A Mobile Video Question Answering System for E-learning

Yi-Ting Huang[1], Ching-I Chung[1], Chi-Cheng Tsai[1], Chia-Hsing Shen[1], Yu-Chieh Wu[2] and Jie-Chi Yang[1]
[1]*Graduate Institute of Network Learning Technology,* [2]*Department of Computer Science and Information Engineering, National Central University, Taiwan*
[1]*{coral, cecilia, ken, arcadia, yang}@cl.ncu.edu.tw,* [2]*bcbb@db.csie.ncu.edu.tw*

Abstract

In this paper, the authors proposed a mobile video QA (Question Answering) system for online annotation and ubiquitous multimedia learning. Under the wireless and mobile environment, learners could interact with the system through natural language questions rather than traditional query-like text retrieval. The handheld device is not as well as traditional keyboard input, while the annotation is quite inefficient and inconvenient. To solve this, the authors designed a simple browsing and clicking model to enable learners annotating and auto-feedback as e-mail to learner's mail box. A pilot study was conducted to evaluate the preliminary experimental results for learning. The experimental results showed that the proposed system effectively engaged learners in multimedia video QA in handheld device where all subjects agreed that the system is positive and useful.

1. Introduction

Wireless technology, mobile products, and personal digital assistance (PDA) augment human experiences, and also change the outdoor activities of students. There are many applications for handheld devices that are used in educational activities, such as class response system, mobile computer supported collaborative learning, mobile game, context aware, and outdoor learning [2][5][7][9]. Learners can explore knowledge through a handheld device (for example PDA) in inquiring way [3][4]. As soon as a learner feels curious, handheld device is just the medium of persisting learning experience. Without the constraint of time and space, learners will not lose any opportunity to learn in seamless learning space.

Nowadays, when learners have some questions, the normal approach is to find out the answers through web-page search engines, such as Google and Yahoo. However, those techniques are designed to retrieve relevant documents and web-pages instead of exact answers. It is difficult to pinpoint accurate key words in the full document when a learner never heard of his question. What are the right keywords they could get the right response? It is hard for a beginner to look for. In addition, the wireless communication of mobile device is not mature enough for users to browse quickly. More accurate response is necessary. Therefore, an question answering for exact answer extraction is indispensable.

With the development and innovation of technology, the diversity of multimedia format is getting numerous. Multimedia learning plays gradually an important role in learning. In fact, video contains rich human knowledge, but it is difficult to discover in the present tool. On the other hand, J. S. Bruner [1] considered the learning structure need coordinate the students' cognitional structure on the psychology view. Multimedia content including vivid vision and voice stimulate the sight and auditory sense of learners, which connect strongly to human awareness. When multimedia technology is involved in instruction, it could imitate long-term observed lecture content, simplify complex lecture content, or contrast all kinds of cause and effect, etc. Learners can understand abstract concept easily in such multimedia environment than traditional instructional strategy.

This paper intends to apply the characteristics of the wireless application and improve the issue of the present search ability. A mobile-based video QA system was designed to integrate with multimedia content question answering, retrieval, and annotation in the wireless and mobile environment. A learner can ask questions through wireless connection. After video question answering process, a list of ranked answers together with text and videos were represented. Learners can directly annotate the returned answers through clicking and browsing. The system subsequently sends the annotated results as email to learners. In this way, the retrieved multimedia content could be further browsed and traced in desktop PCs.

The paper is organized as follows. Section 2 offers examples of system usage. Section 3 describes the system architecture. Section 4 presents a pilot study. The conclusion and future work are given in section 5.

Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)
0-7695-2916-X/07 $25.00 © 2007 IEEE

2. User scenarios

In this section, the authors present three examples to explain the benefits of using the system.

Scenario 1: Preview

Tom is a student and he always spends one hour to school. Tom often prepares his homework before taking a class. Today, he had a question about a new term of science. He had to look up the dictionary, but he still did not understand what it means. Then he took his handheld device, and queried what he did not know. For a while, he got the answers with vivid video, photo, and abstract. The system solved his problem immediately. Tom could continue his preparing a lesson.

