

論文 / 著書情報
Article / Book Information

論題(和文)	音声対話を用いた情報検索システムの検討
Title(English)	
著者(和文)	山口晃一郎, 古井貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	日本音響学会 1998年春季講演論文集, Vol. , No. 3-6-19, pp. 119-120
Citation(English)	, Vol. , No. 3-6-19, pp. 119-120
発行日 / Pub. date	1998, 3

◎山口晃一郎 古井貞熙(東工大)

1. はじめに

以前我々は、連続音声認識のための効率的なネットワーク構造を提案し、情報検索システムに有効であることを示した^[1]。今回は、このネットワーク構造を利用して、以前とは異なるタスクに適用した。さらに対話システムを導入し、利用者の要求をできるだけ満たすよう試みた。また、種々の対話戦略に対して評価方法を設定し、いずれの戦略が優れているかを検討できるようにした。

2. 音声認識システム

以前我々は、電話番号案内システムにおいて、単語辞書を DAWG 構造を用いて作成した^[2]。この DAWG 構造は探索空間を削減し、認識速度、認識性能の双方を向上させることができる。この構造はネットワークへの単語の追加や削除を一単語単位で高速に行うことができるという特徴も持っている。

今回、我々が対象としたタスクは店舗情報案内システムである。このタスクは認識対象として、区名、業種という2つのキーワードクラスを持つ。業種は食品関係に限定した。認識対象の連続音声はこれらのキーワードに加えてその前後に入るノンキーワードの組合せからなる。キーワード部分の認識結果から、該当する店舗の住所、電話番号を出力する。「えーと、渋谷区のお菓子屋さんお願いします。」といったものが例文として挙げられる。ノンキーワードの部分は以前のシステムの語彙に現システム特有の表現を追加した。DAWG 構造の特徴からこの追加は容易に実現することができる。アルゴリズムとしては、まず追加する単語の接頭辞を DAWG ネットワーク中で探索し、存在すればマージを行い、存在しなければ新たな状態として加える。そして、追加する単語の残りの部分を可能なだけマージする。

図1において ku は区名を、mise は店名を表すキーワードモデル。g1、g2、g5 はそれぞれ文頭、文中、文末に挿入される可能性のあるノンキーワード(フィラー)モデル。g3 と g4 は今回新たに導入したノンキーワードモデルである。g3 は店名につく接頭語「お」を表し、g4 は店名につく接尾語「さん」等を表す。

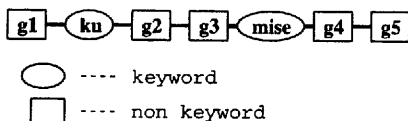


図1. メイン文法

3. 音声認識実験

今回のシステムの単語辞書は 306 単語のキーワードを持つ。ノンキーワード 1019 単語を新たな集合としてネットワークに加えた。メイン文法において、キーワード単語は区名、店名の順で出現するものとするが、区名のみの脱落は認める。また、店名は言い直しに対処するため二回までの出現を許すものとする。

3.1 学習および評価データ

学習用音素 HMM は以前使用したのと同じものを使用する。評価セットは男性話者 8 名の 104 文章で、すべての発声は計算機室に設置されたマイクを使って収録された。

3.2 認識実験結果

表1に認識実験の結果を示す。top は1位に、top n は n 位までの候補に正解がある確率を示す。以前のタスクとは、辞書の大きさ、メイン文法の制限、評価データ等の違いから比較は難しいが、ほぼ同程度の性能と言える。

表1. 認識性能

		正解率		
候補		top	top 5	top 10
店舗情報タスク	区名	87.5	91.3	92.3
	店名	80.8	94.2	98.1
	平均	84.2	92.8	95.2
電話番号タスク	平均	86.2	96.7	

4. 対話システム

人間と機械が対話することによって、システムの不完全な部分を補い、最終的にはシステム利用者の要求を満足する回答を導き出す。システムの入力には音声とタッチ入力の併用、システムからの出力はディスプレイを想定した。今回は対話戦略として4つのパターンを考案した。いずれのパターンでも、区名をまず決定し、その後に店名を決定するという流れをとっている。また、メイン文法の性質から区名の部分が認識されない場合がある。その場合はまず、区名をもう一度発声してもらい、区名だけについて認識を行うものとする。

パターン1 … まず、1位を画面に出力し、正解があるかどうか質問する。1位の候補の中に正解がなかった場合は、2位から10位までの候補をすべて出力し、その中から正解を選んでもらう。この中にも正解がない場合は、もう一度発声してもらい認識を行う。

* A study of information retrieval systems using spoken dialogue.

