Answer Generating Methods for Community Question and Answering Portals

Haoxiong Tao[1], Yu Hao[2], and Xiaoyan Zhu[3]

Department of Computer Science and Technology, Tsinghua University, China
taohaoxiong@gmail.com,
haoyu@mail.tsinghua.edu.cn,
zxy-dcs@tsinghua.edu.cn

Abstract. Community question answering (cQA) portals have accumulated numerous questions and their answers over time. Community users can search questions in cQA portals, but the returning answers often contain information which is redundant or irrelevant to the questions. Relying on the similar questions and their answers from the cQA portals, we propose appropriate answer generating methods for List-type and Solution-type questions (almost half of all questions). The results show that the answer generating methods can improve the answer quality significantly.

Keywords: community question answering, answer generating.

1 Introduction and Related Works

Nowadays, online community question answering (cQA) portals have become a popular way to acquire information. The workflow of online cQA portals is as following. The asker firstly posts a question in the cQA portals, and then other users can answer this question. When there are some answers to the question, the asker can select one answer as "Best Answer" from all the answers. In some cQA portals, the "Best Answer" also can be voted by other users. In recent years, questions and their corresponding answers of cQA portals have become an online knowledge base. Two of the main Chinese online cQA portals are Soso Wenwen (http://wenwen.soso.com) and Baidu Zhidao (http://zhidao.baidu.com). By May 2012, there have been more than 200 million solved questions on Soso Wenwen.

The online cQA portals have following limitations. Firstly, because of the workflow mentioned above, the asker can't get answers in real-time. According to Yang Tang's statistics on Baidu Zhidao[1], it takes about 14 hours for the asker to get the first answer in average. Secondly, the answers are provided by users on Internet. Due to the limitations of single user's knowledge, the quality of many answers is not high, some answers are even wrong. According to the analysis of English cQA portals by Liu et al. [2], about 25% "Best Answers" are not the best among all answers and at least 52% "Best Answers" are not the unique best answers.

Many cQA portals also support search service like search engine to overcome the unreal-time limitation. Users can search queries in cQA, and cQA will return some

M. Zhou et al. (Eds.): NLPCC 2012, CCIS 333, pp. 249–259, 2012.
© Springer-Verlag Berlin Heidelberg 2012

similar questions and their links to users. But it also has two limitations. Firstly, after searching the queries, users need to click the similar questions' links in order to see the whole answers. Secondly, users always have to spend long time to find useful information because similar questions' answers always contain information which is redundant or irrelevant to the queries.

In order to return high-quality answers to users, previous research mainly focused on two aspects: (1) Trying to predict the quality of cQA answers, and then return the answers with high predicted quality to users [3-4]. To achieve good prediction performance, user profile information is generally needed. (2) Using multi-document summarization (MDS) techniques to summarize answers from different similar questions, and then return the summarized answer to users [5]. Answer generated by MDS techniques is always more comprehensive, but also is less readable.

To improve the answer quality, almost all well-perform systems introduce a question taxonomy [6-9]. The question taxonomy proposed by Fan Bu et al. [10] contains six question types, i.e., List-type, Solution-type, Fact-type, Definition-type, Navigation-type and Reason-type. For List-type questions, each answer will be a single phrase or a list of phrases. For example, "Idioms containing the word horse?" and "List Nobel prize winners in 1990s." are both List-type questions. For Solution-type questions, people ask these questions for solutions, so the sentences in an answer usually have logical orders. For example, "How to treat chronic pharyngitis?" and "How to make pizzas?" are both Solution-type questions. According to the statistics of Fan Bu et al. [10] of the questions of Baidu Zhidao, there are 23.8% questions are List-type questions and 19.7% questions are Solution-type questions. These two types of questions almost consist half of all questions.

Relying on the similar questions and their answers from the cQA portals, we propose appropriate answer generating methods for both List-type and Solution-type questions to generate high quality answers for users. The research framework of this paper is shown in Fig 1, and this paper's work is mainly focusing on the "Answer Generating Module".

