Determination of the Number of Candidates Using Recognition Scores for N-best Based Speech Interface

Kook Cho and Yoichi Yamashita
Dep. of Computer Science, Ritsumeikan University
1-1-1, Noji-Higashi, Kusatsu-shi, Shiga, 525-8577
Japan
{cho,yama}@slp.is.ritsumei.ac.jp

ABSTRACT
Recently statistical techniques have greatly improved the performance of speech recognition. The correct result is not always obtained when only the most probable recognition result is shown to the user. Therefore, some speech interface systems display several candidates of speech recognition to the user, and they ask him/her to choose the correct answer from the candidates. In such a speech interface system based on the N-best speech recognition, the determination of the number of candidates to be shown becomes an important problem. When many recognition candidates are displayed the probability that the correct answer is included will becomes high, but the user needs much time and much effort to search the correct answer. This paper describes a technique of determining the number of candidates dynamically using the distribution of the recognition scores for the N-best speech recognition. The proposed method reduces the number of candidates to be shown without degrading the speech recognition rate.

KEY WORDS
Speech recognition, N-best, Recognition score, Heuristics

1. Introduction

Recently the expectation for the speech interface that uses speech recognition has risen because stochastic methods improves the speech recognition rate [1][2][3][4][5].

The correct result is not always obtained when only the most probable recognition result is shown to the user. In this case, another probable result is repeatedly shown to a user until the correct result is obtained, and for the user, it becomes a big burden. In order to mitigate such a burden, multiple recognition results are displayed to the user, and he/she is asked to choose the correct one among the candidates. Such speech interface systems based on N-best speech recognition have been studied [6][7][8][9][10].

In a speech interface system based on the N-best speech recognition, the determination of the number of candidates to be shown becomes an important problem, it usually decides the number of candidates to display beforehand. When many recognition candidates are displayed the probability that the correct answer will be included becomes high, but the user needs much time and much effort to search the correct answer. Thus, in order to reduce the number of candidates displayed without decreasing the rate by which a correct answer is included in a candidate, it is necessary to determine the number of candidates dynamically based on a recognition result, without deciding the number of candidates beforehand.

This paper describes a technique of determining the number of candidates dynamically using the distribution of the recognition scores for the N-best speech recognition. In addition, the validity of the proposed method is verified.

2. Analysis of recognition score

The characteristics of the score distribution in the recognition result of N-best are investigated. Because the recognition score is dependent on the length of the sentence, the recognition score is normalized by the length of the sentence.

2.1 Task

The task of speech recognition is a homepage retrieval of the laboratory. The number of vocabularies of the task grammar and a word dictionary is changed for three tasks. First, in the task 1, the vocabulary sizes of a word dictionary are 94. As for the task grammar 28 patterns of the sentence are prepared. At the task 2, the vocabulary sizes of a word dictionary are 184. As for the task grammar 41 patterns of the sentence are prepared. In the task 3, the vocabulary sizes of a word dictionary are 1360. As for the task grammar 45 patterns of the sentence are prepared. Moreover, the utterance sentence prepared the sentence of various lengths.

2.2 Analysis data

The 20 speakers' utter 20 sentences and they were used to analyze the recognition score. Speech recognition was performed using Julian rev.3.3 (standard) [11], which sets the number of the candidates N-best to 30.

The probabilities that the correct answer is included in the displayed N-best candidates will be called the rate of correct answer. The rates of correct answer of 30-best recognition are 100%, 95.75%, and 42% for the task 1, 2, and 3, respectively. Moreover, we find that although the number of the candidates in N-best was set as 30, 30

candidates of the sentence were not always generated in fact. The number of average of generated candidates is 9.79, 27.8, and 24.0 for the task 1, 2, and 3, respectively.

2.3 Analysis of heuristics 1

2.3.1 Score difference between adjacent candidates

When the difference of the score between the n-th candidate and the (n+1)th candidate is large, a possibility that the correct answer is included after the (n+1)th candidate becomes low. In this case, the (n+1)th candidate or later is not displayed. This is called heuristics 1.

