Automatic Multimedia-based Question-Answer Pairs Generation in Computer Assisted Healthy Education System

Yang-Yen Ou*, Shu-Wei Chuang, Wei-Chun Wang, and Jhing-Fa Wang
Department of Electrical Engineering, National Cheng-Kung University, Tainan, Taiwan

2022 10th International Conference on Orange Technology (ICOT) | 978-1-6654-5419-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICOT56925.2022.10008119

***Abstract*—Automatic question-answer pairs generation is becoming more and more important, reducing the workload and time of manual question generation. We developed the automatic question-answer pairs generation system for patient education, which can be converted into question-answer pairs in matching video content according to different patient education videos. These question-answer pairs can fully help patients understand pre-and post-operative patient education, by collecting a large number of existing patient education videos in the hospital. The retrieval-based question answering system can query the questions and answers database established by the question-answer pair generation system. The system we propose is mainly divided into three parts. The first part is the text generation (TG) module, which processes the video into text and provides the data required by the subsequent modules. The second part is the answer extraction (AE) module, which aims to extract entities and nouns in the text as candidate answers. The third part is a BART-based question generation (QG) module, which generates corresponding questions by inputting sentences including answers.**

***Index Terms*—Question-Answer Pairs, Question Generation, Named Entity Recognition, Question Answering System, BART**

I. Introduction

In recent years, the medical staff is exhausted physically and mentally and the workplace is faced with a large number of confirmed new crowns, resulting in increased care pressure and burden. Since medical manpower is extremely precious, to effectively reduce the workload of medical staff, there are many service robots used in hospitals on the market. This kind of transportation robot is a robot that can deliver medicines, small medical devices, documents, specimens, blood, blood samples, X-rays, dressings, prescriptions, office supplies, and other small items, as well as medium or large volumes such as medical waste. However, there are very few in specific areas of professional knowledge, such as the necessary and highly repetitive patient education work. Because it is difficult to collect a large amount of professional field data, resulting is impossible to establish a question-answer pairs database in a specific field. One of the most critical challenges in question-answer pairs database is the scarcity of labeled data since getting question-answer pairs with human-annotated text is very expensive. Our system[1] can thus improve this problem based on context or automatically generated high-quality question-answer pairs from large amounts of unstructured text.

[1]This work was supported in part by the National Science and Technology Council of Taiwan, R.O.C., under Grant MOST 111-2221-E-006 -177 -MY2

Nowadays, AI technology is booming, and different industries have their own smart customer services, such as bank smart customer service, and shopping guide smart customer service. In order to train a customer service robot in a specific field, it is necessary to use high-quality and a large number of business scenario question-and-answer pairs as a corpus for training, but sorting out the question-answer pair corpus will consume a lot of labor costs. The question generation (QG) task is to automatically generate appropriate questions from text passages. The most straightforward approach is to generate answer-aware questions [1]. In answer-aware question generation, the model displays an answer and a paragraph and asks to generate a question for that answer by considering the paragraph context. Although there are many papers available for the QG task [2], it is still not as mainstream as QA. One of the reasons is that most of the early papers used complex models and processing pipelines and no pre-trained models were available. Several recent papers [3] [4] [5] [6], especially UniLM and ProphetNet, have few SOTA pretrained weights available for QG, which seem rather complicated to use. Also, since answer-aware models need answers to generate questions, answer extraction needs something that can extract answers from the text. This can be done using various methods such as ner, noun phrase expansion, etc.

We collect a large number of medical and health education videos to build a database of question-answer pairs and cooperate with the retrieval-based question answering system to achieve a FAQ for health education, to help medical staff reduce their workload. This system is mainly divided into three parts. The first part is the text generation module, which processes the video into text and provides the data required by the subsequent modules. The second part is the answer generation (AG) module, which aims to extract entities and nouns in the text as candidate answers. The third part is a BART-based question generation (QG) module [7], which generates corresponding questions by inputting sentences including answers.

