ISSN: 2321-7782 (Online)
Volume 3, Issue 3, March 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Multimedia Answer Generation for Community Question Answering

Pooja Karanjawane[1]
Department of IT,
RMDSSOE,
Pune, India

Sanket Karanjule[2]
Department of IT,
RMDSSOE,
Pune, India

Karishma Shaikh[3]
Department of IT,
RMDSSOE,
Pune, India

Pooja Sharma[4]
Department of IT,
RMDSSOE,
Pune, India

Abstract: Community question answering (cQA) have become popular over the past few years. Using (cQA) one can post questions and answers related to that questions and information seekers are able to find information related to the questions. Existing cQA forums usually provide only text format answers, which are not much informative enough for some questions. In this paper, we propose a idea that is able to enrich text format answers in cQA with appropriate multimedia data .Our approach mainly consist of three main component: answer medium selection, multimedia data selection and presentation and query generation for multimedia search. This approach automatically determines corresponding multimedia data to be added with textual answer. It then collects data automatically from the web to support the text based answer. After this large QA datasets are processed and then added to pool. Different from a lot of MMQA to provide directly answer question with image and video data our approach can deal up with various complex question data sets.

Keywords: QA, medium selection, question answering, re-ranking

I. INTRODUCTION

Question Answering:(QA) is a computer science which includes the fields of information retrieval and natural language processing (NLP), which deals with the task of building systems that automatically provides precise answer questions posed by humans in a natural language. A QA implementation is usually a computer program that may construct its answers by querying a structured database of knowledge. In other words, MMQA systems can extract answers from an unstructured collection of documents on web. Some examples of document collections used for MMQA systems include: Local collection of reference texts from web, internal organization documents and web pages, Compiled newswire reports, a set of Wikipedia pages, a subset of World Wide Web pages. MMQA research attempts to deal with mainly two types of questions which include: Closed-domain Questions: This type of question answering includes questions under a specific domain (for example; medicine and automotive maintenance), and can be seen as an easier task. Reason behind this is Natural Language Processing systems can exploit domain-specific knowledge frequently formalized in ontology. Alternatively, closed-domain might refer to a situation where only a limited type of questions are accepted, such as questions asking for descriptive rather than procedural information. Open-domain Question: This type of question answering includes questions about nearly anything, and can only rely on general ontology and world knowledge. On the other hand, these systems usually have much more data available from which to extract the answer.

II. RELATED WORK

The early QA system started in 1960s which mainly focused on specific domain. Based on question and answers we can summaries QA pair in open domain. Text based QA system started in early 1990s developed by TREC. cQA came into approach along with web 2.0. Currently all the cQA such as WikiAnswers, Askmeta filter and yahoo answers support only text based answers, which may not provide sufficient information to user. In 2003 videoQA system came into existence which supported the text based QA system. After 2004 using video QA system optical character reorganization and automatic speech reorganization system was formed up by Y-C-Wu and Y-S-lee. Regarding Multimedia QA was very sparse and was restricted up to open domain.

The amount of digital information present over the web is very large & searching information related to specific Question become and essential task. The research in this field started in early 1980's. In early 1990's there was rapid development in contain analysis technology hence it become complex to tackle the audio and video retrieval problems. Multimedia search can be categorized in two parts content based search and text based search. Content based retrieval system has several limitations like difficulty in finding visual query, high computational cost. Hence keyword based search engine can be used for media search. Currently media search engines are built upon text associated information. Hence we get many irrelevant results so for this we use Re-ranking Technique which improve search relevancy.

III. ANSWER MEDIUM SELECTION

This is the first component of our scheme which determines whether we need to and which type of medium we should add to enrich the textual answers extracted. Some questions, such as "When did second world war begin", pure textual answers are enough. Whereas, questions like "Who is president of India", is better to add images for answering the question. Questions like "how to cook beef?" need a video answer.

Answer medium selection is a QA classification task, where given a question and textual answer, we categorize it into one of the following four classes:

(a) Only Text, where original textual answers are sufficient;

(b) Text + Image, where image information needs to be added;

(c) Text + Video, which means that only video information needs to be added;

(d) Text + Image + Video we add both image and video information.

The task of best type of answer medium selection is more challenging as we are dealing with real data on the web, including complex and multi-sentence questions and answers. To extract rules for connecting QA texts and the best answer medium types. This can be done with two steps. First is to analyse question, answer, and multimedia search performance and second is to learn a linear SVM model for classification based on the results generated[2].Translation of a complex question into simpler questions would identify ambiguities and treat them in context or by interactive clarification in this model. Our multimedia system consists of principled solutions which have been implemented in a domain independent manner and which produce answers that are reasonably relevant, informative, and conversational in style. Such a system makes it possible to begin to study users interacting with a question-answering application.

