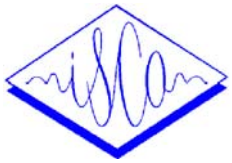

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

Improving Recognition Correct Rate of Important Words in Large Vocabulary Speech Recognition

Yasuo Shirosaki , Hideaki Kikuchi , Katsuhiko Shirai

School of Science and Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555 Japan
shiro@shirai.info.waseda.ac.jp

10.21437/Eurospeech.1999-131

ABSTRACT

The speech recognition technique has used in the various fields, with the progress of it and the needed functions are different for purposes. For example, speech dictators need to recognize whole sentences, but spoken dialogue systems need to recognize important words rather than whole sentences.
In this paper, we propose a method of modifying language model for improving recognition correct rate of important words. This method multiplied probabilities of unigram and bigram of important words by constant weight and made important words to easy to be recognized. And we examine differences of recognition result in changes of a set of important words. As a result, the correct rate of important words increased when we multiplied probabilities of important words by weight.

1 Introduction

A statistical language model are extensively used in speech recognition tasks various methods of modifying language model are studied to improve recognition result[1][2][3]. In this paper, we propose a method of modifying language model for improving recognition correct rate of important words in large vocabulary speech recognition. And we examine differences of recognition result in changes of a set of important words.
At first, we investigated the influence of change of language model on recognition result and we selected the best language model of them. Then we construct some specified word sets and change probability of unigram and bigram of specified words by adding constant weight. We studied the differences of correct rate and accuracy of specified words in change of weight. And also we studied the error rate that unspecified words were recognized as specified words. At last, we applied the proposed method to the spoken dialogue system for academic paper retrieval and confirmed the efficiency of it.

2 Modifying Language Model

In this paper, we suppose only unigram and bigram of words as language models. In this section, the method of modifying language model will be described.

2.1 Method of Modifying Unigram

When an unigram probability of word w_i is $p(w_i)$ and number of words is N,

$$\sum_{i=1}^{N} p(w_i) = 1 \tag{1}$$

The method of modifying language model is expressed as following equation:

$$p'(w_i) = \begin{cases} p(w_i) \times x & \text{(important word)} \\ p(w_i) & \text{(unimportant word)} \end{cases} \tag{2}$$

In this equation, x is a weight for important words and $p'(w_i)$ is an unigram probability multiplied by a weight x. After the calculation of the equation (2), unigram probabilities will be normalized to fill the equation (1) by the following equation:

$$p_{new}(w_i) = \frac{p'(w_i)}{\sum_{j=0}^{N} p'(w_j)} \tag{3}$$

: $p_{new}(w_i)$ is new modified unigram probability.

2.2 Method of Modifying Bigram

When a bigram probability of word (w_i, w_j) is $p(p_j|p_i)$ and number of words is N, sum of $p(w_j|w_i)$ is 1.

$$\sum_{j=1}^{N} p(w_j|w_i) = 1 \tag{4}$$

But, unfrequent combinations of two words were cut-off when making a language model with word bigram. So, in fact, sum of $p(w_j|w_i)$ is smaller than 1 as next equation:

$$\sum_{j=1}^{N} p(w_j|w_i) = m_i (0 <= m_i <= 1) \tag{5}$$

and a back-off weight of word $w_i(bo\_wt_i)$ plus m_i is 1.

$$\sum_{j=1}^{N} p(w_j|w_i) + bo\_wt_i = 1 \quad (6)$$

When the word w_j is an important word, we changed the probability $p(w_j|w_i)$. The method of modifying language model is same as it of modifying language model of word unigram.

$$p'(w_j|w_i) = \begin{cases} p(w_j|w_i) \times x & \text{(important word)} \\ p(w_j|w_i) & \text{(unimportant word)} \end{cases} \quad (7)$$

3 Used Data

3.1 Phoneme Model

The used phoneme model was trained as three thousand states triphone and eight mixture from speech database collected by reading ten thousand sentences of the article database[4]. Features used in training were 12 mel-frequency cepstral, 12 delta cepstral, and delta power.

3.2 Language Model

In this part, we investigate the influence of change of language model on recognition result and see the language model for the test.
We made the word dictionaries and the language models(word bigram) of five thousand frequently appeared words in the following three types of article databases: A) Articles of a year(20,854,007), B) Articles of four years(88,650,913), C) Articles removed non-sentences from B)(65,347,098)(The parenthesized numbers indicates number of morphemes.)[5].
As the result of recognition by using above mentioned three model, the correct rate of whole words are A)63.35%, B)72.85%, and C)76.02%. In the following test of modifying language model, we used C) because C) has a highest correct rate among three types.

3.3 Test Data

Test set was one hundred speech by twenty-three speakers from the above mentioned speech database but un-overlapped.

4 Experimental Results

In a practical system, it is expected that important words are specified according to tasks. But in this paper, we examined task-independent methods of specifying important words.

- Word Set 1 A hundred words frequently appeared in the word dictionary.
- Word Set 2 A hundred nouns (except proper nouns) frequently appeared in the word dictionary.
- Word Set 3 A hundred nouns (include proper nouns) frequently appeared in the word dictionary.
- Word Set 4 A hundred words unfrequently appeared in the word dictionary.

The results for each word set are shown in Table 1, and we investigate the results in following part.

4.1 Word Set 1

Not only whole correct rate but also specified word correct rate decreased in multiplying by a weight. And the accuracy decreased, too.
In this test, the specified words have large probabilities before modifying, so we can suppose that the words except specified words were recognized to specified words and both accuracy of whole and specified words decreased.

