Vol.2019-NL-240 No.11
2019/6/14

高齢者発話予測システムの検討

魏 琪[1,a)]　若山 龍太[1,b)]

概要：高齢者介護の現場において、人材不足、外国人労働者参入などに伴い高齢者とのコミュニケーションの重要性が高まっている。近年のディープラーニング技術の進化により音声認識システムの音声認識性能は大幅な向上を見せているものの、高齢者音声に対する音声認識精度については、青年層や壮年層に対する音声認識精度と比較するとより低いという現実がある。本研究は、高齢者音声認識精度を高めることを目標とし、音声認識結果から発話予測を行うシステムを開発する。Word2Vec モデルを使い、音声認識の結果、平仮名などの組み合わせから、音声認識結果に基づく発話予測を行う。

A consideration on the conversation prediction system for senior citizen

1. はじめに

高齢化と核家族化が進む中、「高齢者介護」は社会全体の課題になっている。2025 年ごろまでに日本の人口の 1/3 は 75 歳以上に達し、日本は超高齢化社会に突入と言われている。介護現場の介護職員不足に対し、外国人労働者の受け入れが進む一方、コミュニケーションの壁が大きな課題となっており、高齢者とのスムーズなコミュニケーションは介護において重要な要素となっている。

近年、音声認識技術の進化、高齢者向けの音声認識システム開発のプロジェクトが進められている [1] [2] [3]。しかしながら、高齢者音声に対する音声認識精度は 20 代〜50 代の音声と比較して認識率が低下する傾向があると言われている。高齢者は加齢に伴い調音器官の筋肉が衰え、発音が変化し、音声が不明瞭になると言われている [4]。また、人口の 1 割近くに、疾患に伴う発音障害が発生すると報告されている [5]。

このような背景から、高齢者に向けた音声認識システムの認識精度向上のため、高齢者の音声データを収録して、高齢者音声コーパスを作成することが試みられた [6] [7] [8]。しかしながら、高齢者音声の収集、特に発音障害、発音不明瞭な自然発話の収集が課題となっている。我々の研究では、音声認識システムが出力する音声認識結果から正しい意味の言葉を予測するシステムを開発した。

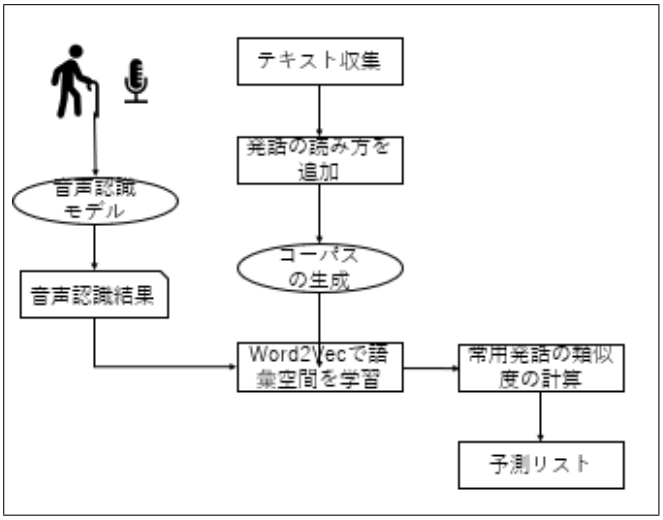

図 1　発話予測システム

2. 発話予測モデル

本研究のシステム構成を**図 1** に示す。収集したデータに平仮名を変換したデータを組み合わせ、Word2Vec モデル作成する。音声認識システムが出力する音声認識結果の平均ベクトルを計算し、事前に準備した常用会話リストから会話の類似度を計算、発話予測を行う。なお音声認識システムは当社が開発した VCRM システムを利用する。

[1] Hmcomm 株式会社
〒 105-0012 東京都港区芝大門 2-11-1　富士ビル 2F
[a)] qi.wei@hmcom.co.jp
[b)] ryota.wakayama@hmcom.co.jp

表 1　発話バリエーションルール

No.	発話バリエーション作成ルール
1	単語の漢字、ひらがな、カタカナの表現は統一する。
2	バリエーション文の最後は「。」で終端する。
3	文の途中には「、」は入れない。
4	フィラー（「えー」、「えっと」等の非言語）は使用しない。
5	長音記号は全角「ー」で統一し、「〜」は使用しない。
6	「？」「！」「/」等の記号や、絵文字は使用しない。