II. Proposed System

A. System Overview

The patient education video is obtained through the text generation module to obtain the video content text and restore the punctuation marks. The recipient uses the answer generation (AG) module to find suitable answers as candidates, and

TABLE I
THE EXAMPLE OF PUNCTUATION RESTORATION

Original Text	Processed Text
大家好我是奇美醫院護理師小林今天要向大家介紹空腸造口灌食的方式因為阿熊先生目前還不能從嘴巴進食必須先由空腸造口灌食來補充營養所以現在要跟你們衛教如何從空腸造口灌食過程中有任何疑問都可以隨時提出來詢問後續若有換人照顧的話也請你要將注意事項作交接喔	大家好我是奇美醫院護理師小林，今天要向大家介紹空腸造口灌食的方式，因為阿熊先生目前還不能從嘴巴進食必須先由空腸造口灌食來補充營養，所以現在要跟你們衛教如何從空腸造口灌食，過程中有任何疑問都可以隨時提出來詢問，後續若有換人照顧的話也請你要將注意事項作交接喔。

finally enters a sentence including the answer. The question generation (QG) module produces corresponding problems, and the overall system architecture diagram is shown in Fig. 1.

B. Text Generation Module

We use the audio converter to convert the collected educational videos into audio files, and automatically convert the voice into text through google automatic speech recognition (ASR). Since there are no correct paragraph sentences in the text, the generated question is irrelevant to the answer. We will improve this with punctuation recovery handling [8]. Add punctuation marks in text paragraphs to separate all sentences to facilitate subsequent task processing. Examples of text generation are shown in TABLE I.

C. Answer Extraction Module

We adopt the method of answer-aware question generators, so we need to find out the entities in the text that are suitable as candidate answers. To analyze the entities from the text, we use the tool of ckip tagger [9] to do segment words, part-of-speech tagging, and named entity recognition. Common types of entity recognition include 1.CARDINAL, 2.QUANTITY, 3.TIME, and regard them as candidate answer categories, which provides the question generation (QG) module for question generation. The example of named entity recognition is shown in TABLE II.

D. Question Generation Module

Before generating a question, we have to find the sentence that contains the answer, so use punctuation in the text and the paragraph search module to find the sentence to input to the

TABLE II
THE EXAMPLE OF NAMED ENTITY RECOGNITION

Named Entity Type	Examples
CARDINAL	翻身前準備2-3個枕頭或棉被
QUANTITY	灌食完畢倒入20-30毫升温開水沖進管內剩物
TIME	水溶性優碘消毒完之後過30秒有消毒的作用

question generation (QG) module. Next, we use a pre-trained language model BART-based question generation module [10] to generate questions relative to the answers and finally write the answers and questions back to the database. The question answering system can query the database generated by the question-answer pair generation module. The example of named entity recognition is shown in TABLE V. The input format highlights the sentence with ¡hl¿ tokens if it has an answer.

E. Question Answering System

Frequently asked question (FAQ) belongs to a retrieval-based question answering system. The question answering system matches the user's questions with the question set of the database. User questions are often short text sentences. The library has one fixed answer and question, as well as multiple extended questions and keywords. What the model needs to solve is to find the most acceptable answer to the user among the standard questions for a given query.

III. EXPERIMENTAL RESULTS

A. Experiment of Named Entity Recognition for Proposed System

To fine-tune the NER model, we process the dataset with a BIO joint labeling approach. The B label represents the beginning of a word, the I label represents the middle of a word, and the O label represents not a word. We have a total of 7 classes of labels B-CARDINAL, I-CARDINAL, B-QUANTITY, I-QUANTITY, B-TIME, I-TIME, and O. For evaluation, NER uses the following metrics:

The precision is the ratio between correctly identified positive values and all identified positive values. The precision metric shows the number of correctly labeled predicted entities.

$$\frac{\text{number of true positive}}{\text{number of true positive + number of false positive}} \tag{1}$$

The recall is the ratio between the predicted true positives and what was actually tagged. The recall metric reveals how many of the predicted entities are correct.