2.3.2 Analysis result

The difference between the scores of the n-th candidate and the (n+1)th candidate is investigated for each rank of the correct answer. The result of the task 3 which investigated the difference between the scores of the 1st and 2nd candidate is shown in Figure 1. The horizontal axis shows the rank of the correct answer and the vertical axis shows the difference between the scores of the 1st and 2nd candidate. The total number of dots is 400 in all. In addition, "0" of a horizontal axis means that there was no correct answer in 30 candidates.

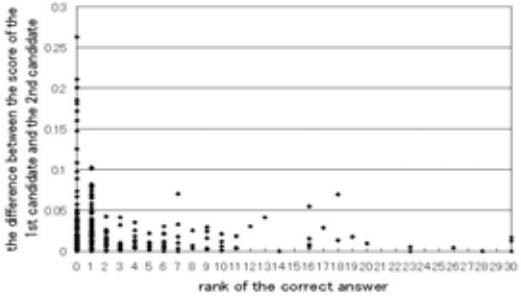


Figure 1: The difference between the scores of the 1st candidate and the 2nd candidate. (Task3)

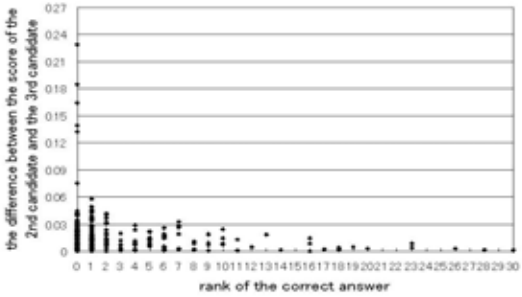


Figure 2: The difference between the scores of the 2nd candidate and the 3rd candidate. (Task3)

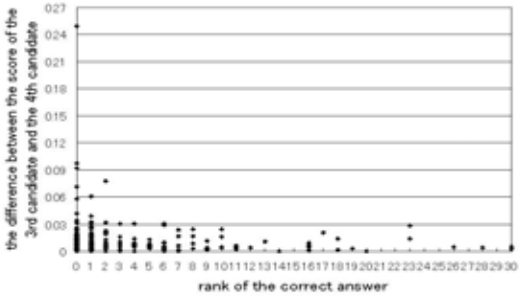


Figure 3: The difference between the scores of the 3rd candidate and the 4th candidate. (Task3)

In the task 3, when the difference of the score between the 1st and 2nd candidate is 0.05 or more, most of correct answers are in the 1st candidate.

Next, the result of the task 3 which investigated the difference of the score of the 2nd and 3rd candidate is shown in Figure 2. When the difference of the score between the 2nd and 3rd candidate is 0.03 or more, most of correct answers are before the 2nd candidate.

Next, the result of the task 3 which investigated the difference of the score of the 3rd and 4th candidate is shown in Figure 3. When the difference of the score between the 3rd and 4th candidate is 0.03 or more, most of correct answers are before the 3rd candidate.

Moreover, the same tendency was found in the task 1 and 2.

2.4 Analysis of heuristics 2

2.4.1 The difference of the score to the 1st candidate

When the difference of the score between the 1st candidate and the n-th candidate is large, it can expect that the probability (the rate of the correct answer) that the correct answer is included after the n-th candidate is low. No correct answers are found after the n-th candidate. This is called heuristics 2. Let the value of the score difference between the 1st candidate and the n-th candidate be represented by θ_1.

2.4.2 Analysis result

The average number of candidates and the rate of the correct answer when changing θ_1 were investigated. The analysis result of the task 1 is shown in Figure 4. Thereby, when θ_1 is larger than 0.12, it turns out that the rate of the correct answer keeps high, although the number of candidates gets fewer according to the decrease of θ_1. Moreover, the same tendency was found also for the task 2 and 3.

