International Journal of Advanced Information in Engineering Technology (IJAIET) ISSN: 2454-6933
Vol.3, No.4, April 2016

Beyond text QA Multimedia Diverse Relevance Ranking based Answer Generation by Harvesting Web

Sumna V P.
Dept. of Computer Science and Engineering
MEA Engineering College
Perinthalmanna, Kerala, India

Afsar P.
Dept. of Computer Science and Engineering
MEA Engineering College
Perinthalmanna, Kerala, India

***Abstract*— Community QA (cQA) has emerged as an extremely popular alternative to acquire information online, owning to the following facts. cQA forums usually provide only textual answers, which are not informative enough for many questions. To overcome these problems previous studies proposed three steps: First, information seekers are able to post their specific questions on any topic and obtain answers provided by other participants. Second ,comparison with automated QA systems. Third, over times, a tremendous number of QA pairs have been accumulated in their repositories. But the major problem is the lack of diversity of the generated media data. The adopted a method to remove duplicates, but in many cases more diverse results may be better. In our proposed work, we will further improve the scheme investigate Multimedia search diversification methods to make the enriched media data more diverse. Proposed diverse relevance ranking scheme is able to take relevance and diversity into account by exploring the content of images and their associated tags. First, it estimates the relevance scores of images with respect to the query term based on both the visual information of images and the semantic information of associated tags. Then, estimate the semantic similarities of images based on their tags. Based on the relevance scores and the similarities, the ranking list is generated by a greedy ordering algorithm which optimizes average diverse precision, a novel measure that is extended from the conventional average precision.**

Keywords—cQA; relevance score; Greedy ordering

I. INTRODUCTION

As rapid growth of internet information searching has become normal activity in people's daily life. The usage of search engine for information retrieval is increasing enormously over the years. Traditional search engines provides ordered list of web page results based on keyword based matching. This approach provides confusion to the users for vast quantity of information returned in response by search engines. In addition to this the user also finds it difficult to get the exact answer among the returned response by search engine. As a result Community question answering (CQA) emerged as an alternative to the users where the users can get the answers provided by other participants [1]. It not only provides answers to the users but also acts as a platform to users where they can share their answer, discuss their opinion and rate the answer. Moreover it generates flexibility to the user in order to get the best answer and also allow users to accumulate more question answer pairs for preservation and retrieval of answered questions in cQA repositories [2]. For example Wiki Answer, one of the most well-known cQA systems, hosts more than 13 million answered questions distributed in 7,000 categories.

But the disadvantage in the existing community question answering (cQA) forums mostly provides only textual answers which are not informative for many questions Textual answer itself is not sufficient enough to the user for understanding and memorize the content for various questions. In fact users generally post URLs (Uniform resource locator) for linking textual answers with corresponding images and videos .This confirms the importance of multimedia [3] content. On the other hand the increasing growth of multimedia content such has image and video over the web has been witnessed the importance of multimedia search. For example, YouTube serves 100 million distinct videos and 65,000 uploads daily, and the traffic of this site accounts for than 20% of all web traffic and 10% of whole internet , comprising 60% of video watched online. The photo-sharing containing site Flickr contained more than 4 billion images. This clearly shows the importance of multimedia content.

In this paper we propose a model that enrich the textual answers with the corresponding media data in cQA. It comprises of mainly three components:

1. Query generation for searching the multimedia content. In order to find the multimedia content we need to find the informative word which helps to retrieve the corresponding content.
2. Answer medium selection: For a given question it predicts which type of media data should be added .Here we will categorize mainly into four types they are: text, text+ image, text+ video ,text+ Image+ video. It means this approach will decides which type of data should be added for enriching the textual answers. Here we are not considering audio as most of the content can be represented in text form.
3. Multimedia data selection and presentation: Based on the generated informative word we will search and vertically collect corresponding media data such as video and image using multimedia search engines. Multimedia search engines are those which will provide only media data .For example YouTube for videos, Google images, Picasa and Flickr.