表 2　発話フレーズの書き起こしサンプル
発話フレーズ: あ〜いたよ（体が疲れた時）

書き起こし発話
もう疲れた。
あーしんどいねえ。
身体が悲鳴を上げとる。
もう身体がいっぱいです。
疲れたよ。
身体が限界です。
かったるくてねえ。
もうきついねえ。
もう動けないよ。
骨が折れるねえ。

2.1　VCRM システム

本研究の音声認識は当社の VCRM システムを利用する。VCRM はハイブリッド型 DNN-HMM モデルを使用し、基本となる音響モデル構築のため国立国語研究所で構築された日本語話し言葉コーパス (CSJ) を用いた。また、環境雑音、残響の影響を低減させるため CSJ に白色雑音、人工残響を重畳したデータを用いたマルチコンディション学習による音響モデルを構築し、これをベースラインとして用いた。また、ベースラインとなるモデルに、介護施設から収集した音声データおよび書き起こしテキストを用いて追加学習を行った。

2.2　データ収集

本研究では、介護設施の協力を得て高齢者の常用会話 75 件をリストアップし、それぞれの発話フレーズを**表 1** のルール沿って同じ意味の発話バリエーションを作成した。**表 2** は発話フレーズから発話バリエーション作成のサンプルを示す。この操作で、合わせて 539 件の発話バリエーションを作成した。

2.3　Word2Vec モデル

Word2Vec [9] は Tomas Mikolv らが考案したニューラルネットワークを用いた skip-gram モデルと呼ばれる言語モデルにより単語の分散表現を計算する手法の一つで、入力された単語の前後の単語を予測するようにニューラルネットワークを学習する。学習後、文章中の単語を任意次元のベクトルに変換し、意味的に似ている単語は空間上の近い位置に配置され、単語同士の演算や単語の類似度の導入が可能となる。

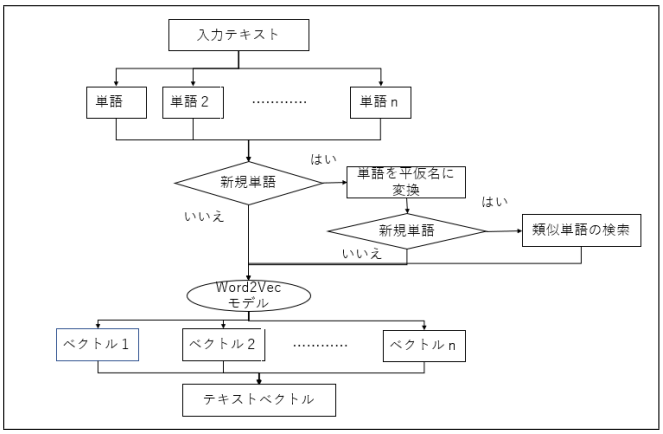

図 2　入力テキストベクトルの計算

本研究は 2.2 の方法により収集したデータセットを MeCab [10] を使用して単語分割を行う。そして、読み方を加えて、意味的に類似している発話同士を組み合わせ、Word2Vec 訓練用データセットを作成する。作成したデータセットは Word2Vec モデルとして学習する。

2.4　文章類似度

本研究は、高齢者が明瞭に発音できない言葉を予測することを目指す。2.3 のように単語の分散表現をもとに発話を表現し、発話の類似度を計算する。

発話は単語のベクトルに基づき発話ベクトルを計算する。二つの発話の類似度は、各発話のベクトルからコサイン類似度を計算する。ただし、高齢者の音声認識結果の中には、誤認識された言葉が多い場合や、2.2 で作成するデータセットの中に含まれない単語もある。単語のベクトルの計算手順を**図 2** に示す。単語が Word2Vec 中に含まれない場合は平仮名に変換し、平仮名のベクトルを計算する。もしも変換した平仮名も Word2Vec モデル中に含まれない場合は、単語と平仮名をトレニーグテキストの単語に一番似てる単語に変換してベクトルを計算する。

以上の手法を用いて、音声認識結果の平均ベクトルを計算する。そして、事前準備した常用発話リストの発話の類似度を計算し、上位 3 位の発話を列挙する。二つの発話 p, q のベクトル表現を v_p,v_q とすると、発話類似度は下の式のように表される。

$$g_{sim}(p,q) = \sin(v_p, v_q) \tag{1}$$

ここで、$\sin(v_p, v_q)$ は二つベクトルの類似度であり、本研究ではコサイン類似度を用いている。g_{sim} の値が大きいほど、二つ発話の類似性が高いことを表す。

