Interactive and Informative Community Question Answering with Multimedia Support

Deepak Kumar R[1],Sivaraman S[2],Arun Prasad K[3]
[1] [2] Department of Information Technology
Agni college of Technology
[3]Research Scholar, St.Peters University
[1]*deepakrdk@outlook.com*, [2]*shivram93@live.com*, [3]*arunprasadchamy@gmail.com*

International Conference on Green Technologies for Power Generation and Communications-ICGPC 2014

Abstract: Generally community us to seek information for our queries and doubts. It enables the community members to post questions and answers to that. However, existing community question answering forums usually provide only textual answers, which are not informative enough for many questions. To improve the understanding and provide additional information we propose a novel scheme which allows community members to post the multimedia content i.e., Images and Videos. Unlike several MMQA researches, this scheme deals with more complex questions rather than answering to question with image and video data.

Keywords: *Community question answering, Answer medium selection, Multimedia search, Multimedia data selection and presentation.*

INTRODUCTION

Question-answering (QA) is a technique to answer question posted in natural language automatically. In keyword-based search systems, it greatly facilitates the communication between humans and computer by naturally stating users' intention in plain sentences. It also avoids tedious browsing of a vast quantity of information contents returned by search engines for the correct answers. However, fully automated question answering still faces challenges that are not easy to tackle, such as the deep understanding of complex questions and the sophisticated syntactic, semantic and contextual processing to generate answers. It is found that, in most cases, automated approach cannot obtain results that are as good as those generated by human intelligence.

Community question answering has emerged as a popular alternative to acquire information online, owning to the following facts. First, information seekers are able to post their specific questions on any topic and obtain answers provided by other participants. By leveraging community efforts, they are able to get better answers than simply using search engines. Second, in comparison with automated question answering systems, community question answering usually receives answers with better quality as they are generated based on human intelligence. Third, over times, a tremendous number of question answering pairs have been accumulated in their repositories, and it facilitates the preservation and search of answered questions.

Despite their great success, existing community question Answering forums mostly support only textual answers, as shown in Figure 1. Unfortunately, textual answers may not provide sufficient natural and easy-to-grasp information. Figure 1 (a) and (b) illustrate two examples. For the questions "*What is Bluetooth and how does it work*" and "Do anybody know how to make pizza", the answers are described by long sentences. Clearly, it will be much better if there are some accompanying videos and images that visually demonstrate the process or the object. Therefore, the textual answers in community question answering can be significantly enhanced by adding multimedia contents, and it will provide answer seekers more comprehensive information and better experience. In fact, users usually post URLs that link to supplementary images or videos in their textual answers.

For example, for the questions in Figure 1 (c) and (d), the best answers on Y!A both contain video URLs. It further confirms that multimedia contents are useful in answering several questions. But existing community question answering forums do not provide ad- equate support in using media information. In this paper, we propose a novel scheme which can enrich community-contributed textual answers in community question answering with appropriate media data. Figure 2 shows the schematic illustration of the approach. It contains three main components: (1) Answer Medium Selection, (2) Multimedia Search, (3) Multimedia data selection and presentation.

(1) Answer medium selection. Given a question paper pair, it predicts whether the textual answer should be enriched with media information, and which kind of media data should be added. Specifically, we will categorize it into one of the four classes: text, text + image, text + video, and text + image + video. It means that the scheme will automatically collect images, videos, or the combination of images and videos to enrich the original textual answers.

(2) Multimedia search. In order to collect multimedia data, we need to generate informative queries. Given a question answering pair, this component extracts three queries. The most informative query will be selected by a three-class classification model.

(3) Multimedia data selection and presentation. Based on the generated queries, we vertically collect image and video data with multimedia search engines. We then perform reranking and duplicate removal to obtain a set of accurate and representative images or videos to enrich the textual answers. It is worth mentioning that there already exist several re- search efforts dedicated to automatically answering questions with multimedia data, i.e., the so-called Multimedia Question Answering (MMQA). For example, Yang *et al* proposed a technology that supports factoid question answer in news video. Yeh et al. presented a photo-based question answer system for finding information about physical objects. Li *et al* proposed an approach that leverages YouTube video collections as a source to automatically find videos to describe cooking techniques. But these approaches usually work on certain narrow domains and can hardly be generalized to handle questions in broad domains. This is due to the fact that, in order to accomplish automatic MMQA, we first need to understand questions, which is not an easy task.

