The performance prediction method on sentence recognition system using a finite state automaton

著者	伊藤 彰則
journal or publication title	IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994. ICASSP-94
volume	1994
number	1
page range	397-400
year	1994
URL	http://hdl.handle.net/10097/46624

doi: 10.1109/ICASSP.1994.389272

THE PERFORMANCE PREDICTION METHOD ON SENTENCE RECOGNITION SYSTEM USING A FINITE STATE AUTOMATON

Takashi Otsuki[1] *Akinori Ito*[2] *Shozo Makino*[3] *Teruhiko Otomo*[1]

[1]Faculty of Engineering, Yamagata University, 4-3-16 Jounan, Yonezawa, 992 Japan
[2]Education Center for Information Processing, Tohoku University, Kawauchi, Aoba, Sendai, 980 Japan
[3]Graduate School of Information Sciences, Tohoku University, 2-1-1 Katahira, Aoba, Sendai, 980 Japan

ABSTRACT

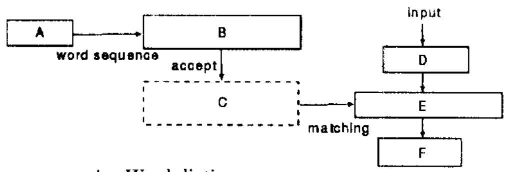
This paper presents the performance evaluation method on sentence recognition system which uses a finite state automaton. The relationship between word recognition score and sentence recognition score can be predicted using the number of word sequences at a short distance from the sentence. But it is not clear that how we get this number when the finite state automaton is used as linguistic information. Therefore, we propose the algorithm to calculate this number in polynomial time. Then we carry out the prediction using this method and the simulation to compare with the prediction, and it is shown that our method is usable when the quality of the word lattice is good.

1. INTRODUCTION

Present speech recognition systems are not perfect so that there are many recognition errors in their recognition results: sequences of linguistic units such as phonemes, syllables and words. To correct these errors, many kinds of linguistic information are used in the speech recognition systems. For example, bigram and trigram of phonemes or words, a finite state automaton, context free grammar, etc. And it is important to know the ability of error correction of these linguistic information because it is useful for evaluation and design of the speech recognition system.

The ability of error correction of the linguistic information depends on how it can eliminate the sequences that do not exist in the task, in other words, how it can reduce the complexity of the task. To measure this ability, first we must decide what indicates the complexity. K. Abe et al. used Hamming distance (we will call it "distance" for convenience) as the indicator of the complexity, and derived the theoretical formula to get the lower limit of word recognition score on the recognition system using a dictionary. Where the word recognition score can be predicted based on the recognition score of linguistic units and the number of word pairs with short distance in the dictionary in case of no segmentation error and uniform word occurrence probability [1]. The definition of $D(S,T)$: the distance between sequences $S = s_1 s_2 \cdots s_L$ and $T = t_1 t_2 \cdots t_L$ is given by

$$D(S,T) = \sum_{i=1}^{L} f(s_i, t_i) \qquad (1)$$



A : Word dictionary
B : Finite state automaton (FA)
C : Word sequences accepted by FA
D : Word recognition unit
E : Optimum word sequence selection
F : Sentence recognition result

Figure 1. Model of the sentence recognition system using FA.

$$f(a,b) = \begin{cases} 0 & \text{if} \quad a = b \\ 1 & \text{else} \end{cases} \qquad (2)$$

T. Otsuki et al. proposed the algorithm to calculate the number of phoneme sequences at a short distance from the word in polynomial time when the transition information between phonemes is used instead of the dictionary, and proposed the more accurate formula to predict word recognition score[3].

M. M. Sondhi et al. introduced the perplexity to measure the complexity of the task[2]. When the entropy of the task is H, the perplexity F_p is defined as follows.

$$F_p = 2^H \qquad (3)$$

S. Nakagawa et el. proposed the evaluation method for continuous speech recognition systems, and they made clear the relationship between task perplexity and sentence recognition score[4]. But there is a problem that the perplexity can not reflect the similarity between sentences.

Therefore, in this paper we use the distance between word sequences as the indicator of the complexity, and propose the performance prediction method on sentence recognition system using a finite state automaton.

