

深層学習を用いた日本手話音節の適格性解析

高藤 朋史 三輪 誠 佐々木 裕 原 大介

豊田工業大学

{sd18421, makoto-miwa, yutaka.sasaki, daisuke}@toyota-ti.ac.jp

1 はじめに

日本手話は語彙が少なく、新たな概念や表現が難しい語彙に対応するため、新しい手話が毎年提案されている ([3] など)。しかし、そのような手話の中には、ろう者が使用しづらい、意味が通じづらいと感じる、日本手話として不適格な音節が含まれている。このため、通訳などの場面で情報伝達が難しい、日本手話初心者が誤った手話を学習する、といった問題が生じている。原因は、音声言語における音節に対応する手話の表現 (手話における音節) が日本手話の表現規則に反しているためだと考えられる。しかし、日本手話の表現規則は、現在明らかになっていない。そのため、音節の適格性の判断はろう者の直感に頼る必要があり、既に作られた全ての音節、今後新たに作成される音節の適格性を人手で判断するのはコストが大きい。したがって、音節が日本手話として適格か不適格かを自動的に判断するシステム、しいては、音節の不適格性を判断するための規則の発見が必要とされている。

八幡ら [2] は、音節を人手でコード化し、適格性・不適格性をタグ付けしたデータを対象に、音節の適格性を解析した。具体的には、コードを元にした2値特徴の組み合わせ特徴を用いて、ロジスティック回帰モデルの分類器を作成した。しかし、八幡らの研究では、特徴の組み合わせは2つにとどまっておらず、特徴間の関係を十分考慮できておらず、より高度に特徴の関係を考慮した音節の表現を用いた分類器の作成が必要である。

本研究では、より適切な音節の表現を行うことで、音節の適格性分類精度を向上することを目的とし、この適切な音節の表現のために深層学習を用いた音節の表現モデルを提案する。評価では、ロジスティック回帰を用いたモデルに比べて分類正解率・F値の向上を確認し、提案手法の有効性を確かめることができた。一方で、組み合わせ特徴を用いた従来研究に比べて、分類正解率が低くなるという結果となった。

2 関連研究

2.1 コーディングによる手話の記述

本研究では、新日本手話コーディングマニュアル [4] に従ってコード化された手話音節を使用する。音節はコード化により、音節タイプ・手型・位置・動き・接触・手のひらの向き・中手骨の向き、の7カテゴリーにおける全27種類によって記述される。

日本語-手話辞典 [5]、新しい手話シリーズ [3] に掲載されている音節を対象にコード化を行い、それぞれの音節は、複数名のろう者により適格な手話音節 (正例) か不適格な手話音節 (負例) かの分類情報が付与されている。

2.2 ロジスティック回帰による手話音節適格性解析

八幡ら [2] は、コード化された手話音節のデータを使用し、日本手話音節の適格性解析を行った。分類器にはロジスティック回帰モデルを採用している。データの扱いとして、コード化された手話音節から、さらに詳細な特徴を抽出している。詳細な特徴の抽出により、コード化された手話音節を、27種類の特徴から891種類に細分化した。そして、その細分化した特徴について、各特徴を組み合わせた組合せ特徴を抽出し、利用している。この結果、八幡らは87.0%の分類正解率を得ている。

3 提案手法

本研究では、深層学習を用いて手話音節を表現し、手話音節の適格性を分類する手法を提案する。コード化された手話音節のデータから、2次元的に扱うことが可能な手型特徴を畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) で特徴抽出、

