

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 20 日現在

機関番号：14603

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500115

研究課題名(和文) 機械学習および統計モデルに基づく音声対話システムの応答生成の研究

研究課題名(英文) Study on response generation for spoken dialogue systems based on machine learning and statistical models

研究代表者

川波 弘道 (Kawanami, Hiromichi)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：80335489

交付決定額(研究期間全体)：(直接経費) 3,900,000円

研究成果の概要(和文)：音声情報案内システムが広く使われるようになった。しかし用例ベースのシステムの応答性能と応答文の多様性の点では改善の余地がある。そこで、応答性能向上に向けて機械学習によるトピック分類を、応答文の多様化に向けて統計的翻訳技術の援用を検討した。

自動翻訳技術を用いた応答文の多様化については変換元の質問文と変換先の応答文との対応付けの点で課題があることが示唆されたが、機械学習によるトピック分類では、SVMやpboostによる分類、またそれらの識別器の出力を用いた二段階識別の有効性が示された。

また、今後開発が期待されるパーソナルなシステムでのそれらの実現に向けて個人向け雑談ロボットの開発も行った。

研究成果の概要(英文)：Speech-oriented information guidance systems have put into practice. Many of them introduce an example-based response generation technique. However, improvement of response accuracy and variety of response text expression are expected for a more useful interface. In this study, we propose machine learning approach to response selection and statistical machine translation approach to produce variation of response expression.

The proposed machine learning classifier which introduces two-stage classification using outputs of first stage classifiers, SVM, pboost etc. overcomes a conventional example-based method. On the other hand, it remains precise analysis to translate/convert a question to an answer because of difficulty of association between question and a answer.

In addition, we developed a prototype of a chat robot for upcoming demand on a personal spoken dialogue system.

研究分野：音声情報処理

キーワード：音声インタフェース 音声対話システム 情報検索

## 1. 研究開始当初の背景

音声を用いた情報案内システムは、その言語で音声コミュニケーションをすでに行っているユーザにとっては新たに習得すべき負担が少ないことから、機器操作に慣れていないユーザが情報にアクセスする方法として期待されてきた。音声情報案内システムは MIT が 1989 年に開発した Voyager (電話による地理案内) やその応用である Pegasus (航空券予約), Jupiter (天気案内) で広く知られるようになった。日本語音声でも文発声で入力可能なバス運行案内システム (京大) や観光案内 (NICT), インターネット Voice Search (Google) が実現された。

本課題の開始時点で筆者らは公共のコミュニティセンターに用例ベースの音声情報案内システム (引用文献) を 9 年にわたり運用していたが、その頃からインターネットを介したモバイルデバイス上のサービスが一般に普及し始めた。筆者らのシステムも含め、それらのシステムは、用意されたシナリオに基づくシステム主導の情報提供を行うのではなく、主にユーザからの幅広い質問に対して何らかの応答ができるよう、事前に用意した質問 (クエリ) 用例とのマッチングで応答生成を行うものであった。

この手法はシンプルではあるが、筆者らのシステムに関していえば、300 種類程度の応答を行い、ユーザ発話に対して大人から子供まで 70~60% の応答正解率を達成していた。しかしながら特に子供に対しては応答性能の改善の余地があった。

また、応答文はユーザの話し方に合わせてテキスト表現を変えるといったことはしておらず、提示する情報ごとに (大人と子供別に用意した) 定型文を出力していた。そのため、長く使い続けていると「同じ事しか言わない」ことでユーザの利用への関心を低下させている可能性があった。より有用なシステムとするためには応答文において多様性を実現することも検討課題であった。

## 2. 研究の目的

そのような状況を鑑み、本提案では音声情報システムの応答性能の向上 (【課題 1】) と応答文の多様性を実現 (【課題 2】) する手法を研究する。本研究では対話管理は研究対象とせず、一問一答式応答の枠組で性能改善を図る。我々が運用するシステムでもそれを採用しているが、一問一答形式はユーザが情報を即時に得られること、最も単純な対話の形であるため結果的にシステムとユーザの間で状態把握の上で混乱が生じないことなどの利点がある。また一般的にタスクの追加が容易である。