ISSN: 2321-7782 (Online)

IV. Question Based Classification

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 WikiAnswers. We first employ the method in [3] multiple sentence questions to extract the core sentence from each question given. This classification is accomplished with two steps.

We categorize questions based on interrogatives i.e. some starting words and ending words, which can directly find questions that should be answered with text. Questions can mainly be categorized into the following classes based on interrogative words: yes/no class such as “Do we have MIS lecture tomorrow?”, Choice class such as “Which continent is bigger, Asia or Europe?”, Quantity class such as “When was first mobile phone made?”, Enumeration classes such as “Name any three cities of vidarbha?”

Description class such as “What are the ways to reduce pollution” Given a question, we first judge whether it should use only textual answer based on the interrogative word, If not, we further perform a classification with a Naive Bayes classifier For the rest questions, we perform a classification using a naive Bayes classifier.

For Naive Bayes classifier, we extract a set of text features, including bigram text features, head words, and a list of class-specific related terms. Head word is defined as the word specifying the object that a question seeks for. Head words play an major role in determining answer medium. As, in question “what year did world war begin”, the head word is year, based on which we can judge that the sought-after answer is a simple date. Therefore, it is reasonable to use textual answer medium.

V. Answer Based Classification

Like question, answer can also deliver important information. For example, for the question “how do you cook chicken”, we may find a textual answer as “cut it up, put in oven proof dish”. From this we can judge that the question can be better answered with a video clip as the answer describes a dynamic process of cooking. Answer classification deals with extraction of bigram text features and verbs from answers. Verbs in an answer will be useful for judging whether the answer can be enriched with video content or not. If a textual answer contains many complex verbs, it is more likely to describe a dynamic process and thus it has high probability to be well answered by videos. Therefore, verb can be an important. Based on the bigram text features and verbs, we also build a Naive Bayes classifier with a set of training data, and then perform a four-class classification with the model.

VI. Media Resource Analysis

Appropriate answer medium is not the conclusion, 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 mediums. For the question such as “How do I export Internet Explorer browser history”, it should be answered using video content. But in fact video resources related to this topic on the web are hard to find. Thus, here we only introduce our method for search performance prediction. Search performance can be predicted on the fact that, most frequently, search results are good if the top results are quite coherent, which defines a clarity score given by Kullback-Leibler (KL) divergence[4] which is :

$$\text{Clarity}\ (C_i) = \sum_{w \in Vcs} P(w|\theta q) \log_2 \frac{P(w|\theta q)}{P(w|\theta Ci)}$$

Here Vci is the entire vocabulary of the collection. i=1, 2, 3 represent text, image and video. The function on p in formulae is the query and collection language models. The Clarity value becomes smaller as the top ranked documents approach a random sample from the collection.

VII. Query Generation

Apart from collection of relevant image and video data from the web there is need to generate appropriate queries from text QA pairs before performing search on multimedia search engines. This can be done with two steps among which the first step is extraction of queries. Frequently search engines do not work well for queries that are long and verbose and textual question answers are usually complicated [5]. Thus need to extract a set of informative keywords from questions and answers for querying comes in picture. Second step consist of query selection. We can generate different queries: one from question, one from answer, and one from the combination of question and answer both. Most informative query depends on the QA pairs.

Three queries are generated from each QA pairs. In first step we convert the question to a query, i.e., we convert a grammatically correct interrogative sentence into one of the syntactically correct declarative sentences [6]. Whereas in second step, we identify several key concepts from verbose answer which will have the major impact on effectiveness of the sentence [7]. In the end we combine the two queries that are generated from the question and the answer respectively and thus three queries are obtained, and the next step is to select one of those three. Following features are adopted for this purpose:

POS Histogram is used as it reflects the characteristic of a query and is motivated by several observations. For the queries that contain a lot of complex verbs it will be difficult to retrieve meaningful multimedia results. So we use POS tagger to assign part-of-speech to each word of both question and answer respectively.

Secondly we use search performance prediction. The reason for this is, for certain queries, existing image and video search engines cannot return results that are expected. KL divergence is adopted here which measures a clarity score for each query. So that we can generate 6-dimensional search performance prediction features in all note that there are three queries and search is performed on both image and video search engines. Thus, for each QA pair, we can generate 42-dimensional features. Based on the extracted features, we train an SVM classifier with a labelled training set for classification, i.e., selecting one from the three queries.

VIII. Search and Presentation

Problem of the existing re-ranking methods is that they use query-independent global visual features for re-ranking method which overlooks the fact that many queries are actually person-related. Our system suggests using facial features instead of global visual features for re-ranking the search results of queries that are related to person. Also our statistics show the same. In this work we suggest a query-adaptive re-ranking approach which first decides whether a query is person-related or non-person-related.

