

畳み込みニューラルネットワークを用いた吃音症の分類

岡優希 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

吃音症には大きく分けて3種類の症状があるが、音声からそれらの症状を自動的に分類することができるのと吃音症の治療や吃音者に利用しやすい音声認識を用いた製品の開発などを行う際に有効であると考えられる。吃音の音声認識を行う手法としては、逆行DP(ダイナミックプログラミング)を用いる手法[1]が提案されている。この手法では、同じ音の繰り返しに着目し、吃音だと思われる部分を検出している。この手法では、DPを用いた手法[2]に比べて認識率が11%から58%まで向上することが確認されている。しかし、この手法では音の繰り返しが顕著に表れないような症状を検出することは難しい。また、もともと音の繰り返しがある「おおさか」「おおいた」などの単語を誤って吃音だと認識してしまうことがあることもわかっている。

それに対し、ディープラーニングの手法の1つである畳み込みニューラルネットワーク[3]は画像認識だけでなく音声認識の分野においても優れた成果をあげ、注目されている。畳み込みニューラルネットワークは多層の階層構造を持つニューラルネットワークであり、入力データに含まれる特徴を学習によって自動的に抽出できることが知られている。吃音者の話し方には音の繰り返し以外にも様々な特徴があると考えられるため、吃音の分類においても畳み込みニューラルネットワークを用いることでそのような特徴を学習によって抽出し、分類に活かすことができると考えられる。

本研究では、畳み込みニューラルネットワークを用いた吃音症の分類を提案する。

2 吃音症

吃音症は、発話障害の1つであり、話し言葉が滑らかに出てこないことをさす。吃音の代表的な症状は、連発、伸発、難発の3つに分けられる。

(1) 連発 (繰り返し)

「み、み、みかん」などのように同じ文字を連続して発声する状態

(2) 伸発 (引き伸ばし)

「みーかん」などのように語頭の音を引き伸ばす状態

(3) 難発 (ブロック)

「ッ... みかん」などのように発声できずに言葉が詰まってしまう状態

3 畳み込みニューラルネットワークを用いた吃音症の分類

提案システムでは、音声データをスペクトログラムに変換したものを入力として、畳み込みニューラルネットワークを用いて吃音症のどの症状にあたるかを分類する。

3.1 構造

提案システムで用いる畳み込みニューラルネットワークでは、0.5秒分の音声信号に対応するスペクトログラムを入力とし、入力された音声信号が連発、伸発、難発、症状なしのいずれに相当するかを出力するように学習を行う。畳み込みニューラルネットワークは図1に示すような構造のものを用いる。

