

音声対話と全文検索を利用した電子ガイドシステム(2)

2E-5

—音声認識部—*

山田雅章 伊藤史朗 酒井桂一 小森康弘 藤田 稔

キャノン(株) 情報システム研究所

1 はじめに

我々は、音声対話を用いて「旅行」に関するガイダンスを行なうシステム TARSAN の研究開発を進めている。TARSAN では全文検索を用いて CD-ROM 文書から情報を抽出するため、対象となるデータの規模が大きくなる。音声認識部もこれに合わせて大規模データに対応する必要が生じる。

本稿では、大規模データに対応した音声認識の実現手法として、動的な文法を用いた方法について述べる。

2 音声認識部の構成

概要

音声認識部の概要を図1に示す。

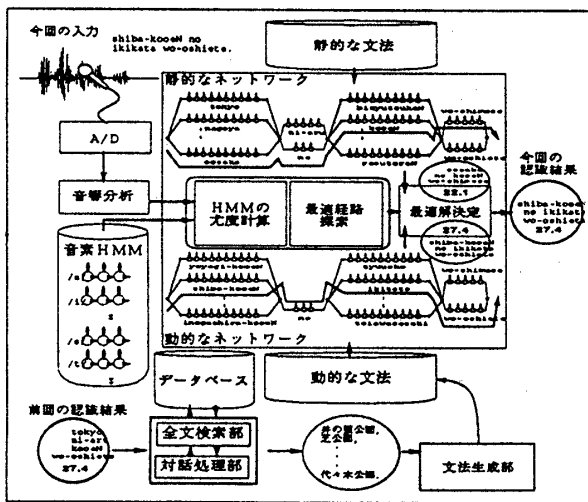


図1: 音声認識部

音声認識部には、正規文法表現から展開した音素HMMによる有限状態オートマトンを文法として用いている。

動的な文法の使用

自然な対話を進めるには、地名や検索対象(例えば温泉)の名前といった固有名詞を認識することが必要となる。しかし、これらをそのまま認識対象として文法に

取り込むと、語彙数が大きくなりすぎて認識率・認識時間に悪影響が生じる。これを解決するために対話状況や検索結果を用いて文法を動的に変更する方法をとった。これにより、次発話において発声してほしい文や発声されると予測される文を受理することができるようになる。文法を変更するのに必要な情報は対話管理部[2]が生成する。

また、上記の動的に変更される文法以外に、どんな対話状況においても入力可能な文を受理する静的な文法を用いている。

高速化

音声認識の速度を向上させるために、ビームサーチおよび複数の計算機によるパラレル化を行なっている。

複数の計算機によるパラレル化は、HMMの出力確率の計算・オートマトン上の探索の双方において行なっている。HMMの出力確率の計算は、計算機の演算能力に応じて入力音声パラメータ列を各計算機に割り振ることでパラレル化している。オートマトン上の探索は、動的・静的文法のそれぞれを別々の計算機上で行なうことでパラレル化している。文法を動的・静的の2つに分割することにより、文法の動的変更および正規文法からオートマトンへの展開に要する計算コストを小さくすることができる。これらの文法はそれぞれ更に分割可能であるが今回は分割しなかった。これは、今回用いた文法の場合ではオートマトン上の探索に要する計算時間が約1~2秒であり、計算機間の通信に要する時間を考慮すると、文法の分割によるオートマトン上の探索時間を高速化する効果が薄いためである。

また、動的な文法の生成・オートマトンへの展開は、合成音声による検索結果の読み上げの最中に行なっている。これにより音声認識開始時のオーバーヘッドを軽減している。

対話処理部との接続

本システムでは入力インタフェースとして音声入力とキーボード入力を用意している。ユーザは音声入力とキーボード入力を自由に使い分けることができる。これを実現するために、音声認識の結果を漢字仮名混じり文に変換して対話処理部に送っている。

3 ハードウェア構成

音声認識部のハードウェア構成を図2に示す。

音声認識部には NeXTstation Turbo(33MHz clk) 2台と NeXTcube(25MHz clk) 1台を用いている。各 NeXT

*An Electronic Guidance System with Speech Conversation and Full Text Retrieval (2) - Speech Recognition - Masayuki YAMADA, Fumiaki ITOH, Keiichi SAKAI, Yasuhiro KOMORI, Minoru FUJITA (Information Systems Research Center, Canon Inc.)

