

一次元 CNN によるモーラ長を用いた非日本語母語話者の音韻的特徴の分析

Analysis of Phonological Features of Non-Native Speakers of Japanese Using Mora Length with One-Dimensional CNN

一色雄太 黄宏軒

Yuta Isshiki Hung-Hsuan Huang

福知山公立大学情報学部

Faculty of Informatics, The University of Fukuchiyama

Abstract: 現在の日本は、人口減少渦にあり、それにより労働人口の減少も起こっている。その中でソフトウェアエージェントによる日本語学習が可能になれば多くのより外国人が日本国内で働くことが可能になると考えた。その中で、エージェントによる日本語能力の分類のために、本論文では日本語母語話者の話す日本語と非日本語母語話者の話す日本語にどのような差が存在するのかを分析する。

1 はじめに

日本の人口は、2008年の1億2808万人をピークに徐々に減少し、生産年齢人口は2018年の7,545万人から2065年には4,529万人まで減少すると内閣府は推計している [1]。女性の社会参画や定年の引き上げにより、労働人口は漸減であるが、有効求人倍率を見ると平均で1を超えている。こうした労働者不足の現状は、経済の成長を妨げることになり、改善すべき事項である。また、厚生労働省によれば、現在、約170万以上の外国人労働者が労働人口の不足分の一部を補っている [2]。

しかし、現状の外国人労働者には複数の問題点がある。その一つに、技能実習生を中心とする外国人労働者の日本語能力が低いことが挙げられる。実際に、国際協力人材機構が実施したアンケートでは、実習実施者の37.3%、監理団体の35.3%が特定技能制度と異なり、実習生の日本語能力が低いと回答している。また同アンケートで、65.7%の実習実施者や57.4%の監理団体が技能実習生のメリットが労働力人口の不足を補うことに役立っていると考えている [3]。これらから、これからの労働力問題を解決することの一つに外国人労働者の日本語力を向上させることがあるという考えに至った。

また日本語力を向上させる方法として、本論文では音声に注目をする。音声に注目する理由としては、会話が必要なサービス業、特に少子高齢化により介護の求人が増えると予測されるからである [4]。

このとき、本研究では、日本語能力を向上させる手段としてソフトウェアエージェントを用いることを目

標としている。このソフトウェアエージェントの概要としては、日本語学習者との会話において、会話音声から学習者の日本語能力を推定し、学習者に応じた学習方法を提案するというものとしている。ソフトウェアエージェントを用いる理由としては、日本語教育に対して人員を割く必要がなくなるほかにも、教育現場において人間ではないロボットであっても教育効果があると大久保ら [5] が期待しているからである。

本論文におけるリサーチクエスションは以下の2点である。1点目は、各モーラは前後のモーラとどのような関係を持っているのかである。2点目は、非日本語母語話者の話す日本語はどのような特徴を持っているのかである。この2点について分析を行う。

本研究は非日本語母語話者の日本語学習者の支援を目的としたソフトウェアエージェント作成を目標としている。その際、ソフトウェアによる評価を適切に行うために、音声の長さ情報を利用して分析を行っている。そのために、共同作業を行っており、発言の内容がある程度統制されており、日本語母語話者と日本語非母語話者の差異を調査可能なグループディスカッションのデータコーパスを収集した。本論文では、そのデータコーパスを用いて非日本語母語話者の発音に違和感を覚える要因の一つに音素長があると仮定して、対象の位置に関係なくパターン認識を行うことが可能で、直接特徴を抽出することが可能である一次元の畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) [6] を用いた分析結果を報告する。



図 1: 会話実験の様子

2 データコーパス

本研究は、非日本語話者が最終的に日本国内での就労を目標としている。そのためにも、データコーパスは実際に働く環境に即した環境としている。具体的な内容は、日本人多数の環境で、実際に仕事で行われるような一つの話題テーマに関するグループディスカッションである。