II. RELATED WORKS

A. Text QA to multimedia QA

In 1960"s QA system has been started to investigate and it mainly focused on expert system in particular domains. The text based QA has attained popularity in the year 1990"s. Depending upon the type of questions and expected answers, QA can be summarized into following classes: open domain-Based QA[2], Restricted-domain QA[3], Definitional QA and List QA[1]. The automatic QA still faces some difficulties in answering complex questions and cQA is an alternate approach. The existing cQA like Yahoo! Answers, Wiki Answers, and meta filter supports pure text-based answers alone which is not sufficient for users. To overcome this problem multimedia search has been introduced which adds images and videos along with text. Multimedia QA system relies on video optical character recognition (VOCR) and automatic speech recognition (ASR) [5].

B. Multimedia search

The amount of digital information stored on the web has been grown in vast; hence extracting the desired information is an important task. To overcome this problem multimedia search has been introduced which is classified into two categories: text-based search and content-based search [11]. Text-based search is based on the text queries and term based specification, where it matches the text with media in the web to retrieve the media data, and in order to improve the performance some machine learning methods are used which automatically annotate the entities for gathering information. Various social websites like Flicker, Face book uses the manually annotated media entities along with text-based search method which faces some technical issues. To overcome this problem content-based search is used which filters the information by analyzing the content in the media instead of metadata. The content-based retrieval has some limitations such as high computational cost, difficulty in finding the visual queries and large gap between the visual description and users semantic Expansion .Multimedia search re-ranking algorithm is used for improving the search relevance by extracting the visual information of images and videos [4].

C. Multimedia search Re-ranking

The Re-ranking algorithm has been categorized into two techniques: pseudo relevance feedback [13] and graph based re-ranking[12].Pseudo relevance feedback is used to collect the relevant samples and that samples are assumed to be irrelevant. A classification or ranking model is used for ranking the samples and provide feedback by labeling the results as relevant or irrelevant. Graph classification method is based on the two premises. First one is variance between the initial ranking list and processed ranking list should be small. Second one is visually similar samples should be ranked very nearer. This algorithm constructs a graph where the images or videos represent the vertices and pair wise similarities represent the edges. These two approaches rely on the visual similarities between two media entities. It should measure the resemblance like color, texture, shape and so on; Almost query adaptive technique is used for estimating the similarities. Let us consider an example as finding out the person. Here we need to identify the similarities of facial characteristics. Based on this identification, queries are classified into two classes namely person related query or non-person related query.

III. PROPOSED SYSTEM

A. system Architecture

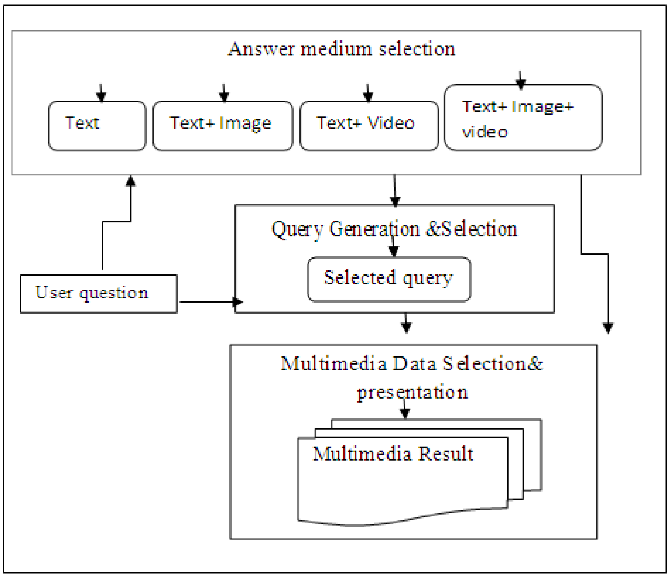


Fig1.System Architecture

B. System Module

Our system consist of three modules answer medium selection, query generation &selection and multimedia data selection & presentation.