Our proposed approach in this work does not aim to directly answer the questions, and instead, we enrich the community- contributed answers with multimedia contents. Our strategy splits the large gap between question and multimedia answer into two smaller gaps, i.e., the gap between question and textual answer and the gap between textual answer and multimedia answer. In our scheme, the first gap is bridged by the crowd-sourcing intelligence of community members, and thus we can focus on solving the second gap. Therefore, our scheme can also be viewed as an approach that accomplishes the MMQA problem by jointly exploring human and computer. Figure 3 demonstrates the difference between the conventional MMQA approaches and an MMQA framework based on our scheme. It is worth noting that, although the proposed approach is automated, we can also further involve human interactions. For example, our approach can provide a set of candidate images and videos based on textual answers, and answerers can manually choose several candidates for final presentation.

RELATED WORK

A.From Textual question answer to Multimedia question answer

The early investigation of question answering systems started from 1960s and mainly focused on expert systems in specific domains. Text-based question answering has gained its research popularity since the establishment of a QA track in TREC in the late 1990s. Based on the type of questions and expected answers, we can roughly summarize the sorts of question answering into Open- Domain question answering, Restricted-Domain question answering, Definitional question answering and List question answering. However, in spite of the achievement as described above, automatic question answering still has difficulties in answering complex questions. Along with the blooming of Web 2.0, community question answering becomes an alternative approach. It is a large and diverse question- answer forum, acting as not only a corpus for sharing technical knowledge but also a place where one can seek advice and opinions. However, nearly all of the existing community question answering systems, such as Yahoo!Answers, WikiAnswers and Ask Metafilter, only support pure text- based answers, which may not provide intuitive and sufficient information. An early system named Video question answering was presented in.

This system extends the text- based question answering technology to support factoid question answering by leveraging the visual contents of news video as well as the text transcripts. hua et al. Proposed a generalized approach to extend text-based question answering to multimedia question answering for a range of factoid, definition and "how-to" questions. Their system was designed to find multimedia answers from web-scale media resources such as Flicker and YouTube. However, literature regarding multimedia question answering is still relatively sparse. As mentioned in Section I, automatic multimedia question answering only works in specific domains and can hardly handle complex questions. Different from these works, our approach is built based on community question answering. Instead of directly collecting multimedia data for answering questions, our method only finds images and videos to enrich the textual answers provided by humans.

B. Multimedia Search

Due to the increasing amount of digital information stored over the web, searching for desired information has become an essential task. The research in this area started from the early 1980s by addressing the general problem of finding images from a fixed database. With the rapid development of content analysis technology in the 1990s, these efforts quickly expanded to tackle the video and audio retrieval problems. Generally, multimedia search efforts can be categorized into two categories: text-based search and content-based search. The text-based search approaches use textual queries, a term- based specification of the desired media entities, to search for media data by matching them with the surrounding textual descriptions. To boost the performance of text-based search, some machine learning techniques that aim to automatically annotate media entities have been proposed in the multimedia community. Further, several social media websites, such as Flickr and Facebook, have emerged to accumulate manually annotated media entities by exploring the grass root Internet users, which also facilitates the text based search. The keyword-based search engines are still widely used for media search. However, the intrinsic limitation of text-based approaches make that all the current commercial media search engines difficult to bridge the gap between textual queries and multimedia data, especially for verbose questions in natural languages.

C. Multimedia Search Reranking

As previously mentioned, current media search engines are usually built upon the text information associated with multimedia entities, such as their titles, ALT texts, and surrounding texts on web pages. But the text information usually does not accurately describe the content of the images and videos, and this fact can severely degrade search performance. Reranking is a technique that improves search relevance by mining the visual information of images and videos. Existing reranking algorithms can mainly be categorized into two approaches, one is pseudo relevance feedback and the other is graph-based reranking. The pseudo relevance feed- back approach, regards top results as relevant samples and then collects some samples that are assumed to be irrelevant. A classification or ranking model is learned based on the pseudo relevant and irrelevant samples and the model is then used to rerank the original search results. It is in contrast to relevance feedback where users explicitly provide feedback by labeling the results as relevant or irrelevant.