2. THEORY OF SENTENCE RECOGNITION SCORE PREDICTION

Fig.1 shows the model of the sentence recognition system using a finite state automaton. Sequence of word candidates which comes from the word recognition unit is matched

0-7803-1775-0/94 $3.00

with the word sequences accepted by the finite state automaton, and the word sequence with maximum likelihood is selected as the sentence recognition result. The error correction ability of the finite state automaton determines the relationship between word recognition score and sentence recognition score.

The relationship between word recognition score and sentence recognition score can be predicted using the number of acceptable word sequences at a short distance from the sentence using the method we proposed[3]. When the sentence $X = x_1 x_2 \cdots x_L$ is uttered, assume the word lattice I is given by

$$I = \begin{bmatrix} I_{11} & I_{21} & \cdots & I_{L1} \\ I_{12} & I_{22} & \cdots & I_{L2} \\ \vdots & \vdots & \ddots & \vdots \\ I_{1M} & I_{2M} & \cdots & I_{LM} \end{bmatrix} \quad (4)$$

$$I_{ij} \sim \begin{cases} N(\mu, \sigma^2) & \text{if} \quad x_i = w_j \\ N(0, \sigma^2) & \text{else} \end{cases} \quad (5)$$

where L is the length of the sentence, M is the number of words in the task and $w_1, w_2, \cdots, w_M$ are words in the task. Then the word recognition score α is given by

$$\alpha = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \left\{ \Phi\left(\frac{x}{\sigma}\right) \right\}^{M-1} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx \quad (6)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{ -\frac{t^2}{2} \right\} dt \quad (7)$$

And the sentence recognition score β is given by

$$\beta = 1 + \sum_i (-1)^i C_i \lambda_i \quad (8)$$

$$\lambda_i = \Phi\left(-\frac{\sqrt{i}\mu}{\sqrt{2}\sigma} \right) \quad (9)$$

$$C_i = \frac{1}{N} \sum_X \binom{N_1(X)}{i} \quad (10)$$

where N is the number of sentences in the task.

But it is hard to get $N_1(X)$ because the number of acceptable sequences glows up rapidly with L and M. Therefore, in order to predict the relationship between word recognition score and sentence recognition score in case of using a finite state automaton as linguistic information, we propose the algorithm to calculate the number of acceptable sequences at a short distance from the sentence in the task in polynomial time.

3. ALGORITHM TO CALCULATE $N_1(X)$

First, we define the following symbols.

- $c_1, c_2, \ldots, c_M$: M kinds of word categories
- $X = x_1 x_2 \cdots x_L$: Sentence which consists of L words where ith word is x_i
- X_l : Partial sequence of X which consists of $x_1 x_2 \cdots x_l$
- $N_d(X)$: The number of word sequences at a distance of d from X

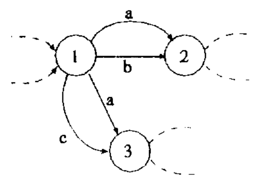


Figure 2. An example automaton

- $\Lambda(c)$: The number of words belonging to category c
- $V(x)$: The category which word x belongs to
- $(Q, \Sigma, \delta, q_0, F)$: Finite state automaton where Q is the group of states, Σ is the group of categories, δ is the state transition function, q_0 is the initial state and F is the group of final states
- U : The number of states
- Ω_l : Group of word sequences that cause l times transition from q_0
- Π_l : Group of states after l times transition from q_0
- Ω_l^q : Group of word sequences that cause l times transition from q_0 to q
- $N_d(X_l, \Omega_l^q)$: The number of word sequences in Ω_l^q at a distance of d from X_l
- $\Psi(q)$: Group of categories that cause transition at state q

  $$\Psi(q) = \bigcup_{q' \in Q} \{c | \delta(q, c) = q'\}$$

  In the example automaton shown in fig.2, $\Psi(1) = \{a, b, c\}$
- $\Phi(p, q)$: Group of categories that cause transition to state q at state p

  $$\Phi(p, q) = \{c | \delta(p, c) = q\}$$

  In the example automaton shown in fig.2, $\Phi(1, 2) = \{a, b\}$
- $\eta(q, c)$: Group of states after category c is input to state q

  $$\eta(q, c) = \{q' | \delta(q, c) = q'\}$$

  In the example automaton shown in fig.2, $\eta(1, a) = \{2, 3\}$

Since sentence X is acceptable to the FA,

$$X_l \in \Omega_l \quad (11)$$

and

$$N_0(X_l, \Omega_l^q) = \begin{cases} 1 & \text{if} \quad V(x_l) \in \bigcup_{p \in Q} \Phi(p, q) \\ 0 & \text{else} \end{cases} \quad (12)$$