1. Answer Medium Selection

In this model the given question is judged whether it requires any media data or it requires only textual answer. Here we will categorize mainly into four types such as text, text+ image, text +video, text +image+ video based on the given question [5]. It is not sufficient for the user to understand if we provide only textual answers. For example, for the question "how to cook chicken stew" we may find the answer as "clean it and cut it into ...."This clearly indicates that it can be better understandable if we provide video answers. "when india got republic day?"in this type of questions textual answer alone is sufficient. Whereas "Who is the father of india?" In this case textual answer together with image provides sufficient information to the user. Here images will enrich the textual answers. In some situation questions may also need all things together i.e. text+ image+ video which include both videos and image along with text. The Answer medium selection is classified into three methods as follows [12]:

1. Question based classification- In this approach the questions are classified based on two steps. First one is to

identify the interrogative words (yes/no) and then it will directly answer the question. Second one is to classify the interrogative words using the naïve bayes classifier. To build nave bayes classifier we extract text features like the bigram text features head words and class specific related words.Table1 shows the four class classification model based on class specific related words.

2. Answer based classification-In this approach verb and bigram words are extracted from the Question. The verbs will help to identify how to enrich the textual answers i.e. either by means of image or video.
3. Media resource analysis by learning- In this module first define the defines a clarity score for a query4 based on the relative entropy (or Kullback-Leibler (KL) divergence) between the query and collection language models

$$clarity_q(C_i) = \sum_{w \in V_{ci}} P(w|\theta_q) log_2 \frac{P(w|\theta_q)}{P(w|\theta_{c_i})} \qquad (1)$$

Where C_i is the entire vocabulary of the collection C_i, and i=1 , 2, 3 represent text, image and video, respectively. The terms p(w\Θq)and p(w|Θci)are the query and collection language models, respectively. The Clarity value becomes smaller as the top ranked documents approach a random sample from the collection (i.e., an ineffective retrieval).

The query language

$$P(w|\theta_q) = \frac{1}{\mathcal{Z}} \sum_{d \in \mathcal{R}} P(w|D)P(q|D) \qquad (2)$$

where p(q|D) is the query likelihood score of document D.

$$\mathcal{Z} = \sum_{D \in \mathcal{R}} P(w|D)P(q|D) \qquad (3)$$

$$P(q|D) = \prod_{w \in q} P(w|D) \qquad (4)$$

Perform medium selection by learning a four-class classification model based on the results of question-based classification, answer-based classification, and media resource analysis. For question-based classification, we have four scores, i.e., the confidence scores that the question should be answered by "text", " text+ image ", "text+ image+ video ", and " text+ video". Similarly, for answer-based classification we also have four scores. For media resource analysis, we have three scores SVM with linear kernel.

2. Query Generation for Multimedia Search

This method is used for generating the queries before performing the search in multimedia search engine. Queries will helps to retrieve the most relevant images and videos from the web. First step is query extraction i.eto extract informative keywords from questions and answers [6]. Second step is query selection i..e either from the questions or the answers or the combination of both the question and answers. This query selection method includes the features of POS histogram and search performance prediction [1]. This Query generation for multimedia search supports only the 42-dimensional search prediction for each QA pairs.

Categories	Class Specific Related words
Text	Name, population, period, times, country ,height ,website , birthday, age ,date ,rate, Distance, speed, religions, number, etc.
Text+ image	Color , 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.

Table1. Representation of four classification model

3. Multimedia Data Selection And Presentation

This method uses the queries which are generated earlier to perform the searching of image and video content from the web search engine [11]. Here search engine uses the graph-based re-ranking method to identify whether the query belongs to either person related query or non person related query. The face detection method is used to identify the most returned images and from that we can decide the person related query. The person related queries extract the 256-D Local binary pattern features [12] from the largest faces of images or video frames. The non-person related query requires extracting the 428-dimensional visual features and then it is to be re-ranked.

4. Diverse Relevance Ranking (DRR)

Diverse Relevance Ranking (DRR) scheme is used for social image search. It is able to rank the images based on their relevance levels with respect to query tag while simultaneously considering the diversity of the ranking list. The scheme works as follows. First, we estimate the relevance score of each image with respect to the query term as well as the semantic similarity of each image pair. The relevance estimation incorporates both the visual information of images and the semantic information of their associated tags into an optimizations framework, and the semantic similarity is mined based on the associated tags of images.