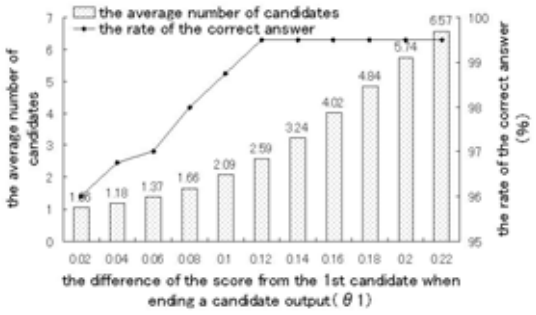


Figure 4: Change of the rate of the correct answer and the average number of candidates for heuristics 2.

2.5 Analysis of heuristics 3

2.5.1 The n-th candidate Score

When the n-th candidate score is small, it is expected that a possibility that the correct answer is included after the n-th candidate is low. No correct answers are found after the n-th candidate. This is called heuristics 3. Let the value over the n-th candidate score be represented by θ_2.

2.5.2 Analysis result

The average number of candidates and the rate of the correct answer when changing θ_2 were investigated. The

analysis result of the task 2 is shown in Figure 5. Thereby, when θ_2 is larger than -27.0, it turns out that the rate of the correct answer keeps high, although the number of candidates gets fewer according to the decrease of θ_2. Moreover, the same tendency was found also for the task 1 and 3.

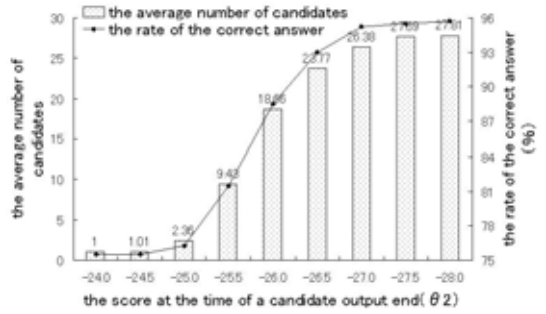


Figure 5: Change of the rate of the correct answer and the average number of candidates for heuristics 3.

3. Decision of the number of recognition candidate displays

3.1 Rule set

The following rule set is used for the determination of the number of candidate displays.

Task 1,

[rule 1] When the difference of the score between the 1st and 2nd candidate is 0.03 or more, only the 1st candidate is displayed.

[rule 2] When the difference of the score between the 2nd and 3rd candidate is 0.01 or more, no candidates after the 2nd rank are displayed.

[rule 3] When the difference of the score between the 1st and n-th candidate is 0.10 or more, no candidates after the (n-1)th rank are displayed.

[rule 4] When the n-th candidate score is -26.5 or less, no candidates after the (n-1)th rank are displayed (n= 2, 3, 4, --, 30).

Task 2,

[rule 1] When the difference of the score between the 1st and 2nd candidate is 0.05 or more, only the 1st candidate is displayed.

[rule 2] When the difference of the score between the 2nd and 3rd candidate is 0.03 or more, no candidates after the 2nd rank are displayed.

[rule 3] When the difference of the score between the 3rd and 4th candidate is 0.02 or more, no candidates after the 3rd rank are displayed.

[rule 4] When the difference of the score between the 4th and 5th candidate is 0.02 or more, no candidates after the 4th rank are displayed.

[rule 5] When the difference of the score between the 1st and n-th candidate is 0.12 or more, no candidates after the (n-1)th rank are displayed.

[rule 6] When the n-th candidate score is -27.0 or less, no candidates after the (n-1)th rank are displayed (n= 2, 3, 4, --, 30).

Task 3,

[rule 1] When the difference of the score between the 1st and 2nd candidate is 0.05 or more, only the 1st candidate is displayed.

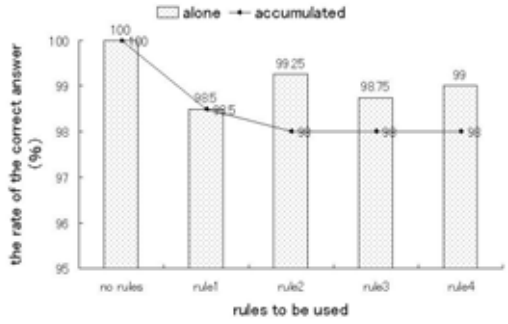
[rule 2] When the difference of the score between the 2nd and 3rd candidate is 0.03 or more, no candidates after the 2nd rank are displayed.