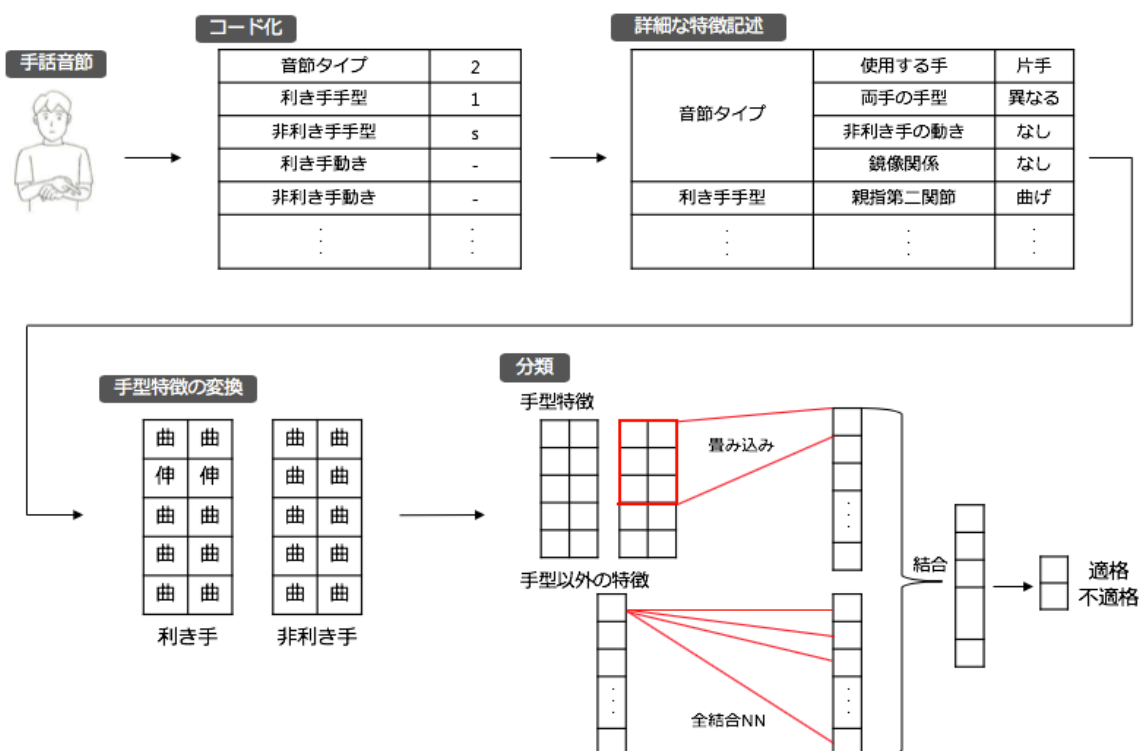


図 1: 提案手法の概要

手型以外の特徴を全結合型 NN で表現し、手話音節の高度な組み合わせ特徴を表現する。手法の概要図を図 1 に示す。手法の流れとして、コード化された手話音節から詳細な特徴抽出を行った後、手型の特徴についてを CNN 用いて、手型以外の特徴については全結合型 NN を用いて、それぞれの表現を得る。この 2 種類の特徴から計算された表現を結合したものを線形分類に通すことで分類を行う。本節では、この特徴抽出と深層学習モデルについて、以降説明を行う。

3.1 特徴抽出

データセットからの特徴の抽出は次のように行う。新日本手話コーディングマニュアルに沿って記述された、音節を表現する 27 要素のベクトルから、手型を表す 2 要素 (利き手手型, 非利き手手型) を抽出する。その後、手型を表現する 2 要素、手型以外を表現する 25 要素についてそれぞれ詳細な特徴を抽出する。手型は、各指の第二関節、付け根関節の曲がり具合を伸び、緩み、曲げの 3 値で定義し、 5×2 の行列で表現する。手型以外を表現する詳細な特徴は 670 次元の、0 もしくは 1 を要素にもつ数値ベクトルを用いる。これは、後述のデータセット上に現れる全 670 種の手型以

外の詳細な特徴について、ある詳細な特徴を音節が含む場合 1, 含まない場合 0 を記述した数値ベクトルである。

3.2 深層学習モデル

提案する CNN と全結合型 NN を用いた分類モデルの全体像を図 2 に示す。さらに CNN のモデルの詳細について図 3 に示す。

提案手法の CNN は指の各関節の曲がり具合を指 1~3 本ずつ考慮するモデルである。入力された情報をフィルタサイズ 1×2 , 2×2 , 3×2 で畳み込みを行い、活性化関数 ReLU を通してそれぞれの表現を得る。これはそれぞれ指 1 本, 2 本, 3 本の情報を同時に見ることに相当する。この畳み込み層で隣接した指ごとにまとめた情報を $16 \times 1 \times 1$ のマックスプーリングによりプーリングを行なう。畳み込み、プーリングを両手の手型に関して行う。最後に、各フィルタで得られた出力を結合し、12 次元の出力を得る。この際、片手手話であった場合、NN へ値を入力する際に非利き手の全ての要素を 0 とする。

