

認知カトレーニングを目的とした事例ベース雑談音声対話システムの研究開発 (162309001)

Development of example-based spoken chat system for cognitive training

研究代表者

北岡教英 徳島大学

Norihide Kitaoka Tokushima University

研究分担者

泓田正雄[†] 橋本昌伸^{††} 太田健吾^{†††}

Masao Fuketa[†] Masanobu Hashimoto^{††} Kengo Ohta^{†††}

[†]徳島大学 ^{††}株式会社ウィッツ ^{†††}阿南工業高等専門学校

[†]Tokushima University ^{††}Witz Co., Ltd. ^{†††}National Institute of Technology, Anan College

研究期間 平成 28 年度～平成 30 年度

概要

音声対話は認知カトレーニングに有効である。高齢者が雑談などを通じてロボットなどとコミュニケーションをとる機会を増やすため、雑談対話を行える対話システムの構築を目指した。そのために、一般の成人とは異なる高齢者の音声を高精度に認識できる DNN・HMM 音声認識システムの構築、雑談を頑健に継続して行えるための事例ベース対話システムを開発し事例を自動収集法、どのような応答を返せばよいか臨機応変に考えて応答し対話を継続させる方法、を主に研究し、これらを統合した対話システムを構築した。本稿ではこれらのうち最初の二つについて述べる。

1. まえがき

少子高齢化社会において、高齢者が介護者などの手を煩わせることなく認知力のトレーニングを行えるための雑談音声対話システムの構築を目標とする。これを実現するために行った研究のうち、一般成人とは異なる音響特徴を持つ高齢者の音声の認識性能の向上、特定の話題に対して様々な雑談応答を返すための話題に適応した応答生成手法、について述べる。

2. 研究開発内容及び成果

2.1. 高齢者音声認識の高精度化

現在最新の音声認識システム、特に音響的な特徴を捉える部分である音響モデルの構築のためには、それらを機械学習で獲得するための音声データベースが不可欠である。我々は研究期間を通じて、名古屋・徳島・長崎・山形の各地域において、高齢者施設の協力を得て、高齢者の音声を収集した。読み上げる音声の内容は、ATR503 文と呼ばれる音素のバランスがとられた文で、これを約 50 文ずつの 10 セットに分けて、一人当たり 1 セットを読み上げてもらって音素バランスの取れたデータベースとした。結果として、平均年齢 79.2 歳の 201 名による発話をデータベース化できた。

収録した音声データを用いて Deep Neural Network と Hidden Markov Model を用いた音響モデル (DNN・HMM モデル) の学習を行った。音声認識のテストデータとするために、学習用の文の読み上げを行った後にテスト用の文を読み上げ、さらに被験者と実験者との雑談を実施し、それを収録して音声区間を切り出して対話音声のテストデータセットも作成した。研究の知見から、対話音声用の音響モデル構築には、対話音声に近い自由発話の音声データベースである CSJ データベースを用いて高齢者音声を用いて DNN の転移学習を行った。転移学習のベースとして、読み上げ音声である JNAS (成人音声)、S・JNAS (比較的高齢の音声) を用いた場合とも比較した。音声認識評価結果を表 1 に示す。

表1 高齢者の読み上げ音声および対話音声の認識結果 (単語誤り率)

学習データ		JNAS	S・JNAS	CSJ
テストデータ	読上	21.35	20.71	16.81
	対話	69.04	70.45	51.69

読み上げ音声に対しては、単語誤り率で 17% を切り、認識率に換算して 80% を大きく上回ることができた。対話音声についてはまだ検討が必要であるが、予備的な実地試験を見る限り、ロボットに対する高齢者の話し方は、読み上げ音声と対話音声の中間で読み上げに近い丁寧な話し方であることから、70~75% 程度の認識率が得られるものと考えられる。

2.2. 事例ベース対話システムの事例ベース自動収集

ユーザ発話を入力として、対話事例の中から入力に近い発話例を見つけ、それに対応する応答を返すのが事例ベース音声対話システムである。この事例がユーザに合っていないければ、適切な応答が返せない。そこで、事前に適応しておくことで特定ユーザに頑健に応答できるシステムを構築・提案してきた。その概要を図 1 に示す。

事例ベース雑談音声対話システムの事例データベースとして、ツイッターの対話的ツイートを用い、収集した 11 万事例からランダムに 3000 対を抽出し、それらから適切

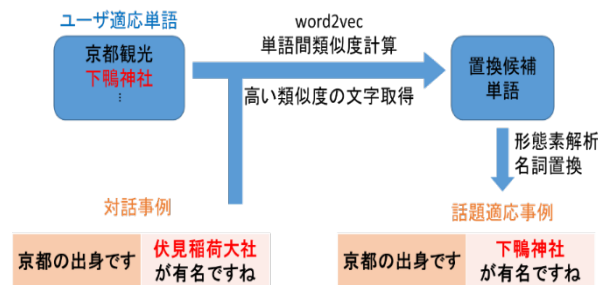


図 1: 単語類似度を利用した対話事例の話題適応手法
対話事例を人手で判断して 162 対話対を選んだ。これ

らの中に含まれる単語と置換する単語は、ユーザのプロファイル情報を Web 検索にかけ、検索された単語の中から、TF-IDF (Term Frequency-Inverse Document Frequency) 値が大きい単語を選んだ。これらの単語と、事例中の単語で、単語の分散表現 (Word2vec) のコサイン類似度基準で類似する単語を置換することによる対話事例のユーザ適応法を適用した結果、406 事例を生成でき、そのうち 196 事例が適切な生成事例であると判断された。

