

## 音声質問応答システムのユーザ発話を用いた質問応答データベース拡張支援

成松 宏美\* 中野 幹生† 船越 孝太郎† 長谷川 雄二† 辻野 広司†

\* 電気通信大学 大学院情報システム学研究所 † (株) ホンダ・リサーチ・インスティテュート・ジャパン

## 1 はじめに

近年、街案内や情報検索などに音声対話が導入されてきている。その一つである質問と応答のペアからなるデータベースを用いる音声対話システム (Question-Answering Spoken Dialogue Systems; QASDS) は、システムの動きが予測しやすく、構築が容易であり、現在多くの研究がなされている [1, 2, 3]。しかしこれらの研究は学習データに大量のユーザ発話が必要である。そこで我々は大量のユーザ発話を収集するための初期システムとして簡易な質問応答システムを開発し [4]、コンテンツ作成者が様々なトピックのシステムを容易に実現することを可能にした。本システムは、質問と応答のペアからなる質問応答データベース (QADB) を用いる。QADB の質問部には予想される質問を文書で記述し、マッチングに用いるキーフレーズを指定する。音声認識用統計言語モデルは記述した質問文例から作成し、認識結果と指定されたキーフレーズを比較することで、例文にマッチしたペアの応答を返す。しかし、あらゆるユーザ発話をコンテンツ作成者が予測することは難しく、未知語であった場合に、応答ができない、認識誤りによりユーザの意図通りの応答ができない等の問題がある。実際にシステムを使ってもらって得た発話をもとにデータベースを拡張することでシステムの性能を向上することができるが、そのためには、コンテンツ作成者が全ての発話を聞いたり書き起こしたりする必要がありコンテンツ作成者の労力がかかってしまう。

本研究では、システムが自動でデータベース外発話であることを検出し、データベースに追加すべき発話のリストを提示することで、全ての発話を聞いたり書き起こしたりするコンテンツ作成者の手間を省き、データベース拡張を容易に可能にすることを旨とする。分類器を用いてデータベース外質問を検出する従来手法 [5] に対して、データベース外である確率を示すスコアの低い順にコンテンツ作成者に提示をすることで、システムの能動学習を可能にし、作業効率の向上を目指す。

## 2 提案手法

提案手法では、機械学習によって構築した分類器を用いて、システムが自動でデータベース外発話を検出し、その発話がデータベース外である確率を示すスコアを算出する。分類器は、データベースから作成した言語モデル ( $LM_{db}$ ) で認識した結果から得られる特徴量と Julius ディクテーションツールキット [6] 付属の大語彙の言語モデル ( $LM_{lv}$ ) を用いて認識した結果から得られる特徴量 (特徴量の詳細は [5] を参照) を利用する。

得られた特徴量は、発話が同じキーフレーズをもつ例

文にマッチしたかの確からしさ、すなわち選択された応答がユーザの意図通りかどうかや、キーフレーズの認識の正しさなどを示す。また、 $LM_{lv}$  を用いて認識した結果と  $LM_{db}$  を用いて認識した結果との組み合わせにより得られる特徴量は  $LM_{db}$  での認識の確からしさを示す。これは、発話の信頼度を求める技術に似ており、発話の中に QADB にあるキーフレーズ以外の名詞句やフレーズなどキーフレーズとなり得るものが存在するか否かを推定することができる。もしそのようなフレーズが存在すれば、その発話はデータベース外発話である可能性が高いと言える。

## 3 実験と評価

## 3.1 実験データ

データマイニングツールキット weka [7] を用いて、発話から抽出した特徴量を用いた Naive Bayes Classifier [8] を学習した。用いたデータは、先行研究 [5] で用いたものと同じで、以下に概要を示す。

- 発話データ: 街案内に関する QASDS で収集した発話のうち 25 人  $\times$  25 発話 = 625 発話
- 言語モデル: 初期データベースの例文から作成した言語モデル 語彙数 460 語 + 5000 語<sup>1</sup> ( $LM_{db}$ )、と Julius ディクテーション実行キット付属の言語モデル [6] 語彙数 60250 ( $LM_{lv}$ )

言語モデルの学習には palmkit<sup>2</sup> を用い、音声認識には Julius [6] を用いた。

分類器の学習には、発話データのうちノイズや区間検出誤りを除いた 515 発話を 2 つの言語モデル  $LM_{db}$ 、 $LM_{lv}$  を用いて認識した結果から得られる特徴量を使用した。初期データベースが応答を見つけれられたかどうか、認識結果に相槌を含むかどうか、キーフレーズの出現回数など、32 個の特徴を抽出し、特徴選択により残った 25 個の特徴量を用いた。各特徴量の寄与率 (一つの特徴量を除いたときの評価値の変化率) を、検出の F 値がどれくらい低下したかを評価することで求めた。このときの Naive Bayes の閾値は 0.5 とした。以下に、分類に有効な特徴量上位 5 つを示す。

- $LM_{db}$  と  $LM_{lv}$  の認識結果の第一候補における (音響スコア/発話長) の差
- $max_i$  (キーフレーズ  $i$  の認識結果全候補での出現回数/認識結果の全候補に含まれる単語数)
- $\frac{1}{n} \sum_{i=1}^n$  (応答選択に使用したキーフレーズ  $i$  の認識結果全候補での出現頻度/認識結果全候補に現れるの単語数) ( $n$  は応答選択に用いるキーフレーズの数)
- $min_i$  (応答選択に用いたキーフレーズ  $i$  の信頼度)

<sup>1</sup> キーフレーズの湧き出し誤りを防ぐため、使用頻度の高い 5000 語を追加した。

<sup>2</sup> <http://palmkit.sourceforge.net/>

\* Graduate School of Information Systems, The University of Electro-Communications

† Honda Research Institute Japan Co., Ltd.