[rule 3] When the difference of the score between the 3rd and 4th candidate is 0.03 or more, no candidates after the 3rd rank are displayed.

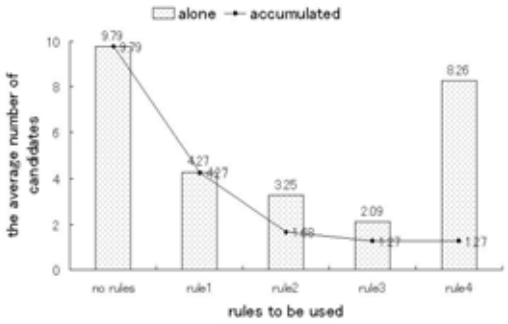
[rule 4] When the difference of the score between the 1st and n-th candidate is 0.14 or more, no candidates after the (n-1)th rank are displayed.

[rule 5] When the n-th candidate score is -27.0 or less, no candidates after the (n-1)th rank are displayed (n= 2, 3, 4, --, 30).

When each rule set is used, the result of task 1, 2, and 3 is shown in Figure 6, 7, and 8, respectively. "alone" is a result in the case of using one rule in each rule set independently. "accumulated" is as a result of rule set application, and is the result of applying the rule whose conditions compares the rule sequentially from the rule 1 and corresponded first. When using each rule by the alone or the accumulated, the result that investigated how the rate of the correct answer would change is shown in (a) of Figure 6, 7, and 8. As a result of using a rule by the accumulated, the rate of the correct answer is 98%, 93.5%, and 40.5% for the task 1, 2, and 3, respectively. Although some rates of the correct answer have fallen compared with the case where 30 candidates are displayed, there is no big difference.

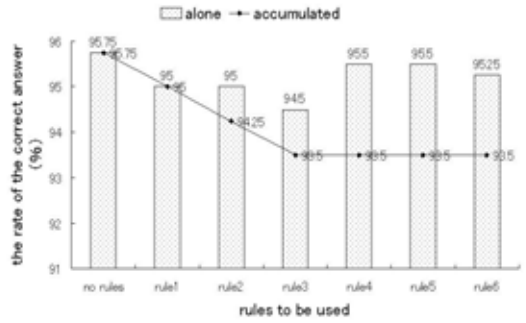


(a) Change of the rate of the correct answer.

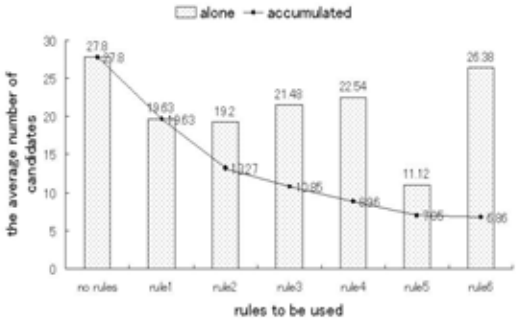


(b) Change of the average number of candidates.

Figure 6: The rate of the correct answer and the average number of candidates. (Task1)

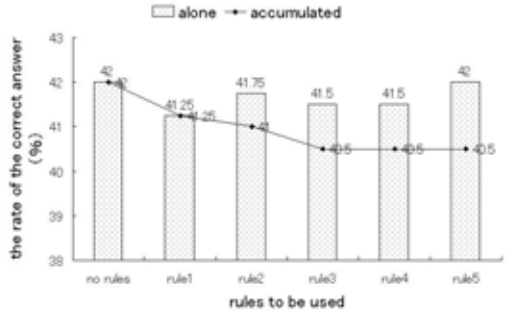


(a) Change of the rate of the correct answer.

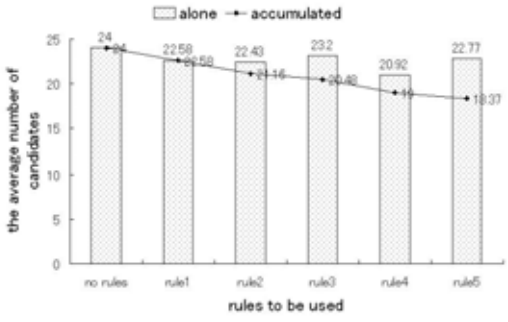


(b) Change of the average number of candidates.