3.　システムの評価

3.1　言い換え音声の発話予測

まず、発話の言い換え内容の予測性能を評価する。2.2

表 3 評価結果サンプル

（1）テストケース1
発話：「お腹が空いた」
誤り発話：「おがかが空いた。」

ベースライン	発話予測システム
×	○

（2）テストケース2
発話：「お腹が空いた」
誤り発話：「おなっがすいた。」

ベースライン	発話予測システム
×	○

（3）テストケース3
発話：「お茶ください。」
誤り発話：「おじゃください。」

ベースライン	発話予測システム
×	○

（4）テストケース4
発話：「飯が美味しい。」
誤り発話：「めひがおしひい。」

ベースライン	発話予測システム
×	×

のデータセットから、書き起こしテキストからランダムに50件をテストデータとして選出し、そのテストデータを発話内容予測モデルに入力したとき、回答候補上位3件中に正解発話フレーズが含まれていたら真、そうでなければ偽として算出したものを、ここでは正答提示率と定義している。

評価・検証の結果、追加のチューニングを実施することにより100%の確率で発話内容を予測できることが示された。

3.2 誤認識音声の発話予測

誤認識音声の発話予測評価のため、ベースラインシステムを作成する。ベースラインシステムは2.2のデータセットからWord2Vecモデルを作成したもので、新規の単語の平仮名と検索機能を含まないシステムである。

表 3 に構音障害の誤り傾向のあるテキストを入力したときの予測結果を示す。入力された単語が、学習したWord2Vecモデルの未知語となる場合は、当該単語の平仮名などを組み合わせて学習済みの類似単語を検索するよう発話予測システムを構成することで、一定の予測精度向上が見られた。一方、誤認識が多いの場合は、予測はまだ難しいことがわかる。

4. まとめと今後の展望

本研究では、高齢者音声に対する音声認識精度を向上させるために、音声認識結果を自然言語処理手法を用いて予測するシステムの開発・評価を行った。音声認識自体の精度高めるためには高齢者音声データベースの構築が一般的な方法であるが、高齢者音声データベースの構築は困難を伴う場合が多い。一方で、音声認識の結果を言語的に学習し、音声認識結果の精度を高めることも一つの方向である。本研究では学習していない言葉の検索はルールベースに基づくものであったため、今後、ディープラーニング等の手法を用いて発話予測システム構築に取り組む予定である。

謝辞 本研究の高齢者常用発話音声の収集にご協力頂いた医療法人玉昌会に感謝を申し上げる。また、本研究は革新的研究開発推進プログラム ImPACT の助成を受けた。

参考文献

[1] 朗 馬場, 伸一 芳澤, 実一 山田, 晃伸 李, and 清宏 鹿野. 高齢者音響モデルによる大語彙連続音声認識. **電子情報通信学会論文誌**. *D-II*, **情報・システム**, *II-パターン処理* = *The transactions of the Institute of Electronics, Information and Communication Engineers. D-II*, 85(3):390–397, mar 2002.

[2] 隆宏 井手, 光徳 水町, and 良久 中藤. D-14-11 高齢者の音響的特徴と音声認識性能との関係性の検討 (d-14. 音声, 一般セッション). **電子情報通信学会総合大会講演論文集**, 2012(1):195, mar 2012.

[3] 信夫 畑岡, 健哉 伊藤, and 圭一郎 大津. 明示的な音声無音区間の削除による高齢者音声認識–長時間無音の存在が音認識率を悪くする. **東北工業大学紀要** *1* **理工学編**, (31):29–34, mar 2011.

[4] 大輔 原田, 光徳 水町, and 勝行 二矢田. 高齢者のめりはりのない声に関する音響的解析. **電子情報通信学会技術研究報告**. *EA*, **応用音響**, 110(285):13–18, nov 2010.

[5] 誠 苅安, 稔 外山, and 登志正 松平. コミュニケーション障害の疫学：音声言語・聴覚障害の有病率と障害児者数の推定. **京都学園大学健康医療学部紀要**, (1):1–12, mar 2016.

[6] 百合絵 入部 and 教英 北岡. 音声認識にむけた超高齢者音声のコーパス構築. **日本音響学会誌**, 73(5):303–310, 2017.

[7] Japanese newspaper article sentences read speech corpus of the aged (s-jnas).

[8] S. Anderson, N. Liberman, E. Bernstein, S. Foster, E. Cate, B. Levin, and R. Hudson. Recognition of elderly speech and voice-driven document retrieval. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 1, pages 145–148 vol.1, March 1999.

[9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 3111–3119, USA, 2013. Curran Associates Inc.

[10] T. KUDO. Mecab : Yet another part-of-speech and morphological analyzer. *http://mecab.sourceforge.net/*, 2005.