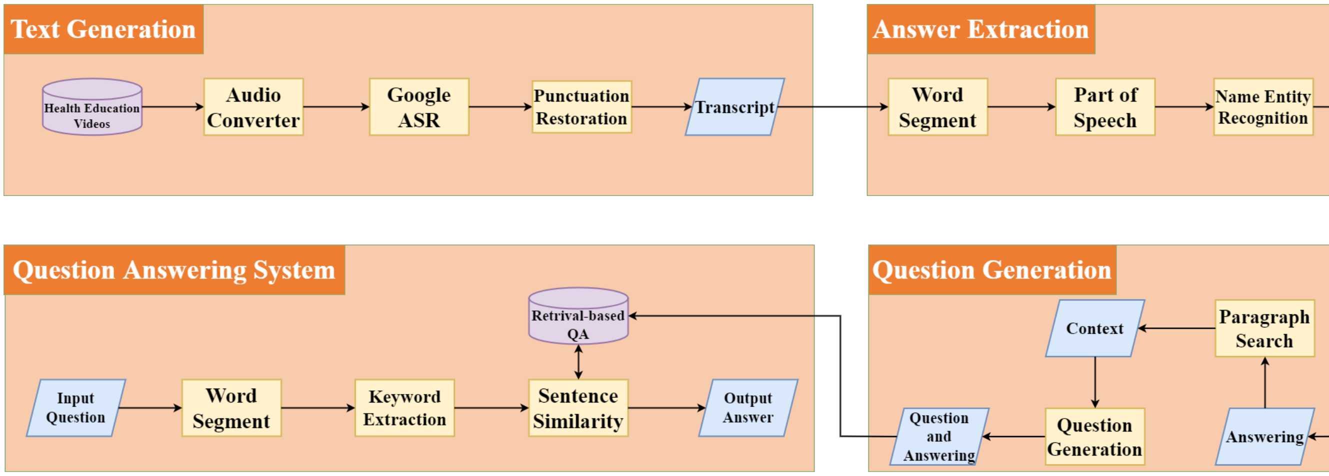

Fig. 1. Overall block diagram

TABLE III
THE EXPERIMENTAL RESULTS OF NAMED ENTITY RECOGNITION

Model	NER(F1)
CKIP BERT Base (without fine-tune)	81.18%
CKIP BERT Base (fine-tune)	83.47%

TABLE IV
THE MOS SCORE OBTAINED BY THE OVERALL SYSTEM IN EACH EVALUATION ITEM

System Description	Average Score
Immediacy	4.5
Accuracy	4.2
Reliability	4.2
Fluency	4.4
Average of whole system	4.32

$$\frac{\text{number of true positive}}{\text{number of true positive + number of false negative}} \tag{2}$$

The F1 score is a numerical value used to express the balance between precision and recall. It is a function of precision and recall.

$$\frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3}$$

The results show that our again fine-tuned model is 2.29% higher than the original model. This is because the number of predicted entities is reduced. As an answer candidate in the field of health education, we only need to identify CARDINAL, QUANTITY, and TIME. The experimental results are shown in TABLE III.

B. Experiment of Question Answering System

To understand the immediacy, accuracy, reliability, and fluency of our question-answer pairs generation apply to question answering system. We use Mean Opinion Score (MOS) to evaluate the experience of users. The MOS is usually represented as a rational number in the range 1 to 5, where 1 is considered the lowest evaluation score and 5 is the highest evaluation score.

The result of MOS was executed with fifteen participants to evaluate the overall performance of the proposed system. Every participant was asked to fulfill the MOS evaluation after interacting with our proposed system. We use four different aspects to evaluate our system and the experimental results are shown in TABLE IV. The average MOS score we get is 4.32, this means that our system can give users a good experience in practical applications.

IV. CONCLUSION

We propose a system for video generation question-answer pairs that will help in building the database. Improving the question of manually labeling question and answer pairs is beneficial to the subsequent use of retrieval-based question answering systems.

We fine-tune the model to do the entity recognition task F1 up to 83.47 and build the FAQ question answering system with the dataset of question-answer pairs. Finally, the average MOS score of our system is 4.32, representing that our system can give users an excellent experience in practical applications.