Here, we introduce the DRR approach. We present it as a general ranking algorithm and leave the two flexible components, i.e., relevance score and similarity estimation of images. The necessity of diversity may seem less intuitive than relevance, but its importance has also been long acknowledged in information retrieval. One explanation is that the relevance of a image with respect to the query should depend on not only the document itself but also its difference with the documents appearing before it. The importance of relevance is clear. In fact, this is usually regarded as the bedrock of information retrieval: if an IR system's response to each query is a ranking of results in order of decreasing probability of relevance, the overall effectiveness of the system to its user will be maximized. The DRR algorithm is actually a greedy approach to optimizing the expected value of the Average Diverse Precision (ADP) measurement. Based on the relevance scores and the similarities, the ranking list is generated by a greedy ordering algorithm which optimizes ADP, a novel measure that is extended from the conventional Average Precision (AP).

IV. RESULT AND OBSERVATION

In this section, we are comparing the performance of the proposed system with the existing system in terms of accuracy rate. To assess efficiency, we measured the accuracy of system. From the end of this experimentation section, we can say that the proposed system has higher efficiency than the existing system.

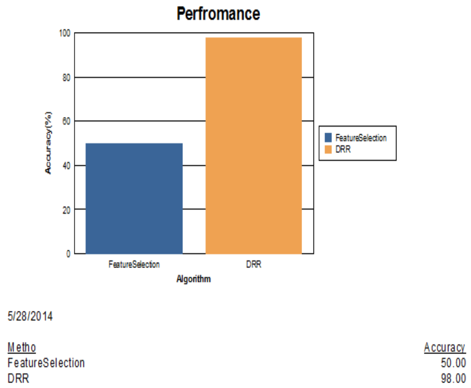
A. Accuracy

Accuracy can be calculated from formula given as follows

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

Figure 4.1 shows the performance of our proposed system.

This graph shows the accuracy rate of existing system such as Beyond Text QA and proposed system such that Beyond Text QA with Diverse Relevance Ranking (DRR) based on two parameters of accuracy and methods such as existing and proposed system. In this graph, x axis will be the methods such as feature selection i.e., medium selection and DRR and y axis will be accuracy rate. From the graph we can see that, accuracy of the system is reduced somewhat in existing system than the proposed system. From this graph we can say that the accuracy of proposed system is increased which will be the best one.



CONCLUSION AND FUTURE WORK

Existing system uses a novel scheme to answer questions using media data by leveraging textual answers in cQA. For a given QA pair, our scheme first predicts which type of medium is appropriate for enriching the original textual answer. Following that, it automatically generates a query based on the QA knowledge and then performs multimedia search with the query. Proposed diverse relevance ranking scheme for social image search, which is able to simultaneously take relevance and diversity into account. It leverages both visual information of images and the semantic information of tags. Finally, query-adaptive reranking and duplicate removal are performed to obtain a set of images and videos for presentation along with the original textual answer.

In our future work, we will further improve the scheme, such as developing better query generation method and investigating the relevant segments from a video.

REFERENCES

[1] Liqiang Nie, Meng Wang, Member, IEEE, Yue Gao, Zheng-Jun Zha, Member, IEEE,and Tat-Seng Chua,Senior Member, IEEE "Beyond Text QA: Multimedia Answer Generation by Harvesting WebInformation"- IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 2, FEBRUARY 2013.

[2] 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

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

[4] T. Yeh, J. J. Lee, and T. Darrell, "Photo-based question answering," in Proc. ACM Int. Conf. Multimedia, 2008.

[5] H.Yang,S.Neo,L.Chaisorn,T.Chua Available : http://citeseerx.ist.psu.edu/viewdoc/

[6]H. Cui, M.-Y. Kan, and T.-S. Chua, "Soft pattern matching models for definitional question answering," ACM Trans. Inf. Syst., vol. 25, no. 2, pp. 30–30, 2007.

[7] R. C. Wang, N. Schlaefer, W. W. Cohen, and E. Nyberg, "Automatic set expansion for list question answering," in Proc. Int. Conf. Empirical Methods in Natural Language Processing, 2008.

[8] L. A. Adamic, J .Zhang , E. Bakshy, andM. S. Ackerman, "Knowledge sharing and Yahoo answers: Everyone knows something," in Proc. Int. World Wide Web Conf., 2008.

