RESEARCH ARTICLE OPEN ACCESS

Enhanced Multimedia Information Retrieval on Web Using Text Based Query

Meenakshi Alamelu S, Mannar Mannan J
Department of Information Technology
Anna University Regional Campus, and Coimbatore
Tamil Nadu -India

ABSTRACT

Community question answering (CQA) services have accomplished quiet popularity over the past 10 years. It helps user to get information from a comprehensive set of well-answered questions. Text based Information retrieval system cannot exposes enough attention to understand concept corresponding to user query. It will take more time to read and understand and also there is a linguistic problem prevent the user to read the document or the user may not have enough time to read the entire document to overcome(address) these issue, a multimedia based QA system needed. Texted, audio, video and image based information given to any corresponding query. The semantic based text document rank has been proposed for rank the document and the audio analysis is done corresponding to the query concept by analysis of the metadata of audio file and its internal representation. The video has been ranked corresponding to query is done by dividing the frames. The image has been done corresponding to the query by edge detection method. The performance of this work overcomes the limitations of existing in present work.

Keywords:- Community question answering (CQA), Query analysis, medium selection, question answering..

I. INTRODUCTION

The central concern of multimedia information retrieval (MIR) is easily stated: given a collection of multimedia data (i.e., a complex information object, with components of different kinds, such as text, images, video and sound, all in digital form), find those that are relevant to information need of the user. CQA is 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. 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.

But the disadvantage in the existing community question answering (CQA) forums mostly provides only textual answers which are not informative for many questions Fully automated QA still faces challenges that are not easy to tackle, such as the deep understanding of complex questions. The sophisticated syntactic, semantic and contextual processing to generate answers is needed to address above said issues.

Reading paragraph and understand concept from the given answer for corresponding query is a time consuming process. The reading may tire while reading paragraph. The linguistic problems are also make complicated to understand the answer. Textual answers cannot provide sufficient answer for all type of queries in the forums, and also cannot be efficient.

In this paper, we propose a model that enriches the textual answers with the corresponding media data in CQA. It does not aim to directly answer the questions, and instead of, it's enriching the community-contributed answers with multimedia contents. A proposed work is textual answers in CQA with appropriate multimedia data. Output image would compare with existing image dataset and Also, proposing the audio data using wave (.wav) analysis and metadata search. The approach is built based on enhanced multimedia information.

II. RELATED WORK

The image content by using colors, textures or shapes based. Similarity measures are used to determine how similar or dissimilar in the given query image and the image database collections [14] A novel approach MMQA (Multimedia answer generation) by which user can get answer in textual as well as media form. The approach will have three components: (1) Answer medium selection (2) Query generation for multimedia search, (3) Multimedia data selection & Presentation. Initially it will predict whether it's necessary to add multimedia data along with textual answer and if it will require then which type of class data should be added. Data can be of following classes: text, text + image, text + image + video. Then media data will be add to enhance original textual answer. After this we have to generate informative queries and second component will extracts three queries from the question, answer, the QA pair. According to query third component will collect images and video and then final answer representation is done [1].

Improve textual answers in CQA with appropriate multimedia data. MMQA consists of three components in its scheme; those are Answer medium selection, Query generation for multimedia search, and Multimedia data selection and presentation. This approach automatically figures out which type of multimedia information should be added to get an elaborated textual answer. For this, it then automatically collects data from web to enrich the answer. By processing a huge set of question- answer pairs and adding them to a dataset, his approach can set up a novel multimedia question answering (MMQA) approach as users can find multimedia answers by comparing their questions with those in the dataset. His approach MMQA is different from rest of the MMQA research efforts as it not only provide direct question answers with image & video data but also is based on community-contributed textual answers and hence it is able to give answers of more complex questions [2].

It point out that it is important to map the query to a few relevant concepts instead of search with all concepts. In addition, we show that solving this problem through both text and image inputs are effective for search, and it is possible to determine the number of related concepts by a language modeling approach [3].

The main contributions of this approach to multimedia information retrieval literatures include: (a) development of a truly bilingual video QA system, (b) presentation of a robust bilingual passage retrieval algorithm to handle no-word boundary languages such as Chinese and Japanese,(c) development of a large-scale bilingual video QA corpus for system evaluation, and (d) comparisons of seven top-performing retrieval methods under the fair conditions[4].

He proposes to exploit semantic attributes for image search reranking. Based on the classifiers for all the predefined attributes, each image is represented by an attribute feature consisting of the responses from these classifiers. A hyper graph is then used to model the relationship between images by integrating low-level visual features and attribute features. Hyper graph ranking is then performed to order the images. Its basic principle is that visually similar images should have similar ranking scores. In this work, we propose a visual-attribute joint hyper graph learning approach to simultaneously explore two information sources. A hyper graph is constructed to model the relationship of all images[5].