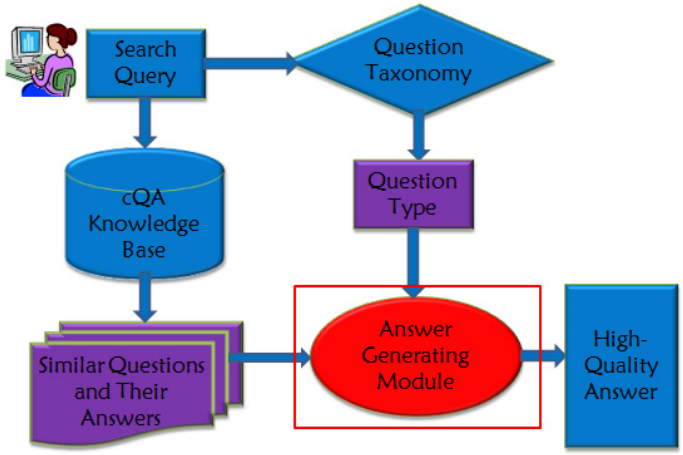

Fig. 1. Research Framework

The remainder of this paper is organized as follows: Section 2 introduces a clustering based answer generating method for List-type questions. Section 3 introduces a visible list based answer generating method for Solution-type questions. In the last section we conclude this work.

2 List-Type Questions

2.1 Answer Generating Method

According to the definition of List-type questions, each answer will be a single phrase or a list of phrases [10]. Table 1 shows an example of List-type question, "Idioms containing the word horse?" Many answers are a single phrase or a list of phrases. For example, the answer "Sensual pleasures: Describe the life is very extravagant" in the "Other Answers" is a single phrase which can answer the question. In this paper, such single phrase which can answer the List-type question is denoted as answer point.

By analysis of the answers of List-type questions, we find there are two characteristics as follows: (1) "Best Answer" often don't contain all answer points, which means "Other Answers" contain some additional answer points which are not in "Best Answer". But it takes long time for users to find the additional answer points in "Other Answers" because "Other Answers" often contain lot of information, among which some is redundant or irrelevant to the question. Take the question in Table 1 as an example, in the "Other Answers", "Sensual pleasures" is redundant because we have "Sensual pleasures : Describe the life is very extravagant" already. (2) Answer points which are high-quality or relevant to the question are often appear in more than one answers.

Base on the above two characteristics, we propose an answer generating method for List-type questions which is based on the clustering of answer points. Firstly, each answer is split into one or several answer points. Secondly, similar answer points are clustered into one category. Then for each category, a representative answer point is selected from each category for output. The answer generating method is shown in Table 2.

2.2 Method Result and Analysis

For the question shown in Table 1, the method shown in Table 2 is used to generate answer. The method output is shown in Table 3. The number at the end of each output means the number of answer points the corresponding category has after clustering. Take the first output as an example, its corresponding category has 3 answer points, two same answer points "Sensual pleasures: Describe the life is very extravagant" and another answer point "Sensual pleasures" with the same idiom.

From the above example, we find that compared to "Best Answer", the answer generated by the method in Table 3 contains more answer points, or more information. Furthermore, the outputs are ranked by answer point number each category has. The top ranking outputs have more similar answer points, therefore with high

probability they have high quality and credibility. Because the outputs are ranked, we can also control the length of answer. Therefore, the answer generating method we proposed can improve the answer quality for List-type questions.

On the other hand, we also want to point out two points for the above answer generating method, for which further research is needed. (1) For the step 1 shown in Table 2, each answer should be split into answer points at first. For this paper, the answer is

Table 1. Example of List-type Questions

Question：含有马字的成语(Idioms containing the word horse?)[1]
Best Answer：
马不停蹄，马齿徒增，马到功成，马翻人仰，马革裹尸，马工枚速，马首是瞻
Other Answers[2]：
【一马当先】作战或做事时，不畏艰难，勇敢地走在他人前面。
【千军万马】形容士兵众多，声势壮大。
【天马行空】喻才思豪放飘逸。
【声色犬马】形容生活非常糜烂。(Sensual pleasures: Describe the life is very extravagant)
声色犬马(Sensual pleasures)
一马当先
千军万马
寒蝉仗马
声色犬马 声：歌舞；色：女色；犬：养狗；马：骑马。形容剥削阶级荒淫无耻的生活方式。
走马观花
走马：骑着马跑。骑在奔跑的马上看花。原形容事情如意，心境愉快。后多指大略地观察一下。
马放南山：比喻不再作战，天下太平。
马马虎虎：草草率率，随随便便。形容办事草率。
……

[1] People ask this Chinese question for a list of idioms which containing the Chinese character "马" (means "horse" in English). Among the "Best Answer" and "Other Answers", we can see almost each phrase contains the Chinese character "马". If translate the Chinese answers into English, the answers will not contain character "马" or "horse" any more, which will make this example confusing. So we just list the Chinese answers in Table 1.

