

音声対話エージェントによる生駒市コミュニティセンターの案内システム

西村 竜一, 西原 洋平, 鶴身 玲典, 李 晃伸, 猿渡 洋, 鹿野 清宏

奈良先端科学技術大学院大学 情報科学研究科

1 はじめに

音声対話ロボット [1] など音声インタフェースに対する注目は高まっている。しかし、いまだに一般の人々が気軽に利用できる音声対話システムは少ない。このため、ユーザとシステムのインタラクションの記録データの収集が不十分であった。音声インタフェースの実用化をすすめていく上で、フィールドテストによる実際のデータの収集は必要不可欠である。

われわれは、生駒市協力のもと「生駒市北コミュニティセンター ISTA はばたき」の音声情報案内システム「たけまるくん」を開発した(図1)。本システムは、同センター内に常設され、来訪者がいつでも利用できる音声インタフェースを持った案内システムである。また、本システムで来訪者の発話音声の収集を行い、システムの改良に利用することができる。

本発表では、システムの概要および発話データの収集結果について述べる。

2 システムの概要

本システムは、センターの館内施設や観光情報、周辺情報など案内を行うためのシステムである。一問一答形式の音声対話機能を持ち、来訪者の質問に対して、合成音声とたけまるエージェント(図2)を用いて案内を行う。また、関連する Web ページや地図などのイメージを表示することもできる。想定している質問は、以下のようなトピックに関するものである。

1. センター館内の案内(部屋や施設の場所)
2. 業務内容(手続き方法や開館時間)
3. 周辺案内(駅、バス停、郵便局)
4. 奈良・生駒の観光情報
5. たけまるくん自身に関する情報

図3にソフトウェアモジュール構成を示す。

サーバは、各モジュールの出力と動作状態を保持するプログラムである。モジュールはサーバ経由で他のモジュールの出力結果と動作状態を知ることができる。

音声合成モジュールは、Text-To-Speech による応答音声を出力する。赤外線センサを持つ人検知センサは音声認識の録音開始のトリガに利用できる。ソフトウェアエージェントはたけまるエージェントの表示、Web ブラウザは関連 Web ページの提示を行うモジュールである。たけまるエージェントのアニメーションの作成にはマクロメディア社の Flash を利用した。現在、38の動作パターンを持つ。また、たけまるエージェントは、音声認識の録音中のうなずき動作で、ユーザへ発話のタイミングを知らせることができる。

3 用例テキストに基づく応答の生成

応答の生成は、あらかじめ用意した応答候補の中から、一つの応答を選ぶことで行う。現在、202の応答候補が

The Ikoma Community Center Guidance System with Speech Dialogue Agent.

Ryuichi NISHIMURA, Yohei NISHIHARA, Ryosuke TSURUMI, Akinobu LEE, Hiroshi SARUWATARI, Kiyohiro SHIKANO (Graduate School of Information Science, Nara Institute of Science and Technology)

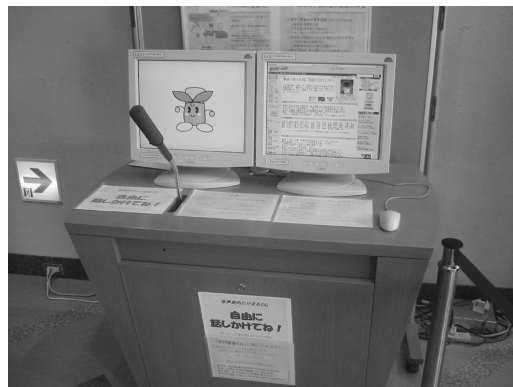


図1: システム外観

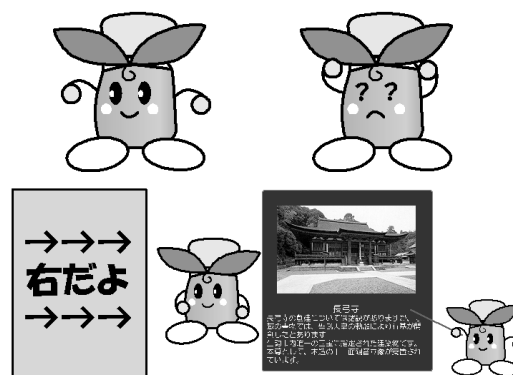


図2: たけまるエージェントの例

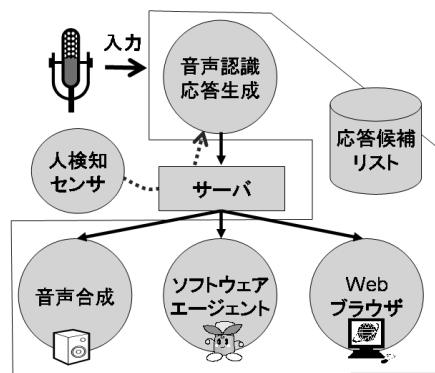


図3: ソフトウェアモジュールの構成

登録されている。

応答の選択は、音声認識結果とあらかじめ準備した用例テキストとの形態素列の一致数のスコア計算による(図4)。

用例テキストは、あらかじめ想定した質問文の形態素解析済みテキストである。全ての用例テキストには、応答候補の中から応答として最も相応しいものを定義している。音声認識結果に対して、用例テキストとの形態素の一致数を求め、対応する応答候補のスコアに加算する。カウントには名詞、動詞、形容詞など自立語のみを用いた。全ての用例テキストに対してカウントし、最も高いスコアを持つ応答候補を応答結果とする。複数の応答候補のスコアが同じ時、それらの中からランダムに選