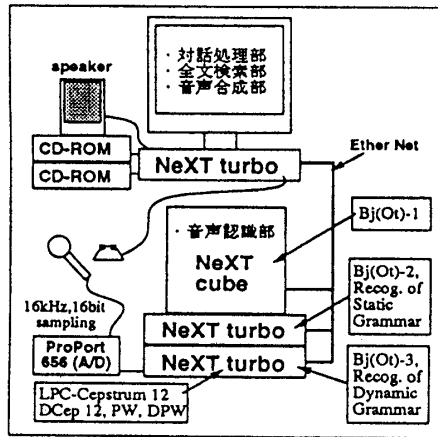


図 2: 音声認識部

間を Ethernet で接続し、Mach の message 機能を用いて通信を行なう。音声は外付けの A/D コンバータから入力する。

4 実験

実験条件

A/D変換は16bit, 16kHz サンプリングで行なう。パラメータは LPC-cepstrum, power, delta cepstrum, delta power の 26 次元, シフト幅 10ms の音声データを用いた。音素 HMM には, 音響学会データベース (男性 30 名, 約 4,500 文) を用いて連結学習した 3 状態 3 ガウス分布の連続型 HMM を 32 種類用いている。評価には, 本タスク用に作成した音声データベース (男性 18 名, 本タスク用の基本 132 語からなる 250 文, 計 4,500 文, 平均発声時間 2.49 秒) を用いた。タスクの語彙数は, 静的文法で約 130 語, 動的文法で最大約 140 語である。

認識速度実験

HMM の出力確率の計算に用いる計算機 (NeXTstation Turbo 33MHz) を 1~5 台に変えてパラレル化による効果を測定した。実験には静的な文法のみを用いた。表 1 に結果を示す。表中の認識時間は, 上記評価用データベースの音声パラメータを入力してから認識結果が出力されるまでに実際に要した時間の平均値である。なお, 認識率は約 95%であった。

次に, 実際に TARSAN を用いて対話を行なった際の発声時間と音声認識に要する時間を表 3 に示す。併せて, 静的・動的文法の語彙数も示す。実験は図 2 に示した構成で表 2 の対話内容で行なった。本実験は, 秒単位で 20 回測定し平均値をとった。被験者は 1 名である。

以上の実験より, 次のことがいえる。

- 複数台の NeXT を用いたパラレル処理による高速化は, NeXT 4 台が限界である。
- 認識に要する時間は 2~5 秒であり, これは, 発声時間の約 1.2~2 倍程度である。

表 1: NeXT の台数と認識速度の関係

| NeXT の台数 (台) | 1 | 2 | 3 | 4 | 5 |
|--------------|------|------|------|------|------|
| 認識時間 (sec) | 5.67 | 3.67 | 3.17 | 2.84 | 2.87 |

表 2: 対話例

| No. | 話者 | 対話内容 |
|-----|-----|---|
| 1 | Usr | 林間にあるゴルフ場を知りたい。 |
| | Sys | 住所を指定して下さい。 |
| 2 | Usr | 千葉県です。 |
| | Sys | 千葉市に 4 件, 野田市に 4 件, 印旛村に 2 件, 下総町に 2 件など, 合計で 26 件あります。 |
| 3 | Usr | 野田市では。 |
| | Sys | 千葉カントリー倶楽部梅郷コース, 千葉カントリー倶楽部野田コース, 紫カントリー倶楽部あやめコース, 紫カントリー倶楽部すみれコースです。 |
| 4 | Usr | 立地条件が河川敷では。 |
| | Sys | 他の条件, 所在地が野田市, を残しますか。 |
| 5 | Usr | はい。 |
| | Sys | 大利根チサンカントリークラブ, 野田市パブリックゴルフ場です。 |
| 6 | Usr | 大利根チサンカントリークラブの電話番号は。 |
| | Sys | 大利根チサンカントリークラブの電話番号は (0471) 38-2121 です。 |
| 7 | Usr | ありがとう。 |
| | Sys | またどうぞ。 (動的な情報の認識) |

表 3: 対話における認識時間

| 対話 No. | 発声時間 (sec) | 認識時間 (sec) | 静的文法 語彙数 | 動的文法 語彙数 |
|--------|------------|------------|----------|----------|
| 1 | 3.07 | 4.05 | 127 | 0 |
| 2 | 1.79 | 2.1 | 127 | 0 |
| 3 | 1.77 | 3.2 | 127 | 46 |
| 4 | 2.79 | 3.3 | 127 | 92 |
| 5 | 0.93 | 1.95 | 127 | 92 |
| 6 | 3.83 | 4.4 | 127 | 90 |
| 7 | 1.60 | 2.8 | 127 | 90 |

5 おわりに

本稿では, 動的な文法を用いた音声認識手法について述べた。本手法により, 認識率や認識速度を損なうことなく, 大規模データに対応した音声認識を行なうことができる。今後は,

- NeXT 用信号処理ボード ISPW (IRCAM Signal Processing Workstation) による高速化。
- HMM の高精度化による認識精度の向上。
- 不要語・未知語や雑音の処理による耐久性の向上。

を行なっていく。

参考文献

- [1] 藤田 他: 音声対話と全文検索を利用した電子ガイドシステム (1) - システム概要 -, 情報処理学会 46 回全国大会予稿。
- [2] 酒井 他: 音声対話と全文検索を利用した電子ガイドシステム (3) - 対話処理部 -, 情報処理学会 46 回全国大会予稿。
- [3] 伊藤 他: 音声対話と全文検索を利用した電子ガイドシステム (4) - 検索処理部 -, 情報処理学会 46 回全国大会予稿。