Ranking Algorithm:

The basic aim is to check whether a query is person-related or non-person related. A text-based rule, analyzing the textual QA information, which is not easy to accomplish the task by simply analyzing the textual terms in a statement. Take an example for the query BSB, it is not easy to judge that it is the abbreviation of "Backstreet Boys" which is person-related, but from the image search results, we can find that most returned images contain several faces and thus we can determine it is a person-related. We can even choose to match each query term with a person list, such as a celebrity list, but it will not be easy to find a complete list. Also it will be difficult to keep the list updated. So, we adopt a method that analyses image search result specifically for each image in the ranking list, we perform face detection and then extract 7-dimensional features, including the size of the largest face area, the number of faces, the ratio of the largest face size and the second largest face size, and the position of the largest face and average the 7-dimensional features of the top 150 search results and it forms the features for classification of query.

If a query is related to person, we perform face detection for each video key-frame and image and if key-frame or an image does not contain faces, it will be not considered in re-ranking. If faces are found in key-frames or images, we extract the 256-D Binary Pattern features from the largest faces of images or video frames.

When it comes to non-person-related queries, we extract 428-dimensional global visual features, including 225-D block-wise colour moments generated from 5-by-5 fixed partition of the image, 128-D wavelet texture, and 75-D edge direction histogram. After re-ranking, visually similar images or similar videos may be ranked together. Thus, duplicate removal step is performed to avoid information redundancy, and check the ranking list from top to bottom. If an image or video is similar or close to a sample that appears above it, we remove it. More specifically, we remove the image or video if there exists that satisfies .Here we empirically set to 0.8 throughout the work. After duplicate removal, we keep the top 10 images and top 2 videos (keeping which kind of media data depends on the classification results of answer medium selection). When presenting videos, we not only provide videos but also illustrate the key-frames to help users quickly understand the video content as well as to easily browse the videos.

IX. System Design

The very first thing to be done in our system is that you need to have the question data sets to be collected After this the very first thing is answer medium selection in this from the given QA pairs it predicts whether the textual answer is sufficient or multimedia data should be added with the textual answer. It means it will automatically collect video images or combination to enrich the original textual answer.

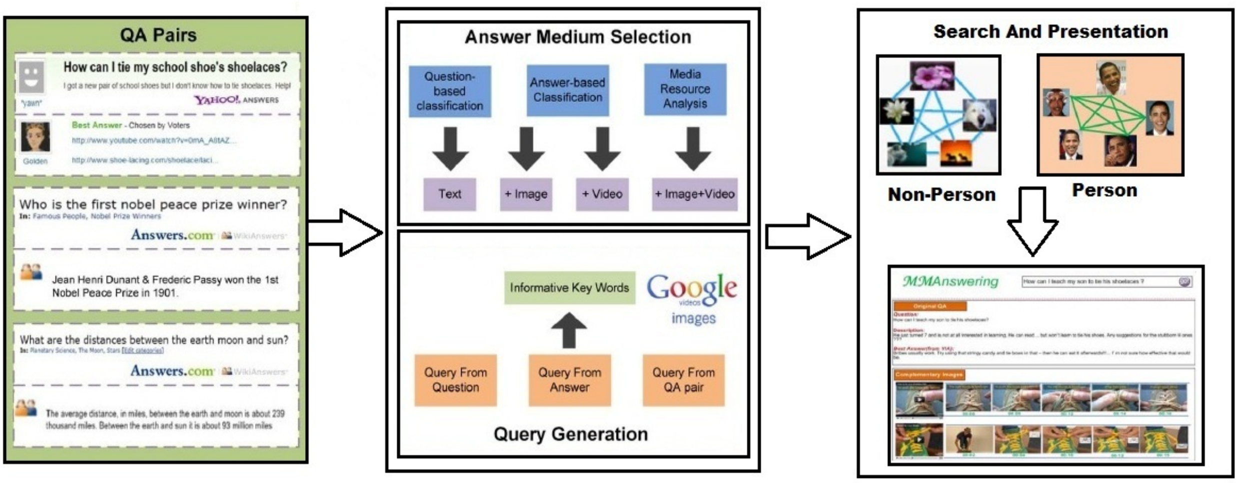

Fig. 1 System Architecture

Next is query generation in this we need to generate the queries from the given QA pairs before multimedia search is being performed. It can be done in two phases First is query extraction and next is query selection in this from the generated three queries the best query is selected. Three query are generated one from question one from answer and one from question answer pair.

Finally graph based re-ranking and duplication removal is done because there are many irrelevant images and videos results we get that should be removed in order to get the appropriate answer for question.