本論文で用いるコーパスは、図 2 のような日本語を母語とする 3 人と日本語を勉強している中国語を母語とする 1 人の 4 人による日本語での 15 分間のグループディスカッションとしている。グループディスカッションでは、発言の内容がある程度統制されるために共同作業を行う内容であるため、日本語母語話者と日本語非母語話者の差異は調査可能である。中国語を母語としている参加者は日本に 1 年以上滞在しており、日本語能力試験 N2 レベルに合格することを条件としている。その 7 人の話者を分析し、日本語を母語とする参加者と比較する。また、性別による結果の偏りを防ぐため、各グループは男女同数の参加者で構成されている。

今回、中国語母語話者を選んだ理由は、2022 年の時点で、在留資格を有しているうちの 21.2 % が中国語を母語としていることや、サービス業に就いている外国人労働者の 27 %、小売業の 33 % を占めているからである [7]。

今回用いたデータコーパスでは、日本語母語話者 3 人、中国語母語話者 1 人が 15 分間ディスカッションを行っているデータが 7 グループで各グループ 2 セッションが行われており、合計で日本語母語話者が 630 分間、中国語母語話者が 210 分間のデータとなり、日本語母語話者が 121,901 モーラ、中国語母語話者が 12,561 モーラのデータが得られた。なお、cnn を行うタイミングで不均衡データとなりモデルの偏りを防ぐために SMOTE を使用した。

3 分析手法

本論文では、収集した中国語母語話者と日本語母語話者の音声をモーラ単位で分割して分析をする。モーラとは、韻律学における単位で、単音節 1 つ分に相当する音の長さのことである。本論文では、収集した会話音声に対して、Julius¹の音素セグメンテーションキットを用いて分割おり、a, a:, b, by, ch, d, e, e:, f, g, gy, h, hy, i, i:, j, k, ky, m, my, N, n, ny, o, o:, p, q, r, ry, s, sh, t, ts, u, u:, w, y, z の 38 種類に分けられている。N は撥音、q は促音を表している。

本研究におけるリサーチクエスチョンは、以下の 2 点である：

1 点目は、各モーラは前後のモーラとどのような関係を持っているのかである。我々はこれまで行ってきた分析の中で日本語母語話者と非日本語母語話者の間のモーラの長さのばらつき具合に差が存在していることが分かっている [8]。その中で、音声を分類する段階で、前後のモーラにどれだけの影響を受けているのか/前後のモーラへどれだけの影響を与えているのかを分かっておくことでより効率的なソフトウェアエージェントによる分類が可能になるのではないかと考えた。本論文では、どれだけの影響を与えている、与えられているのかを見るために、機械学習モデルのパラメータ値の調整を分手法とした。本論文でチューニングを行ったパラメータは、フィルターサイズとフィルター数である。パラメータ調整時には、2 節で示したデータコーパスのうち中国語母語話者をそれぞれ一人ずつテストデータとする leave-one-person-out 交差検証を用いて、各テストデータの再現率の算術平均をこのモデルの再現率として、再現率が最も高いモデルを適切なパラメータであるとしている。leave-one-person-out 交差検証を用いる理由としてはモデルの汎化性能を重視するためである。

2 点目は、非日本語母語話者の話す日本語はどのような特徴があるのかである。非日本語母語話者の話す日本語の特徴を得ることで、実際にソフトウェアエージェントを用いて分類を行うときに必要な箇所のみで分類のための演算を行うことが予想され、効率的な分類が期待できると考えたからである。この分析では、先述したチューニングされたパラメータを用いた機械学習モデルを用いる。ここでは、パラメータがチューニングされたモデルに 2 節で示したデータコーパスを全て用いて学習を行い、畳み込み層のフィルタの値を用いて分析を行う。

¹<https://julius.osdn.jp/index.php>

4 結果の考察

調整したパラメータは、フィルタサイズ (3, 5, 7) とフィルタ数 (16, 24, 32, 40, 48), 一つ目の結合層の出力の次元 (16, 32, 40, 48) で 60 通りで検証を行った. epoch 数は 30 で行った. その結果, フィルタサイズが 5, フィルタ数が 24, 結合層の次元が 16 のものが一番高い再現率 57.4 %を出しており, 2 要素での分類であるためチャンスレベルである 50 %よりも高い結果となった. ここから, モーラごとに分けて分類を行う場合は 5 つずつ分けることが効果的であると推測される.

