Chinese Sentences Composition for People with Dysarthria

Ren-Sheng Jhong
Dept. of CSIE
National Taipei University
New Taipei City, Taiwan R.O.C
leomustdie93107@gmail.com

Chun-Yao Hsu
Dept. of CSIE
National Taipei University
New Taipei City, Taiwan R.O.C
justin20067@gmail.com

Yue-Shan Chang
Dept. of CSIE
National Taipei University
New Taipei City, Taiwan R.O.C
ysc@mail.ntpu.edu.tw

Hung-Shing Huang
Dept. of CSIE
National Taipei University
New Taipei City, Taiwan R.O.C
msn9110@gmail.com

I. INTRODUCTION

Automatic Speech Recognition (ASR) [5] is a challenging task that requires considerable effort. As well-known that traditional machine learning and deep learning approaches used in the ASR always require large amounts of training data, which may be up to several tens million or hundreds million training data set. This will be a big challenge for the people with dysarthria because their speech intelligibility is poor and cause them to seldom communicate with others.

Although many deep learning methods have been developed to compose meaningful sentences, such as BERT, there are still some problems for compositing Chinese sentence [1] for dysarthria patient. Most of characters cannot be accurately recognized. Therefore, in order to compose meaningful Chinese sentences [3], in this paper we propose a framework and algorithms to compose Chinese sentences. The steps are presented as follows: First of all, we collected the sentences that user said before to construct the N-Gram corpus [2] through CKIP (Chinese Knowledge and Information Processing) [1], especially the 2 to 4 grams of Chinese words, which is used as the corpus for 2-4 grams recognition. Here one gram is a single Chinese word. Secondly, the top 50 recognized characters of a certain voice by our previously speech recognition system [4] are used to generate N-Gram Candidate Matrix (NG-CM) through the N-Gram corpus. Thirdly, we also construct the History Successive Word Dictionary (HSWD) through the sentences user said before. The HSWD show the relationship between terms (words). Finally, we develop algorithms to exploit NG-CM and HSWD to compose candidate Chinese sentences for user.

While a user speaks unclear and unintelligibility sentence, the acoustic model (AM) [4] recognizes the voice and generate 50 possible words for a certain voice. After the AM generate all candidate words of certain voices, the system will, based on the N-gram repository, generate N-gram candidate matrix for the candidate words. Next, the sentence composition module will utilize the HSWD and N-gram candidate matrix to compose 10 candidate sentences to the user. And then user can select the correct sentence and feedback to the HSWD and N-gram repository for improving the composition accuracy in the future.

A. System flow

In order to further illustrate the function of the whole system, this section describes the flow of the system, as shown in Fig. 1. We use the speech recognition function in [4], this system will first convert the voice of dysarthria patients into spectrogram images through MFCC (Mel-scale Frequency Cepstral Coefficients). Since the sentences spoken by the patients include many voices of words, we exploit a named voice tailoring module to cut each voice into a single voice file. Finally, the AM is utilized to recognize the voice file into 50 candidate voices of the word. The reason for generating 50 candidate voices is that the accuracy of this AM at top 50 can reach 95%. If the accuracy can be improved in the future, the candidate voices of the word can be reduced. For details of this part, please refer to [4].

Due to the characteristics of Chinese, each Chinese word is only a voice, and each term may contain 2-4 words, for example, the " Taipei" is one word in English, but it is a term including two words "台(Tai)" and "北(Pei)". When this recognize the voice may be "台(Tai)" followed by "北(Pei)", it should be the term "台北". Therefore, we can first store "台北(Taipei)" into 2-gram in the N-gram repository. Here we explain the N-gram candidate matrix mechanism. Assuming that the sentence has 6 words, denoted as V, the 1-Gram means only contain one word. The AM will split and recognized the voices into V_1, V_2,…,V_6. The 2-gram is to group 2 voices into together, such as V_1V_2, V_2V_3 … V_5V_6, while 3-gram is $V_1V_2V_3$ … $V_4V_5V_6$, and 4-Gram is $V_1V_2V_3V_4$, $V_2V_3V_4V_5$, $V_3V_4V_5V_6$. The system combines the generated N-Gram Candidate Matrix with the HSWD (historical successive word dictionary) for the Sentence Composition Module to form a complete Chinese sentence. How to build N-gram repository and N-gram candidate matrix will be described in the following section.

After the system recognizes all of candidate voices by means of the AM module, the system will use the N-gram repository to generate the N-gram candidate matrix (NG-CM), especially 2 to 4 grams. The step will adopt the hash table

[1] https://ckip.iis.sinica.edu.tw/

function to map N recognized candidate voices to N-gram repository. If the mapping is success, the N candidate voices will be recognized a term and will be stored into the NG-CM to be as a candidate term of the sentence. For example, if a two candidate voices "台(Tai)" followed by "北(Pei)", and the grouped two voices maps to N-gram repository success, the term "台北(Taipei)" will be stored into NG-CM. After the NG-CM is generated, the system will call a sentence composition module and exploit HSWD to compose the candidate sentences. The algorithm will be presented in next subsection.