Figure 7: The rate of the correct answer and the average number of candidates. (Task2)



(a) Change of the rate of the correct answer.



(b) Change of the average number of candidates.

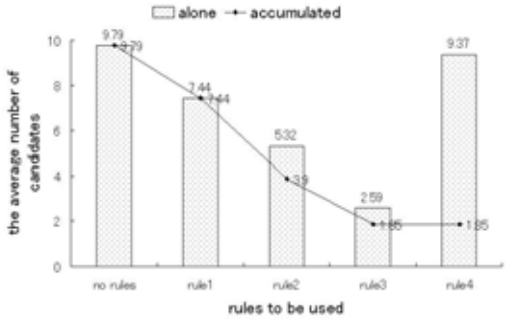
Figure 8: The rate of the correct answer and the average number of candidates. (Task3)

When using each rule by the "alone" or the "accumulated", the result which investigated how the average number of candidates would change is shown in (b) of Figure 6, 7, and 8. When each rule is used by the accumulated, the average number of candidates decreases. The average number of candidates is 1.27, 6.86, and 18.37 for task 1, 2, and 3, respectively. This is decreasing more sharply than the case where 30 candidates always display.

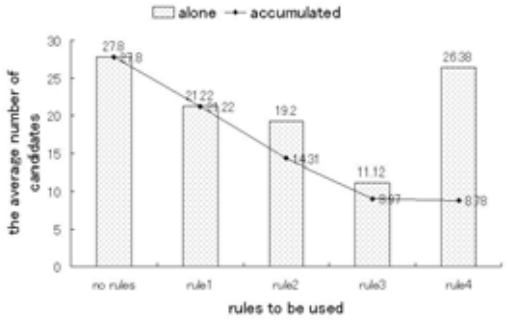
Proposed rules reduces the number of candidates to be shown, without degrading the rate of the correct answer compared with the case where 30 candidates are always displayed. Therefore, it can be said that the recognition score of the N-best candidate is effective in the determination of the number of displayed candidate.

3.2 Unification-izing of rule

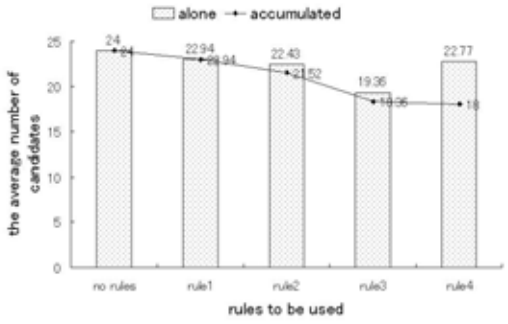
The rule sets mentioned in 3.1 are developed for each task. Then, the rule set independent of tasks was considered by merging the rule sets for three tasks.



(a) Task 1.



(b) Task 2.



(c) Task 3.

Figure 9: The average number of candidates with the common rule set.

[rule 1] When the difference of the score between the 1st and 2nd candidate is 0.06 or more, only the 1st candidate is displayed.

[rule 2] When the difference of the score between the 2nd and 3rd candidate is 0.03 or more, no candidates after the 2nd rank are displayed.

[rule 3] When the difference of the score between the 1st and n-th candidate is 0.12 or more, no candidates after the (n-1)th rank are displayed.

[rule 4] When the n-th candidate score is -27.0 or less, no candidates after the (n-1)th rank are displayed (n= 2, 3, 4, --, 30).

When each rule is used by the alone or the accumulated, change of the average number of candidate of task 1, 2, and 3 is shown in Figure 9. When each rule of this rule set was used by the accumulated, the rate of the correct answer is 99.5%, 94.75%, and 40% for task 1, 2, and 3, respectively. The average number of candidates is 1.85, 8.78, and 18, for task 1, 2, and 3, respectively. The rule set reduces the number of candidates to be shown without degrading the rate of the correct answer.