Scenario 2: A learning tool

Tom arrived at school. His first class was History today. In class, the teacher told a history story about Napoleon. Tom is very interested in this history, and he adores Napoleon. Thus he wondered more information about Napoleon, like "Where was Napoleon born?" He asked the question using handheld device, and then got top-10 results including multimedia information. Tom browsed the results, meanwhile, bell was ringing, and the other class was beginning. Tom annotated some items which he considered that was important to keep, and mailed to his mailbox.

Scenario 3: Anytime, anywhere

After the class has finished, the teacher asked Tom to visit an astronomical exhibition. Tom is unfamiliar with astronomy, but he enjoyed learning any fresh subject. Due to the lack of eclectic construction in exhibitions, Tom could not consult advanced information anytime, such as "What the main substance is Sun composed of?" However, he use the system to find the answer, then he could understand how Sun be composed of. It is not only presented with text, but also multimedia showed. Tom could make sense more than traditional method, and situated where he was.

3. System architecture

3.1. System overview

The system is composed of server and client. Figure 1 shows the system architecture, included Video Question Answering Subsystem, Mail Subsystem, Connection Subsystem, and Mobile Client.

The Video Question Answering Subsystem contains four models: Chinese Word Segmentation Model, Initial Retrieval Model, Video Processing Model, and Ranking Model. Chinese Word Segmentation Model is used to tokenize the text into a series of words. Then the Initial Retrieval Model serves at finding the roughly related videos for answer extraction. In this paper, one of the state-of-the-art document retrieval models (Okapi BM-25) [6] was used to rank top-1000 video clips. The Ranking Model simultaneously receives both the top-1000 relevant videos, and the content information from the Initial Retrieval Model and the Video Processing Model. The goal of Video Processing Model is to recognize the contents from the input raw videos. Mail Subsystem sends the learner's annotation file to his mail automatically. Connection Subsystem coordinates the data flow between the server and clients in wireless environments. It transfers from client requests to the server and receives the data from the server, converts the question into multimedia information to the Mobile Client.

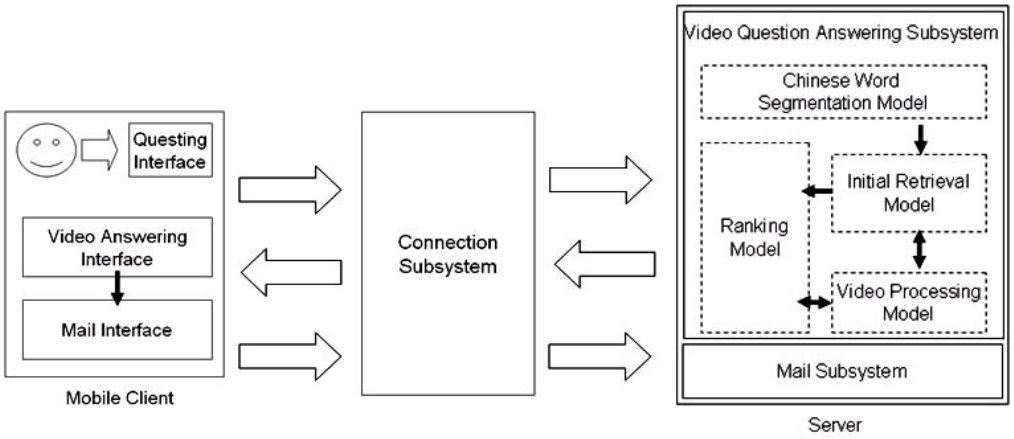

Figure 1. System architecture

The Mobile Client has three interfaces: Questing Interface, Video Answering Interface and Mail Interface. At first, a learner can input his question on Query Interface. Secondly, Video Answering Interface shows the browsing interface which contains top-10 retrieved answers, the captured frames, corresponding video clips and text descriptions. Meanwhile, the learner can directly watch the video on his PDA. If the learner wants to keep answers, he can annotate what he wants. Finally, Mail Interface associates with Mail subsystem on server send the learner information. Figure 2 shows an example of query interface (A) and the retrieved answers (B).