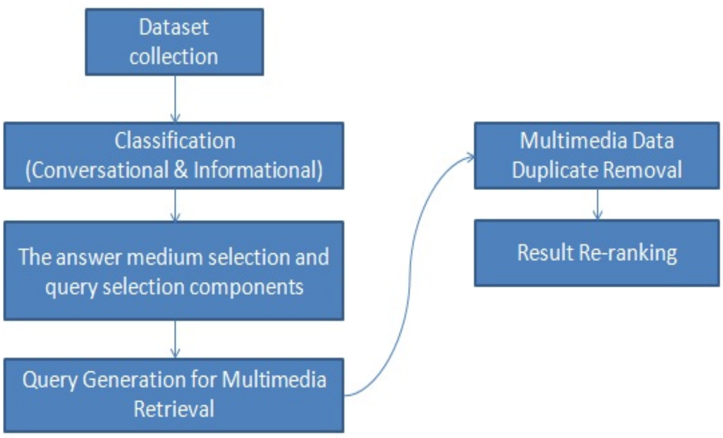

Fig. 2 Process Flow

The very first thing we have to do is data set collection from different existing cQA & Multimedia From existing Search Engines like YouTube, flicker, Google Etc.

After the data set collection is done. The data set is the classified in two type like informative & conversational. After this answer medium selection and query selection procedure component are applied on asked question. After the answer medium selection is done query generation algorithm are used to generate the query for the multimedia data. Which generate a query based on answer, based on question & based on QA pair. From the three query one query will be selected which gives more relevant information for corresponding question. Query Selection will be done on the basis of POS (Part Of Speech) feature which we are extracting from the both question and answer.

From the generated query we collect related Multimedia Data (Image/Video). Generated result may contain more irrelevant data as it based upon text based indexing. Therefore Re-ranking is applied which reorder the textual answer that we have previously obtained. Here we used graph based re-ranking method as below

$$r_{(k)}^{j} = \alpha \sum_{i \epsilon Bj} r_{(k-1)}^{i} P_{ij} (1-\alpha) r_{(0)}^{j}$$

After the re-ranking technique [8] is performed all the similar images & videos are ranked up to gather and finally multimedia duplicate removal is applied on this to avoid information redundancy.

X. Conclusion

This will improve the quality of answer provided by cQA and will be used for the betterment of answers for user. Few disadvantages are that our system may fail to work if the question entered is too complex and contains much complex verbs. Another problem that may occur is it may lack in diversity.

We will further improve the answer medium selection and query selection performance. We will also investigate methods to boost the relevance of the finally selected images and videos, as irrelevant multimedia content may degrade user experience. We also plan to conduct a more comprehensive user study on a larger dataset.

Acknowledgement

We wish to thank Prof. D. A. Meshram for their valuable guidance and inspiration for this paper. We like to express a deep sense of gratitude from the bottom of our heart to our parents & well-wishers', without whom it was impossible for us to reach at this stage.

References

1. Nie, Meng Wang, Yue Gao, Zheng- Jun Zha, and Tat-Seng Chua "Beyond Text QA: Multimedia Answer Generation by Harvesting Web Information", IEEE Multimedia Transaction, 2013.
2. J. Zhang, R. Lee, and Y. J. Wang, "Support vector machine classifications for microarray expression data set", in Proc. Int. Conf. Computational Intelligence and Multimedia Applications, 2003.
3. A. Tamura, H. Takamura, and M. Okumura, "Classification of multiple-sentence questions", in Proc. Int. Joint Conf. Natural Language Processing, 2005.
4. S. Cronen-Townsend, Y. Zhou, and W. B. Croft, "Predicting query performance", in Proc. ACM Int. SIGIR Conf., 2002.
5. L. Nie, S. Yan, M.Wang, R. Hong, and T.-S. Chua, "Harvesting visual concepts for image search with complex queries", in Proc. ACM Int. Conf. Multimedia, 2012.
6. E. Agichtein, S. Lawrence, and L. Gravano, "Learning search engine specific query transformations for question answering", in Proc. Int. World Wide Web Conf., 2001.
7. M. Bendersky and W. B. Croft, "Discovering key concepts in verbose queries", in Proc. ACM Int. SIGIR Conf., 2008.
8. X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, "Bayesian video search re-ranking", in Proc. ACM Int. Conf. Multimedia, 2008.

AUTHOR(S) PROFILE

Pooja Karanjawane, Pursuing a degree in IT from RMDSSOE, warje

Sanket Karanjule, Pursuing a degree in IT from RMDSSOE, warje

Karishma Shaikh, Pursuing a degree in IT from RMDSSOE, warje

Pooja Sharma, Pursuing a degree in IT from RMDSSOE, warje