

音声認識システムにおける
認識精度の推定法

6N-6

北井 幹雄 吉田 孝

(NTT 電気通信研究所)

1. まえがき

人と機械が音声で会話をしながら処理を進めていく音声認識システムにおける適応的な会話手順の実現法を検討している¹⁾。本稿では、音声認識装置が出力する認識候補の正誤の推定に、認識時のマッチング距離の統計情報を用いることの効果について述べる。

2. 適応的な会話手順

電話したい相手の名前を入力すれば接続する音声ダイヤルの運用実験を行い、操作性、保守性に関する諸課題を検討してきた^{2) 3)}。そこでの会話手順の例を図1に示す。

音声認識性能の制限から認識候補の確認は必須であるが、利用者の操作時間短縮の観点から、認識候補の確認を極力省略することが重要である。候補の確認の省略には、候補を正解と推定して次の処理に進む場合(ケース1: 図1の太線①)と候補を不正解と推定して利用者に再発声を要求する場合(ケース2: 図1の太線②)がある。

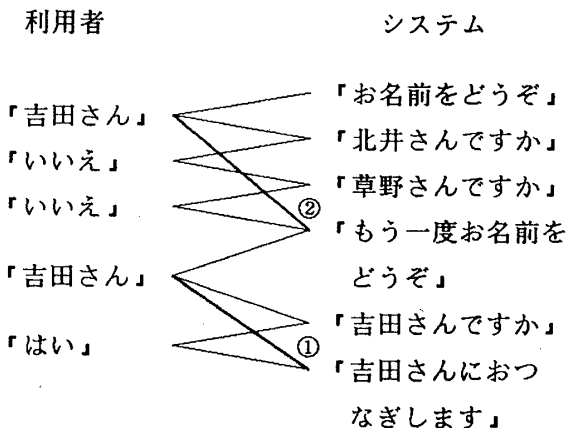


図1. 会話手順

3. 音声ダイヤルでの認識候補の正誤の推定

運用実験時の認識候補の正誤の推定は、認識時のマッチング距離に対し、認識候補を正解あるいは不正解と推定するしきい値を予めシステムに固定的に決めて置いて、認識候補の入力語とのマッチング距離をしきい値と大小比較することで行った(固定しきい値)。

この方法では、認識候補の正誤の推定の誤りはなかったが、認識の第一候補を正解と推定して確認を省略したのが全体(運用時に利用数の多かった4人に対する約7ヶ月間の認識装置の認識結果の1051データ)の約14.5%、認識不成功と推定して利用者に直ちに再発声を要求したのが全体の0%しかなかったので改善の必要がある。1051データのうち、第一候補が正解であるデータは902個、正解を含んでいないデータは118個(残りは第二候補が正解であるデータの41個)あったので改善は可能である。

運用実験で得られた利用者毎の認識時のマッチング距離の分布(平均と分散をガウス分布で近似した例を図2に示す)から、利用者毎に認識候補の正誤を推定するしきい値を変えることで、認識候補の確認省略を向上できることが明らかになった²⁾ので、以下に述べる。

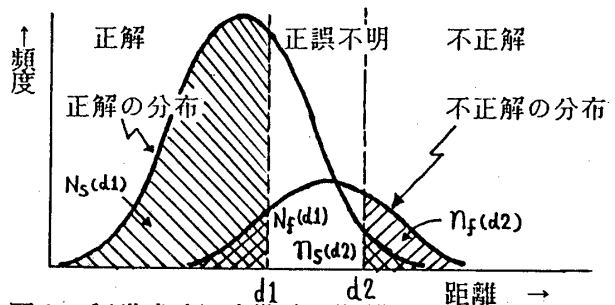


図2. 認識成功、失敗時の距離の分布

Evaluation of Voice Recognition Accuracy

Mikio KITAI, Takashi YOSHIDA

NTT Electrical Communication Laboratories

4. 統計情報を用いた認識候補の正誤の推定

図2の様な認識正解、不正解時の距離の分布に対し、推定誤り率を α とした時に、(1)式を満足する距離の最大値 $d1$ を、認識候補を正解と推定するしきい値とする。同様に推定誤り率を β とした時に、(2)式を満足する距離の最小値 $d2$ を認識候補を不正解と推定するしきい値とする(可変しきい値)。

$$\alpha > \frac{Nf(d1)}{Ns(d1)+Nf(d1)} \quad (1)$$

$$\beta > \frac{ns(d2)}{ns(d2)+nf(d2)} \quad (2)$$

但し、 $Ns(d1)$ は距離が $d1$ 以下であった認識正解時のデータの総数、 $Nf(d1)$ は距離が $d1$ 以下であった認識不正解時のデータの総数である。また、 $ns(d2)$ は距離が $d2$ 以上であった認識正解時のデータの総数、 $nf(d2)$ は距離が $d2$ 以上であった認識不正解時のデータの総数である。

先と同じ運用実験時の4人の1051個のデータに対し、認識候補の正誤推定の誤り許容率を0% (0%あるいは1%)とし、距離の分布からのしきい値の決定をデータが50 (100)個以上得られてから行った場合の、認識候補の確認省略率、再発声要求率などを表1に示す。

但し、認識時の距離のデータが50あるいは100個を越えるまでは、運用実験時と同じしきい値を使って認識候補の正誤の推定を行った。

第一候補を正解と推定して正しく候補の確認を

省略した割合は45~55%で、固定しきい値時の14.5%に比べてかなり向上した。しかし、誤って正解と推定して確認を省略した割合は、距離の分布によるしきい値の決定を開始するのに必要なデータ数を100とした場合は全体の1%以内に抑えられているのに対し、必要なデータ数を50とした場合は2.3%となった。音声ダイヤルでは誤って正解と判断されると誤接続となり、再ダイヤルの必要があるので後者のように失敗率が大きくなることは問題が大きい。

一方、第一候補を不正解と推定して正しく再発声を要求した割合は0.2%しかなかった。誤り許容率 β を振っても効果はなかった。

以上の結果より、認識時の距離の分布を認識候補の正誤を推定するしきい値の決定に利用すれば、認識候補の確認の省略に効果があることが分かった。また、不正解の認識候補のリジェクトには、ここで挙げた方法では効果がないことが分かった。

5. あとがき

距離の分布から決定されるしきい値による、認識候補の正誤の推定の有効性が確認された。

認識対象とする単語セットによっては、入力音声のパワーや発声長を認識候補の正誤推定に使用すれば、更に推定の確度が向上する見通しが得られており、今後の課題である。

1. 距離の分布を利用した識候補語の正誤の推定 (利用者4人、1051データ)

項目	必要データ数			
	50	100	100	
	誤り許容率	0%	0%	1%
CASE 1	正しく確認を省略した数	504 48.0%	407 38.7%	431 41.0%
	誤って確認を省略した数	24 2.3%	5 0.5%	8 0.8%
CASE 2	正しく再発声を要求した数	2 0.2%	2 0.2%	2 0.2%
	誤って再発声を要求した数	2 0.2%	1 0.1%	1 0.1%

(文献)

- 1) 吉田他「音声ダイヤル方式の検討 (3)」信学技報 SE84-138, (1985).
- 2) 吉田「音声認識用テンプレートの自動取得」61・信学総全大No.1954
- 3) 吉田他「音声ダイヤルにおける操作時間分析」60・信学総全大No.187