Consider one sequence Y in Ω_{l-1}. After category c is input and the current state moves to q, this sequence Y becomes a partial sequence of $\Lambda(c)$ sequences. In case of $c = V(x_l)$, the distance between X_l and one sequence with last

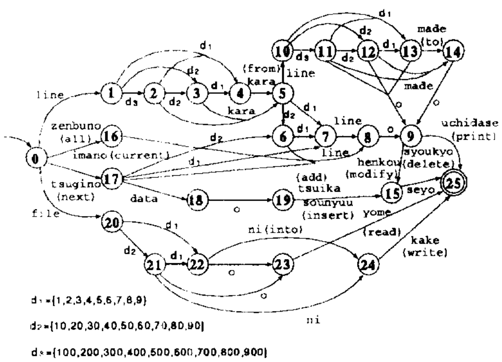


Figure 3. FA accepting text editor command

word of x_l in $\Lambda(c)$ sequences is equal to the distance between X_{l-1} and sequence Y. And the distance between X_l and remainder $\Lambda(c)-1$ sequences is one more than the distance between X_{l-1} and sequence Y. On the other hand, in case of $c \neq V(x_l)$, the distance between X_l and $\Lambda(c)$ sequences is one more than the distance between X_{l-1} and sequence Y. Therefore $N_1(X_l, \Omega_l^p)$ can be calculated as follows.

$$N_1(X_l, \Omega_l^p) = \sum_{q \in \Pi_{l-1}} \sum_{c \in \Phi(q,p)} f(q, c, l-1) \qquad (13)$$

$$f(q,c,l) = \begin{cases} N_1(X_l, \Omega_l^q) + (\Lambda(c)-1) N_0(X_l, \Omega_l^q) & \text{if } c = V(x_{l+1}) \\ \Lambda(c) N_0(X_l, \Omega_l^q) & \text{else} \end{cases} \qquad (14)$$

Finally, the number of sequences at a distance of 1 from sentence X is given by

$$N_1(X) = \sum_{q \in \Pi_{L-1}} \sum_{c \in \Psi(q)} \sum_{p \in \eta(q,c) \cap F} f(q, c, L-1) \qquad (15)$$

Fig.4 shows the algorithm to calculate $N_1(X)$. The computation amount in step 2 is $O(LMU^2)$ per sentence.

4. PREDICTION OF SENTENCE RECOGNITION SCORE

Using this algorithm and eq.(6) (8), we predict the relationship between word recognition score and sentence recognition score using the finite state automaton which accepts Japanese text editor commands shown in fig.3. This task has three categories of digits d_1, d_2, d_3 and 18 Japanese words.

First, we calculate N_1 on 380 sentences accepted by this automaton using the proposed algorithm to get C_i (see eq.(10)). Then we calculate α and β with same σ and μ. By moving σ or μ, we can get the relationship between word recognition score and sentence recognition score.

In addition, we carry out the simulation to evaluate the accuracy of the prediction. First, we make word lattice according to eq.(5), giving normal random value as word