ANSWER MEDIUM SELECTION

The first component of our scheme is answer medium selection. It determines whether we need to and which type of medium we should add to enrich the textual answers. For some questions, such as "When did America become allies with Vietnamese", pure textual answers are sufficient. But for some other questions we need to add image or video information. For example, for the question "Who is Pittsburgh quarterback for 2008", it is better to add images to complement the textual answer, whereas we should add videos for answering the question "How to install a Damper pulley on a neon". We regard the answer medium selection as a question answering classification task. That means, given a question and textual answer, we categorize it into one of the following four classes: (a) only text, which means that the original textual answers are sufficient; (b) text + image, which means that image information needs to be added; (c) text +video, which means that only video information needs to be added; and (d) text + image + video, i.e., we add both image and video information. There are some existing research efforts on question classification. Li and Roth developed a machine learning approach that uses the SNoW learning architecture to classify questions into five coarse classes and 50 finer classes. They used lexical and syntactic features such as part-of-speech tags, chunks and head chunks together with two semantic features to represent the questions. Zhang and Lee used linear SVMs with all possible question word grams to perform question classification. Arguello et al.

Investigated medium type selection as well as search sources for a query. But there is no work on classifying question paper pairs according to the best type of answer medium. This task is more challenging as we are dealing with real data on the web, including complex and multi-sentence questions and answers, and we need to extract rules to connect QA texts and the best answer medium types. We accomplish the task with two steps. First, we analyze question, answer, and multimedia search performance. Then, we learn a linear SVM model for classification based on the results.

TABLE I

REPRESENTATIVE INTERROGATIVE WORDS.

Interrogative Word	Category
be, can, will, have, when, be there, how + adj/adv	Text
what,wher, which, why, how to, who, etc	Need further classification

TABLE II

REPRESENTATIVE CLASS-SPECIFIC RELATED WORDS.

Categories	Class-Specific Related Word List
Text	name, population, period, times, country, height, website, birthday, age, date, rate, distance, speed, religions, number, etc
Text + Image	colour, pet, clothes, look like, who, image, pictures, appearance, largest, band, photo, surface, capital, figure, what is a, symbol, whom, logo, place, etc.
Text + Video	How to, how do, how can, invented, story, film, tell, songs, music, recipe, differences, ways, steps, dance, first, said, etc
Text + Image +Video	president, king, prime minister, kill, issue, nuclear, earthquake, singer, battle, event, war, happened, etc

Class-Specific Related Word List name, population, period, times, country, height, website, birthday, age, date, rate, distance, speed, religions, number, etc, colour, pet, clothes, look like, who, image, pictures, appearance, largest, band, photo, surface, capital, figure, what is a, symbol, whom, logo, place, etc. Since many questions contain multiple sentences (actually our statistics on Y!A show that at least 1/5 of the questions contain at least two sentences, and the number is around 1/10 for Wiki Answers) and some of the sentences are uninformative, we first employ the method in to extract the core sentence from each question. The classification is accomplished with two steps. First, we categorize questions based on interrogatives (some starting words and ending words), and in this way we can directly find questions that should be answered with text. Second, for the rest questions, we perform a classification using a naive Bayes classifier.
How to, how do, how can, invented, story, film, tell, songs, music, recipe, differences, ways, steps, dance, first, said, etc. president, king, prime minister, kill, issue, nuclear, earthquake, singer, battle, event, war, happened, etc. We also extract a list of class-specific related words in a semi-automatic way. We first estimate the appearing frequency of each phrase in the positive samples of each class. All the phrases that have the frequencies above a threshold (we empirically set the threshold to 3 in this work) are collected. We then manually refine the list based on human's expert knowledge. Examples of class-specific related words for each class are shown in Table II.

B. Answer-Based Classification

Besides question, answer can also be an important information clue. For example, for the question "*how do you cook beef in gravy*", we may find a textual answer as "*cut it up, put in oven proof dish* ...” Then, we can judge that the question can be better answered with a video clip as the answer describes a dynamic process.

C. Media Resource Analysis

Even after determining an appropriate answer medium, the related resource may be limited on the web or can hardly be collected, and in this case we may need to turn to other medium types. For example, for the question "*How do I export Internet Explorer browser history*", it is intuitive that it should be answered using video content, but in fact video resources related to this topic on the web are hard to find on the current search engines.