Given a complex query, our scheme first detects the noun-phrase based visual concepts and crawls their top ranked images from popular image search engines. Next, it constructs a heterogeneous probabilistic network to model the relatedness between the complex query and each of its crawled images. The network seamlessly integrates three layers of relationships, i.e., the semantic-level, cross-modality level as well as visual-level. These mutually reinforced layers are established among the complex query and its involved visual concepts, by harnessing the contents of images and their associated textual cues. Based on the derived relevance scores, a new ranking list is generated. Extensive evaluations on a real-world dataset demonstrate that our model is able to characterize the complex queries well and achieve promising performance as compared to the state-of-the-art methods [6].

For retrieving the media information such as image and video we introduce a semantic web Technology. This approach automatically determines which type of media data should be added for the textual answer. It then automatically collects data from the web to enrich the answer. By processing a large set of QA pairs and adding them to a pool, our approach can enable a novel multimedia question answering (MMQA) approach a-s users can find multimedia answers by matching their questions with those in the pool. Different from a lot of MMQA research efforts that attempt to directly answer questions with image and video data, our approach is built based on community-contributed textual answers and thus it is able to deal with more complex questions[7].

A set of features for representing texture and instrumentation. In addition a novel set of features for representing rhythmic structure and strength is proposed. The performance of those feature sets has been evaluated by training statistical pattern recognition classifiers using real world audio collections. Based on the automatic hierarchical genre classification two graphical user interfaces for browsing and interacting with large audio collections have been developed [8].

Classify to music into three broad categories: rock, classical and jazz. We discuss the feature extraction process and the particular choice of features that we used- spectrograms and mel scaled cepstral coefficients (MFCC). We use the texture-of texture models to generate feature vectors out of these. Together, these features are capable of capturing the frequency-power profile of the sound as the song proceeds. Finally, we attempt to classify the generated data using a variety of classifiers. we discuss our results and the inferences that can be drawn from them[9]

Human supervision is introduced to learn the model weights offline, prior to the online reranking process. While model learning requires manual labeling of the results for a few queries, the resulting model is query independent and therefore applicable to any other query. The experimental results on a representative web image search dataset comprising 353 queries demonstrate that the proposed method outperforms the existing supervised and unsupervised reranking approaches [10]

In theory, speech recognition technology can make any spoken words in video or audio media subject to text indexing, search and retrieval. This article describes the News-on-Demand application created within the Info media Digital Video Library project and discusses how speech recognition is used for transcript creation from video, time alignment of closed-captioned transcripts, a speech query interface, and audio paragraph segmentation. The results show that speech recognition accuracy varies dramatically depending on the quality and type of data used, but the system is quite useable with only moderate speech recognition accuracy [11].

A novel query suggestion scheme named Visual Query Attributes Suggestion (VQAS) for image search with QBE. Given a query image, informative attributes are suggested to the user as complements to the query. These attributes reflect the visual properties and key components of the query. By selecting some suggested attributes, the user can provide more precise search intent which is not captured by the query image. The evaluation results on two real-world image datasets show the effectiveness of VQAS in terms of retrieval performance and the quality of query suggestions [12].

This paper extends our approach to perform event-based QA by uncovering the structure within the external knowledge. The knowledge structure loosely models different facets of QA events, and is used in conjunction with successive constraint relaxation algorithm to achieve effective QA. Our results obtained on TREC-11 QA corpus indicate that the new approach is more effective and able to attain a confidence-weighted score of above 80% [13].

He makes a survey about text and content based image retrieval system. Image retrieval is performed by matching the features of a query image with those in the image database. It can be classified as text-based and content-based. The text-based Image retrieval applies traditional text retrieval techniques to image annotations. The content-based Image retrieval apply image processing techniques to first extract image features and then retrieve relevant images based on the match of these features. Feature extraction is the process of extracting image features to a distinguishable extent to extract[14].

III. CORE WORK

The proposed approach is that, it does not aim to directly answer the questions, and instead of, it's enriching the community-contributed answers with multimedia contents. A proposed work is textual answers in cQA with appropriate multimedia data. Output image would compare with existing image dataset and Also, proposing the audio data using wave (.wav) analysis and metadata search. The approach is built based on enhanced multimedia information. It is worth nothing that, although the proposed approach is automated, the new technique can also further involve human interactions. For example, this approach can provide a set of candidate images and videos based on textual answers, and answers can manually choose several users for final presentation.