REFERENCES

[1] A. Kumar, A. Kharadi, D. Singh, and M. Kumari, "Automatic question-answer pair generation using deep learning," in *2021 Third International*

TABLE V
EXAMPLES OF QUESTION GENERATED USING THE QUESTION GENERATION MODULE

Answer-aware Question Generation	
Health Education Paragraph: 大家好我是奇美醫院護理師小林，今天要向大家介紹空腸造口灌食的方式， …… …… 先將灌食袋導管管夾關閉後到入50毫升溫開水，將灌食袋掛於點滴架上高於造口40-50公分，打開導管管夾排空空氣於彎盆內，將配方奶置入袋內將冰塊袋置放於外袋以防配方奶腐臭，反折廔管移除綠色尿管塞子將管口與廔管接上，將灌食袋滴室架於機器滴室架上輕拉軟管處向下順著滾輪旋轉在管路置放處，按壓灌食機器on控制鈕開機，打開灌食袋導管管夾後按壓start控制鈕，機器滾輪開始轉動即開始灌食。 灌食完畢倒入20-30毫升溫開水沖進管內剩物避免殘存在導管內的食物腐敗，按壓機器後的控制鈕以暫停機器轉動，反折造廔管移除灌食袋管口將綠色尿管塞子塞住造廔口，將灌食袋從重灌食機器上移除以清水洗淨灌食袋後晾乾以備下次灌食使用。空腸造口管灌食注意事項:1.灌食過程每4小時需以灌食空針灌注50毫升溫開水沖洗造廔管預防管路阻塞、2.需適時檢視冰塊袋是否有融化需更換情形、3.攪拌配方奶粉時需攪拌均勻避免結塊導致管路阻	
Example 1	
Input Format:	先將灌食袋導管管夾關閉後到入[HL]50毫升[HL]溫開水
Output Question:	灌食袋導管管夾關閉後到入多少毫升溫開水
Example 2	
Input Format:	將灌食袋掛於點滴架上高於造口[HL]40-50公分[HL]
Output Question:	灌食袋的高度高於造口多少公分
Example 3	
Input Format:	灌食完畢倒入[HL]20-30毫升[HL]溫開水沖進管內剩物避免殘存在導管內的食物腐敗
Output Question:	灌食完畢到入水量多少毫升
Example 4	
Input Format:	灌食過程每4小時需以灌食空針灌注[HL]50毫升[HL]溫開水沖洗造廔管預防管路阻塞
Output Question:	灌食過程每4小時需以灌食空針灌注多少毫升

Conference on Inventive Research in Computing Applications (ICIRCA), 2021, pp. 794–799.

[2] Y. Sun, D. Tang, N. Duan, T. Qin, S. Liu, Z. Yan, M. Zhou, Y. Lv, W. Yin, X. Feng, B. Qin, and T. Liu, "Joint learning of question answering and question generation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 5, pp. 971–982, 2020.

[3] D. B. Lee, S. Lee, W. T. Jeong, D. Kim, and S. J. Hwang, "Generating diverse and consistent qa pairs from contexts with information-maximizing hierarchical conditional vaes," *arXiv preprint arXiv:2005.13837*, 2020.

[4] B. Yao, D. Wang, T. Wu, T. Hoang, B. Sun, T. J. Li, M. Yu, and Y. Xu, "It is ai's turn to ask human a question: Question and answer pair generation for children storybooks in fairytaleqa dataset," *CoRR*, vol. abs/2109.03423, 2021. [Online]. Available: https://arxiv.org/abs/2109.03423

[5] H.-T. Su, C.-H. Chang, P.-W. Shen, Y.-S. Wang, Y.-L. Chang, Y.-C. Chang, P.-J. Cheng, and W. H. Hsu, "End-to-end video question-answer generation with generator-pretester network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4497–4507, 2021.

[6] X. Ma, Q. Zhu, Y. Zhou, and X. Li, "Improving question generation with sentence-level semantic matching and answer position inferring," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 8464–8471.

[7] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[8] paddlepaddle, "punctuation recovery," http://www.paddlepaddle.org/.

[9] P. Li, T. Fu, and W. Ma, "Remedying bilstm-cnn deficiency in modeling cross-context for NER," *CoRR*, vol. abs/1908.11046, 2019. [Online]. Available: http://arxiv.org/abs/1908.11046

[10] P. Huang, "Transformer qg on drcd," https://huggingface.co/p208p2002.