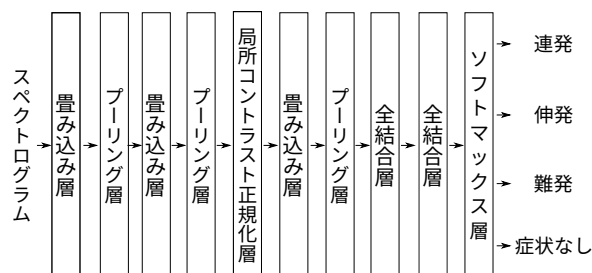


図1: 提案システムで用いる畳み込みニューラルネットワークの構造

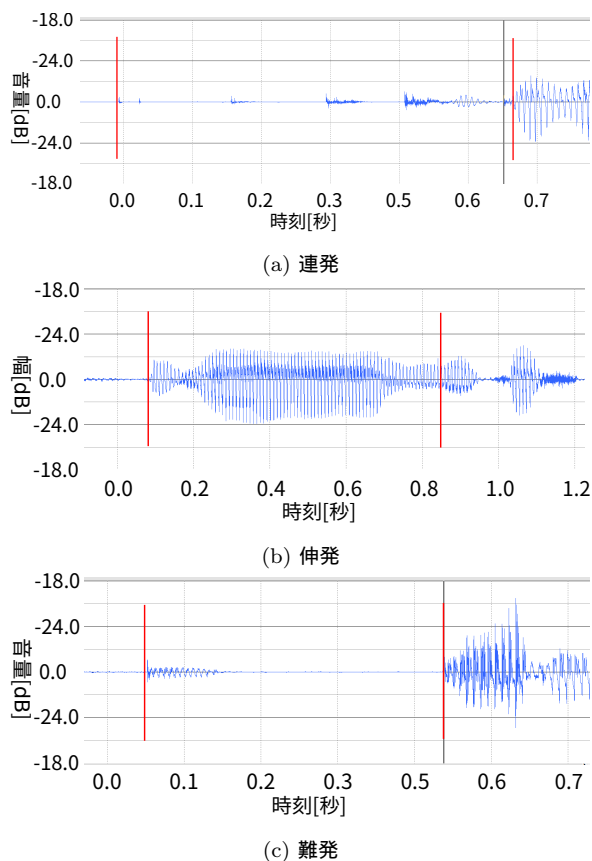


図 2: 各症状に対応する波形の例

3.2 学習データの作成

音声データに対して以下のような手順で学習データを生成する。

Step 1 : 症状に対応する範囲の特定

音声データを波形を見ながら聴き、症状に対応する範囲を手作業で特定していく。図 2 に実際の波形とそのデータに対して症状に対応する範囲を特定した例を示す。図 2 において、2 本の赤い線で囲まれた部分がそれぞれの症状に対応する部分である。

Step 2 : 領域に対するクラスラベルの付与

Step 1 で特定した症状に対応する範囲の領域に対して「連発」、「伸発」、「難発」の 3 種類のクラスラベルを付与する。いずれの症状に該当する領域とも特定されなかった領域に対しては「症状なし」のクラスラベルが付与されることになる。なお、「症状なし」のクラスラベルが付与される領域には音声自体が存在しない、無音の部分も含まれることになる。

Step 3 : n_1 秒単位のスpekトラムの生成と教師信号の決定

提案システムでは、開始位置を n_2 秒ずつずらした n_1 秒分の音声データに対応するスペクトログラムを入力とし、それに対して症状の分類を行う。音声データを開始位置を n_2 秒ずつずらした n_1 秒分のデータに分割し、それぞれのスペクトログラムを求める。また、それぞれのデータに対応する部分に Step 2 で付与したクラスラベルに基づいて教師信号を割り当てていく。データに対応する範囲が、Step 1 で分割した領域の 1 つにのみ対応している場合には、その領域に付与されたクラスラベルに対応するクラスが教師信号として割り当てられる。データに対応する範囲が、Step 1 で分割した複数の領域にのみ対応している場合には、 n_1 秒間の中で最も長く含まれている領域に付与されたクラスラベルに対応するクラスを教師信号として割り当てる。

4 症状の分類

学習を行った畳み込みニューラルネットワークにスペクトログラムを入力し、クラス分類を行う。学習終了後に実際にこのシステムを使う場合には、(1) 発話の中に 3 つの症状がどれくらい含まれていたのかと (2) 症状として分類された部分の中で各症状が現れた割合をユーザに示す。

5 計算機実験

吃音症の症状の含まれる 7 人 (男性 4 人、女性 3 人) 分の発話データを用いて、提案システムにおいて症状の分類を行う実験を行った。症状により分類精度に違いはあるものの、 F 値として 0.7 程度の精度で分類が行えることを確認した。

参考文献

- [1] 柳田益造: “吃音の音声認識,” 情報処理学会第 65 回全国大会, 2003.
- [2] 成川晃訓, 柳田益造: “吃音のモデルとその実験的検証”, 電子情報通信学会技術研究報告, SP, Vol.98, No.423, pp.1–8, 1998.
- [3] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, Vol.86, No.11, pp.2278–2324, 1998.