

# 音声情報案内システムにおける質問応答データベース 最適化手法の検討

木田 学<sup>†</sup> 川波 弘道<sup>†</sup> 猿渡 洋<sup>†</sup> 鹿野 清宏<sup>†</sup>

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科 〒630-0192 奈良県生駒市高山町 8916-5

E-mail: <sup>†</sup> {manabu-k, kawanami, sawatari, shikano}@is.naist.jp

**あらまし** 本稿では、実環境下で収集した音声データからの音声情報案内システムの再設計手法を提案する。音声情報案内システムとは、音声を入力インターフェースとし、ユーザに知りたい情報を提供するシステムである。しかし、必ずしも全ての入力音声に対して正しい応答を返しているとは限らず、ユーザに誤った情報を提供している場合がある。本手法では、実環境下で収集したデータから質問応答正解データベースを構築し、応答生成システムの性能を評価する尺度である応答正解率を算出した。これにより、収集データを基にしたシステムの再設計が可能となる。応答正解率を基に既存システムの問題点を検証し、悪影響のある品詞の検出と質問応答データベースの最適化を実験的に評価した。その結果、応答正解率は、大人が 75.5%から 81.3%に 5.8%向上し、子供が 47.8%から 61.0%に 13.2%向上した。

**キーワード** 音声対話, 応答正解率, 応答生成, 質問応答データベース, 品詞情報

## Optimizing the Question-Response Database of a Speech based Guidance System

Manabu KIDA<sup>†</sup> Hiromichi KAWANAMI<sup>†</sup>

Hiroshi SARUWATARI<sup>†</sup> Kiyohiro SHIKANO<sup>†</sup>

<sup>†</sup> Graduate School of Information Science, Nara Institute of Science and Technology,

8916-5 Takayama-cho Ikoma-shi, Nara, 630-0192, Japan

E-mail: <sup>†</sup> {manabu-k, kawanami, sawatari, shikano}@is.naist.jp,

**Abstract** In this paper an approach to optimize the question-response database of a speech based guidance system is proposed and evaluated. Such a system uses speech as input, and offers information for the user demands. However the system response may not always be as expected by the user. We propose a technique to improve the response generation system by building up question-response database using speech data collected in a real environment. Problems of existing system are further investigated. The detection of the parts of speech with negative effects and the optimization of the question-response database are evaluated experimentally. As a result, the response accuracy improved from 75.5 to 81.3% for adults and from 47.8 to 61.0% for children.

**Keyword** Speech dialogue, Response accuracy, Answer generation, Question-response database, Parts of speech

### 1. はじめに

実環境での音声認識技術の応用として、音声対話インターフェースが実用化され注目されつつある。音声を入力インターフェースとして扱う利点として、キーボードやマウス、タッチパネルなどのインターフェースよりも使いやすく、人に優しいインタフ

ェースであることが挙げられ、今後ますますその需要は増加すると考えられる。

しかし、既存の対話システムはユーザの要求を完全に満たす段階までは至っていないのが現状である。その要因は様々考えられるが、利用実態に基づく調査・検討が必要不可欠である。人と機械

表1 「たけまるくん」で収集された有効発話

2002.11~2004.10の有効発話データ 98,852発話			
大人		子供	
頻度2以上	頻度1	頻度2以上	頻度1
14,282発話(1,399種類)	5,021発話	52,655発話(5,073種類)	26,894発話

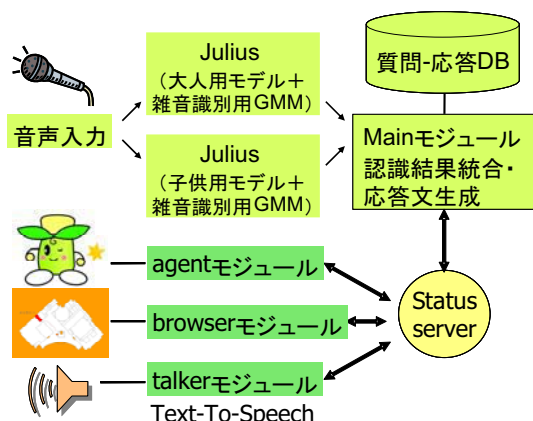


図1 「たけまるくん」のシステム構成図

の対話データを分析することがシステム自体や要素技術の改良の過程では必須である[1]。対話システムに対する満足度は利用したユーザにしかわからず、実環境下での導入例も少ない。そのため、ユーザ音声データベースが不十分であり、実環境下における対話システムの性能評価ができないという問題点に着目した。ここで、自然言語処理のみでなく音声を入力としたシステムとしての応答生成の総合的な評価を行なう必要がある。また、応答性能に悪影響を与えている部分を特定することが必要である。