[9] G. Zoltan, K. Georgia, P. Jan, and G.-M. Hector, Questioning Yahoo! Answers, Stanford InfoLab, 2007, Tech. Rep.

[10] H. Yang, T.-S. Chua, S. Wang, and C.-K. Koh, "Structured use of external knowledge for event-based open domain question answering," in Proc. ACM Int. SIGIR Conf., 2003.

[11] S. K. Shandilya and N. Singhai, "Article: A survey on: Content based image retrieval systems," *Int. J. Comput. Appl.*, vol. 4, no. 2, pp. 22–26, 2010.

[12] S. K. Shandilya and N. Singhai, "Article: A survey on: Content based image retrieval systems," *Int. J. Comput. Appl.*, vol. 4, no. 2, pp. 22–26, 2010.

[13] Y. Liu, T. Mei, X.-S. Hua, J. Tang, X. Wu, and S. Li, "Learning to video search rerank via pseudo preference feedback," in *Proc. Int. Conf. Mul-timedia & Expo*, 2008.

[14] X. Li and D. Roth, "Learning question classifiers," in *Proc. Int. Conf. Computational Linguistics*, 2002.

[15]J. Zhang, R. Lee, and Y. J. Wang, "Support vector machine classifica-tions for microarray expression data set," in *Proc. Int. Conf. Computa-tional Intelligence and Multimedia Applications*, 2003.

[16] J. Arguello, F. Diaz, J. Callan, and J. F. Crespo, "Sources of evidence for vertical selection," in *Proc. ACM Int. SIGIR Conf.*, 2009.

[17]A. Tamura, H. Takamura, and M. Okumura, "Classification of mul-tiple-sentence questions," in *Proc. Int. Joint Conf. Natural Language Processing*, 2005.

[18]Z. Huang, M. Thint, and Z. Qin, "Question classification using head words and their hypernyms," in *Proc. Int. Conf. Empirical Methods in Natural Language Processing*, 2008.

[19] Z.-J. Zha, M. Wang, Y.-T. Zheng, Y. Yang, R. Hong, and T.-S. Chua, "Interactive video indexing with statistical active learning," IEEE Trans. Multimedia, vol. 14, no. 1, pp. 17–27, 2012.

[20] I. Ahmad and T.-S. Jang, "Old fashion text-based image retrieval using FCA," in Proc. ICIP, 2003.

[21] M. Wang, X. S. Hua, R. Hong, J. Tang, G. J. Qi, Y. Song, and L. R. Dai, "Unified video annotation via multi-graph learning," IEEE Trans. Circuits Syst. Video Technol., vol. 19, no. 5, pp. 733–749, 2009.

[22] M. Wang, X. S. Hua, T. Mei, R. Hong, G. J. Qi, Y. Song, and L. R. Dai, "Semi-supervised kernel density estimation for video annotation," Comput. Vision Image Understand., vol. 113, no. 3, pp. 384–396, 2009.

[23] J. Tang, R. Hong, S. Yan, T. S. Chua, G. J. Qi, and R. Jain, "Image annotation by KNN-sparse graph-based label propagation over noisytagged web images," ACM Trans. Intell. Syst. Technol., vol. 2, no. 2, pp. 1–15, 2011.

[24] J. Tang, X. S. Hua,M. Wang, Z.Gu, G. J.Qi, and X. Wu, "Correlative linear neighborhood propagation for video annotation," IEEE Trans. Syst., Man, Cybern. B, vol. 39, no. 2, pp. 409–416, 2009.

[25] 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.

[26] S. K. Shandilya and N. Singhai, "Article: A survey on: Content based image retrieval systems," Int. J. Comput. Appl., vol. 4, no. 2, pp. 22–26, 2010.

[27] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and X.-S. Hua, "Bayesian video search reranking," in Proc. ACM Int. Conf. Multimedia, 2008.

[28] A. P. Natsev, M. R. Naphade, and J. TešiĆ, "Learning the semantics of multimedia queries and concepts from a small number of examples," in Proc. ACM Int. Conf. Multimedia, 2005.