Mining Wikipedia and Yahoo! Answers for Question Expansion in Opinion QA

Yajie Miao and Chunping Li

Tsinghua National Laboratory for Information Science and Technology (TNList)
School of Software, Tsinghua University,
Beijing 100084, China
yajiemiao@gmail.com, cli@tsinghua.edu.cn

Abstract. Opinion Question Answering (Opinion QA) is still a relatively new area in QA research. The achieved methods focus on combining sentiment analysis with the traditional Question Answering methods. Few attempts have been made to expand opinion questions with external background information. In this paper, we introduce the *broad-mining* and *deep-mining* strategies. Based on these two strategies, we propose four methods to exploit Wikipedia and Yahoo! Answers for enriching representation of questions in Opinion QA. The experimental results show that the proposed expansion methods perform effectively for improving existing Opinion QA models.

Keywords: Opinion QA, Question Expansion, Wikipedia, Yahoo! Answers.

1 Introduction

Question Answering (QA), which aims to retrieve answers to human-generated questions automatically, is an important research area in text mining and information retrieval. Many of the methods in this area have been proposed mostly for the task of answering fact-based questions, e.g., "When was the Kyoto Protocol adopted?". However, in many cases, users are more interested in the opinions towards specific events or objects. Questions querying about opinions or attitudes are defined as *opinion questions*, e.g.,"How do the Chinese regard the human rights record of the United States?".

The existing methods for Opinion QA focus on utilizing sentimental information to obtain desirable results. However, a key problem for Opinion QA is that the information needs expressed by an opinion question is much more complicated than a fact-based question. The lexical elements (i.e., words) in the opinion questions are usually unable to express such needs completely. One way to address this problem is to enrich the representation of an opinion question with information from some external knowledge repositories.

In this paper, we exploit Wikipedia and Yahoo! Answers to expand the questions in Opinion QA. We adopt two mining strategies, i.e., broad-mining and

M.J. Zaki et al. (Eds.): PAKDD 2010, Part I, LNAI 6118, pp. 367–374, 2010.

deep-mining in both Wikipedia and Yahoo! Answers, and propose four expanding methods: *Wiki-broad*, *Wiki-deep*, *Yahoo-broad* and *Yahoo-deep*. Experiments show that all of these four methods boost the performance of the state-of-the-art Opinion PageRank [4] model. Also, we observe that Wiki-deep is the most effective method for question expansion in Opinion QA.

The rest of the paper is organized as follows. Section 2 reviews previous work. Section 3 formulates the proposed expansion methods. In Section 4, we present and discuss the experimental results. We have the concluding remarks and future work in Section 5.

2 Previous Work

Opinion QA is still a new area in QA research. Stoyanov et al. [1] trained a *subjectiveness classifier* which filters out the objective sentences from the answer candidates and therefore improves the quality of the answers to opinion questions. Kim et al. [2] proposed that the opinion holders in the opinion question and its answers should be the same. Based on this, they improved the performance of Opinion QA by identifying opinion holders in the sentences. In the TAC 2008 Opinion QA track [3], most participants found answers to opinion questions through combining linearly the topic and opinion weights of answer candidates. Li et al. [4] proposed the Opinion PageRank and Opinion HITS models for answering opinion questions. In both models, the topical relevance and sentimental information are combined in a unified graph-based framework.

There has been a growing amount of research on employing Wikipedia to enhancing traditional text mining tasks. Gabrilovich et al. [5] proposed a method to improve text classification through enriching document representation with Wikipedia concepts. Banerjee et al. [6] proposed to improve clustering of short texts by using Wikipedia concepts as additional features. Hu et al. [10] proposed two mapping strategies for enriching document representation with Wikipedia concept and category information. The enriched documents are used for the task of text clustering.

Wikipedia and Yahoo! Answers have also been applied for Question Answering. Ye et al. [7] proposed to summarize a series of definitions from Wikipedia, which serve as answers to definition questions. Wang et al. [8] proposed an *analogical reasoning* approach for measuring the linkage between questions and their answers. Through exploiting the previous Yahoo! Answers data, their approach can build links between a new question-answer pair. Wang et al. [9] proposed a *syntactic tree matching* method for retrieving similar questions from Yahoo! Answers when given a new question. However, these works have made no attempts to use Wikipedia or Yahoo! Answers for question expansion.

3 Question Expansion

In this section, we first formulate a method for generating topic words for opinion questions. Then, we present our methods for question expansion.