また、対話事例に複数の固有名詞が含まれ、それらを同時に置換しないと適切な事例とならない場合に対処するため、単語の分散表現の演算を用いた複数単語同時置換による適応法を提案した。これは、例えば、

東京-日本+フランス⇨パリ

となる単語の分散表現の特徴を用いて、図 2 のように、複数単語 (この場合、「高知」と「四万十川」) を同時に置換 (この場合、「大阪」と「淀川」へ) する。この方法により、55 の不適切例が生成されたものの、さらに 30 の成功例を得ることができた。



図 2 複数単語の同時置換

2.3. 雑談音声対話システムの応答種別選択

非タスク指向型の雑談対話システムにおいて対話をより長く継続するために、ユーザの発話に対する適切な応答種別を自動的に判定する手法を提案した。

雑談対話データについて説明する。提案手法の学習と評価のために、高齢者施設の協力のもとで、高齢者とインタビューアの 1 対 1 の雑談対話データを収集した。この対話では、雑音の少ない環境下において、インタビューアの 10 個の質問に対し、高齢者が語り手となって自由に雑談する。7 名の高齢者を語り手とする 3,062 発話を収集して分析した結果、インタビューアの発話は 9 種類 (相槌 (はい、うん)、明るめ相槌 (そうなんですか、なるほど)、暗め相槌 (そうなんですか)、話題を広げる、盛り上げる、話題交替、笑い、共感、何もしない (ユーザの発話待ち)) に分類されることがわかった。

評価実験を行った。高齢者の発話をこれらの 9 種のいずれかに分類し、ラベルの付与を行った。さらに、機械学習機への入力とするために、ユーザ (高齢者) 発話の単語を Word2vec により 200 次元の分散表現にした。これらについて、

- (1) 分散表現の並びを LSTM-RNN で直接 9 種に識別
- (2) LSTM-RNN への入力の結果最終的に得られるベクトルを基に SVM で識別
- (3) Word2vec の単語ベクトルを平均して SVM で識別
- (4) LSTM-RNN 識別器の事後活率を用いて SVM で識別を実施した。さらにこれらにオープンソースソフトウェア openSMILE (open-Source Media Interpretation by Large feature-space Extraction) を用いて、The INTERSPEECH 2010 Paralinguistic Challenge で使用された音響特徴量と統計量を併せた場合も実験した (表 2)。Word2vec と音響特徴量を併せた場合に最も良い結果を得ており、多くの場合に適切な王尾等を選べるようになった。平成 30 年度にはユーザ発話の単語の並びをより

高精度に反映する手法を検討して効果を得ているので、それも統合して対話システムに導入する。

表 2 応答種別選択手法の評価結果

特徴量	識別機	適合率	再現率	F値
Word2vec による分散表現	LSTM	0.401	0.410	0.406
LSTM エンコードベクトル	SVM	0.284	0.358	0.317
Word2vec 分散表現		0.379	0.481	0.424
LSTM 事後確率		0.191	0.414	0.261
LSTM エンコードベクトル+音響		0.452	0.477	0.464
Word2vec 分散表現+音響		0.533	0.546	0.539
LSTM 事後確率+音響特徴量		0.469	0.512	0.490

3. 今後の研究開発成果の展開及び波及効果創出への取組

これらの成果は小型ロボットの音声対話システムとして実装も行った。今後は、このシステムを早急に実証実験する。既に福祉関連企業から問い合わせがあり、協力が始まった。企業とも協力の上で、実際の高齢者施設に持ち込んで対話実験を行う予定である。

さらに、高齢者を対象とした音声対話、特に雑談による癒しを得るシステムの要望は他からも寄せられており、需要が高いことも分かってきた。今後は企業間での利害関係に注意しつつ、それぞれの企業への技術の移転を促進していく。そのために、我々の一部が参画して技術移法人も設立した。この法人によるコンサルタント業務を通じた技術の展開を行うとともに、別途研究費を獲得することにより、基礎研究も進展させる。実際に、文部科学省科学研究費補助金基盤研究を獲得し、今年度から実施することとなった。

4. むすび

高齢者音声認識の高精度化、事例ベース音声対話システムの高度化などの各テーマに対して、成果を得た。今後は研究成果そのものの実証実験および技術移転と、技術研究開発を共に実施する体制を整え、音声対話に基づく人間-機械協奏時代に本委託研究の存在感を放ち、分野全体での研究開発促進に資することを目指す。

【誌上发表リスト】

- [1] Ryota Nishimura, Daisuke Yamamoto, Takahiro Uchiya, Ichi Takumi, "Web-based environment for user generation of spoken dialog for virtual assistants," EURASIP Journal Audio, Speech, Music Process., pp. 1-13 (2018 年 11 月)
- [2] 西村良太、長尾拓海、一万田郁仁、北岡教英, "高齢者の音声知覚特性に基づいた音声の明瞭化加工法の研究," 日本知能情報ファジィ学会誌, Vol. 30, No. 6, pp. 840-845 (2018 年 12 月)
- [3] Norihide Kitaoka, Shuhei Segawa, Ryota Nishimura, Kazuya Takeda, "Recognizing emotions from speech using a physical model," Acoustical Science and Technology, Vol. 39, Issue 2, pp. 167-170 (2018 年 2 月)