前述のように、従来我々は質問用例データベースとのマッチングによる応答選択を採用していた。小規模なシステムであれば人手

によるクエリ拡張で性能改善は可能であるが、数百種類の応答を持つシステムでは人手による用例追加方法の工夫による改善は限界がある。そこで、運用で得られた発話データに簡単なラベル付けをするのみで機械学習により質問の種類を識別できる方法を検討する。機械学習による分類は当時自然言語処理において盛んに研究されていたが、ユーザのシステム入力発話を用いた研究はあまり行われていなかった。また、大量のユーザ発話コーパスが利用できることを活かして人手を使わないデータも学習に利用する半教師あり学習による手法も検討する。

また、我々のシステムも含め、一般的に情報案内システムは正しく応答することが重要であるため、あらかじめ用意したテキストを出力した。ユーザフレンドリーなシステムとするため、ユーザが求めている情報そのものは同じであっても、ユーザの語彙や言い回しの違いに合わせて応答する表層文を変える方法を検討した。これも人手による規則化ではやはり多様化には限界があるため、本研究では統計的自動翻訳技術に着目し、質問文を応答文に変換するアプローチを検討する。

## 3. 研究の方法

### 【課題 1 : 応答性能向上】

機械学習を用いた応答生成を行うに当たって、人手によるコストを低減するため、ユーザ発話に対して人手で応答文の対応付けのみを行ったデータベースを用意し、学習を行うこととした。学習に用いる特徴量として、ユーザ発話の自動認識結果の Bag-of-Words を用いることとした。ただし、応答文ごとに対応付けられるユーザ発話の多様性や数は、応答文によって多寡があるため、同じトピックの応答文を人手でまとめ 27 トピックに分類し、トピック分類の識別器を作成することとした。トピックを個々の応答文の選択はトピック分類の後段として行う構成とし、トピック分類の精度向上を目的とした。ただし、学習データ数の問題から最終的には、上位階層の 15 トピックによる分類実験を行った。

分類器として、SVM (Support Vector Machine), pboost (PrefixSpan based Subsequence Boosting), 最大エントロピー法を検討した。また、それらの出力結果を組み合わせ、Stacked Generalization のアプローチによる多段分類についても検討した。

### 【課題 2 : 応答文の多様化】

本来の統計的機械翻訳では翻訳元言語のテキストと翻訳先言語のテキストの対からなるテキストコーパスを用いて、言語間翻訳モデルと翻訳先言語の言語モデルを確率モデルで表現する。これを応用し、質問文を応答文に翻訳 (生成) する。言語間翻訳では 2 つのテキストは同一の意味内容を持つため、基本的に語句を 1 対 1 で対応させることが

可能であるが、質問文から応答文への変換は直接的な対応づけが成立しないケースが多い。しかしながら質問で用いられるフレーズとその応答で用いられるフレーズに何らかの対応は存在すると考え、フレーズベースの翻訳モデル(引用文献)をもとに、応答生成への導入の可能性を調査した。

それぞれの課題の評価基準として、応答性能についてはトピック選択正解率を従来の類似度に基づく質問用例選択手法と比較し、応答文の多様化については、まず生成された応答文が正しく情報を含んでいるかの観点で従来手法と比較し、加えて言語の自然さを主観的に評価することとした。

#### 4. 研究成果

##### 【課題1：応答性能向上】

機械学習によるトピック分類に2段階分類を導入することで大人、子供発話とも分類性能を改善された。15のトピックの分類タスクにおいて、1段目の識別に音声認識結果のBOWを特徴量としてSVM, pboost, 最大エントロピー法による識別を行い、それらの出力結果を特徴量としてSVMによる2段目の識別を行う、Stacked Generalization手法を導入した。加えて、人手による応答文ラベル付与を行っていないユーザ発話コーパス内データにトピックの自動付与を行い、開発データの識別性能の向上に寄与するデータのみを学習データに追加する、半教師あり学習にも着手した。コーパスの一部を使用した実験を行った範囲では有意な性能向上には結び付かなかったが、子供音声についてはデータを拡大することで向上する可能性が残されている。

また、機械学習による音声の分類の有効性が示唆されたことで、追加的に、システム入力データの不要入力(雑音と不要発話)の棄却実験も行った。特徴量として自動音声認識結果のBOWに加えて、あらかじめ音声データと雑音データを用いて作成した音響的特徴のGMM(Gaussian Mixture Models)のモデル尤度を用いることで、GMM尤度に基づく従来手法を超える棄却率を実現した。