3.1 Topic Word Generation

Wikipedia and Yahoo! Answers can receive queries from users and return a list of relevant (Wiki) articles or (Yahoo) questions. However, their retrieval modules are design for the query which consists of several keywords. If the query is sentences in natural language, Wikipedia and Yahoo! Answers are quite likely to return no relevant results. Therefore, the opinion questions should not be submitted to Wikipedia and Yahoo! Answers directly. To address this problem, we first generate topic words for the questions and use these topic words as the query.

Usually several questions can be about the same topic, though they are pertaining to various aspects. For a topic T, there are n questions which make up the *question set* $Q_T = \{q_1, q_2, ..., q_n\}$. All the words in the n questions are ranked according to their frequencies of appearing in Q_T. The top K non-stop words are chosen as the topic words for T. For instance, in the MPQA dataset [1], there are totally 4 questions under the topic "kyoto". With the above procedures, we can get topic words such as "Kyoto", "US", "Protocol", etc.

3.2 Expansion with Wikipedia

Wikipedia is a huge document corpus which contains more than three millions articles. In addition, Wikipedia undergoes constant development, so its breadth and depth steadily increase over time. The topic words generated in Section 3.1 are combined into the Wikipedia query. After searching in its database, Wikipedia returns relevant Wiki articles, which are ranked according to their relevance to the query.

Wiki-broad. We adopt a broad-mining strategy for exploring the ranking list. The M most relevant articles are selected out as the Wikipedia article set WA. In Wikipedia, each article has a title which summarizes the most essential ideas. From WA, we only extract the titles to form the title set WT. Also, the *redirect titles* in WA, which show explicitly the articles which redirect to the ones in WA, are also included in WT. All the non-stop words in WT are extracted to form the title word set TW. If a word appears more than once in TW, it is considered to be a single word (rather than multiple words). Then the set TW is viewed as the *expansion set* for the questions.

Wiki-deep. In addition to text contents, Wikipedia also constructs links between articles. These links provide valuable information about the relation between concepts[1]. If there is a link from the article p_1 to the article p_2, we can conclude that p_2 presents relevant information about the concept of p_1. Therefore, p_2 can be further exploited to extend the contents of p_1. This is the basic idea for our Wiki-deep method. In Wiki-broad, the top M articles are selected from the retrieval results. However, in Wiki-deep, we only focus on the first article which is also the most relevant one. The first paragraph of a Wiki article

[1] In Wikipedia research, the title of a Wikipedia article is usually referred to as a concept.

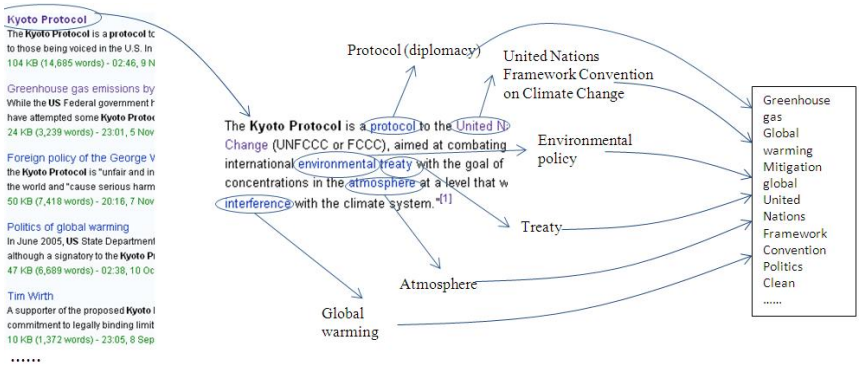

Fig. 1. An example of Wiki-deep

usually serves to summarize the definition of the concept or the main points of the article. All the Wiki articles, which the links in the first paragraph of the most-relevant article point to, are selected to form the link article set LA. Then, all the non-stop words in the titles of the articles in LA and in the title of the most-relevant article are extracted to form a word set. Duplicate words are considered to be a single word. This set is used as the expansion set for the questions. Figure 1 gives the process of Wiki-deep for the questions whose topic words are "Kyoto", "US" and "Protocol" (see Section 3.1).

3.3 Expansion with Yahoo! Answers

Besides Wikipedia, we also use Yahoo! Answers as external knowledge for question expansion. The topic words generated in Section 3.1 are combined into the query which is submitted to Yahoo! Answers. With the APIs provided by Yahoo! Developer Network[2], we get a list of Yahoo questions which are also ranked according to their relevance to the query. For each question, Yahoo! Answers returns various forms of information, e.g., *subject* (a brief statement of the question), *content* (a more detailed statement of the question), *chosen-answer* (the best answer chosen by users), etc.

