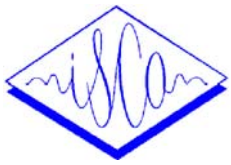

EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

ANALYSIS OF INTERACTIVE STRATEGY TO RECOVER FROM MISRECOGNITION OF UTTERANCES INCLUDING MULTIPLE INFORMATION ITEMS

Yasuhisa NIIMI, Takuya NISHIMOTO and Yutaka KOBAYASHI

Department of Electronics and Information Science,
Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto, 606 JAPAN
e-mail: {niimi,nishi,koba}@dj.kit.ac.jp

ABSTRACT

This paper proposes and analyzes mathematically an interactive strategy to recover from misrecognition of utterances including multiple information items through a short conversation with a speaker. First the speech recognizer in a dialogue system recognizes an utterance and evaluates the reliability of each item contained in it. The dialogue system accepts only those items of which the reliability is high, while it rejects the items which are unreliably recognized, or confirms the content of them. The paper, given the performance of the recognizer, derives two quantities P_{ac} and N, which can describe the performance of the dialogue system using this interactive strategy: P_{ac} is the probability that all information items included in user's utterance are conveyed to the system correctly, and N is the average number of turns taken between the user and the system until all the items are accepted.

1. INTRODUCTION

A number of studies on spoken dialogue systems have been conducted based on the stochastic speech recognition[1,2]. However, it is still difficult to correctly recognize casual utterances often observed in human conversations. Many interactive methods have been reported to recover from recognition errors. However, most of them have been empirical, and not involved the mathematical analysis [3]. We have proposed and mathematically analyzed a dialogue control strategy to relieve speech recognition errors[4,5]. This strategy makes a dialogue system accept those utterances which are reliably recognized, but reject those or confirm the content of those which are unreliably recognized. It has, however, dealt with utterances equally regardless of an amount of information contained.

In this paper we pay attention to the number of information items included in an utterance. By information item we mean, for example, keywords necessary to build a query command in the information retrieval task. We expand the dialogue control strategy mentioned above so as to involve an effect of the number of items under the assumption that the reliability in recognizing items can be evaluated independently of their location in an utterance. Then we, given the performance of the recognizer, derive two quantities P_{ac} and N, which can describe the performance of the dialogue system using this interactive strategy: P_{ac} is the probability that all information items included in user's utterance are conveyed to the system correctly, and N is the average number of turns taken between the user and the system until all the items are accepted. In other words, the paper gives a quantitative relation among the complexity (the number of items) of utterances which can be used in dialogues, the performance of a dialogue system and that of the speech recognizer used in it.

Section 2 gives a brief description of an expanded strategy and assumptions made on the performance of a speech recognizer. Section 3 derives two formulae which can be used to evaluate the performance of a dialogue system. Finally section 4 explains how to apply the derived formulae to design a dialogue system.

2. INTERACTIVE STRATEGY TO RECOVER FROM MISRECOGNITION

2.1. Interactive strategy

In this section we propose an interactive strategy to recover from speech recognition errors. Consider the following dialodue situation to imagine how the proposed strategy works. A speaker in front of the dialogue system produces an utterance containing several information items, for example, three items. The dialogue system recognizes this utterance and computes the reliability of each item. Suppose the system decided to confirm the first and second items and to reject the third item depending on the values of thier reliabilities. To confirm the first and second items the system produces an utterance like "Is the first item XXX and the second YYY ?". The speaker is supposed to respond to this confirmation by a simple utterance like "Yes and no.", which means that the first item is correct while the second is not. We will call "yes" and "no" response items below. The system recognizes this simple response and computes the reliabilities of the two response items. If it accepts both items, then the system assures that the first item has been recognized correctly while the second has not. It already knows that the third item should be rejected. So to prompt the speaker to reinput the rejected items, it produces an utterance like "Please speak again of the second and third items." This cycle will continue until all the items are accepted.