4.2 Word Set 2

The recognition errors to specified words increased, and the accuracy decreased. About the correct rate, though the whole correct rate decreased, the correct rate of specified words increased. Comparing with test 1, the decrease of whole accuracy was small because there were few short words like particles in the specified words.

4.3 Word Set 3

The specified words were similar to test 2, so the result was similar, too. The decrease of ths accuracy of specified words was larger than whole accuracy, either.

4.4 Word Set 4

The result of this test, It was different from test 1 - 3. Both the correct rate and the accuracy of whole and specified words increased. It is because that weighting on words which have small probabilities improved correct rate and accuracy of them. The words of which the original probability was small and the weighting on words which have small probabilities does not influence on increase of mis-recognition.

4.5 Discussion

From these results, we can find next tendencies about the correct rate and the accuracy. In increaseing weights, the correct rate of whole words, the accuracy of whole words, and the accuracy of specified

Word Set	weight	whole words		important words	
		correct rate(%)	accuracy(%)	correct rate(%)	accuracy(%)
1	1	73.92	53.39	71.33	66.39
	5	70.61	46.19	70.35	58.65
	10	69.29	43.87	69.03	54.70
	20	68.96	41.72	70.51	52.22
	50	66.23	36.59	70.18	43.99
	100	63.33	32.70	69.36	36.24
2	1	73.92	53.39	81.16	75.36
	5	72.43	52.32	85.51	71.01
	10	72.76	52.98	88.41	71.01
	20	71.69	52.57	88.41	56.52
	50	70.94	52.40	88.41	44.93
	100	68.79	50.75	91.30	27.54
3	1	73.92	53.39	82.05	75.64
	5	73.10	53.15	85.90	69.23
	10	72.35	52.81	85.90	64.10
	20	70.61	51.41	85.90	51.28
	50	69.62	51.16	88.46	41.03
	100	67.22	49.34	87.18	35.90
4	1	73.92	53.39	74.14	72.41
	5	75.17	54.80	79.31	76.72
	10	74.75	54.64	79.31	76.72
	20	74.34	54.30	80.17	75.00
	50	73.68	54.06	80.17	73.28
	100	74.34	54.39	83.62	75.00

Table 1: results of 4 word sets

words decreased, and only the correct rate of specified words increased.
When weighting important words, it was supposed that weighting words which have small probability is better than weighting words that have large probability in the accuracy. As to the correct rate, not only specified words but also whole words increased.
If the weight was made too large, recognition errors to specified words and insertion errors increased.

5 Application to Spoken Dialogue Retrieval System

At first, we made a language model for spoken dialogue retrieval system from 83 sentences of dialogue corpus. The conditions of retrieval are author and keyword. We replaced authors' names to "Author" and keywords to "Keyword". We made authors' list and keywords' list, and distributed probabilities of "Author" to words on authors' list and "Keyword" to words on keywords' list(language model (BASE)). Based on this model, we made another language model: (MODIFIED) probabilities of author were multiplied by weight(20).
Test set was 47 speech by one speaker and the sentences were same to above mentioned 47 sentences. We classified 47 sentences to three types: 1) informing author names as a condition of retrieval(14 sentences), 2) informing keywords as a condition of retrieval(15 sentences), 3) others(18 sentences).
The recognition results of these language models are shown in Table 2. In the result of recognizing sentence-type 1) by the modified language model, there is 2.5 point increase of accuracy though there is no decrease of recognition rate of whole words.

6 Conclusion

In this paper, we investigated influences of change of a set of specified words. And we confirmed the some tendencies in increasing weights. Because the number of specified words is dependent on the task in practical system, we should investigate how many words we can specify to important words.
We investigated the relation between weight and correct rate. From the result of experiment which weighting unfrequently appeared words, we can say that weighting on words which have small probabilities is efficient for improvement of recognition rate of not only specified words but also whole words. In this paper, the weight was constant in modifying the each language model. But changing weight according to original probabilities of specified words as important words may be efficient for improvement of recognition rate of not only specified words but also whole words. We will research a method of controlling weights to

language model	sentence type			
	1)author	2)keyword	3)others	all
BASE	65.41	58.20	51.88	58.84
	61.01	55.56	50.38	55.93
MODIFIED	66.04	58.73	48.12	58.21
	63.52	56.08	45.11	55.51

correct rate(%) / accuracy(%)

Table 2: results of 3 language models

maximize recognition rate.
Also, we applied the proposed method to the spoken dialogue retrieval system. In the result of recognizing sentence-type 1) by the modified language model, there is 2.5 point increase of accuracy though there is no decrease of recognition rate of whole words. In the future, we will develop a method of controlling weight on probabilities of a language model with dialogue control.

References

[1] E.Segarra, L.Hurtado, "Construction of Language Models Using the Morphic Generator Grammatical Inference(MGGI) Methodology" Proc. of Eurospeech97,pp2695-2698,1997

[2] Shuwu ZHANG, Taiyi HUANG, "An Integrated Language Modeling with n-gram model and WA model for Speech Recognition" Proc. of Eurospeech97,pp2699-2702,1997

[3] Philip Clarkson, Ronald Rosenfeld, "STATISTICAL
LANGUAGE MODELING USING THE CMU-CAMBRIDGE TOOLKIT" "Statistical Language Modeling Using the CMU-Cambridge Toolkit" Proc. of Eurospeech97,pp2607-2710,1997

[4] T.Kawahara, A.Lee, T.Kobayashi, "Evaluation of Japanese Dictation ToolKit -1997 version-" IPSJ.98-SLP-21-10,1998(in Japanese)

[5] K.Ito, A.Ito, T.Utsuro, "Common Platform of Japanese Large Vocabulary Continuous Speech Recognition Research -Developement of text corpus-" IPSJ.97-SLP-18-2,1997(in Japanese)