Yahoo-broad. The broad-mining strategy is adopted for expanding questions with the retrieved Yahoo questions. The N most relevant questions in the ranking list are selected and their subjects form the set YS. We only use the subjects in order to cover more Yahoo questions. Then, all the non-stop words in YS are used as the expansion set for the opinion questions. Similarly, duplicate words are considered to be a single one.

Yahoo-deep. Also, we propose the Yahoo-deep method to mine the retrieved Yahoo questions. In this method, we only focus on the most relevant question retrieved from Yahoo! Answers, e.g., "Why dont the US government ratify Kyoto protocol?" in Fig. 2. The subject, content and chosen-answer of this

[2] http://developer.yahoo.com/answers/

Fig. 2. An example of Yahoo-deep

most-relevant question are concatenated together as the expansion of this question. All the non-stop words in the concatenation are extracted to form the expansion set for the opinion questions. In this method, by exploiting more details about the most-relevant Yahoo question, we mine the Yahoo! Answers archive at the deeper level. Figure 2 shows an example for the process of Yahoo-deep.

With each of the above methods, we get the expansion set for the opinion questions. Note that the questions under one topic have the same expansion set.

4 Experiments

4.1 Experimental Setup

In [4], the experimental results show that the Opinion PageRank model outperforms all the systems in TAC 2008. Therefore, Opinion PageRank is currently one of the most effective methods for Opinion QA. In our experiments, we use Opinion PageRank as the Opinion QA method.

The MPQA dataset [1] is used as the benchmark in this study. MPQA contains 15 opinion questions and has been widely used in Opinion QA research. We adopt the evaluation metrics used in the TAC Opinion QA track [3]. For each opinion question in MPQA, annotators have given a list of answer segments. Each segment is assigned a confidence value which shows its relative importance. The *Recall* of the answers is calculated as $Recall = r/R$, where r is the sum of the confidence of segments in the answers, and R is the sum of the confidence of segments for the question. The *Precision* of the answers is calculated as

$$Precision = 1 - ((l - A)/l). \tag{1}$$

where l is the number of non-whitespace characters in the answers, A is the allowance of the answers, and $A = 100 * a$ (a is the number of segments in the answers). The final *F-score* is calculated with the TAC official value β=3 [3], which means Recall is three times as important as Precision.

$$F - score = \left(\beta^2 + 1\right) * Recall * Precision / \left(\beta^2 * Precision + Recall\right) \quad \beta = 3. \tag{2}$$

The overall Recall, Precision and F-score are the average of their corresponding values over the 15 questions.

4.2 Performance Evaluation

We expand each opinion question in MPQA with Wiki-broad, Wiki-deep, Yahoo-broad and Yahoo-deep respectively. For each method, the expanded questions are "answered" by Opinion PageRank and the retrieved answers are evaluated. We take the *no-expansion* method, in which the original questions are inputted into Opinion PageRank without any expansion, as our baseline.

The parameters are set in the following way. The parameter K, which denotes the number of selected topic words, is set to 3. The parameters M (Wiki-broad) and N (Yahoo-broad) are both set to 10. Figure 3 gives the results for the various methods. In the figure, we can see that all of the four expansion methods outperform the no-expansion method which adopts no question expansion operations. When using the Wiki-deep method, Opinion PageRank performs best (F-score: 0.1872) and achieves around 12.5% improvements over no-expansion. This demonstrates that our methods indeed take effects in improving the performance of Opinion Question Answering. Also, we notice from the figure that Wiki-deep is the most effective expansion method when considering F-score. From Section 3.2, we know that when using Wiki-deep, we put more emphasis on links between articles than textual contents. These links are able to represent the relation between concepts at the semantic level. Therefore, a question expanded by Wiki-deep can embody the information needs of the original question more accurately and comprehensively.

Another observation from Fig. 3 is that expansion with Wikipedia (Wiki-broad and Wiki-deep) obtains higher F-score than expansion with Yahoo! Answers (Yahoo-broad and Yahoo-deep). This is partly because the contents in Yahoo! Answers are generated by users in a free way. Therefore, the expansion sets generated