表 1: 収集音声の年齢層と性別ごとの発話数

年齢層		男性	女性	性別不明	合計
a)	幼児	22	240	100	362
b)	低学年子供 (小学校 3 年生ぐらいまで)	1100	2165	762	4027
c)	高学年子供 (中学生ぐらいまで)	122	115	15	252
d)	成人	2869	718	15	3602
e)	高齢者	7	1	0	8
x)	年齢層の判断ができなかったもの	0	5	29	34
合計		4120	3244	921	8285

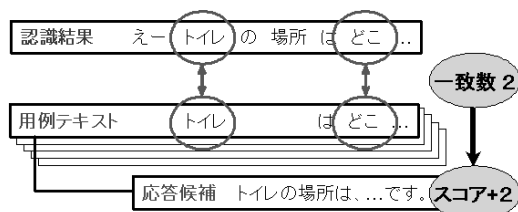


図 4: 用例テキストベースのスコア計算

択した。なお、アンケートや市役所業務記録などから作成した 2,309 文の用例テキストが登録されている。

本手法では、用例テキストの数を増やすことで、発話の言い回しや語句の違いなどの表現様式の違いに対して柔軟に対処できるのが特徴である。

4 音声認識モジュールの構成

音声認識プログラムには、Julius[2] を使用した。音響モデルは、展示会場雑音を重畳した PTM triphone HMM モデル (性別非依存) である。

また、言語モデルは、下記の 2 種類のテキストから 3-gram 言語モデル (単語数 40k) を学習し、マージ (融合割合 1:1) したものを用意した [3]。これをベース言語モデルと呼ぶ。

1. 検索サイト経由で集めた生駒市関連および生駒市ホームページ内の Web ページ, 108 万文, 総単語 3126 万個, 異なり単語 21 万個
2. 人手で収集した想定質問文テキスト, 6,488 文, 総単語 5 万個, 異なり単語 3,231 個

さらにベース言語モデルに文法を適用した文法適用言語モデルを作成した [4]。適用に用いたのは、異なり単語数 441 の人手で作成した本システム向け文法である。

音声認識結果の出力には N-best 出力を用いた。正解単語が 1-best 結果では出力されない場合でも、2-best 以降の本来の正解単語をスコア計算に含めることができる。

5 発話音声の収集とテストセットの作成

本システムの運用開始からは 2002 年 12 月 26 日まで休館日を除く 40 日間で、雑音のみの入力や正しく音声収録されていないものも含めて、22,613 回の発話データ (約 1.3GB, 録音時間 708 分) を収集した。

また、人手による収集データの整備をすすめている。この作業では、削除可能な雑音の切り出しによる除去、発話内容のテキストへの書き起こし、音声のみからの主観による話者の性別と年齢層のラベル付けを行っている。これまでの整備済みデータは 16,070 個であり、そのうち発話を明瞭に聞き取れる比較的クリーンなものは 8,285 個であった。この 8,285 発話の年齢層および性別の分布を示した表 1 から、本システムでは男女を問わず広い年齢層から発話データの収集ができることがわかる。

表 2: 音声認識実験の結果

言語モデル	単語正解率 (%)	単語正解精度 (%)
ベース	81.5	76.9
文法適用	84.4	78.9

次に、本システムの性能指標を調べるため、8,285 発話から成人男女の発話による計 500 文 (男性 280 文, 女性 220 文, 総単語数 3,223 個) を選択し、テストセットデータとした。

6 実験

評価実験では、音声認識性能と本システムがどの程度正しく応答できるかの性能指標を調べた。

音声認識実験の結果を表 2 に示す。文法適用言語モデルはベース言語モデルよりも高い認識精度を示すことがわかる。今回は雑音の比較的少ないテストセットで 80% 以上の性能を得たが、実際の運用を考えると雑音のまじった音声での評価と対策が必要である。また、ユーザに子供が多いため、子供音声での評価も今後の課題である。

次に、応答性能について調べた。応答生成プログラムに、100-best の音声認識結果を入力し、結果を集計した。応答内容が満足なものかは人の主観により判断した。言語モデルには実際の運用に使用している文法適用言語モデルを用いた。

実験の結果、満足な応答を得られたものは、500 文中 348 文で 69.6% の応答正解率であった。

7 まとめ

生駒市北コミュニティセンターの音声情報案内システム「たけまるくん」について述べた。また、本システムを用いて、実際の発話データを収集できることを示し、システムの性能指標を調べた。他の条件での評価や収集データを用いたシステムの改良が今後の課題である。今後も本システムの運用を通じて、様々な観点から音声インタフェースの検討を行っていききたい。

謝辞 白本 和久氏, 米田 秀一氏をはじめとする生駒市職員のみなさまに多大なるご支援をいただいた。深く感謝します。

参考文献

- [1] R. Nisimura et al. : "ASKA: Receptionist Robot with Speech Dialogue System," *IROS2002*, pp.1314-1317, 2002
- [2] A. Lee et al. : "Julius - An Open Source Real-Time Large Vocabulary Recognition Engine," *EUROSPEECH2001*, pp.1691-1694, 2001
- [3] 長友他: "相補的バックオフを用いた言語モデル融合ツールの構築," 情処誌, Vol.43, No.9, pp.2884-2893, 2002
- [4] 西村他: "生駒市コミュニティセンター音声情報案内システムの開発と運用," 情処研報, 2003-SLP-45-6, 2003