本稿では、我々の研究室で構築した音声情報案内システム「たけまるくん」[2]の対話モデルにおいて、約2年間収集した発話データに対して正解応答を付与し質問応答正解データベースを構築した。この質問応答正解データベースを用いて、悪影響のあるシステム部を特定し、応答生成システムの再設計をすることでよりユーザが満足する音声対話システムを構築する手法を提案する。雑音や背景会話などの無効発話を除く有効発話のみを対象に評価尺度として応答正解率を用いた。応答正解率を基にして、悪影響のある品詞の検出と質問応答データベースの最適化を実験的に評価し、その有用性について示す。

## 2. 情報案内システムにおける応答生成

### 2.1. 情報案内システム「たけまるくん」

本研究で用いた情報案内システム「たけまるくん」[2]について説明する。システムの構成図を図1に示す。「たけまるくん」システムは、2002年11月から奈良県生駒市の北コミュニティセンターに常設され、施設や付近の観光案内、たけまるくん自身への質問に対して情報案内サービスを行っている。対話形式は、一問一答型で、対話制御は行っていない。そのため、完全に間違った応答を返した場合でも、次の発話の応答生成に影響はない。音声認識エンジンには大語彙連続音声認識エンジン Julius[3]を用いている。大人と子供でそれぞれ異なる音韻モデル及び言語モデルを利用した並列デコーディング処理[4]によって、音響尤度を基準として大人/子供の判定を行なっている。さらに、笑い声、背景雑音などを音声長や雑音 GMM[5]で検出することによりユーザの発話のみを抽出している。応答文は、大人が242種類、子供が255種類である。現在までに、2002年11月から2004年10月までの約2年間の全入力データに対して、発話内容の書き起こし文、発話から推定されるユーザの年齢層、雑音のタグなどのラベルを付与している。この収集発話データの分析を表1に示す。

応答生成システムは、想定される質問文とその答えとなる応答文の QA ペアを持つ質問応答データベース(大人:4,081文、子供:10,367文)の質問に対して、Juliusによる10-bestの認識結果と形態素マッチングを行ない、最も類似度の高い QA ペアを探し出すシステムである。形態素マッチングによる類似度は、QA ペアの質問文それぞれに対して認識結果の単語がいくつ含まれているかをスコアとして計算し、そのスコアを質問文の形態素の数で除算したものである。

### 2.2. 応答生成の課題と提案法

応答生成システムの設計手法として、対話事例の収集データを質問応答データベースに加えることで応答生成性能を向上させる手法が提案されている[6]。実システムでは評価されているが、「たけまるくん」における応答生成の問題点として、

- i. 質問応答データベースに不必要なデータが混入し精度の劣化を起こしている

表2 学習データとテストセットに用いた発話数

	大人		子供	
	期間	発話数	期間	発話数
学習データ(頻度2回以上)	2003.08以外	13,552	2003.08以外	48,608
テストセット(頻度1回も含む)	2003.08	1,091	2003.0801~0809	1,938

表3 現状の「たけまるくん」の応答正解率

年齢層	入力(QAペア)	学習データ	テストセット	テストセットの主観評価
大人	書き起こし文(4,081)	93.9	88.9	93.1
	音声ファイル(4,081)	83.8	75.5	81.2
子供	書き起こし文(10,367)	69.2	65.1	74.0
	音声ファイル(10,367)	59.8	47.8	58.2

ii. 形態素の品詞情報を考慮していない

という二つの要因が考えられる。これらの問題を解決する提案法を以下に述べる。

まず、現在の応答生成システムは、人手によって作成される質問応答データベースに強く依存している。質問応答データベースの中に、形態素マッチングによる類似度が高くなりやすい質問データがあると、まったく違う意図の質問を吸い込む可能性がある。こうした悪影響のあるデータを応答正解率の変動値から検出し、除外することで応答正解率の向上について検証する。