手型以外の特徴については、各特徴にベクトルを付与し、ベクトルの各要素の平均を取ることで一つの手

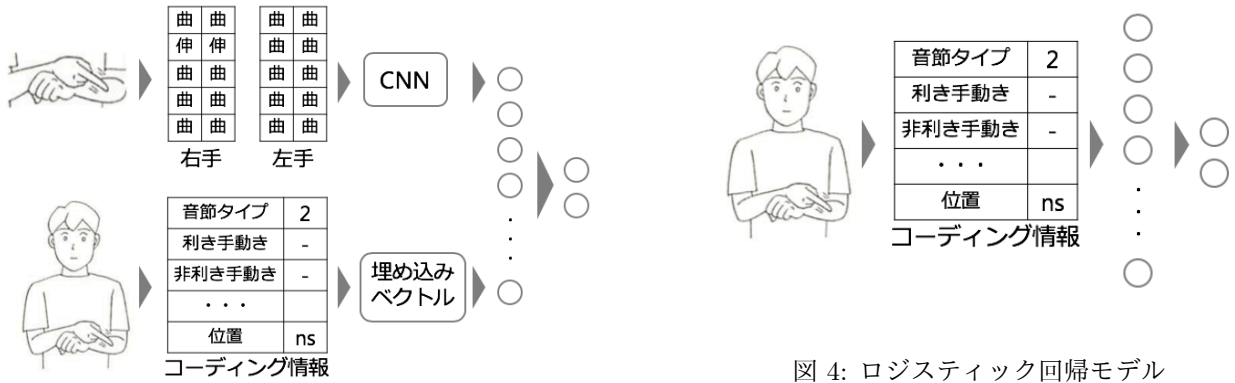


図 4: ロジスティック回帰モデル

図 2: CNN モデル

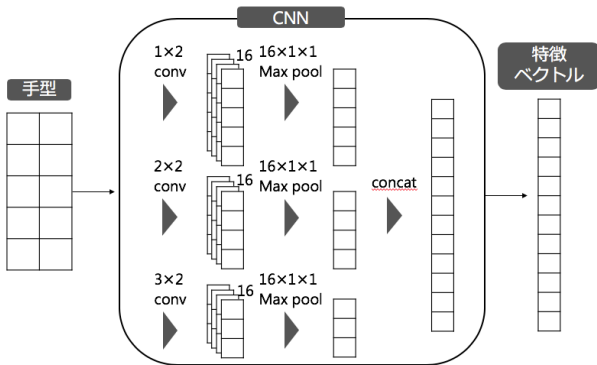


図 3: CNN 詳細

話音節を表現する。これにより、670次元の0もしくは1が要素の数値ベクトルから20次元の特徴ベクトルへ変換する。

最後に、左右の手型の特徴、手型以外の特徴を結合した44(=12+12+20)次元のベクトルを単層NNを通しsoftmax関数に入力し、適格・不適格を確率として出力する。それぞれ得られた出力を単層の全結合型のNNの入力とし、softmax関数を通すことで分類結果を得る。

4 実験

4.1 データセット

データセットとして、『わたしたちの手話 新しい手話シリーズ』[3]および『日本語-手話辞典』[5]に掲載されている音節を、新日本手話コーディングマニュアルに沿ってコーディングしたものを用いる。実験には、正例2,291データ、負例600データの全2,891データを用いた。このデータについて、[2]に従って、学習データ・開発データ・評価データへの分割を行なった。各データ内の正例・負例の数を表1に示す。図に

示したとおり、どの分割においても正例・負例の比率が一定となるような分割数となっている。

各音節への適格・不適格の付与は、25名のろう者が日本手話としての適格性を3段階で評価したものを用いている。[2]に従って、各音節について、ろう者の各評価にスコアを定義し、ある音節に対する25名のスコア合計が4を超えたものを不適格な音節として利用した。評価とスコアの対応を表2に示す。