[2] Because there are too many answers to this question, we only list some representative answers instead of all answers. The URL of this question is http://zhidao.baidu.com/question/100494388.html

Table 2. Answer Generating Method for List-type Questions

Question q is a List-type Question, its similar questions have n answers, denoted as $A_1, A_2, ..., A_n$.

1. Split each answer A_i into a set of answer points $S^{(i)} = \{s_1^{(i)}, s_2^{(i)}, ..., s_{n_i}^{(i)}\}$ by line or by sentence. Initially, each answer point $s_j^{(i)}$ is an independent category, denoted as $t_j^{(i)} = \{s_j^{(i)}\}$. The set of all categories is denoted as T.
2. The similarity between two categories t_1 and t_2 is defined as:

$$\text{sim}(t_1, t_2) = \max_{s_i \in t_1, s_j \in t_2} \text{sim}(s_i, s_j) \quad (1)$$

in which $sim(s_i, s_j)$ is defined as the cosine similarity of s_i and s_j based on words.

3. Set a threshold τ. If there exist two categories in T with the similarity larger than τ, then cluster these two categories. Repeat this process until no categories can be clustered.

4. After step 3, suppose there are k categories in T, i.e., $T = \{t_1, t_2, ..., t_k\}$. Then for each category t_i, select an representative answer point s_i, so that:

$$s_i = \max_{s_j}(\sum_{s \in t_i} sim(s, s_j)) \quad (2)$$

5. For each category t_i, the output is s_i and m_i, in which m_i is the number of answer points t_i has. Rank all the outputs by the answer point number from largest to smallest.

split by line or sentence, which works for most situations. But in some situations, the answer is not well formatted, which makes it hard to split answer into answer points. (2) For the step 3 shown in Table 2, the threshold is set to be a constant value by author for this paper. But different List-type questions vary a lot, so another research work is to find appropriate threshold for different List-type questions.

3 Solution-Type Questions

3.1 Visible List

Solution-type questions concentrate on solution, and the sentences in answers always have logical orders. The answer generating method based on clustering for List-type questions is not suitable for Solution-type questions, because the logical orders between sentences will be disrupted. Usually, the answer for a Solution-type question

Table 3. Example of the Method Result

Question：含有马字的成语(Idioms containing the word horse?)
Best Answer：
马不停蹄，马齿徒增，马到功成，马翻人仰，马革裹尸，马工枚速，马首是瞻
Method Output[3]：
【声色犬马】形容生活非常糜烂(Sensual pleasures: Describe the life is very extravagant)(3)
兵强马壮 形容军队实力强，富有战斗力(3)
【驷马难追】话一说出口，难以收回(3)
走马观花（2）
后多指大略地观察一下(2)
【原班人马】同一批人员(2)
……

describes a solution to the question, so it often contains visible list. Visible list is a list explicitly indicates the sequence of all steps using mark numbers. Table 4 gives an example question from Baidu Zhidao, its answer contains several visible lists.

We choose 1179 solved Solution-type questions from Baidu Zhidao, and acquire similar questions and their answers from cQA. As the questions are solved, there is "Best Answer" for each question. According to the characteristics of visible list, we use mark number matching method to extract visible lists from all answers. Among all questions, there are 358(30%) questions' answers having visible lists, and the average length of their "Best Answer" is above 1400 words, which is definitely too long. In comparison, the average length of visible lists is around 600 words, which is much more concise. Take the question in Table 4 as an example. Compared to the "Best Answer", which contains more than 6500 words, List 3, List 4, List 5 can give a more brief answer to the question.

Table 4 shows that there could be more than one visible list in the answer. From the example, we can find some visible lists are good answers to the question, such as List 3, List 4, and List 5, while some are not good answers to the question, such as List 1 and List 2. The remainder of this section is organized as follows. We will propose a method to select the best list from all the visible lists in the second subsection. In the third subsection, we will evaluate this method and the quality of its result.

3.2 Select the Best List

In our dataset, there are 196 out of 358 (55%) questions have more than one visible list in their similar questions' answers. In this subsection, we will concentrate on how

[3] Here only list some top outputs, because the whole method output is too long.