また、形態素マッチングを行なう際に、必ずしも音声認識結果のすべての形態素を扱う必要はない。ユーザの質問の意図は、ある特定の品詞に強く反映されていると考えられる。実環境下での話し言葉を対象に、対話システムの応答生成に不必要な品詞情報を検出し、応答生成時に除外して形態素マッチングを行なうことにより応答正解率の向上について検証する。

本実験で用いた学習データとテストセットを表2に示す。学習データは、2003年8月以外のデータで頻度が2以上のデータ(大人:13,552発話、子供:48,608発話)である。テストセットは、大人は2003年8月(1,091発話)、子供は2003年8月1日から9日まで(1,938発話)のすべての発話である。

### 2. 3. 質問応答正解データベースの構築

質問応答正解データベースとは、あるタスクにおいて想定される質問文とその質問に対する正しい応答文を応答正解データとするQAペアを持つデータベースである。

本稿では、一つの質問文に対して正解の応答文をただ一つとし、質問応答正解データベースの構築を行った。実環境下では、一つの質問に対して複数の正解といえる応答文が存在すると考えられるが、その場合、一つの質問文に対して複数の応答文を付与する必要がある、膨大なコストがかかることになってしまう。そのため、質問に対してユーザを満足させるとされる応答文を応答正解データとして付与した。実環境下で収集された

2002年11月から2004年10月までの約2年間の有効発話(98,852発話)のうち頻度が2以上の大人及び子供の書き起こし文に対して応答正解データを付与した。ただし、表2に示すテストセットの期間に対しては、有効発話すべてに応答正解データを付与した。

### 2. 4. 応答正解率

本稿では、応答生成システムに書き起こし文もしくは音声ファイルを入力したとき、質問応答正解データベースの応答正解データと一致する応答文を生成した割合を応答正解率として、以下の式で定義した。

$$\text{応答正解率(\%)} = \frac{\text{正解データと一致したデータ数}}{\text{入力データ数}} \times 100 \dots (1)$$

### 3. 実験

提案手法の有効性を検証するため、既存の応答生成システムと提案法の応答正解率の比較実験を行った。

#### 3. 1. 実験条件

「たけまるくん」によって収集した発話データから学習データとテストセットについて応答正解率の算出を行った。実験は大人及び子供それぞれの書き起こし文と音声ファイルの入力について行った。

基準となる現状の「たけまるくん」の応答正解率を表3に示す。また、複数の応答が正解として存在するので、テストセットに対して第3者が主観的に評価した結果も表3にあわせて示す。

書き起こし文を正解文として、テストセットの音声ファイルの単語認識率を表4に示す。

表3からテストセットにおける応答正解率は大人の音声で75.5%、子供の音声で47.8%と、大人と子供で大きく差があることがわかる。この要因として、大人は施設の案内や周辺案内など情報案内に関する質問が多いのに対して、子供は「たけまるくん」自身への質問が多いことが考えられる。情報案内タスクに比べ、「たけまるくん」自身への

表4 テストセットの単語認識率

	単語正解率(CORR)	単語正解精度(ACC)
大人	83.9	80.2
子供	73.8	69.0

表5 質問応答データベースで除外されたQAペアの一例

質問文	応答文
何+{ナニ/ナン}+14/0/0 が+ガ+61/0/0 ?+?+77/0/0	ごめんなさい、わすれました。
たけまる+タケマル+6/0/0	私は、たけまるです。
どこ+ドコ+14/0/0 です+デス+74/56/1 か+カ+70/0/0	ここは、北コミュニティーセンターです。
.....	.....