また, 図 2 はフィルタサイズを 5, フィルタ数を 24, 結合層の次元を 16 に調整を行いデータコーパスを用いて学習を行った時の畳み込み層のフィルタである. この時図の明るいところが高い特徴を示している箇所, 暗いところがあまり特徴を示さない箇所である. この時各フィルタが中国語母語話者への判定に使われる重みを表 1 で示している. この重みより, F5, F6, F13, F15 が中国語話者の特徴を表していると推測され, また, 逆に F21, F22, F23 は中国語話者の特徴を表していないと推測することができる. ここから, それぞれのフィルタの重みを見ると, F13, F15 より, モーラを 5 つずつ分割したとき前半の 3 つに中国語母語話者らしい特徴があることが分かる. また, F21, F22, F23 より後半 2 つには中国語母語話者らしい特徴がないことも分かる. ここから, 中国語母語話者は話すときに今発音している前の発音から影響を受けており, 直後の発音の影響を受けづらいたことが推測される.

5 まとめ

本論文では, 非日本語母語話者のためのエージェント作成に向けて, 効率よく分類を行うために次元 cnn を用いて日本語母語話者と非日本語母語話者の違いの分析を行った. その結果, leave-one-person-out 交差検証より, モーラごとに分けて分類を行う場合は 5 つずつ分けることが効果的であると推測される. また, 5 つずつに分けた中でも, フィルタから中国語母語話者は, 日本語母語話者に比べて現在話しているモーラから直前の発音の影響を受け, 直後の発音の影響を受けづらいたことが分かった.

今後の研究では, 前半の 2 つでどのようなモーラがどのような発音をすることによって今発音しようとしている音声に影響を受けるのかを調べ, よりエージェントが効率よく日本語の能力の分類を行えるように特徴をまとめていく.

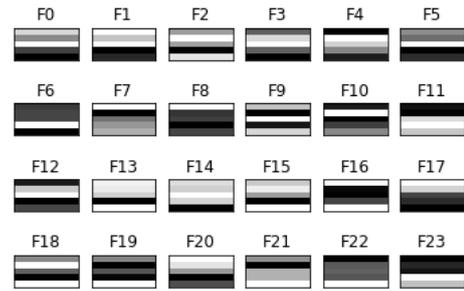


図 2: 畳み込み層から得られたフィルタ

表 1: 各フィルタの重み (filter は各フィルタの名称)

filter	重み	filter	重み	filter	重み
F0	2.269	F8	5.767	F16	-4.842
F1	-4.058	F9	-5.662	F17	1.730
F2	3.215	F10	-3.796	F18	1.060
F3	2.323	F11	-4.341	F19	-0.082
F4	-2.102	F12	-5.939	F20	-4.576
F5	7.410	F13	5.321	F21	-10.321
F6	4.697	F14	-0.137	F22	-9.079
F7	2.199	F15	5.610	F23	-12.687

参考文献

- [1] 内閣府: 少子化対策白書 (2022)
- [2] 厚生労働省: 「外国人雇用状況」の届出状況 (2022)
- [3] 国際人材協力機構: 技能実習・特定技能制度見直しに関するアンケート調査結果 (2022)
- [4] 厚生労働省: 2025 年に向けた介護人材にかかる需給推計 (確定値) について (2015)
- [5] 大久保友幸, 兼光有沙, 永井美雪, 田部井賢一: 育向けソーシャルロボットの現状, 東京都立産業大学院大学紀要, No. 14, pp. 91-96 (2020)
- [6] Y. LeCen, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel: Back-propagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, Vol.1, No.4, pp.541-551 (1989)
- [7] 厚生労働省: 「外国人雇用状況」の届出状況まとめ (令和 4 年 10 月末現在) (2023)
- [8] 一色雄太, 黄宏軒: 日本語学習者の音韻的特徴の分析ヒューマンコミュニケーション基礎 (HCS) 研究会 2023 年 8 月研究会, 電子情報通信学会 (2023)