To mathematically analyze this interactive dialogue control strategy, we assume the followings for the perfor-

mance of the speech recognizer in a dialogue system.

(1) The speech recognizer can compute the reliability $R(I_i)$ in recognizing each item I_i contained in an utterance.

(2) $R(I_i)$ is positive and the reliability is as high as $R(I_i)$ is small. An example of R having this property is $-logP(I|A)$ where $P(I|A)$ is the posterior probability of an item I given A which is the acoustic data stream of I[6].

(3) R is independent of the content and the location of an item in an utterance.

(4) Only substitution errors of items occur in the speech recognition, but insertions and deletions do not occur. This assumption is not practical, but is made for mathematical simplicity.

Under these assumptions the proposed interactive recovery strategy can be stated formally as follows.

(1) A speaker produces an utterance including n information items.

(2) For each item I_i the dialogue system computes $R(I_i)$ and then takes one of the following three actions.

c1 If $R(I_i) \leq \theta_1$, the dialogue system accepts an I_i. Let α be the probability that an item is accepted, and p the probability that the accepted item has been recognized correctly. α and p are generally dependent on θ_1, α being proportional but p inversely proportional to θ_1.

c2 It confirms the content of I_i, if $\theta_1 < R(I_i) \leq \theta_2$. Let β be the probability that the confirmation is made, and q be the probability that the confirmed item has been recognized correctly.

c3 It rejects I_i if $R(I_i) > \theta_2$. The probability for this is $\gamma = 1 - \alpha - \beta$. We call these four probabilities α, β, p and p recognizer's parameters below.

(3) The dialogue system produces an utterance to confirm all the items it has decided to confirm.

(4) The speaker is supposed to respond to this confirmation simply using "yes" and "no".

(5) The dialogue system accepts only those affirmative responses r_i's of which $R(r_i)$ is less than or equal to θ_3 and rejects the negative responses and the responses which are unreliably recognized. Let δ be the probability that a response item is accepted and s the probability that the accepted response has been recognized correctly.

(6) The dialogue system lets the speaker to know what items have been rejected and prompts him or her to speak again of them.

(7) The speaker produces the directed utterance.

This cycle from (2) to (7) will continue until all the items are accepted.

2.2. The recognizer parameters

In this section we consider how to estimate the four recognizer parameters α, β, p and q. First recognizing many items included in training utterances by a speech recognizer to be used in the dialogue system, we have a recognition result, "correct" or "incorrect", and its reliability measure $R(I)$ for each item I. Then we can create two histograms of $R(I)$ for the correct recognition and the incorrect recognition. Let $NT(x,y)$ and $NF(x,y)$ denote the accumulated frequency of the correct and the incorrect recognitions respectively for which $x < R(I) \leq y$. By using these notations, the four recognizer parameters α, β, p and q can be defined as follows:

$$\left.\begin{array}{lcl} \alpha & = & N(0,\theta_1)/N(0,\infty) \\ \beta & = & N(\theta_1,\theta_2)/N(0,\infty) \\ p & = & NT(0,\theta_1)/N(0,\theta_1) \\ q & = & NT(\theta_1,\theta_2)/N(\theta_1,\theta_2) \end{array}\right\} \quad (1)$$

where $N(x,y) = NT(x,y) + NF(x,y)$.

3. ANALYSIS OF THE STRATEGY

3.1. Simplified version of the strategy

First we analyze a simplified version of the interactive recovery strategy stated in the previous section. The strategy is simplified in that the dialogue system uses only rejection as a means of recovering speech recognition errors. Accordingly the steps (3), (4) and (5) of the strategy are omitted.

The operation of a dialogue system using the simplified strategy can be described by a Markov process as shown in Fig. 1, in which a state u_k indicates a situation where a speaker is to produce an utterance containing k items, and the state u_0 indicates the situation where all the items contained in an utterance have been accepted. A transition from a state u_k to a state u_l means that receiving an utterance containing k items the dialogue system accepts $(k-l)$ items and tells the speaker not to accept the rest items. In particular, a self-loop where k is equal to l means that none of items can be accepted. If a speaker produces an utterance containing n items at the outset, the process in which this utterance is completely accepted is represented by a sequence of state transitions originating from the state u_n, traveling several states and ending at the state u_0. This sequence will be called a transition sequence below.