QUERY GENERATION FOR MULTIMEDIA SEARCH

To collect relevant image and video data from the web, we need to generate appropriate queries from text QA pairs before performing search on multimedia search engines. We accomplish the task with two steps. The first step is query extraction. Textual questions and answers are usually complex sentences. But frequently search engines do not work well for queries that are long and verbose.

Therefore, we need to extract a set of informative keywords from questions and answers for querying. For each question answer pair, we generate three queries. First, we convert the question to a query, i.e., we convert a grammatically correct interrogative sentence into one of the syntactically correct declarative sentences or meaningful phrases. We employ the method in. Second, we identify several key concepts from verbose answer which will have the major impact on effectiveness. Here we employ the method in. Finally, we combine the two queries that are generated from the question and the answer respectively. Therefore, we obtain three queries, and the next step is to select one from them. The query selection is formulated as a three-class classification task, since we need to choose one from the three queries that are generated from the question, answer and the combination of question and answer.

CONCLUSION

Our proposed approach in this work does not aim to directly answer the questions, and instead, we enrich the community-contributed answers with multimedia contents. Our strategy splits the large gap between question and multimedia answer into two smaller gaps, i.e., the gap between question and Textual answer and the gap between textual answer and multimedia answer. In our scheme, the first gap is bridged by the crowd-sourcing intelligence of community members, and thus we can focus on solving the second gap. Therefore, our scheme can also be viewed as an approach that accomplishes the MMQA problem by jointly exploring human and computer the difference between the conventional MMQA approach frameworks based on our scheme. It is worth Noting that, although the proposed approach is automated, we canal so further involve human interactions .For example, our approach can provide a set of candidate images and videos based on textual answers, and answerers can manually choose several candidates for final presentation.

REFERENCES

[1] D. Mollá and J. L. Vicedo, “Question answering in restricted domains: An overview,” Computat. Linguist., vol. 13, no. 1, pp. 41–61, 2007.

[2] Haubold, A. Natsev, A. ; Naphade, M.R., “ Semantic Mutimedia Retrival using Lexical Query Expansion and Model - Based Reranking” Multimedia and Expo, 2006 IEEE International Conference on Multimedia,9-12 July 2006,1761 - 1764.

[3] M. Wang, K. Yang, X.-S. Hua, and H.-J. Zhang, “Towards a relevant and diverse search of social images,” IEEE Trans. Multimedia, vol. 12, no. 8, pp. 829–842, 2010.

[4] S. A. Quarteroni and S. Manandhar, “Designing an interactive open domain question answering system,” J. Natural Lang. Eng., vol. 15, no. 1, pp. 73–95, 2008.

[5] Y.-C. Wu and J.-C. Yang, “A robust passage retrieval algorithm for video question answering,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 10, pp. 1411–1421, 2008.

[6] Y.-C. Wu and J.-C. Yang, “A robust passage retrieval algorithm for video question answering,” IEEE Trans. Circuits Syst. Video Technol.,vol. 18, no. 10, pp. 1411–1421, 2008.

[7]Y.-C.Wu, C.-H. Chang, and Y.-S. Lee, “Cross-Language Video Question/Answering System,” in Proc. IEEE Int. Symp. Multimedia Software Engineering, 2004, pp. 294–301.

[8] Z.-J. Zha, X.-S. Hua, T. Mei, J. Wang, G.-J. Qi, and Z. Wang, “Joint multi-label multi-instance learning for image classification,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008, pp. 1–8