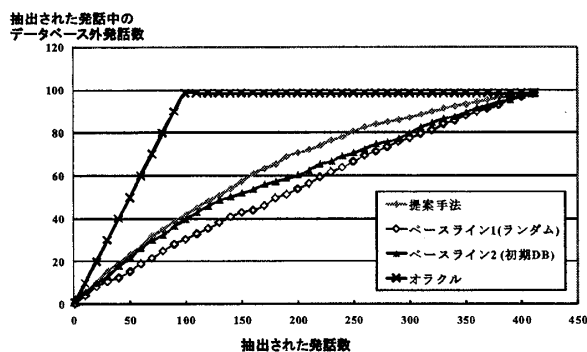


図 1 抽出された発話中のデータベース外発話数

- $min_i$ (キーフレーズ  $i$  の認識結果全候補での出現回数/認識結果全候補に含まれる単語数)

キーフレーズが正しく認識されたかを推定する特徴量が有効であることがわかった。

### 3.2 データベース拡張における作業効率の改善

本手法により作業をどれだけ削減できるかを評価した。

提案手法においては、評価データを 5 人分毎のセットに分け、各評価データセットにつき評価データに含まれない  $n$  話者分のデータで学習した分類器で分類とスコア算出をした。これを全てのデータセットが評価データとなるよう 5 回繰り返す、その平均をとった。以下に示す 3 手法により発話抽出を行った結果を図 1 に示す。

- 提案手法: 分類器により得られたデータベース外であるスコアが高い順に上位  $n$  個の発話を抽出
- ベースライン 1(ランダム): ランダムに  $n$  個の発話を抽出
- ベースライン 2(初期 DB): 初期 DB を用いる手法でシステムがデータベース外と分類した中からランダムに  $n$  個の発話を抽出。もし  $n$  の値が初期 DB によりデータベース外と分類された数よりも大きくなったら、残りはデータベース内発話として分類されている中からランダムに選択
- オラクル: 抽出した発話が全てデータベース外であった場合 (正解)

初期 DB を用いる手法とは、初期データベースの例文から作成した言語モデル  $LM_{db}$  を用いて発話を認識し、得られた認識結果とキーフレーズレベルでマッチする質問文例が初期データベースになれば、データベース外発話とする手法である。グラフの横軸は各手法において抽出した発話数  $n$  を表し、縦軸は各手法で抽出された発話数  $n$  中に含まれていたデータベース外発話数を表す。初めに書き起こした発話に多くデータベース外発話が含まれていれば、作業効率が上がるため、グラフの傾きが大きい程作業効率が高いと言える。評価の縦軸で示しているデータベース外発話数には、同じキーフレーズセットを持つ他の発話が検出されてもカウントせず、データベースに追加すべき発話のみ数えた。例えば { } でキーフレーズを指定すると、『{本屋}は{どこ}にあるの』と『{どこ}に{本屋}があるの』という 2 つの発話は同じキー

<sup>3</sup>初期システムのデータベース外発話検出とは、3.2 節の初期 DB を用いる手法のことを指す。

フレーズを応答選択のキーとするため、重複してカウントしない。

図 1 より、提案手法が最も良いことがわかる。提案手法とベースライン 2(初期 DB) を比較して、提案手法の方がより傾きが大きいため、提案手法においてデータベース外である確率を示すスコアが有効であることがわかった。

## 4 まとめと今後の課題

質問応答のデータベース拡張を支援するための手法として、分類器を用いたスコア付けによるデータベース外発話の自動検出方法を提案した。実験の結果から、提案手法は、初期システムのデータベース外発話検出<sup>3</sup>よりも作業効率を上げることができた。より精度よくデータベース外発話を検出するため、他の有効な特徴量、検出精度を高める言語モデル作成方法について検討する。また、提案手法は、同じドメインの発話を用いて分類器の学習を行っているため、別のドメインのシステムに適用する際には、そのドメインの発話の書き起こしが必要となる。今後、他のドメインの発話で学習した分類器を用いることを検討する。また、本実験は、データベース、分類器ともに固定されているという状況のもとで行ったが、実際のシステム開発状況ではデータベースの常時更新とデータベース外発話の候補を抽出するための分類器の学習が必要となる。今後実際のシステム開発状況に対応した実験も行っていく。

### 参考文献

- [1] 西村他. Julius を用いた学内案内ロボット用音声対話システムの作成. 電子情報通信学会技術研究報告 SP2001-99/NLC2001-64, pp. 93.98, 2001.
- [2] 西村他. 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. 電子情報通信学会論文誌, J87-D-II(3):789.798, 2004.
- [3] T. Cincarek et al. "Development, long-term operation and portability of a real-environment speech-oriented guidance system". IEICE Transactions, 91-D(3):576-587, 2008.
- [4] 成松他. ロボット・エージェント対話行動制御部構築ツール RIME-TK を用いた質問応答機能の実現. 情報処理学会研究報告 SLP-74, 2008.
- [5] 成松他. 音声質問応答システムのユーザ発話を用いた質問応答データベース拡張支援. 情報処理学会研究報告 SLP-77, 2009.
- [6] T.Kawahara et al. Recent progress of open-source LVCSR engine Julius and Japanese model repository. In Proc. IC-SLP, pp.3069-3072, 2004.
- [7] I. H. Witten et al. "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
- [8] G. H. John et al. "Estimating Continuous Distributions in Bayesian Classifiers" In Proc. of the Eleventh Conference on Uncertainty in Artificial Intelligence, pp. 338-345, 1995.