(A) (B)

Figure 2. Application user interface

(A) Questing Interface (The question is "Where was Napoleon born?") (B) Video Question Answering Interface

3.2. Video question answering

The authors developed the Video Question Answering Subsystem on the basis of a robust passage retrieval algorithm in [8]. In this paper, the authors adopt the Discovery videos as learning materials which is in Chinese. Unlike most western languages, there is no explicit word boundary between Chinese words. To solve this, the authors simply tokenize the words according to the atomic characters. This strategy was shown very effective for video QA as against to well-trained Chinese word segmentation tools [8]. Then, the separated words will input the Initial Retrieval Model. Initial Retrieval Model generate relate documents by the traditional search algorithm. Next, the results are ranked by a robust passage retrieval algorithm, and return to users. Initial Retrieval Model and Ranking Model associate with Video Processing Model retrieval multimedia data. The source of Video Processing Module is about 181 Discovery films which is about 130 hours. Our description comes from optical character recognition (OCR) technology. Therefore, learners also can read captions instead of key words as YouTube (http://www.youtube.com).

3.3. Annotation and email

The screen of handheld device is too small to read, and the operation of handheld device is also too difficult to manipulate. For those reasons, the system provides the mechanism of annotation with an email. Learners can annotate what they want in order to collect data. Whenever they get the answers that they asked before, they can just browse them and choose what they want. Also, the system will transmit them to their mailbox (see Figure 3). Therefore, learners can collect the information what they query, and manage their knowledge.

Figure 3. An illustration of email content

4. Pilot study

4.1. Method

A pilot study was conducted to observe the system usage. Ten under-graduated students participated in the study. Before the experiment, the subjects were trained how to use PDA and the system in order to let them familiar with the manipulation of PDA and the system. In the process of the study, the subjects could query questions by themselves at least three. After the activity, the subjects were asked to fill out a questionnaire and interviewed for asking their thoughts about the system. The items for all the questions ranged from 1 (strongly disagree) to 5 (strongly agree). The questions of interview depended on the condition of how the subjects operate the system. Any question could be added if necessary. In addition, how the subjects operated the system is recorded by a camera. How many times the subjects had clicked the photo shoot up to watch video clips (called Watching Rate) and how many results they had annotated to mail (called Annotation Rate) are also recorded.

4.2. Results and discussion

Table 1 shows the average score of the questionnaire. The result shows that learners evaluated high score on all items. Figure 4 shows the percentage of Watching Rate and Annotation Rate of every subject. Those data were divided into three categories for further discussion: the aid of video question answering, the aid of annotation, and the aid of mobile learning.

The aid of video question answering

According to the results of Q1, Q2 and the interview, the authors found that the greater part of subjects considered that video question answering is a better approach for learning, because they liked to get information in a multimedia way. Only one object does not agree that. He answer: "Video question answering skill is different from traditional search. They have distinct purposes, and get distinct information flow". Consequently, he felt that they cannot be compared. On the other hand, according to Watching Rate on the Figure 4, it can be found that learners use the function of watching video clips to understand what they ask. The subject A even watched every result. Thus, it is useful for learner when asking. In the interview, it is found that subjects felt that multimedia content-based results are the faster approach to know what they want than text-based results. In addition, since PDA is not as well as desktop PC, the QA performance is quite important. The higher rank the actual answer achieves the less time user spends on browsing. For the above-mentioned phenomena, video question answering is positive.

Table 1. The average score of the questionnaire

No.	Questions	Mean
Q1	Does the answer to your query which our system provides meets your needs?	4.2
Q2	Is the multimedia information provided by our system more helpful than traditional literal description for you?	4.1
Q3	Is the function of annotation beneficial for you to collect information?	4.2
Q4	Is the function of Email beneficial for you to collect information?	4.4
Q5	Do you want to use our system when you have question in learning?	4.4
Q6	Do you think that our system is beneficial for you to learn?	4.3
Q7	Do you want to own our system in the future?	4.1