It is possible to provide multimedia data set from text based query. Easy understand of concept from text, image, audio and video are enhanced. Image can recognized by its content, the content can analysis by predefined image pattern instead of metadata. For recognizing audio, streaming method is implemented and video uses image pattern.

A. QUERY ANALYSIS

In Query Analysis, 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 first introduce the categorization based on stop word removal Questions can mainly be categorized into the following classes based on removal Here head word is referred to as the word specifying the object that a question seeks. The semantics of head words play an important role in determining answer medium. For instance, for the question "*what is computer*", the head word is "computer", based on which we can judge that the sought-after answer is a simple. Therefore, it is reasonable to use textual answer medium. Questions need "text + image", "text + video" or text + image + video" answers. Therefore, given a question, we first judge whether it should use only based on textual answer. Table I shows the heuristics. 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.

For each QA 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. Second, we identify several key concepts from verbose answer which will have the major impact on effectiveness. 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.

TABLE I
REPRESENTATIVE INTERROGATIVE WORDS

Interrogative Word	Category
be, can, will, have, when, be there, how+adj/adv	Text
what, where, 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.

B. IMAGE, AUDIO AND VIDEO ANALYSIS

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 query. Also, we can choose to match each query term with a person list, such as a celebrity list. we adopt a method that analyzes image search results. Specifically, for each image in the ranking list, we perform edge detection If a query is person-related, we perform edge detection for each image and video key-frame. Visually similar images or videos may be ranked together. Thus, we perform a duplicate removal step to avoid information redundancy.

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. Audio should be processed with metadata search, and its internal representation

IV. EXPERIMENT REULTS

In this section, we present the experimental results. We compare our algorithm with the re-ranking results by directly computing distances. We can see that our algorithm provides much more visually smooth and less noisy results. Our test data are a collection of images associated with several keywords crawled from Microsoft Live Image Search. We manually label the ground-truth of images as either \noise" or \relevant". Relevant images are further classified into Subclasses (images that are visually similar).

We further manually annotate whether each pair of subclasses is semantically relevant. We illustrate the top results before and after re-ranking for two example queries in Figs. 1, respectively, one about object and the other about person. Fig. 1(a) shows the top 7 results before re-ranking about "bird of prey", while (b) shows the top 7 results after re-ranking based on global features. Similarly

After re-ranking, we perform duplicate removal and present the images or/and videos together with the textual answers, depending on the results of answer medium selection. Fig. 2 shows the multimedia answers for 3 example queries.

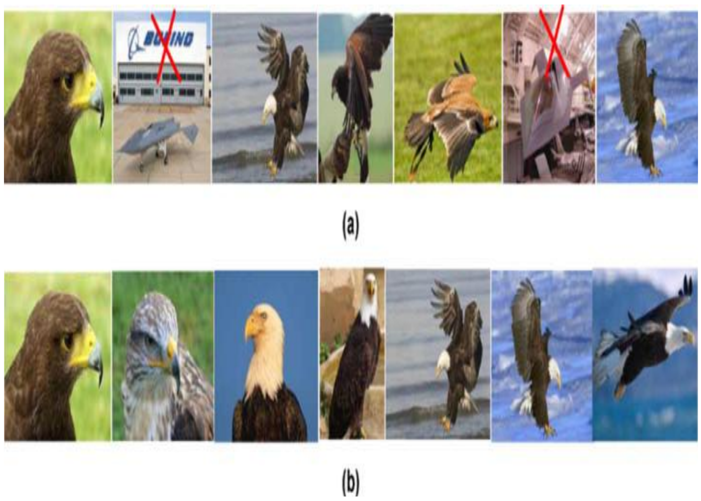

Fig. 1. The non-person-related image search re-ranking results for the query "Bird of prey". (a) The top images before re-ranking; (b) the top images after reranking

Of course, once the user has input a query and the system has determined some number of matches against the audio database, the next logical step is presenting the match results to the user. Here the time-dependent nature of audio data reveals the problem of presentation. For media that are not time12 dependent, such as text or images, the data (or an abbreviated form) is static and can be displayed without any trouble. For time-dependent media such as audio, it is unclear what form should be displayed or presented to the user, since simultaneously playing 20 clips of music (representing 20 query matches displayed at a time) is unlikely to be useful to the searcher. Bainbridge et al. enumerate a number of such problems in presenting retrieved audio, especially when compared to typical functionality supported in presenting retrieved text.