Table 4. Example of Solution-type Question with Visible lists in Its Answers

Question: 慢性咽炎怎么治疗? (How to treat chronic pharyngitis?) [4]
Visible lists in answers:
List 1: Diagnostic criteria:
(1) Illness history: Repeated acute pharyngitis episode of excessive due to the long-term nasal to mouth breathing, alcohol and tobacco...
(2) Symptom: Throat discomfort or pain, or itching, dryness, burning, smoked a sense, foreign body sensation...
(3) Checking: Chronic throat congestion...**List 2: Traditional Chinese Medicine cluster this disease into three types:**
1. Yin Huo-yen type: Throat discomfort, pain, potential faint...
2. Phlegm and blood stasis type: Throat dryness, pain was stinging...
3. Yin-chun and dry type: Throat very itchy, burning, dryness and pain...
List 3: Treatment guide
1、The elimination of all risk factors...
2、Take different approaches to treatment based on different types of pharyngitis...
3、Not to make use of the Panda Hai blunt...
List 4: Three clever methods to treat chronic pharyngitis:
1、Massage: Thumb with the food, the middle finger to rub the sides of the throat 20 to 30 times...
2、Moxibustion: Mild moxibustion or acupressure, every 5 to 20 minutes...
3、Pricking blood therapy: Take the hard of hearing the upper vein...
List 5: Chronic pharyngitis diet modulating
1. Eat foods rich in collagen and elastin...
2. Intake of foods rich in B vitamins...
3. Eat less or not eat fried, spicy, spicy food...

[4] The answers to this question are too long, so we only list some visible lists by omitting detailed information.

to select the best list, namely, the most relevant list to the question. Firstly, we choose five features for every visible list as follows:

- First list

If the list is the first list of the answer, then this feature value is 1, otherwise its value is 0. Denote this feature as *FirstList*.

- The similarity between *guide words* and question

The visible list often contains *guide words*. In the example of Table 4, the first several words of List 1, "Diagnostic criteria", are the *guide words* of List 1. The *guide words* usually summarize the list, so the similarity between *guide words* and question can be used to help evaluate the relevant relationship between visible list and question. We calculate the cosine distance between the *guide words* and the words in question title as a feature, denoted as *GuideSimilarity*. For visible lists without *guide words*, *GuideSimilarity* is set to be a default value.

- Similarity between list content and question

Similar to *GuideSimilarity*, the cosine similarity between list content and question is also used as a feature, denoted as *ContentSimilarity*.

- The ratio of verbs and prepositions in list

The answer to a Solution-type question often gives a solution, so high-quality visible lists often contain much more verbs and prepositions. The word ratio of verbs and prepositions in the content of the list is used as another feature, denoted as *VpRatio*.

- Documents summarization based feature

In the first section, we mentioned that multi-document summarization (MDS) techniques can be used to summarize answers. For Solution-type questions, it's bad to return the summarized answer to users directly, because the logical orders between answers will be disrupted. But the summarized answer often has high information coverage, so we can use it to evaluate the information coverage of visible lists. In detail, suppose the summarized answer contains N sentences, for every visible list, if it contains k sentences out of the N sentences, then it will have a coverage score of k/N. This coverage score is used as a feature, which is denoted as *SummaryScore*.

For the above 5 features, each one is a [0, 1] value. For the 196 questions which had more than one visible lists, we manually label a score to all visible lists (the labeling standard will be introduced in the next subsection), and use them as the training set. For data training, we use Learning to Rank model to get the weight of every feature, and then select the visible list with the highest score as the best list. In this paper, we use the pairwise Ranking Perception model.

3.3 Experiment and Analysis

As mentioned above, we choose 1179 questions from Baidu Zhidao, and extract the visible lists from their similar questions' answers. There are 358 (30%) questions with visible lists in their similar questions' answers, and 196 (55%) of them with more than one visible list, and other 162 (45%) questions with only one visible list. This subsection can be divided into two parts: the first one part is to evaluate the method of selecting the best list; the second part is to evaluate the quality of best list as the answer.

To evaluate the method of selecting best list from all visible lists, at first we manually label a score to the 196 questions with more than one visible list, score 1 for high quality and score 0 for low quality. High quality means the visible list is relevant to the question, is complete and can answer the question, while low quality means the visible list is not relevant to the question, or is not complete enough to answer the question. After labeling all visible lists, there are 69 questions whose visible lists have the same score. As the goal of our experiment is to evaluate the method of selecting best list, these 69 questions have no meaning for this experiment, so we remove them and use remaining 127 questions for the experiment.