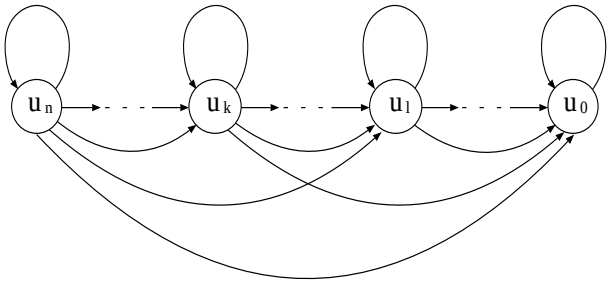

Figure 1: State transition diagram of a dialog system

Now the dialogue system is assumed to be in a state u_k. Let $N(k)$ be the average number of turns taken between

the system and the speaker during a transition sequence from u_k to u_0 and $P_{ac}(k)$ be the probability that all of items is recognized correctly during this transition sequence. The states to which a direct transition can occur from the state u_k are states $u_l (0 \leq l \leq k)$. Since the probability that an item is accepted is α, the probability $p(k,l)$ of a state transition from u_k to u_l is ${}_kC_l\alpha^{k-l}\gamma^l$, where $\gamma = 1-\alpha$. In one of these state transitions except the transition to the state u_0 two utterances are produced; one is an utterance of the speaker and the other is an utterance of the system to ask the speaker to reproduce unaccepted items. Since the average number of utterances produced by the speaker and the system during a transition sequence from u_l to u_0 is $N(l)$, the average number of utterances produced during a transition sequence from u_k to u_0 via $u_l (1 \leq l \leq k)$ is $N(l)+2$. Thus we have the following recursive formula for $N(k)$.

$$N(k) = \sum_{l=1}^{k} p(k,l)(N(l)+2) + p(k,0) \qquad (1 \leq k \leq n) \tag{2}$$

The second term of the right hand side of the above formula corresponds to the direct transition from u_k to u_0, in which the dialogue system does not speak. We can rewrite eq. (2) as follows.

$$N(k) = \frac{1}{1-\gamma^k}\left[\sum_{l=1}^{k-1} {}_kC_l\alpha^{k-l}\gamma^l N(l)+2-\alpha^k\right] \qquad (1 \leq k \leq n) \tag{3}$$

Since $N(1) = 2/\alpha - 1$, we can compute $N(n)$ recursively for any n.

Next we consider $P_{ac}(k)$. $(k-l)$ items are accepted during a state transition from u_k to u_l. The probability $p_{ac}(k,l)$ that all of these items are correctly recognized is p^{k-l}. Using this notation, we have a formula on P_{ac} similar to eq. (2).

$$P_{ac}(k) = \sum_{l=1}^{k} p(k,l)p_{ac}(k,l)P_{ac}(l) \qquad (1 \leq k \leq n) \tag{4}$$

where $P_{ac}(0) = 1$. By the mathematical induction we have

$$P_{ac}(k) = p^k \qquad (1 \leq k \leq n). \tag{5}$$

3.2. Full version of the strategy

In this section we will consider the dialogue system using the full version of the interactive recovery strategy in which all the steps stated in section 2.1 might be taken. The operation of this dialogue system can also be described by the Markov model shown in Fig. 1. The detail of the operation, which is illustrated in Fig. 2, is more complicated than the operation descrived in the previous section.