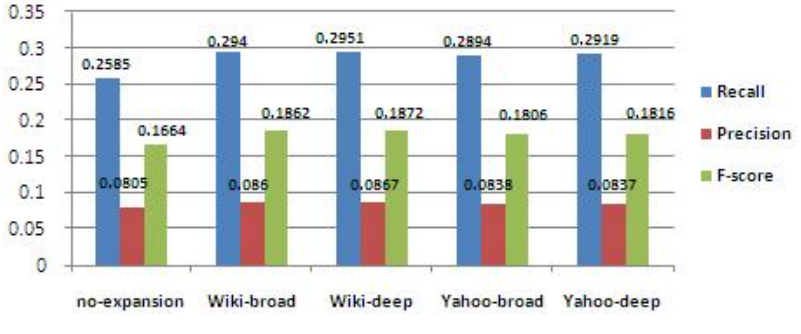

Fig. 3. Performance comparison among the methods

Table 1. P-values in t-tests

Expansion Methods	P-value
Wiki-broad	0.048
Wiki-deep	0.039
Yahoo-broad	0.118
Yahoo-deep	0.086

Table 2. F-score for the combination models

Combination Methods	F-score
Wiki-deep + Yahoo-broad	0.1872
Wiki-deep + Yahoo-deep	0.1872
Wiki-broad + Yahoo-broad	0.1862
Wiki-broad + Yahoo-deep	0.1862

with Yahoo-broad and Yahoo-deep contain noisy words (e.g., "guys", "because", etc.), which contribute little to enriching the representation of the questions. On the contrary, Wikipedia articles are created and edited under strict guidelines, and thus the expansion sets are relatively "purer".

To determine whether these improvements are statistically significant, we perform several single-tailed t-tests. Table 1 shows the P-values of various methods compared with the no-expansion baseline on the F-score metric. Wiki-deep achieves the most significant improvements (the lowest P-value) over no-expansion. Both Wiki-deep and Wiki-broad perform significantly better than the baseline at a 95% confidence level, while the improvements of Yahoo-broad and Yahoo-deep are not significant.

In the above evaluations, we consider the four methods separately. Next, we will investigate whether combining these methods can achieve better results. When combining two methods, we simply merge the expansion sets of the two methods together and get the new set. Table 2 shows the F-score values for the combination methods. From the table, we can see that these four combination methods fail to outperform the best non-combination method, i.e., Wiki-deep. Each combination method achieves the same performance as its corresponding Wikipedia method, e.g., Wiki-deep+Yahoo-broad has the same F-score as Wiki-deep. Moreover, each combination method performs better than its corresponding Yahoo! Answers method. This further proves that expansion with Wikipedia is more effective than that with Yahoo! Answers for opinion questions.

5 Conclusion and Future Work

In this paper, we propose various methods to exploit Wikipedia and Yahoo! Answers for enriching question representation in Opinion QA. The experimental results show that these methods boost the performance of Opinion QA to a

great extent. Also, performance comparison reveals that Wiki-deep is the most effective expansion method.

In our future work, we will consider applying our methods to other types of questions. Also, we will investigate other forms of information, such as Outlines and Infobox in Wikipedia, to enrich sentences in Question Answering.

Acknowledgments. This work was supported by National Natural Science Funding of China under Grant No. 90718022.

References

1. Stoyanov, V., Cardie, C., Wiebe, J.: Multi-perspective Question Answering using the OpQA Corpus. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, pp. 923–930 (2005)
2. Kim, S., Hovy, E.: Identifying Opinion Holders for Question Answering in Opinion Texts. In: Proceedings of AAAI Workshop on Question Answering in Restricted Domains (2005)
3. Dang, H.T.: Overview of the TAC 2008: Opinion Question Answering and Summarization Tasks. In: Proceeding of Text Analysis Conference (2008)
4. Li, F., Tang, Y., Huang, M., Zhu, X.: Answering Opinion Questions with Random Walks on Graphs. In: Proceedings of the 47th Annual Meeting of the Association of Computational Linguistics, pp. 733–745 (2009)
5. Gabrilovich, E., Markovitch, S.: Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge. In: Proceedings of the 21st National Conference on Artificial Intelligence, pp. 1301–1306 (2006)
6. Banerjee, S., Ramanathan, K., Gupta, A.: Clustering Short Texts using Wikipedia. In: Proceedings of the 30th ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 787–788 (2007)
7. Ye, S., Chua, T., Lu, J.: Summarizing Definition from Wikipedia. In: Proceedings of the 47th Annual Meeting of the Association of Computational Linguistics, pp. 199–207 (2009)
8. Wang, X., Tu, X., Feng, D., Zhang, L.: Ranking Community Answers by Modeling Question-Answer Relationships via Analogical Reasoning. In: Proceedings of the 32th ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 179–186 (2009)
9. Wang, K., Ming, Z., Chua, T.: A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services. In: Proceedings of the 32th ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 187–194 (2009)
10. Hu, X., Zhang, X., Lu, C., Park, E.K., Zhou, X.: Exploiting Wikipedia as External Knowledge for Document Clustering. In: Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 389–396 (2009)