表6 質問応答データベース最適化後の「たけまるくん」の応答正解率

年齢層	入力(QAペアの数)	学習データ	テストセット	テストセットの主観評価
大人	書き起こし文(4,054)	98.8	91.1	94.1
	音声ファイル(4,028)	93.1	81.3	85.4
子供	書き起こし文(10,287)	93.3	80.3	87.8
	音声ファイル(10,145)	75.0	61.0	70.2

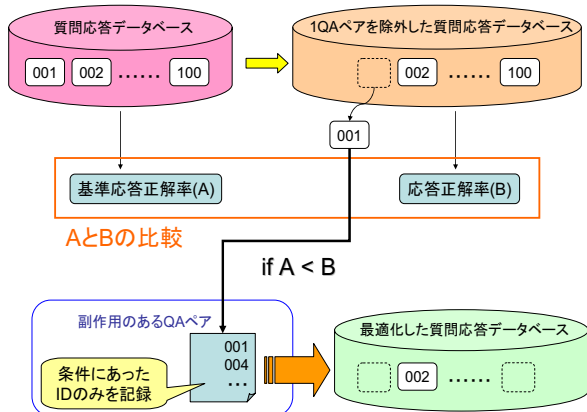


図2 質問応答データベース最適化手法のフローチャート

質問タスクは質問の幅が広く応答生成が難しいためと考えられる。また、書き起こし文と音声ファイルによる応答正解率の違いも子供の方が大きいですが、これは表4の単語認識率の違いによるものと考えられる。

### 3. 2. 質問応答データベースの最適化実験

大人子供の質問応答データベースの最適化のフローチャートを図2に示し、手順を以下で説明する。

- i. 質問応答データベースからQAペアを一つ除外する
- ii. 除外した質問応答データベースを用いて学習

データの応答正解率を算出する

- iii. 表3の学習データの応答正解率と比較し、上回ったものはそのINDEXをリストに記録し、QAペアを質問応答データベースに戻す
- iv. i.~iii.の作業を全QAペアについて繰り返す
- v. リストに記録されているINDEXを質問応答データベースから除く

除外することによって応答正解率が上昇したQAペアは悪影響を与えることのほうが多く、低下したQAペアは必要なQAペアと考えられる。

実験では、大人及び子供それぞれの書き起こし文と音声ファイルについて悪影響のあるQAペアを検出する。検出されたデータを除外することで質問応答データベースの最適化を行なう。

除外されたデータの一例を表5に示す。最適化された質問応答データベースを用いたときの応答正解率を表6に示す。また、テストセットに対して第3者が主観的に評価した結果も表6にあわせて示す。

表6が示すように質問応答データベースを最適化することにより応答正解率が大人の音声で5.8%、子供の音声で13.2%の向上が見られた。また、第3者による主観評価でも大人の音声で4.3%、子供の音声で12.0%の向上した。このことから、不必要なQAペアのデータが適切に除外され、対話システム全体として精度が向上したことがわかる。除外されたQAペアの数を表8に示す。除外されたQAペアの数は元々の質問応答データベースの数の1~2%程度である。除外された質問文は、2、3個の形態素から構成されたものが多く、キーワードとなる単語が1個しか含まれずユーザの意図がはっきり

表7 書き起こし文入力の最適化と音声ファイル入力の最適化の比較実験の結果

年齢層	入力(学習データの種類:QAペアの数)	学習データ	テストセット
大人	書き起こし文(音声:4028)	97.2	93.3
	音声ファイル(書き起こし文:4054)	84.5	76.1
子供	書き起こし文(音声:10,145)	88.2	80.8
	音声ファイル(書き起こし文:10,287)	72.0	59.5

表8 最適化によって除外されたQAペア

年齢層	入力(QAペア)	除外されたQAペアの数
大人	書き起こし文(4,081)	27
	音声ファイル(4,081)	53
子供	書き起こし文(10,367)	80
	音声ファイル(10,367)	222

表9 不要と判断された品詞の一例

品詞情報	例
名詞-非自立-助動詞語幹	「よう」「よ」
名詞-接尾-人名	「君」「さん」「ちゃん」
記号-読点	「、」

表10 不要と判断された品詞情報を考慮した「たけまるくん」の応答正解率

年齢層	入力(QAペアの数)	学習データ	テストセット
大人	書き起こし文(4,081)	94.1	88.9
	音声ファイル(4,081)	83.8	75.6
子供	書き起こし文(10,367)	72.4	66.6
	音声ファイル(10,367)	60.2	48.1