By Kohichiro Yamaguchi and Sadaoki Furui (Tokyo Institute of Technology)

パターン2 … 1位から10位までの候補をすべて出力し、その中から正解を選んでもらう。この中にも正解がない場合は、もう一度発声してもらい認識を行う。

パターン3 … 1位から10位までの候補を1つずつ出力し、その都度、正解かどうかを質問する。候補すべてを質問しても正解がない場合は、もう一度発声してもらい認識を行う。

パターン4 … 1位の候補を1つ出力し、正解かどうかを質問する。正解でない場合は、もう一度発声してもらい認識を行う。

4.1 対話システムの評価方法

様々な対話戦略を評価するために、対話に対してコストを導入する。各々の戦略に対してコストを予想して、戦略の善し悪しを決定する。このコストとして、以下のものを導入した^[3]。

W_i … コンピュータの指示、問を聞く(見る)コスト。

$W_s * (Sp + Cp)$ … 人間が発話し認識結果を待つコスト。 Sp は人間の発声時間(Speech time)、 Cp はコンピュータの計算時間(Computing time)。

$W_c * <\#choices>$ … タッチ入力における選択肢数に対するコスト。 $<\#choices>$ は選択肢数。

4.2 評価結果

各々のパターンに対して認識率を基に対話のコストの期待値を計算すると次のよう近似できる。

パターン1 … $(2 + p_2 + p_4)W_i + (p_3 + p_4)W_s + (4.1 + p_1 - 2p_3 + 2.7p_2)W_c$

パターン2 … $(2 + p_4)W_i + (p_3 + p_4)W_s + (5.5 - 2p_3)W_c$

パターン3 … $(2 + p_1 + p_2 + p_4 + \alpha)W_i + (p_3 + p_4)W_s + (4 + 2p_1 + 2p_2 - 2p_3 + \alpha)W_c$

パターン4 … $(2 + p_1 + p_2)W_i + (p_1 + p_2 + p_3)W_s + (4 - 2p_3)W_c$

ここで、 $p_1(p_2)$ は区名(店名)において1位に正解が現れない確率、 $p_3(p_4)$ は区名(店名)において10位までに正解が現れない確率、 α は区名と店名の両方において、2位から10位までそれぞれの順位について正解が現れない確率の合計値を表す。ただし、これらの確率は同じ認識システムを使って得られる数値であるので、独立しているものではない。どれかが良くなれば、他の数値も良くなると考えられる。今回は、

W_i, W_s, W_c 各々の係数について $p_1 = kp_2 (k \geq 0)$ とし、誤認識率 p_2 の関数としてグラフに表してみた(図2)。図中、 $\hat{p}_2$ は今回の実験での認識率を表す。この図から、各々のコストのどれを低く設定したいかによってどのパターンを選べば良いかがわかる。

今回の認識率では、機械の発話回数を少なくしたければパターン2を、人間の発話回数を少なくしたければパターン4以外を、選択肢数を少なくしたければパターン4を、平均的にコストを少なくしたければパターン1を、といった選択が可能となる。現システムにパターン1を用いた時のシステムからの平均発話回数は2.2回、利用者の平均発声回数は1.1回、平均キータッチ回数は2.1回となる。将来、認識率が上がった場合はパターン4が優れたものになることも導き出される。

5. まとめ

本論文では、DAWG構造を用いた大語彙連続音声認識システムを、以前とは異なるタスクに適用し、新たなシステムを作成した。そして実験を行うことによってこのシステムの認識性能を確認した。

また、タッチ入力を併用する対話システムを構築した。さらに、対話システムの評価法を導入し、様々な対話戦略を比較できるようにした。実際、いくつかの具体的な対話戦略を考案し、それぞれを比較することによって各々の対話の善し悪しを検討した。

謝辞

店舗情報データを提供していただいたNTT情報開発株式会社、様々なアドバイスをいただいたNECの花沢健様、ならびに東工大古井研究室の皆様へ感謝致します。

参考文献

- [1] 花沢健, 南泰浩, 古井貞照, "連続音声認識のためのネットワーク構造を用いた効率的探索手法", 日本音響学会講演論文集2-6-5, 平成9年3月。
- [2] 花沢健, "会話音声認識のための言語処理モデルの研究", 東京工業大学修士論文, 平成9年3月。
- [3] E. Levin, R. Pieraccini, W. Eckert, "Learning Dialogue Strategies within the Markov Decision Process Framework", IEEE ASRU97 Workshop, pp72-79, 1997。

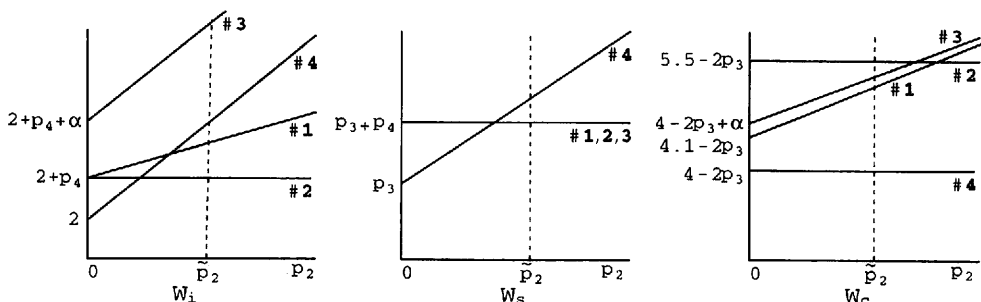


図2. 誤認識率に対する各コストの係数の変化 (# : 対話パターンの番号)