Diversifying Information Needs in Results of Question Retrieval

Yaoyun Zhang, Xiaolong Wang, Xuan Wang, Ruifeng Xu, Jun Xu and Shixi Fan
Key Laboratory of Network-oriented Intelligent Computation
Harbin Institute of Technology, Shenzhen Graduate School, China
{xiaoni5122, xuruifeng.hitsz, hit.xujun}@gmail.com
{wangxl, wangxuan}@insun.hit.edu.cn

Abstract

Information need is an important factor in question retrieval. This paper proposes a method to diversify the results of question retrieval in term of types of information needs. *CogQTaxo*, a question hierarchy is leveraged to represent users' information needs cognitively from three linguistic levels. Based on a prediction model of question types, three factors, i.e., scores of IR model, question type similarity and question type novelty are linearly combined to re-rank the retrieved questions. Preliminary experimental results show that the proposed method enhances the question retrieval performance in information coverage and diversity.

1 Introduction

Most current question retrieval system attempts to fetch questions semantically similar to the query question (Jeon et al, 2005), together with the accepted answers from a large question-answer pair archive. Previous works focus on reducing the lexicon gap between the query question and retrieved questions (Cao et al, 2010) or recognize the single question type, i.e., the type of information needs(*infoNeed*s) in the query, and confine the types of retrieved questions to be the same as the query (Lytinen and Tomuro, 2002). Normally, the retrieved questions are ranked according to the semantic similarity to the query question.

However, Taylor (1962) argues that the user may fail to express his *infoNeed*s fully in the question. Besides, given different contextual situations, users may have different intentions, which lead to different *infoNeed*s for the same question (Small and Strzalkowski, 2008).

For an example question q_1, "which bank provides the best credit card?", if the user wants to confirm the bank he knows, the name of the bank is enough for an answer; while the user plans to open a credit card account, he may want to obtain detailed descriptions and comparisons between credit card services of different banks in addition to a single bank name. Furthermore, a play-it-safe user may expect the information source of the answer to be of authority or expertise, while a casual user may expect it to be commonsense that anyone can answer.

Considering these requirement, the following two questions q_2 and q_3 should be provided to the user under a certain context. Nevertheless, such *infoNeed*s are not given explicitly in the q_1.

q_2: Which bank should I choose for credit card, Citi Bank and Bank of America?

q_3: How to choose credit card?

As can be observed, the three questions have different types, which are *entity*, *alternative* and *method*, respectively (Diekema et al, 2003). Apparently, the single-dimensional question taxonomies employed at present are insufficient to model those aspects of users' *infoNeed*s (Pomerantz, 2005). Thus, more comprehensive question taxonomy is needed. The question retrieval results should also be diversified accordingly to fulfill these implicit and context-dependent *infoNeeds*, thus making the results more comprehensive for average users.

Present works (Clark et al, 2009; Santos et al, 2010) mainly target on search result diversification for short queries instead of questions. Their focus is to mine the different interpretations of ambiguous queries or navigations for a broad-sense query. Achananuparp et al. (2010) attempted to diversify the aspects of the answer to complex questions, while they also focus on the short information nuggets returned by search engines.

Based on our knowledge, no previous work has been done on the results diversification for question retrieval. In this paper, we utiliz *CogQTaxo*, a multi-dimensional question taxonomy to model both the explicit and implicit

Proceedings of the 5th International Joint Conference on Natural Language Processing, pages 1432–1436,
Chiang Mai, Thailand, November 8 – 13, 2011.

*infoNeed*s of questions. Based on this, we propose an algorithm to diversify the results of question retrieval in terms of *infoNeed* types. The comparative experimental results show that the proposed algorithm enhances the information coverage and diversity of retrieved questions.

2 *CogQTaxo* - Three Dimensional Question Taxonomy

CogQTaxo is proposed by Zhang et al (2010). It is a framework of three-dimensional question taxonomy by using different levels of linguistic analysis (syntactic, semantic and pragmatic) as the classification criteria.

Let T_i (i=1,2,3) denotes the *i*th dimension of *CogQTaxo*, then:

1. T_1 represents the surface information need (*surfaceIN*), which corresponds to the conventional definition of question types (*QuesTs)*. A question can has one definite type in *surfaceIN*; 14 types are defined for *surfaceIN*, namely *location, person, time, quantity, thing, alternative, definition, comparison, description, procedure, reason, yesNo, abstractEnity* and *other*.

2. T_2 represents implicit information needs (*implicitIN*). *QuesTs* in this dimension are the same as in *surfaceIN*. Nevertheless, it represents the *infoNeeds* which are not expressed explicitly in the question, yet are required to fill the user's information gap. A question has at least one type in *implicitIN*.

