Question answering from frequently-asked question files: Experiences with the FAQ Finder system*. AI Magazine, 18, 57C66.
- Yunbo Cao, Huizhong Duan, Chin-Yew Lin, Yong Yu, and Hsiao-Wuen Hon. 2008. *Recommending Questions Using the MDL-based Tree Cut Model*. In: Proc. of the 17th Int. Conf. on World Wide Web, pp. 81-90.
- Asli Celikyilmaz and Dilek Hakkani-Tur and Gokhan Tur. 2010. *LDA Based Similarity Modeling for Question Answering*. In NAACL 2010 C Workshop on Semantic Search.
- Charibeth Cheng, Cedric Paul Alberto, Ian Anthony Chan, and Vazir Joshua Querol. 2008. *SpellCheF: Spelling Checker and Corrector for Filipino*. Journal of Research in Science, Computing and Engineering, North America, 4, sep. 2008.
- Lynn Silipigni Connaway and Chandra Prabha. 2005. *An overview of the IMLS Project "Sense-making the information confluence: The whys and hows of college and university user satisficing of information needs"*. Presented at Library of Congress Forum, American Library Association Midwinter Conference, Boston, MA, Jan 16, 2005.
- Huizhong Duan, Yunbo Cao, Chin-Yew Lin, and Yong Yu. 2008. *Searching questions by identifying question topic and question focus*. In HLT-ACL, pages 156C164.
- Thomas L. Griffiths and Mark Steyvers. 2004. *Finding scientific topics*. Natl Acad Sci 101:5228C5235.
- Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee. 2005a. *Finding semantically similar questions based on their answers*. In Proc. of SIGIR05.
- Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee. 2005b. *Finding similar questions in large question and answer archives*. In CIKM, pages 84C90.
- Jay J. Jiang and David W. Conrath. 1997. *Semantic similarity based on corpus statistics and lexical taxonomy*. In Proceedings of International Conference on Research in Computational Linguistics, Taiwan.
- Dekang Lin. 1998. *An Information-Theoretic Definition of Similarity*. In Proceedings of the Fifteenth International Conference on Machine Learning (ICML '98), Jude W. Shavlik (Ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 296-304.
- Yandong Liu, Jiang Bian, and Eugene Agichtein. 2008. *Predicting information seeker satisfaction in community question answering*. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08). ACM, New York, NY, USA, 483-490.
- Rada Mihalcea, Courtney Corley, and Carlo Strapparava. 2006. *Corpus-based and knowledge-based measures of text semantic similarity*. In Proceedings of the 21st national conference on Artificial intelligence (AAAI '06), pages 775C780. AAAI Press.
- Raphael Mudge. 2010. *The design of a proofreading software service*. In Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids (CL&W '10). Association for Computational Linguistics, Morristown, NJ, USA, 24-32.
- Franz Josef Och, Hermann Ney. 2000. *A comparison of alignment models for statistical machine translation*. Proceedings of the 18th conference on Computational linguistics, July 31-August 04, Saarbrücken, Germany.
- Franz Josef Och, Hermann Ney. 2003. *A Systematic Comparison of Various Statistical Alignment Models*. Computational Linguistics, volume 29, number 1, pp. 19-51 March 2003.
- Jahna Otterbacher, Gunes Erkan, Dragomir R. Radev. 2009. *Biased LexRank: Passage retrieval using random walks with question-based priors*. Information Processing and Management: an International Journal, v.45 n.1, p.42-54, January, 2009.
- Chandra Prabha, Lynn Silipigni Connaway, Lawrence Olszewski, Lillie R. Jenkins. 2007. *What is enough? Satisficing information needs*. Journal of Documentation (January, 63,1).
- Ellen Voorhees and Dawn Tice. 2000. *The TREC-8 question answering track evaluation*. In Text Retrieval Conference TREC-8, Gaithersburg, MD.
- Kai Wang, Yanming Zhao, and Tat-Seng Chua. 2009. *A syntactic tree matching approach to finding similar*

- questions in community-based qa services.* In SIGIR, pages 187C194.
- Kai Wang and Tat-Seng Chua. 2010. *Exploiting salient patterns for question detection and question retrieval in community-based question answering.* In Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10). Association for Computational Linguistics, Stroudsburg, PA, USA, 1155-1163.
- Hu Wu, Yongji Wang, and Xiang Cheng. 2008. *Incremental probabilistic latent semantic analysis for automatic question recommendation.* In RecSys.
- Xiaobing Xue, Jiwoon Jeon, W. Bruce Croft. 2008. *Retrieval models for question and answer archives.* In SIGIR'08, pages 475C482. ACM.
- Shiqi Zhao, Ming Zhou, and Ting Liu. 2007. *Learning Question Paraphrases for QA from Encarta Logs.* In Proceedings of International Joint Conferences on Artificial Intelligence (IJCAI), pages 1795-1800.