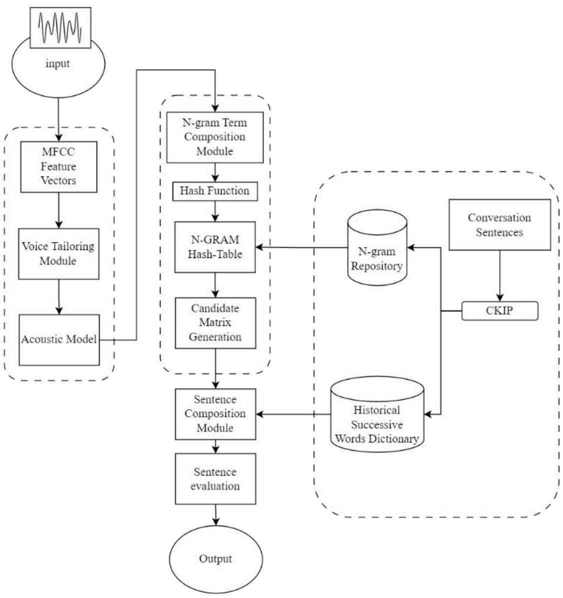

Fig. 1. System follow chart

B. N-gram Candidate Matrix

The N-gram Candidate Matrix (NG-CM) is the important component to compose the Chinese sentence in the work. As shown in Fig. 1, it is built from the result of AM and mapping the recognized grouped candidate voices to the N-gram repository. In the NG-CM, 1-gram candidate matrix is directly built from the AM. And 2-4 grams are built by means of searching the words, which is associated with the grouped candidate voices. Here we introduce two functions can be used to facilitate the process. These are *Search-Words(V, H)* and *Filter(S)*. The *Search-words(V，H)* can be used to search a candidate term regarding to the grouped voices (V) using a determined hashing function (H) from N-gram repository and the *Filter(S)* can be used to eliminate any duplicate or invalid elements from the candidate words list (S). Finally, the candidate term obtained from *Filter(S)* will be put into proper location of NG-CM.

We conduct experiments for evaluating the accumulated WCR for Top 1 to Top 10 generated candidate sentences. We also define a Top N accumulated WCR for evaluating the performance of the experiments. It means that the average accumulated WCR from generated candidate sentences top 1 to N. The result of the experiments is shown in Table I. The experiments directly allow patient to speak a sentence of 3 to 10 words on a tablet and tests the result. The results show that the less words in a sentence, the high accumulated WCR is. Similarly, the high top-N in a sentence, the high accumulated WCR is. In addition, the top 10 accumulated WCR of 10 words sentence is 80 %.

TABLE I. THE AVERAGE ACCUMULATED WCR OF A VARIETY OF EXPERIMENTS

N-word / TopN	3-word	4-word	5-word	6-word	7-word	8-word	9-word	10-word
Top1	78%	78%	70%	67%	63%	70%	63%	56%
Top2	89%	85%	82%	82%	70%	78%	75%	64%
Top3	92%	87%	84%	85%	74%	81%	77%	69%
Top4	92%	87%	86%	86%	75%	82%	78%	74%
Top5	93%	88%	89%	87%	76%	83%	79%	75%
Top6	93%	89%	90%	87%	76%	84%	79%	75%
Top7	93%	90%	90%	88%	76%	85%	81%	76%
Top8	94%	90%	90%	88%	76%	85%	81%	76%
Top9	94%	90%	90%	88%	76%	85%	82%	79%
Top10	94%	90%	90%	88%	76%	85%	82%	79%
Top11~50	97%	92%	93%	89%	80%	86%	84%	80%
No found	3%	8%	7%	11%	20%	14%	16%	20%
AllTop10	94%	90%	90%	88%	76%	85%	82%	79%

This work we have proposed a N-gram based Chinese Sentences Composition method for people with dysarthria. We exploit the N-gram Repository to build the N-gram candidate matrix. In addition, we also propose a HSWD to record the historical successive words. Using the HSWD and N-gram candidate matrix (NG-CM), the sentence composition module can compose the candidate terms retrieved from NG-CM into candidate sentences. We also conducted some experiments to evaluate the performance of the proposed method. The results reveal that the proposed method can compose good enough candidate sentences for the people with dysarthria.

ACKNOWLEDGMENT

This work was finance supported by National Taipei University under granted number 112-NTPU_ORD-F-004.

REFERENCES

[1] Li, Tangqiu; Nyberg, Eric; Carbonell, Jaime G. (2018). Chinese Sentence Generation in a Knowledge-Based Machine Translation System. Carnegie Mellon University. Journal contribution. https://doi.org/10.1184/R1/6621353.v1

[2] Haidar, M. A., & O'Shaughnessy, D. (2012, December). "Topic n-gram count language model adaptation for speech recognition". In 2012 IEEE Spoken Language Technology Workshop (SLT) (pp. 165-169). IEEE.

[3] Khosmood, Foaad, and Robert Levinson. "Combining Corpus-Based Features for Selecting Best Natural Language Sentences." 2011 10th International Conference on Machine Learning and Applications and Workshops. Vol. 2. IEEE, 2011.

[4] B. -Y. Lin, H. -S. Huang, R. -K. Sheu and Y. -S. Chang, "Speech Recognition for People with Dysphasia Using Convolutional Neural Network," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 2164-2169, doi: 10.1109/SMC.2018.00372.

[5] S. Alharbi et al., "Automatic Speech Recognition: Systematic Literature Review," in IEEE Access, vol. 9, pp. 131858-131876, 2021, doi: 10.1109/ACCESS.2021.3112535.