3. T_3 represents users' pragmatic expectations (*pragmaticE*) from the answer. Four binary-valued pragmatic aspects are currently considered: (1) *Specification*: whether the question contains detailed specific information as the context; (2) *Knowledge source*: whether the question requires commonsense or expertise to answer; (3) *Temporal constraint*: whether the answer is time sensitive, i.e., whether the answer should be constraint to a time-frame; (4)*Subjectivity Orientation*: whether the information in the expected answer is subjective-oriented or objective-oriented.

A prediction model is built by Zhang et al (2010) to recognize the types of a question in each dimension of *CogQTaxo*. In this study, *CogQTaxo* is employed to diversify the *infoNeeds* in the results of question retrieval.

3 Diversification Algorithm for Question Retrieval Result

According to the definition of *CogQTaxo*, the three dimensions have different functions in user *infoNeeds* fulfillment, in which *surfaceIN* is fundamental and indispensable from the answer. *implicitIN* provides supportive information and helps to make the answer coverage more comprehensive. Therefore, we use *surfaceIN* and *implicitIN* to diversify the types of *infoNeeds* in retrieval results. The predicted *QuesT* sets in these two dimensions are merged into an extended one, in which the *QuesT*s are equally weighted at present. Meanwhile, the third dimension in *CogQTaxo*, *pragmaticE*, adds pragmatic constraints to the former two.

As displayed in figure 1, our question diversification algorithm is given as follows:

For a input question *p*, the question retrieval system will:

Step 1: Question analysis: The content words (nouns, verbs and adjectives) are extracted from *p* as the question content. Types of *p* in line with *CogQTaxo* are recognized automatically by using the model proposed in (Zhang et al, 2010).

Step 2: Question retrieval: retrieve relevant questions with the information retrieval (IR) model by using question content as the query. The relevance score is denoted as *IRScore*, which is normalized by the highest score of retrieved questions for *p*.

Step 3: Question Reranking with *QuesT* Similarity: Similar to (Lytinen and N. Tomuro, 2002) and (Cao et al, 2010), this step considers the relevance of *QuesT* between *p* and *q* for result ranking. Nevertheless, the question taxonomy deployed here is multi-dimensional. For each question *q* in the retrieved question set, T_iScore is defined as the *QuesT* set distance between *p* and *q* in the *i*th dimension of *CogQTaxo*, i=1, 2, 3. It is calculated by MASI (Passonneau, 2006). Since we

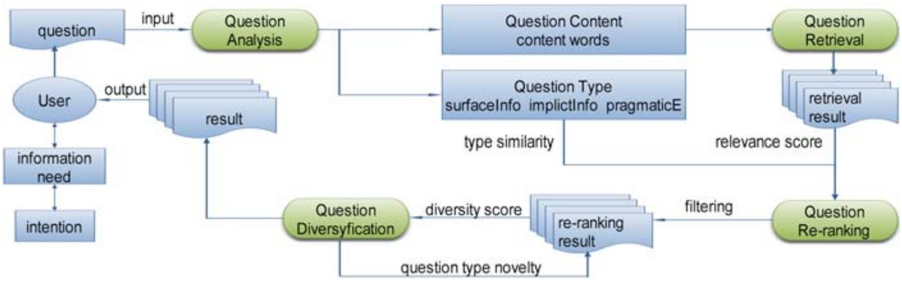

Figure 1 Diversification Procedure of Question Retrieval Results

merge T_1 with T_2, the retrieved results are re-ranked by *rerankScore*, which is defined as:

$$rerankScore = (1 - l_{1+2} - l_3) * IRScore + l_{1+2} * T_{1+2}Score + l_3 * T_3Score$$

where $l_{1+2}, l_3 \in [0,1], l_{1+2} + l_3 \in [0,1]$.

Result questions with the *rerankScore* lower than a threshold λ are filtered.

Step 4: Question *infoNeeds* diversification: This step employs a greedy algorithm to add one question with the largest *infoNeeds* novelty into the final returned question list in each iteration.

Suppose m questions are left in the result set after step 3, we denote *DiverseList* as the list of r questions re-ranked by diversity. For a question q in the m-r remaining result questions, its *QuesTs* novelty is defined as:

$$Novelty(T) = avg(\sum \frac{1}{e^{df_{type_j}+1}}),\ \exists\, type_j \in T^q_{1+2} \cap T^p_{1+2},$$

where df_{typej} is the frequency of $type_j \in T^q_{1+2} \cap T^p_{1+2}$ in *DiverseList*.

Then *diverseScore* is computed as follows:

$$diversScore = (1 - w) * rerankScore + w * Novelty(T),\ 0 \le w \le 1.$$

The question with the highest *diversScore* value is added to *DiverseList*[1].