4.2 実験設定

実験では図2に示した提案手法であるCNNを組み込んだ分類モデルの他に、比較対象として図4に示したロジスティック回帰を用いたモデルも評価した。このロジスティック回帰のモデルは全849種の特徴を含むかどうかを表すベクトルを入力とし分類を行うモデルである。

学習には、オンライン最適化アルゴリズムである

表 1: 評価データの統計

	正例	負例
訓練データ	2,053	538
開発データ	238	62
評価データ	238	62

表 2: タグ付けにおけるスコア

評価	スコア
適格だと思う	0
やや不適格だと思う	1
不適格だと思う	2

Adam [1] を用い、L2 正則化を用いたミニバッチ学習を行った。CNN のハイパーパラメータを表 3 に示す。

4.3 結果と考察

各分類器における評価データでの分類正解率、負例 F 値の結果を表 4 に示す。提案手法である CNN のモデルにおいて、分類正解率 0.856、負例 F 値 0.632 という結果が得られた。これはロジスティック回帰のモデルに対して、分類正解率、負例 F 値において上回る結果となった。この結果より、同等の入力データを扱う場合、手型の特徴という位置情報を持つ特徴を CNN で扱うことが有効であると言える。

一方で、八幡 [2] の手法と比較すると、組合せ特徴考慮した場合、分類正解率で下回り、負例 F 値で上回る結果となった。精度が下回った原因が 2 つ挙げられる。

一つは、線形分類器への入力の差である。本研究では、複雑な組み合わせ特徴を用いて適格性分類をすることを目的とした。そのため、線形分類器への入力の次元数は 44 次元と、八幡 [2] の 174,986 次元と比較して極めて小さい。埋め込みベクトルや CNN からの出力の次元数を増やしたモデルを検討する必要がある。

もう一つの原因は、手型とそれ以外の組み合わせを考慮していない点が挙げられる。全ての特徴間の組み

表 3: CNN モデルにおけるハイパーパラメータ

	値
フィルタサイズ	1×2, 2×2, 3×2
フィルタ数	16
ストライド	1
埋め込みベクトル次元数	20
学習率	0.005
ドロップアウト	なし
ミニバッチサイズ	128
L2 正則化の係数	0.0001

表 4: 実験結果

データ	分類正解率	負例 F 値
CNN	0.856	0.632
ロジスティック回帰	0.832	0.487
八幡 [2]	0.870	0.613

合わせを考慮した八幡 [2] と比べ、本研究では、手型間（指の曲がり具合間）の組み合わせと、手型以外の特徴間の組み合わせのみを考慮している。手型と手型以外の組み合わせを考慮することで、より適格性分類に適切な入力を形成できると考える。

5 おわりに

本研究では、手話音節特徴の高度な組み合わせを考慮した、NN モデル分類器による適格性分類を行う手法を提案した。CNN を用いた手法において、評価データの分類精度 0.856、負例における F 値 0.632 という結果が得られた。

今回、複雑な組み合わせ特徴の有効性を確かめるため、単純なモデルを作成したが、今後、より高次元な組み合わせ、手型と手型以外の組み合わせを考慮するモデルに改善していく。

謝辞

本研究は JSPS 科研費 JP18H00671 の助成を受けたものです。

参考文献

- [1] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *ICLR 2015*, 2015.
- [2] Satoshi Yawata, Makoto Miwa, Yutaka Sasaki, and Daisuke Hara. Analyzing well-formedness of syllables in japanese sign language. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 26–30. Asian Federation of Natural Language Processing, 2017.
- [3] 全日本ろうあ連盟, 日本手話研究所「日本手話確定普及研究部」. わたしたちの手話 新しい手話シリーズ. 全日本ろうあ連盟出版局.
- [4] 原大介. 新日本手話コーディングマニュアル. 2016 (12/11 更新).
- [5] 日本手話研究所. 日本語-手話辞典. 全日本ろうあ連盟出版局.