Step1:Initialize
  $N_0(X_0, \Omega_0^{q_0}) := 1$
  for all $q \in Q$ **do begin**
    for $l = 1$ **to** $L-1$ **do begin**
      $N_0(X_l, \Omega_l^q) := 0$
      $N_1(X_l, \Omega_l^q) := 0$
    end
  end
Step2:1st $\sim (L-1)$**th transition**
  $\Pi_0 := \{q_0\}$ **;States for next expansion**
  for $l := 1$ **to** $L-1$ **do begin**
    for all $p \in \Pi_{l-1}$ **do begin**
      for all $c \in \Psi(p)$ **do begin**
        for all $q \in \eta(p,c)$ **do begin**
          if $c = V(x_l)$ **then begin**
            $N_0(X_l, \Omega_l^q) := 1$
            $N_1(X_l, \Omega_l^q) := N_1(X_l, \Omega_l^q) + N_1(X_{l-1}, \Omega_{l-1}^p)$
              $+(\Lambda(c)-1) \times N_0(X_{l-1}, \Omega_{l-1}^p)$
          else begin
            $N_0(X_l, \Omega_l^q) := 0$
            $N_1(X_l, \Omega_l^q) := N_1(X_l, \Omega_l^q)$
              $+\Lambda(c) \times N_0(X_{l-1}, \Omega_{l-1}^p)$
          end
        end
        $\Pi_l := \Pi_l \cap \eta(p,c)$
      end
    end
  end
Step3:Final state
  for all $p \in \Pi_{L-1}$ **do begin**
    for all $c \in \Psi(p)$ **do begin**
      for all $q \in \eta(p,c) \cap F$ **do begin**
        if $c = V(x_L)$ **then begin**
          $N_1(X) := N_1(X) + N_1(X_{L-1}, \Omega_{L-1}^p)$
            $+(\Lambda(c)-1) \times N_0(X_{L-1}, \Omega_{L-1}^p)$
        else begin
          $N_1(X) := N_1(X) + \Lambda(c) \times N_0(X_{L-1}, \Omega_{L-1}^p)$
        end
      end
    end
  end

Figure 4. An algorithm to calculate N_1

speech recognition tasks", Proc. ICASSP, pp.409-412 (1978)

[3] T.Otsuki, A.Ito, S.Makino and T.Sone, "Performance Prediction of Word Recognition Using the Transition Information between Phonemes or between Characters", Trans. IEICE Japan, **J76-D-II**, 6, pp.1090-1096 (June 1993) (in Japanese)

[4] S.Nakagawa, Y.Ooguro and I.Murase, "An evaluation method for continuous speech recognition systems - Relationship between task complexity and sentence recognition accuracy -", Trans. IEICE Japan, **J73-D**, 5, pp.683-693 (March 1990) (in Japanese)

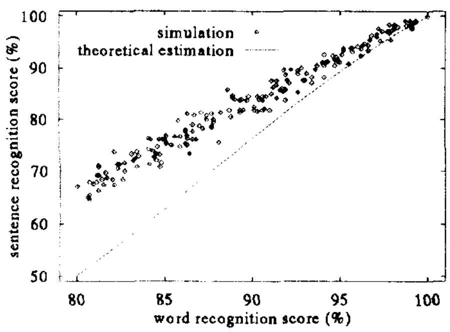


Figure 5. The relationship between word recognition score and sentence recognition score

recognition score, then automaton-driven sentence recognition is applied to this word lattice. This simulation is carried out on 380 sentences used in the prediction.

The results of the prediction and the simulation are shown in fig.5. On condition that the word recognition score is over 90%, it can be said that the curve of the prediction indicates the lower limit of sentence recognition score. But there is difference (about 15% when word recognition score is 80%) between the prediction and the simulation when the word recognition score is under 90%. This is because the approximation in deriving the formula to calculate sentence recognition score (eq.(8)) becomes unreasonable when the quality of the word lattice becomes worse.

5. CONCLUSION

In this paper, we proposed the performance prediction method on sentence recognition system using a finite state automaton. There, we showed the algorithm to calculate the number of word sequences at a short distance from a sentence in polynomial time for predicting the relationship between word recognition score and sentence recognition score. The comparison between the prediction and the simulation showed that our method can be applied to getting the lower limit of the sentence recognition score when the quality of the word lattice is good.

We plan on carrying out the prediction on real sentence recognition system with large task, and deriving the prediction formula usable when the performance of word recognition unit is insufficient.

REFERENCES

[1] K. Abe. K. Hatano and T. Fukumura, "Performance evaluation of character recognition system with dictionary," Trans. IECE Japan, **J52-C**, 6, pp.305-312, (June 1969) (in Japanese)

[2] M. M. Sondhi and S. E. Levinson, "Computing relative redundancy to measure grammatical constraint in