Repeat Step 4 until the *DiverseList* with top n ($m \ge n$) questions are returned to the user.

4 Experiment and Discussion

4.1 Dataset and Experimental Setup

Questions with accepted answers are collected from the Yahoo! Knowledge portal and Baidu Zhidao portal, respectively. After removing redundancy and invalid questions, more than 1,380,000 postings are obtained [2] postings are obtained. The title of the posting is used as the question, while the accepted response is regarded as the answer.

100 questions chosen randomly as the query questions, the other questions are indexed to build the question retrieval system. Only the content words of questions are indexed. In the experiment, we used three IR model, namely Okapi BM25 model, Vector space model and language model, respectively, in which BM25 outperforms. Therefore, only the performance achieved by BM25 is reported in the rest of this section.

Relevance set: The relevance set of the 100 query questions are built by judging the content relevance between the query and the results regardless of the *infoNeeds*. Poolings among the top 10 results by the evaluated methods are conducted. Finally, 2258 relevant questions are collected.

Information need annotation: Three annotators annotate the *QuesTs* of the 100 query questions individually, by following the same instruction as Zhang et al (2010). In this way, three different *infoNeeds* sets of the query questions are generated. The algorithm performance is evaluated on each *infoNeeds* set separately, while the average performance is reported.

Evaluation criteria: We use *MAP_IA*, *MRR_IA* and *P@K_IA* designed by Agrawal et al (2009) as the evaluation metrics. These metrics are originally defined as the weighted arithmetic mean of performance of each subtopic of a query. In this paper, we substitute the subtopics of a query into the potential types of a question. At present we consider all of *QuesTs* as equally weighted. For example, the formula of *MAP_IA* is as follows:

$$MAP\_IA = \frac{\sum_{QuesT_i \in T_1 \cup T_2} W_i MAP(QuesT_i)}{(|T_1| + |T_2|)}$$

Furthermore, the relevance judgment in those metrics between question p and q is not simply bi-

[1] The reported experimental results are derived by l_{1+2}=0.2, l_3 =0.2, w=0.4, λ=0.5, which are obtained in a pilot experiment using 20 randomly selected test queries.

[2] Data used in this paper can be downloaded from Http://qa.haitianyuan.com/cogQTaxo.html

Table 2 Question retrieval results are listed for the query “Which stock is good to buy?”, using BM25 as the retrieval model.

Predicted question type : entity/ description procedure alternative	
BM25 model	BM25 model with question type diversity
Which stock is good to buy?	Which stock is good to buy?
Which stock is good recently?	Which stock is good recently?
Which stock should I buy recently?	What characteristics good stocks have?
Recommend some good stocks to me.	How to identify a good stock?
Is there any good stock to buy?	Which stocks should I buy; good recommendations will be highly rewarded.

Table 3 Information needs diversification performance of the evaluated methods.

	Retrive_M	*Pop_Div*	*SurfaceIN_M*	*implicitIN_M*	*PragmaticE_M*	*LinearC*	*Bow_Div*	*Predict_Div*
MRR_IA	0.343	0.526	0.375	0.371	0.347	0.390	**0.527**	**0.529**
MAP_IA	0.114	0.140	0.138	0.134	0.120	**0.149**	0.058	**0.164**
P_IA @1	0.181	0.211	0.239	0.237	0.213	**0.245**	0.245	**0.262**
P_IA@5	0.192	0.197	0.218	0.215	0.205	**0.230**	0.106	**0.244**

nary valued, as either relevant or not; it is replaced by the similarity between *p* and *q* in *pragmaticE*. As mentioned before, *pragmaticE* add pragmatic constraints to the other two dimensions of *infoNeeds*.

$$infoNeed\_relevance(q) = \begin{cases} T_3Score, & \text{if } relevance(q) = 1 \\ 0, & \text{if } relevance(q) = 0 \end{cases}$$

Evaluated question diversification methods: (1) ***Retrieve_M***: only using the IR model; (2) ***Pop_Div***: Instead of using the *QuesT* prediction model built by (Zhang et al, 2000), the *QuesT*s with the highest relative frequency (larger than 10%), i.e., the most popular *QuesT*s in the top 200 retrieved results by *Retrieve_M* are used as the potential type of *infoNeeds* of the query question; (3) ***SurfaceIN_M***, ***implicitIN_M***, ***PragmaticE_M***: using each of the three dimensions of *CogQTaxo* in the diversification algorithm, individually; (4) ***LinearC***: The first three steps of the diversification algorithm, i.e., without the diversification iteration step; (5) ***Bow_Div***: treating the question as bag-of-words, follows the same procedure without Step 3 in section 3, and only considers the novelty of content words in result questions in Step 4; (6) ***Predict_Div***: the complete proposed diversification algorithm.