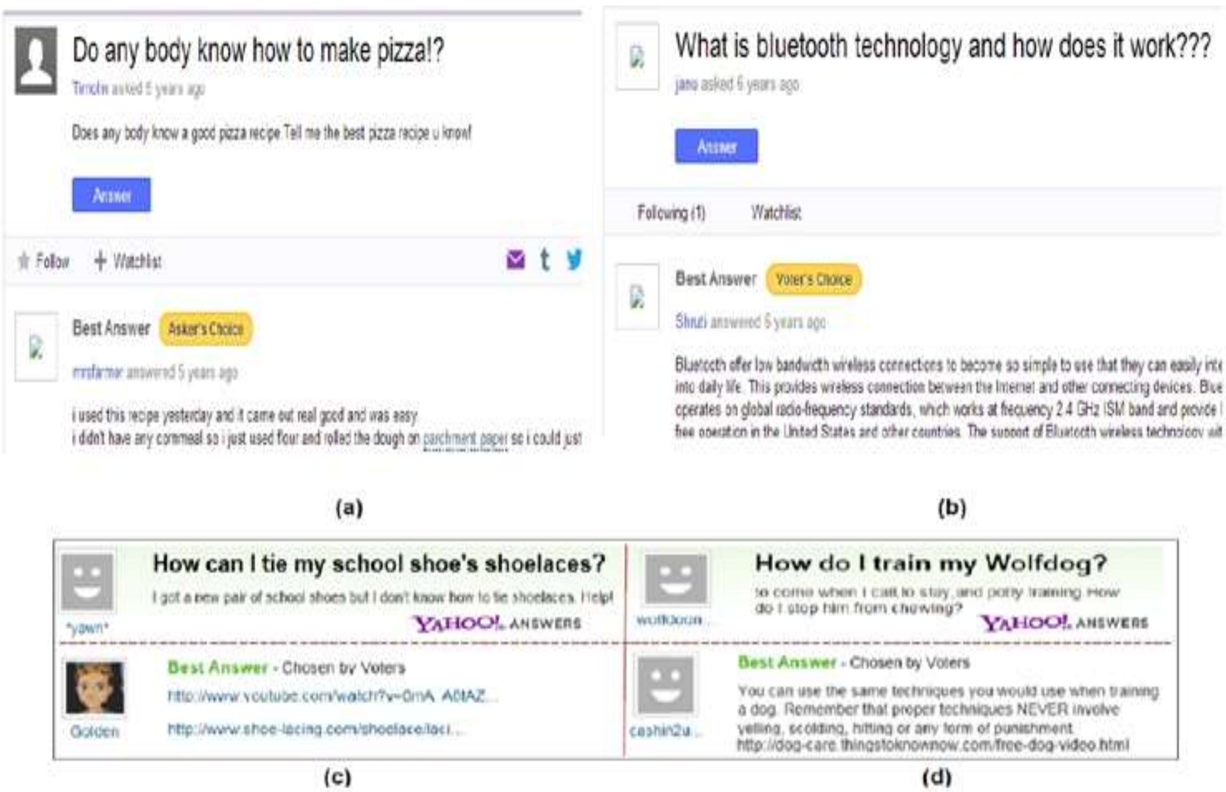

Fig 1. Examples of QA pairs from several popular cQA forums. (a) An example from Yahoo!Answer; (b) an example from Yahoo!Answer; (c) an example from Yahoo! Answer that only contains links to two videos; and (d) another example from Yahoo!Answer.

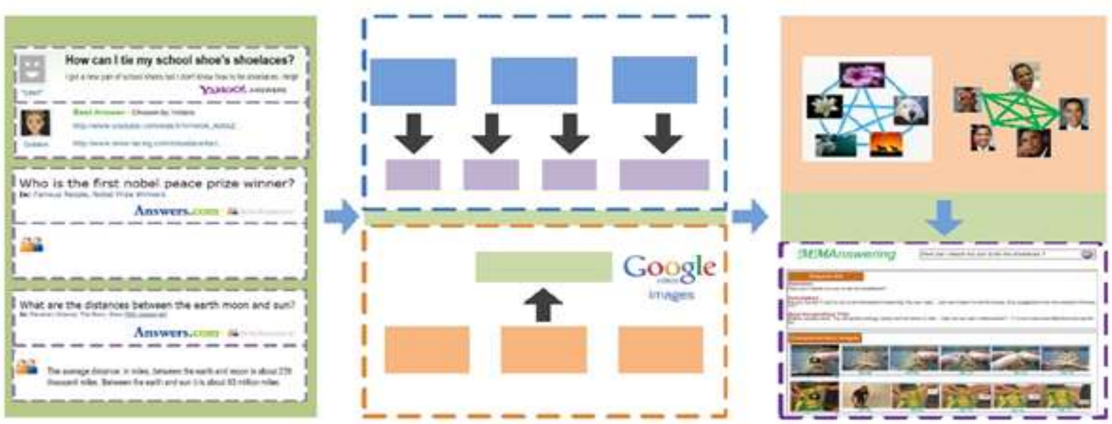

Fig. 2. The schematic illustration of the proposed multimedia answering scheme. The scheme mainly contains three components, i.e., answer medium selection, query generation, and data selection and presentation.