A related research effort is how to browse and navigate through databases of audio. Audio is inherently a stream of time-dependent auditory data, with no standardized structure for interconnecting related points in time in these streams. For text, hypertext provides a structure to indicate relationships between certain parts of the text, both within the same document and between documents.

Fig. 2 Results of multimedia answering for 3 example queries, "the most talented member of NWA", "tie shoelace", and "September 11". Our scheme answers the three questions with "text + image", "text + video", and "text + image + video", respectively.

V. CONCLUSION AND FUTURE WORK

In this paper The motivation and evolution of MMQA, and it is analyzed that the existing approaches mainly focus on narrow domains. We proposed to utilize the audio, the visual information and the images returned by text-based search. In MMQA research that aims to automatically generate multimedia answers with given questions, our approach is built based on the community contributed answers, and it can thus deal with more general questions and achieve better performance. Aiming at a more general approach, this proposed work is, to answer questions using media data by leveraging textual answers in cQA. For a given QA pair, this scheme first predicts which type of medium is appropriate for enriching the original answer. Following that, it automatically generates a query based on the QA knowledge and then performs multimedia search with the query Finally, Image Recognize based on trained dataset .Audio also provided by corresponding to query by analyzing with WAVE (.wav) and metadata search. Finally combining image and audio data, video should be provided.

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. We will also investigate multimedia search diversification methods

REFERENCE

[1] Wang, Yue Gao,"Beyond text QA: Multimedia Answer Generation by Harvesting Web Information," *IEEE Transactions on multimedia*, vol.15, no. 2, 2013.

[2] Handore S. A.1 & Ganeshwade ,"MMQA: Multimedia Answer Generation for Community Question Answering System ,". *IOSR Journal, pp.38-44*

[3] Dong Wang, Xirong Li, Jianmin Li and Bo Zhang,"The Importance of Query-Concept-Mapping for Automatic Video Retrieval,"IEEE Multimedia Magazine.

[4] Yue-Shi Lee, Yu-ChiehWu, Jie-ChiYang," BVideoQA: Online English/Chinese Bilingual Video Question Answering," journal of the American society for information science and technology, 60(3):509–525, 2009.

Junjie Cai, Zheng-Jun Zha, , Meng Wang, Shiliang Zhang, and Qi Tian, *Senior,"* An Attribute-assisted Reranking Model for Web Image Search," ieee transactions on image processing ,vol x,no.xx,2015.

[5] S.Lakshmi Prabha,C.Meenaa , S.Pugazhl Vendhan and S.Raja Ranganathan," enhancing the performance of information etrieval in qa using semantic web technoogy," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014.*

[6] Hrishikesh Deshpande, Unjung Nam, Rohit Singh," MUGEC: Automatic Music Genre Classi_cation," June 14, 2001.Junjie Cai, Zheng-Jun Zha, , Meng Wang, Shiliang Zhang, and Qi Tian, *Senior,"* An Attribute-assisted Reranking Model for Web Image Search," ieee transactions on image processing ,vol. x,no.xx,2015.

[7] Saravana.s," Visual Reranking Of Image Retrieval System," *International Journal of Computer Trends and Technology (IJCTT) – volume 2 Issue 2 Number 4 – Dec 2011*

[8] Alexander, G. Hauptmann , and Michael J. Witbrock," Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval," *Intelligent Multimedia Information Retrieval*, Mark T. Maybury, Ed.. AAAI Press, pps. 213-239, 1997.

[9] Jingwen Bian, Zheng-Jun Zha, Hanwang Zhang, Qi Tian, Tat-Seng Chua," Visual Query Attributes Suggestion,"In ACM Transactions.

[10] M. Surdeanu, M. Ciaramita, and H. Zaragoza, "Learning to rank answers on large online QA collections," in

Proc. Association for Computational Linguistics, 2008.Jacobs and C. P. Bean, "Fine articles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[11] B. Li, Y. Liu, A. Ram, E. V. Garcia, and E. Agichtein,"Exploring question subjectivity prediction in community QA," in *Proc. ACM Int.*

[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] Hui Yang, Tat-Seng Chua, Shuguang Wang, Chun-Keat Koh," Structured Use of External Knowledge for Event-based Open Domain Question Answering,"In ACM Transactions,2003.

[14] T. Karthikeyan, P. Manikandaprabhu, S. Nithya," A Survey on Text and Content Based Image Retrieval System for Image Mining," International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 3, March – 2014

[15] M. Surdeanu, M. Ciaramita, and H. Zaragoza, "Learning Rank answers on large online QA collections," in *Proc. Association for Computational Linguistics*, 2008.Jacobs and C.