論文 / 著書情報
Article / Book Information

論題(和文)	音声認識における効率的なエラー訂正のための言語モデル
Title(English)	Language Model for Efficient Error Correction in Speech Recognition
著者(和文)	ユアンリアン, 篠田浩一, 古井貞熙
Authors(English)	Yuan Liang, Koichi Shinoda, Sadaoki Furui
出典(和文)	日本音響学会2012年春季研究発表会講演論文集, Vol. , No. , pp. 89-90
Citation(English)	2012 Spring Meeting ASJ, Vol. , No. , pp. 89-90
発行日 / Pub. date	2012, 3

Language Model for Efficient Error Correction in Speech Recognition*

Yuan Liang, Koichi Shinoda, Sadaoki Furui
(Tokyo Institute of Technology)

1 Introduction

In recent years speech input interface is becoming popular in smart phone applications [1]. In this interface, speech recognition errors are unavoidable [2]. Users need to correct them manually, which is bothering. Efficient error correction interfaces are strongly demanded.

In most error correction interfaces, when a user finds an error word in the recognition result, he/she first marks it and then either selects the correct word from a candidate list provided by the interface, or inputs the correct word by handwriting (Fig. 1). The candidate list is usually obtained from a word confusion network (CN) [4] generated by a speech decoder.

In this study, we focus on improving the rank of the correct word in the candidate list in order to provide an error correction interface easier to use. We effectively utilize a fact that the other words before/after the error word are correct.

2 Assumptions in Error Correction Interface

Let us consider a situation when a user marks one word as an error word. Our interface asks a user to correct errors from left to right in the recognized sentences and not to go back to the preceding region. Then, we can safely assume that all the words preceding the error word are all correct or already been corrected.

In this study, we do not deal with the case when two successive words are both errors. Our interface waits before starting the error correction process until the time when the user finishes marking the other error word in the succeeding word sequences. Then, we can safely assume that the next word to the error word is correct.

Consequently, we assume that all the previous words and one following word of the error word are correct. We use this to improve the rank of the correct word in CN. We deal with only substitution and deletion errors in this study.

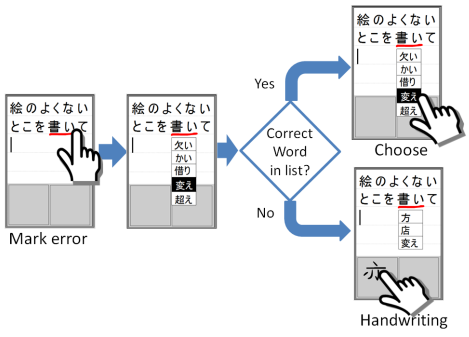

Fig. 1 Error Correction Interface

3 Re-ranking in CN

We utilize a word graph generated by the speech decoder for the re-ranking. Let $w_1, \ldots, w_N$, be a recognized word sequence (sentence), and w_i be an error. First, among the numerous hypotheses in the graph, we select those hypotheses in which $w_1, ..., w_{i-1}$ and w_{i+1} are identical with the recognized result. Then, we reconstruct a new CN from the resulting word graph. We expect the rank of the correct word in this CN is higher than its original rank

4 Experiments

4.1 Experimental setup

We used the Corpus of Spontaneous Japanese (CSJ) [5] for our experiments. 953 academic lectures having 228 hours length were used for training acoustic HMMs. They were three state left-to-right tri-phone HMMs. The transcribed text of 2496 lectures (953 academic lectures and 1543 simulated lectures) was used for training a language model. It consisted of back-off trigrams with a vocabulary of 102k words. The T^3 decoder [6] was used for recognition.

For our evaluation of the re-ranking method, from

*音声認識における効率的なエラー訂正のための言語モデル
ユアン リアン、篠田浩一、古井貞熙 (東工大)

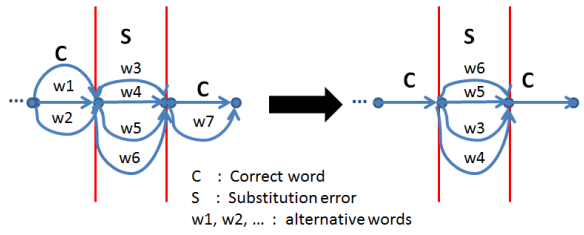

Fig. 2 CN re-ranking for substitution errors

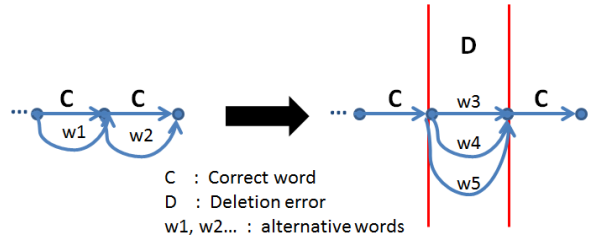

Fig. 3 CN re-ranking for substitution errors

Table 1 The number of correct words whose ranks become better/equal/worse after re-ranking

Better	Equal	Worse
66	986	15

the recognition results of 6563 sentences in the CSJ test set, we extracted 911 sentences in which only one substitution error existed and the other words were all correct, and 156 sentences in which only one deletion error existed and the other words were all correct. The total number of sentences was 1067. In order to compare the ranks of the correct words in CN before and after the re-ranking, we need to know the original rank of the correct word. For a deletion error, we use the rank of the "null" word as the original rank.

4.2 Results

Table 1 shows the comparison of the ranks of correct words in CN before and after the re-ranking. We found that our re-ranking method was effective for improving the ranks of correct words.

We also evaluated our method by Mean Reciprocal Rank (MRR). MRR is a statistic for evaluating any process that produces a list of possible responses to a query, ordered by the probability of correctness:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{r_i} \quad (1)$$

where r_i is the reciprocal of the rank of the correct word in a candidate list. Q is the number of different cases and set to 81 in this experiment. The results in Table 2 show that our method improved the rank of correct words by 22.6%.

Table 2 MRRs before and after the re-ranking

	Before	After
MRR	0.31	0.38

5 Conclusion

This paper has presented a re-ranking method of a correct word in the confusion network for correcting errors in speech recognition. We used not only preceding words to an error word, but also one following word to improve the rank of the correct word. Our evaluation confirmed effectiveness of our method.

In this study we investigated the case where a sentence has only one error word. In future, we are planing to extend our method to the case when multiple errors exist in one sentence. In this study we also assumed that all the preceding words and the following one word are correct. There may be many other kinds of assumptions that can be used for an error correction interface. Studies in this direction is also promising. We also would like to investigate multimodal error correction interfaces implementing the proposed techniques.

References

[1] K. Vertanen *et al.*, "Intelligently Aiding Human-Guided Correction of Speech Recognition", AAAI, 1698-1701, 2010

[2] P. O. Kristensson *et al.*, "Five Challenges for Intelligent Text Entry Methods", AI Mag., 85-94, 2009.

[3] A. Laurent *et al.*, "Computer-assisted transcription of speech based on confusion network reordering", ICASSP2011, 4884-4887, 2011

[4] L. Mangu *et al.*, "Finding consensus in speech recognition: word error minimization and other applications of confusion networks", Computer Speech and Language, 373-400, 2000

[5] S. Furui, "Spontaneous Speech Recognition and Summarization", HLT, pp.39-50, 2005

[6] P. R. Dixon *et al.*, "The Titech Large Vocabulary WFST Speech Recognition System", ASRU, 443-448, 2007