##### 【課題2：応答文の多様化】

課題1と同様に、データ整備コストの低減と実際のシステム運用でのデータと同質であることから、翻訳モデルの翻訳元データには音声認識を用いた実験を行った。自動音声認識を行うと、一つの音声データから順位付きで複数の候補仮説が得られるため、学習と評価において、複数の候補仮説を用いた場合の評価も行った。結果として、質問文データベースの探索による用例ベースの方式の応答性能と同等以上の応答文は生成されず、応答文は多様化するものの言語として不自然な表層文が生成されるという問題が残された。誤りの分析が必要である。ただし、音声

認識結果を用いた翻訳モデル学習は、書き起し文で学習したモデルを用いた場合より適切な応答文が生成された。導入の方法によっては統計的翻訳を有効に活用できる可能性はあると考えられる。

また、本課題の発展として雑談ロボットのプロトタイプの開発にも着手した。研究進捗の過程で音声情報案内システムの需要として、公共サービスとしてのみでなく、高齢者への情報提供や会話の促進、年齢層に依らず個人に特化したパーソナルなシステムへの要求が社会的に大きくなってきた。そこで、ユーザに特化した応答生成を行うシステムの開発を進めることとした。それらは身体性を持つ卓上型ロボットやスマートフォン上のCGエージェントをインタフェースとした雑談システムとしてプロトタイプを作成した。本研究で取り組んだ2つの課題とその成果を組み込んだシステムの実証実験を行うことは研究期限内には到達しなかったが、ユーザがシステムを利用していくとともに応答が自律的にユーザに適応していく音声情報案内システムは実現が望まれる課題であると考えられる。

#### <引用文献>

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, Vol. J87-D-II, No.3, 2004, pp.789-798  
Philipp Koehn et al., “Statistical Phrase-Based Translation,” Proc. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2003, pp.48-54

#### 5. 主な発表論文等

##### [雑誌論文](計2件)

Rafael Torres, Hiromichi Kawanami, Tomoko Matsui, Hiroshi Saruwatari, Kiyohiro Shikano, “Comparison of Methods for Topic Classification of Spoken Inquiries,” Journal of Information Processing, 査読有, vol.21, no.2, 2013, pp.157-167, DOI:10.2197/ipsj.21.157

##### [学会発表](計8件)

Hidekazu Minami, Hiromichi Kawanami, Masayuki Kanbara, Norihiro Hagita, “Chat Robot Coupling with Machine Response and Social Media Comments for Continuous Conversation,” IEEE ICME International Workshop on Multimedia Service and Technologies for E-health (MUST-EH), 2016年07月11日, シアト

ル (アメリカ)

南秀和, 川波弘道, 神原誠之, 萩田紀博,  
“ 継続的対話を目的とした機械対話と人間対話を併用したインタラクションロボット,” 日本ロボット学会学術講演会,  
2014年09月06日, 九州産業大学(福岡県・福岡市)

吉田雄太, 久保慶伍, 川波弘道, 猿渡洋,  
鹿野清宏, “ 音声情報案内システム「たけまるくん」の応答誤り分析とその改善,”  
日本音響学会春季講演会, 2013年03月15日, 東京工科大学(東京都・八王子市)

Kazuma Nishimura, Hiromichi Kawanami,  
Hiroshi Saruwatari, Kiyohiro Shikano,  
“ Response Generation based on Statistical Machine Translation for Speech-oriented Guidance System,”  
APSIPA Annual Summit and Conference 2012, 2012年12月04日, ハリウッド(アメリカ)

〔図書〕(計1件)

Joseph Mariani, Sophie Rossetm Martine Garnier-Rizet, Laurence Devillers (Eds.), Hiromichi Kawanami, Hiroshi Saruwatari, Kiyohiro Shikano, et al., Springer, “ Natural Interaction with Robots, Knowbots and Smartphones -Putting Spoken Dialog Systems into Practice,” 2014, pp.79-85, 261-267, 389-397

〔産業財産権〕

なし

〔その他〕

なし

6. 研究組織

(1) 研究代表者

川波 弘道 (KAWANAMI, Hiromichi)  
奈良先端科学技術大学院大学・情報科学研究科・助教  
研究者番号: 80335489

(2) 研究分担者

鹿野 清宏 (SHIKANO, Kiyohiro)  
奈良先端科学技術大学院大学・情報科学研究科・教授(平成25年4月削除)  
研究者番号: 00263426

猿渡 洋 (SARUWATARI, Hiroshi)  
東京大学・情報理工学系研究科・教授  
研究者番号: 30324974