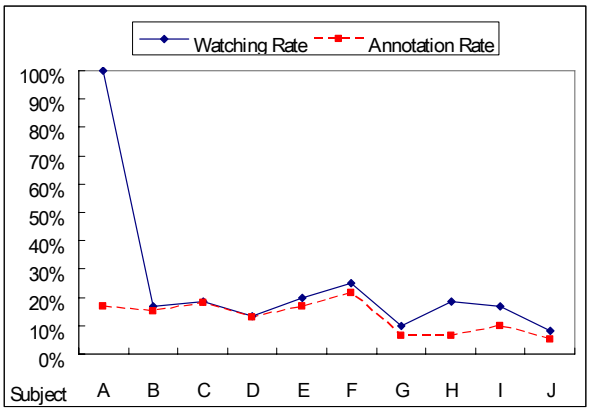

Figure 4. The percentage of Watching Rate and Annotation Rate

The aid of Annotation

According to the results of Q3, Q4, it is found that more than half of subjects liked the functions of annotation and email. They thought it solves the disadvantage of handheld device like the small screen. They thought that it is easier to collect data and manage information. According to the result of interview, a learner would like to annotate the video clips in the top-10 retrieved answers. Videos outside the top-10 usually were not selected. The result show that it has positive correlation between Watch Rate and Annotation Rate. That is why learners consent to the effect of annotation and email. In the study, most learners intend to read every mail and keep it, only few learners do not like receive mail because they are afraid of spam.

The aid of mobile learning

In the subjects' opinions, the system is good for learners to learn. Whenever a learner wants to know something, wherever a learner wants to find something, he will never lose any chance to learn. Due to the portable characteristic of mobile devices, it expands the space and time of learning, it makes learners continue their experience around different environments, and it reaches seamless learning process. Because handheld device is small and convenient, learners can use the system when they are not around desktop. They expressed that they will use when they exhibit, or pick up film, or do their homework, even they are bored. For those reasons, it can find out that the system is helpful.

5. Conclusion and future work

In this paper, the authors proposed a mobile video question answering annotation system, which aims to provide multimedia answers for questions under wireless technology. When input a question on PDA, the system responses the retrieved answers together with videos from the server. Afterwards, video question answering subsystem will deal with the question of the mobile client, and output ten text-based and video-based results for the learner. After getting the responses, the learner can browse answers using up-down buttons and watch video clips with PDA stylus. The learner can annotate notes and the system keeps and send them back as emails to inform learners

As the future work, the authors will integrate the speech recognition techniques, which improve the inconvenience of input. The authors plan to take synchronize connection method, which can minimize the waiting time of transmission. Finally, an effective algorithm to improve OCR errors will be added.

6. References

[1] Bruner, J. S., *The process of education*, Cambridge, MA: Harvard University Press, 1960.

[2] S. Benford, D. Rowland, M. Flintham, R. Hull, J. Reid, J. Morrison, K. Facer and B. Clayton, "Savannah: Designing a location-based game simulating lion behaviour", *Proceedings of Conference on Advances in Computer Entertainment 2004,* 2004.

[3] Y.S. Chen, T.C. Kao, J.P. Sheu, and C.Y. Chiang, "A Mobile Scaffolding-Aid-Based Bird -Watching Learning System", *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'02)*, 2002, pp. 15-22.

[4] S. His, "The electronic guidebook: A study of user experiences using mobile web content in a museum setting", *Proceedings of the IEEE Intl. Workshop on Wireless and Mobile Technologies in Education (WMTE)*, 2002.

[5] H. Ogata, Y. Yano, "Context-aware support for computer-supported ubiquitous learning", *Proceedings of IEEE International Workshop on Wireless and Mobile Technologies in Education*, pp. 27–34.

[6] E. Robertson, S. Walker, and M. Beaulieu, "Experimentation as a way of life: Okapi at TREC", *Journal of Information processing and management*, vol. 36, 2000, pp. 95-108.

[7] J. Roschelle, "Unlocking the learning value of wireless mobile devices", *Journal Computer Assisted Learning*, 2003, pp.260-272.

[8] Y.C. Wu, Y.S. Lee, J.C. Yang, and S.J. Yen, "A New Passage Ranking Algorithm for Video Question Answering", *Advances in Image and Video Technology*, 2006, pp.563-572.

[9] G. Zurita, and M. Nussbaum, "Computer supported collaborative learning using wirelessly interconnected handheld computers", *Computers and Education*, 2004, pp.289-314.