For the remaining 127 questions, if we randomly select a visible list for each question, the probability that the list is high-quality is 51.7%. If we always select the first list, the probability will increase to 63.8%. Use the method mentioned in the previous subsection, we combine different features to select best list. The result is as follows:

Table 5. Result of Selected Visible-lists

Features used to select	High quality probability
All	76.4%
All-*VpRatio*	76.4%(-0.0%)
All-*SummaryScore*	75.6%(-0.8%)
All-*ContentSimilarity*	75.6%(-0.8%)
All-*FirstList*	74.0%(-2.4%)
All-*GuideSimilarity*	69.3%(-7.1%)

Table 5 shows, if we use all features, the probability to select a high-quality list is up to 76.4%, much higher than the method of random selection (51.7%) and selecting first list as the best list (63.8%). The most obvious decrease of high quality probability occurs when we delete *GuideSimilarity* from all features. This indicates that *GuideSimilarity* is a very important feature to select high-quality list.

Up to now, for 162 questions with only one visible list, the only one list could be the answer; for other 192 questions with more than one visible list, the selected best list could also be the answer. In order to evaluate the quality of using visible list as answer, we manually compare the quality of "Best Answer" and visible list for each

question. We mainly focus on the relevance to question, completeness and whether containing redundant information. There are three cases, i.e., visible list is better than "Best Answer", "Best Answer" is better, or they are around the same. The evaluation result is as shown in Table 6:

Table 6. Evaluation Result

Result	Number of Question	Ratio
Visible list is better	91	25.4%
"Best Answer" is better	74	20.7%
Around the same	193	53.9%

From Table 6, we can conclude that using visible list as the answer is better than "Best Answer" on the whole. On the other side, for the questions above, the average length of visible list is 600 words, while the average length is more than 1400 words for "Best Answer", which is more than twice of visible list. Therefore, for Solution-type question, if the similar questions' answers contain visible lists, using the method we proposed to select visible list as the answer, can improve the quality of answer significantly.

4 Conclusion and Future Work

The research on answer generating methods for cQA portals is very important and meaningful. In this paper, relying on the similar questions and their answers from the cQA portals, we have proposed appropriate answer generating methods for List-type and Solution-type questions, which two types consists of almost half of all questions. For List-type questions, the answer generating method is based on the clustering of answer points. For Solution-type questions, the method is based on visible lists. The results show that the answer generating methods we propose can improve the answer quality significantly.

For the answer generating method for List-type questions, we plan to do further research to split the answer into answer points more robustly. For the answer generating method for Solution-type questions, we will introduce more semantic features to improve the semantic relevance between selected list and question. For other types of questions, we will also do further research to generate high-quality answers.

Acknowledgement. This work was carried out with the aid of a grant from the International Development Research Center, Ottawa, Canada, number:104519-006. This work was also supported by the Chinese Natural Science Foundation under grant No.60973104.

References

1. Yang, T.: Question Recommendation and Answer Summarization for cQA portals (Master Thesis), p. 3. DCST of Tsinghua University, Beijing (2010)
2. Liu, Y., Bian, J., Agichtein, E.: Predicting Information Seeker Satisfaction in Community Question Answering. In: SIGIR, pp. 483–490 (2008)
3. Jeon, J., Bruce Croft, W., Park, S.: A Framework to Predict the Quality of Answers with Non-Textual Features. In: SIGIR (2006)
4. Eugene, A., Carlos, C., Debora, D., et al.: Finding High-Quality Content in Social Media. In: WSDM (2008)
5. Liu, Y., Li, S.: Understanding and Summarizing Answers in Community-Based Question Answering Services. In: Proc. of ICCL (2008)
6. Jijkoun, V., Rijke, M.: Retrieving Answers from Frequently Asked Questions Pages on the Web. In: Proc. of CIKM (2005)
7. Hovy, E., Laurie, G., Ulf, H., et al.: Toward Semantics-Based Answer Pinpointing. In: Proc. of HLT (2001)
8. Moldovan, D., Harabagiu, S., et al.: The Structure and an Open-Domain Question Answering System. In: Proc. of ACL (2000)
9. Lytinen, S., Tomuro, N.: The Use of Question Types to Match Questions in FAQFinder. In: Proc. of AAAI (2002)
10. Bu, F., Zhu, X., Hao, Y., et al.: Function-based Question Classification for General QA. In: EMNLP 2010, pp. 1119–1128 (2010)