表11 質問応答DB最適化と不要な品詞情報を併用したときの応答正解率

年齢層	入力(QAペアの数)	学習データ	テストセット
大人	書き起こし文(4,054)	98.7	93.8
	音声ファイル(4,028)	93.0	81.3
子供	書き起こし文(10,287)	94.0	85.2
	音声ファイル(10,145)	74.8	60.7

しない質問文が多かった。

次に、書き起こし文と音声ファイルのどちらの入力が質問応答データベースの最適化に適しているか検証する。書き起こし文及び音声ファイルで最適化した質問応答データベースそれぞれに書き起こし文と音声ファイルの両方を入力として応答正解率を算出し、結果を表7に示す。

表7から書き起こし文より音声ファイルで最適化した質問応答データベースの方が、書き起こし文と音声ファイルの両方で好ましい結果が現れた。音声ファイルの入力では、10-best 全ての認識結果でも出現するような名詞や動詞には高い重み付けがなされる。一方、助詞や記号は10-bestの中で頻度が低いと低い重み付けがなされた状態で類似度計算が行なわれている。そのため、2番目以降の候補のQAペアもより適確な順序になる。1番目の候補のQAペアが除外されたとき、適確な2番目の候補のQAペアが選ばれるため応答正解率が上昇しやすいことが考えられる。

また、子供の質問応答データベースは大人のおよそ2倍近く量があり、不必要データも多くあったことが大人と子供の応答正解率の違いに影響しているのではないかと考えられる。

### 3. 3. 悪影響のある品詞情報の検出実験

3. 2. 節の質問応答データベースのQAペアの代わりに chasen(ver.2.3.3)[7]による形態素解析の品詞番号1から88までの全品詞について、3. 2. 節のi.~v.の手順を行なう。

実験は、大人及び子供それぞれの書き起こし文と音声ファイルを入力として実験を行なった。こ

れらの学習データで共通して悪影響があるまたは変化なしと判断された品詞番号を検出し、不要と判断された品詞をすべて除外して応答生成を行ない応答正解率の向上を検証した。

実験の結果、検出された品詞は、大人の書き起こし文で59、大人の音声ファイルで55、子供の書き起こし文で48、子供の音声ファイルで47であった。この中で共通した品詞情報は、38個であった。不要と判断された38個の品詞の一例を表9に示す。この不要と判断された品詞を除外して応答生成を行なったときの学習データとテストセットの応答正解率を表10に示す。

表10から学習データでは、表3と比較してわずかながら応答正解率が向上することが確認できた。しかし、テストセットになると向上率は子供の書き起こし文のときで1.5%しか向上せず、その他では変化がほとんど見られなかった。この要因として、大人及び子供のそれぞれの入力に対して共通する不要な品詞情報を用いたため、効果を打ち消しあう影響がでたと考えられる。また、3. 2. 節で述べたように音声ファイルの入力では10-bestの認識結果を用いているため、助詞や記号よりも名詞や動詞の重みが高くなっており、品詞情報が応答生成に及ぼす影響は少ないことも原因として考えられる。