4. Experiment result

4.1 Experiment method

Following three methods for displaying the recognition candidates was used to verify the effectiveness of the technique for deciding the number of candidates by using the recognition score. As a result, which method had been used easily most was evaluated. 7 speakers participated in the experiment and the task 2 was used.

(The display method 1)

Only one candidate always is displayed.

(The display method 2)

30 candidates always are displayed.

(The display method 3)

The number of displayed candidate is determined by using recognition scores.

4.2 Experiment result

The number of average utterance and average time to select a correct answer is shown for three display methods in Figure 10. In the three display methods, all subjects answered that it was the easiest to use the case where the number of displayed candidate is determined using recognition scores. The method 3 based on the dynamic determination of the number of displayed candidates displays the correct result with fewer in correct candidates.

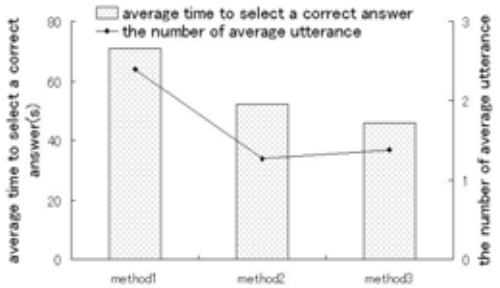


Figure 10: The number of average utterance and average time to select a correct answer.

5. Conclusions

Dynamic determination of the number of displayed candidate using the recognition score of the N-best candidates reduces the number of candidates to be shown without degrading the rate of the correct answer.

Moreover, a result of an evaluation experiment, all subjects answered that it was the easiest to use the case where the number of candidate displays is determined using recognition scores.

6. Acknowledgements

We would like to thank Miss Akiko Miyayama for her assistance during this work.

7. References

[1] Shinya KIRIYAMA, Keikichi HIROSE, "Speech Repy Generation of a Spoken Diglogue System for Literature Retrieval," IPSJ SIG Technical Reports, SLP-27-16, pp105-110, (1999.7).

[2] K. Kondo, C. Hemphill, "Surfin' the World Wide Web with Japanese," Proc. ICASSP'97, vol.2, pp.1151-1154, (1997.4).

[3] Reishi KONDO, Keiko INAGAKI, Kenichi ISO, Yukio MITOME, "Speech Interface for Access to the Newspapers on the Web," IPSJ SIG Technical Reports, SLP-16-8, pp43-48, (1997.5).

[4] Makoto KATSUURA, Satoshi NAKAMURA, Kiyohiro SHIKANO, "Net surfin' by Spoken Keywords," IPSJ SIG Technical Reports, SLP-20-12, pp69-74, (1998.2).

[5] Takeshi INOUE, Yoichi YAMASHITA, "Speech Input Interface for Command-based Systems," the 2001 spring meeting of the acoustical society of japan, 3-P-27, pp217-218, (2001.3).

[6] Takahiro NAKANO, Atsuhiko KAI, Seiichi NAKAGAWA, "Investigation on Speech Interface for form-based Information Retrieval Services on the WWW," IPSJ SIG Technical Reports, SLP-25-1, pp1-6, (1999.2).

[7] Takuya NISHIMOTO, Yutaka KOBAYASHI, Yasuhisa NIIMI, "General Purpose Spoken Dialog System for Database Access on Internet," the 1997 spring meeting of the acoustical society of japan, 2-Q-20, pp179-180, (1997.3).

[8] Mikio YAMAMOTO, Mitsunori TAKAGI, Seiichi NAKAGAWA, "A Menu Guided Spoken Dialog system and Its Evaluation," IEICE Technical Reports, SP93-130, pp17-24, (1994.1).

[9] R. Lau, G. Flammia, C. Pao, and V. Zue, "WEBGALAXY - Integrating Spoken Language and Hypertext Navigation," Proc. of EUROSPEECH'97, vol.2, pp.883-886, (1997.9).

[10] Sunil Issar, "A Speech Interface for Forms on WWW," Proc. of EUROSPEECH'97, vol.3, pp.1343-1346, (1997.9).

[11] http://julius.sourceforge.jp/