4.2 Experimental Results

Table 2 illustrates the top 5 search results of query “Which stock is good to buy?” using *Retrive_M* and *Predict_Div*, respectively. As can be seen, the *infoNeed*s in questions retrieved by *Predict_Div* are more diverse than those retrieved by *Retrive_M*.

Table 3 lists the *infoNeed*s diversification performance achieved by each method, respectively. It is observed that *Predict_Div* outperforms. It is also shown that performance of *Bow_Div* is comparable with *Predict_Div* in *MRR_IA* and *P_IA @1*; however, it is even inferior to *Retrieve_M* in *MAP_IA* and *P_IA @5*. This indicates that the naïve bag-of-word baseline is unable to recall diverse *infoNeed*s of the query, and even deteriorates the performance. *Pop_Div* and *Predict_Div* are comparable in *MRR_IA*. However, in terms of other metrics, *LinearC* and *Predict_Div* are consistently at the top 2 ranks. The reason is that since the predicted types of a question are already diversified by *CogQTaxo*, incorporating it into question re-ranking already enables us to diversify the *infoNeeds* in the results implicitly. Therefore, the explicit diversification step enhances the performance further.

One deficit of the evaluation framework is that the *infoNeed*s of questions in the relevance set are predicted automatically instead of manually annotated; this may result in a bias towards our proposed algorithm. However, since the training set of the question classifier is manually annotated. Thus, it reflects the real user *infoNeed*s distribution. It is assumed that the automatic prediction can also reflect real user *infoNeed*s to some extent. More detailed analysis will be conducted later to examine this problem.

5 Conclusion

This paper proposes a method to diversify the results of question retrieval in term of types of information needs. Comparison results show that the proposed method improves the information need coverage and diversity in retrieved questions.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (No. 61173075 and 60973076).

References

Anne R. Diekema, Ozgur Yilmazel, Jiangping Chen, Sarah Harwell, Lan He, Elizabeth D. Liddy. 2003. What do you mean? Finding answers to complex questions. *Proceedings of the AAAI Spring Symposium: New Directions in question Answering* .

Charles L. A. Clarke, Nick Craswell, Ian Soboroff. 2009. Overview of the trec 2009 web track. *Proceedings of the 18th Text Retrieval Conference.*

Elizabeth D. Liddy. 1998. Enhanced text retrieval using natural language processing. *Bulletin of the American Society for Information Science and Technology*, 24(4): 14–16.

Ingrid Zukerman and Eric Horvitz. 2001. Using Machine Learning Techniques to Interpret WH-questions. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pp. 547–554.

Jeffrey Pomerantz. 2005. A Linguistic Analysis of Question Taxonomies. *Journal of the American Society for Information Science and Technology*, 56(7):715–728.

Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee. 2005. Finding Similar Questions in Large Question and Answer Archives. *Proceedings of Conference on Information and Knowledge Management*, pp. 84–90.

John Burger, Claire Cardie, Vinay Chaudhri, Robert Gaizauskas, Sanda Harabagiu, David Israel, et al. 2001. Issues, Tasks and Program Structures to Roadmap Research in question & Answering (Q&A). *Technical Report, NIST.*

Lytinen and N. Tomuro. 2002. The use of question types to match questions in faqfinder. *AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases.*

Palakorn Achananuparp, Xiaohua Hu, Tingting He. 2010. Answer Diversification for Complex Question Answering on the Web. *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp.375-382

Passonneau, Rebecca J. 2006. Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. *Proceedings of the International Conference on Language Resources and Evaluation*, pp.831–836.

Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. 2009. Diversifying search results. *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pp.5-14.

Robert S. Taylor. 1962. The Process of Asking Questions. *American Documentation*, 13(4):391-396.

Rodrygo L. T. Santos, Craig Macdonald and Iadh Ounis. 2010. Exploiting Query Reformulations for Web Search Result Diversification. *Proceedings of International World Wide Web Conference*, pp.881-890.

Sharon Small, Tomek Strzalkowski. 2008. HITIQA: High-quality intelligence through interactive question answering. *Natural Language Engineering*, 15 (1): 31–54

Xin Cao, Gao Cong, Bin Cui, Christian S. Jensen. 2010. A Generalized Framework of Exploring Category Information for Question Retrieval in Community Question Answer Archives. *Proceedings of International World Wide Web Conference*, pp. 201-210.

Yaoyun Zhang, Xiaolong Wang, Xuan Wang, Shixi Fan. 2010. *CogQTaxo*: Modeling human cognitive process with a three-dimensional question taxonomy. *International Conference on Machine Learning and Cybernetics*, pp.3305 – 3310.