### 3. 4. 品詞情報と質問応答データベースを組み合わせたときの応答正解率の算出

最適化した質問応答データベースに対して品詞情報を考慮した応答生成を行なったときの相乗効果を検証する。

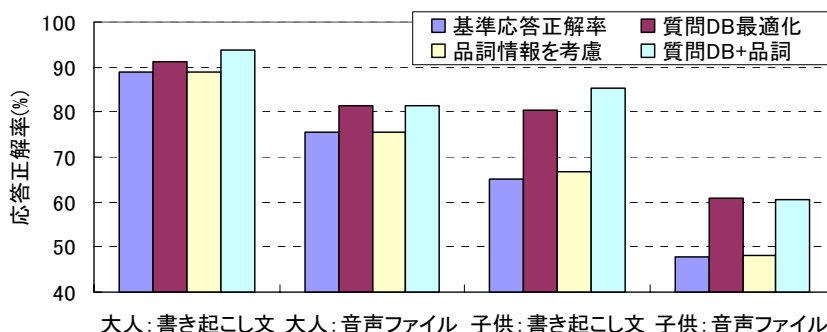


図3 基準応答正解率と提案法による応答正解率

3. 3. 節に示した悪影響のある品詞を形態素マッチングから除外し、最適化された質問応答データベースを用いて応答生成を行なったときの応答正解率を表11に示す。

また、提案法における応答正解率の向上を示すグラフを図3に示す。

表11から大人と子供の両方について書き起こし文では、応答正解率の向上が見られる。しかし、音声ファイルではほとんど向上が見られなかった。品詞情報を考慮した応答生成では、質問応答データベースに悪影響のあるQAペアのデータが少なく、少ないn-bestの認識結果を用いた類似度計算の精度向上が期待できる。今後、単語認識率が向上すれば、応答正解率を向上させることができると考えられる。

#### 4. まとめ

本稿では、約2年間で収集した発話データを元に質問応答正解データベースを作成し、音声情報案内システムの応答生成システムを評価した。質問応答データベースの最適化を行なった結果、応答正解率を大人及び子供の音声ファイル入力においてそれぞれ5.8%、13.2%向上させることができた。形態素マッチングを用いた応答生成において、悪影響のあるデータが少しでも混入していると、応答正解率が大きく低下することがわかる。よって、質問応答データベースを更新する際には、事前に追加するQAペアが悪影響を及ぼす危険性があるかどうかを検証する必要があることが分かった。今回構築した質問応答正解データベースを用いれば、その検証が可能になると考えられる。音声対話システムにおいて10-bestの認識結果を用いた場合、品詞情報を考慮して応答生成を行なっても応答正解率の大きな向上は見られなかった。しかし、最適化後の質問応答データベースを用いて応答生成した場合、書き起こし文を入力として与えたときには応答正解率が大きく向上した。今後、単語認識率が高くなれば、品詞情報を考慮することで、応答正解率がさらに向上することが期待される。

今後の展望としては、品詞情報の導入を重み付けで行なうことで応答正解率の改善を目指す。質

問応答データベースの最適化では、別タスクの音声情報案内システムへの導入でどれだけ効果を発揮できるか試験する必要があると考えられる。

#### 謝辞

この研究は、文部科学省のリーディングプロジェクト「e-Society 基盤ソフトウェアの総合開発」によって行なわれたものである。

#### 参考文献

- [1] A.L Gorin, G. Riccardi, J.H. Wright: "How May I Help You?," Speech Communication, vol.23, pp.113-127, 1997.
- [2] 西村 竜一他, "実環境研究プラットフォームとしての音声情報案内システムの運用", 信学論, Vol.J87-DII, No.3, pp.789-798, 2004.
- [3] Akinobu Lee, Tetsuya Kawahara, Kiyohiro Shikano, "Julius -An Open Source Real-Time Large Vocabulary Recognition Engine", Proc. 7<sup>th</sup> European Conference on Speech Communication and Technology, pp.1691-1694, 2001.
- [4] 西原 洋平, 西村 竜一, 李 晃伸, 猿渡 洋, 鹿野 清宏: "ユーザ層別並列モデルを用いた音声情報案内システム," 日本音響学会講演論文集, 1-8-22, pp.49-50, 2004-3.
- [5] 中村 敬介, 西村 竜一, 李 晃伸, 猿渡 洋, 鹿野 清宏: "実環境音声情報案内システムにおける環境雑音および不要発話の識別," 電子情報通信学会技術研究報告, SP2003-172, pp.13-18, 2004.
- [6] 村尾 浩也, 河口 信夫, 松原 茂樹, 稲垣 康善: "対話事例を利用した音声対話システム," 電子情報通信学会技術研究報告, SP2000-108, pp.79-84, 2000.
- [7] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 高岡 一馬, 浅原 正幸: 日本語形態素解析システム『茶筌』 version2.3.3 使用説明書, 奈良先端科学技術大学院大学, 2003. <http://chasen.aist-nara.ac.jp/>