As shown in Fig. 2, there are three pathes by which an item is accepted by the dialogue system. These are (1) path 'accepted', (2) path 'confirmed-yes-correct' and (3) path 'comfirmed-no-incorrect'. The meaning of the first path is that an item is directly accepted by the condition (c1) in the step (2) of the strategy. The meaning of the second path is that the affirmative response to a confirmed

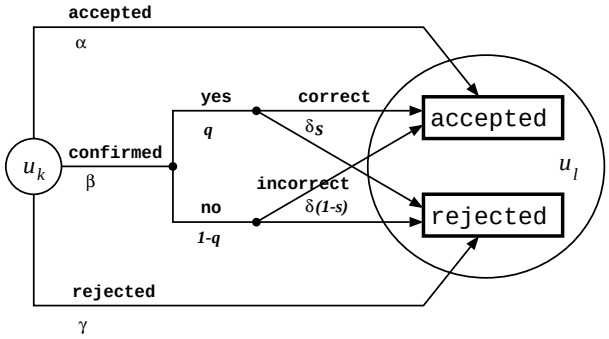

Figure 2: The operation performed during the transition from u_k to u_l.

item is correctly accepted, and the meaning of the third path is that the negative response to a confirmed item is incorrectly accepted as an affirmative response. However, the dialogue system cannot distinguish between the second and third pathes, because it accepts only those of the confirmed items of which the response has been recognized as an affirmative one.

In order to compute the probabilities $p(k,l)$ and $p_{ac}(k,l)$ which have been introduced in the previous section, we first consider the probabilities associated with the three pathes. By referring to the definitions of the recognizer parameters described in section 2, the probability that the first path is taken is α, and the probability that the item accepted by this path has been recognized correctly is p. The probability that either of the second and third pathes is taken is $\beta\epsilon$ where $\epsilon = \delta[qs + (1-q)(1-s)]$, and the probability that the item accepted by either of the second and third pathes has been recognized correctly is $\delta qs/\epsilon$. From these considerations we have,

$$p(k,l) = {}_kC_l(\alpha+\beta\epsilon)^{k-l}(\gamma+\beta(1-\epsilon))^l \tag{6}$$

$$p_{ac}(k,l) = {}_kC_l(\alpha p+\beta\delta qs)^{k-l}(\gamma+\beta(1-\epsilon))^l \tag{7}$$

Now we consider $N(n)$ and $P_{ac}(n)$ for the full version of the interactive recovery strategy. As previously stated $(k-l)$ items are accepted during the state transition from u_k to u_l. These items are accepted by one of the pathes shown in Fig. 2. If at least one of them is accepted by the second or the third path, four utterances are produced during the transition; the speaker must speak at the steps (1)(or (7)) and (4), and the dialogue system at the steps (3) and (6). On the other hand, if all of them are accepted by the first path, only two utterances are produced, because the steps (3), (4) and (5) are omitted in this case. The probability $p'(k,l)$ for this special case to occur can be written as ${}_kC_l\alpha^{k-l}\gamma^l$, where $\gamma = 1-\alpha-\beta$. Furthermore, if the destination of the transition is u_0, the number of the utterances produced in each case reduces by one, because the interaction between the speaker and the dialogue system terminates at the step (5). Thus we have the following recursive formula for $N(k)$.

$$\begin{aligned} N(k) = & \sum_{l=1}^{k}\{(p(k,l)-p'(k,l))(N(l)+4) \\ & +p'(k,l)(N(l)+2)\} + 3(p(k,0)-p'(k,0)) \\ & +p'(k,0) \qquad (1 \leq k \leq n) \end{aligned} \tag{8}$$

Substituting ${}_kC_l\alpha^{k-l}\gamma^l$ for $p'(k,l)$, we can rewrite eq. (8) as follows.

$$N(k) = \frac{\sum_{l=1}^{k-1} p(k,l)N(l) + 4 - (\alpha+\beta\epsilon)^k - 2(\alpha+\gamma)^k}{1-p(k,k)} \quad (1 \le k \le n) \tag{9}$$

Next we consider $P_{ac}(n)$. Eq. (4), the recursive formula for $P_{ac}(k)$, holds for the full version of the interactive recovery strategy. Substituting eqs (6) and (7) for $p(k,l)$ and $p_{ac}(k,l)$ in eq. (4) respectively, we can rewrite it as follows.

$$P_{ac}(k) = \sum_{l=0}^{k-1} \frac{{}_kC_l(\alpha p + \beta\delta qs)^{k-l}(\gamma+\beta(1-\epsilon))^l P_{ac}(l)}{1-(\gamma+\beta(1-\epsilon))^k} \quad (1 \le k \le n) \tag{10}$$

By the mathematical induction we have,

$$P_{ac}(k) = \left(\frac{\alpha p + \beta\delta qs}{\alpha+\beta\epsilon}\right)^k \qquad (0 \le k \le n) \tag{11}$$

4. DESIGN OF A DIALOGUE SYSTEM

Here we relate the mathematical analysis presented in the previous section to specification of a dialogue system. Assume that a dialogue system which uses the simplified strategy be necessary to satisfy the following conditions.

(1) A speaker can produce an utterance including at most n_0 items.

(2) The probability that all of these items are correctly recognized is greater than a constant P_0, that is, $P_{ac}(n_0) > P_0$.

(3) The average number of turns taken between the system and a speaker to accept an original utterance of the speaker is less than a constant N_0, that is, $N(n_0) < N_0$.

Using eq. (3), we can compute $N(n_0)$ for various α, and then determine the lower limit $\alpha(n_0)$ of α necessary to meet the third condition. As assumed in section 2, we can determine $p(n_0)$ for $\alpha(n_0)$, which should satisfy the second condition, that is,

$$P_{ac}(n_0) = (p(n_0))^{n_0} > P_0.$$

If this inequalty holds, then we can construct a dialogue system satisfying the conditions (1) to (3) using the given speech recognizer. Otherwise, we must improve the performance of the speech recognizer, or must release some of the three conditions.

5. CONCLUSION

This paper has presented an interactive strategy to recover from speech recognition errors. Receiving an utterance containing several items, the dialogue system using this strategy accepts those items which have been recognized reliably, while it confirms the content of those items or rejects those items which have been recognized unreliably. We have shown the operation of such a dialogue system can be described by a Markov process, and derived two quantities N and P_{ac} which can be used to evaluate the performance of the system: $N(n)$ is the average number of turns taken between the dialogue system and a speaker until all the items are accepted, and $P_{ac}(n)$ is the probability that all the items is recognized correctly. In other words, we have derived a quantitative relation among the complexity (the number of items) of utterances which can be used in dialogues, the performance of a dialogue system and that of the speech recognizer used in it, and shown how to apply the derived formulae to design a dialogue system.

In this paper only replacement error of items are assumed to occur in speech recognition for simplicity, but this is not the case. So it is left as future work to deal with insertion and deletion of items.

REFERENCES

[1] Zue, V., Glass, J., Goodine, D., Leung, H., Phillips. M., Polifroni, J. and Seneff, S, "The Voyager Speech Understanding System: Preliminary Development and Evaluation," Proc. of ICASSP, pp.73-76 (1990)

[2] Peckham, J., "Speech understanding and dialogue over telephone: an overview of progress in the SUNDIAL project," Proc. of the DARPA Speech and Natural Language Workshop, pp.14-27 (1992).

[3] Cozannet, A. and Siroux, J., "Strategies for oral dialogue control," Proc. of ICSLP, pp.963-966 (1994).

[4] Niimi, Y. and Kobayashi, Y.,"Modeling dialogue control strategies to relieve speech recognition errors," Proc. of EUROSPEECH'95, pp.1177-1180 (1995).

[5] Niimi, Y. and Kobayashi, Y.,"A dialog control strategy based on the reliability of speech recognition," Proc. of ICSLP'96 pp.534-537 (1996).

[6] Young, S., "Detecting misrecognitions and out-of-vocabulary words," Proc. of ICASSP, vol.2, pp.21-24 (1994).