性格情報を用いた音声感情認識の精度向上

Improving the Accuracy of Voice Emotion Recognition Using Personality Information

石田 絹史朗[1)]
指導教員　岩下 志乃[1)]，欅 リベカ[1)]，研究協力者　大竹 正彦[2)]

1) 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻
岩下・欅研究室
2) 東京工科大学 コンピュータサイエンス学部 コンピュータサイエンス学科

本研究では話者によって音声感情認識の精度が低下する問題に対し、性格情報を用いた音声感情認識を提案する。入力音声をメルスペクトログラムと呼ばれる周波数画像に変換し、AlexNet を用いて入力画像と感情の関係を学習する。性格情報の推定には日本語感情表現辞書を用いる。

キーワード：音声感情認識，性格情報，深層学習

１． はじめに

音声感情認識は、音声から話者の感情状態を認識する技術である。人間同士が対話を行う際には、発話内容の理解に加えて話し相手の感情状態を考慮することで対話の円滑化を図っている。今後、人と機械が自然な音声コミュニケーションを行うためには、音声感情認識は必要不可欠な技術と考えられる[1]。

現在は音声感情認識の実用化が進んでいるが、感情の表現の仕方は話者によって異なるため、認識する者にとっては認識精度が低くなることがある。その要因としては性別や性格といった話者の特性が挙げられる。

本研究では、話者の特性のひとつである性格情報を用いた日本語の音声感情認識の精度向上を目的とする。

２．既存研究

Li et al.[2]は、音声、テキスト情報、話者の性格特性の 3 つを考慮した英語の音声感情認識を提案している。この研究では性格特性を推定するために LIWC(Linguistic Inquiry and Word Count)を利用している。これは、テキストに含まれる語彙から心理状態、言語スタイルといったカテゴリーの単語出現頻度を求め、統計的に分析することで、テキストの話者の性格特性を推定する手法である。しかし、LIWC は英語を対象としているため、日本語の感情認識にそのまま利用すると誤差が生じると考えられる。

３．提案手法

本研究では、日本語に特化した単語の分類ツールである日本語感情表現辞書 JIWC(Japanese Linguistic Inquiry and Word Count)[3]を用いた性格推定手法を取り入れた日本語の音声感情認識を提案する。JIWC は 7 つの感情（「悲しい」「不安」「怒り」「嫌悪感」「信頼感」「驚き」「楽しい」）に関する頻出表現を収載している辞書である。これにより、英語での先行研究と同様に、性格特性を考慮しつつ日本語に対応した音声感情認識が可能となる。

処理の流れを以下に示す。初めに入力された音声データをメルスペクトログラムという画像に変換する。メルスペクトログラムとは周波数軸をメル尺度にしたスペクトログラムである。メル尺度とは人間の耳の周波数に対する感覚に基づいた音

響尺度の事である。例として怒りを表すメルスペクトログラムを図 1 に示す。

次に画像認識の精度が高いディープラーニング手法のひとつである AlexNet[4]で入出力を学習させる。入力はメルスペクトログラム（音声の周波数画像）であり、出力は 7 つの感情のうち 1 つである。さらに、入力音声をテキストに変換して JIWC を用いた性格推定を行い、その結果を AlexNet に組み込む。テストデータの音声を入力し、感情認識の精度を評価する。

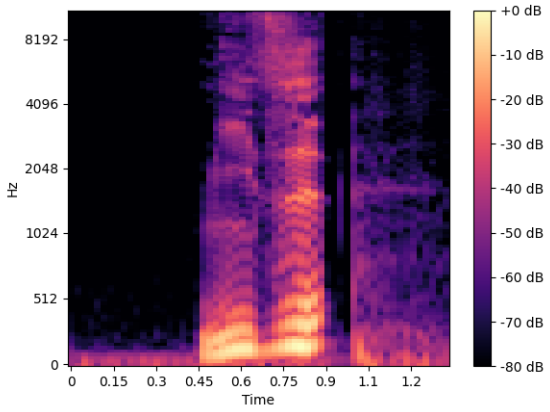

図 1:怒りを表すメルスペクトログラム

４．現時点での進捗

学習のための音声データとして、感情評定値付きオンラインゲーム音声チャットコーパス （OGVC）[5]と広島市立大学感情音声コーパス（HCUDB）[6]から「喜び」「怒り」「悲しみ」の 3 つの感情のデータ 4052 件を用意した。これを 8:2 に分け、訓練データと検証データとした。

学習モデルとして、まずは性格を考慮しない音声感情認識を行えるような AlexNet を作成した。過学習防止のため「early_stopping」（学習の途中打ち切り）を組み込んだ。過学習が起こると訓練データに過度に適合し、新しいデータに対して性能が低下してしまうためである。

現状の環境を用いて、訓練データを 8 割と検証データ 2 割を 5 回行った感情認識精度の結果を表 1 に示す。平均精度は 53.5%となり、精度向上が必要である。

表 1:感情認識精度

回数	精度
1 回目	53.7%
2 回目	61.0%
3 回目	55.6%
4 回目	45.1%
5 回目	52.3%
平均	53.5%

５．おわりに

本研究では性格情報を用いた音声感情認識の提案を行い、性格情報を用いたものとそうでないものとの比較をすることが目的である。今後は学習モデルの改善と、性格特性を組み込む方法を考える。その後、性格情報を用いた場合と用いない場合を比較する事で、性格情報の有用性について評価する。

参考文献

[1] 安藤厚志,“音声感情認識の技術動向”，日本音響学会誌，79 巻 1 号，pp. 72-79，2023

[2] Jeng-Lin Li, Chi-Chun Lee,“Attentive to Individual: A Multimodal Emotion Recognition Network with Personalized Attention Profile”，INTERSPEECH，2019

[3] 柴田大作， 若宮翔子， 伊藤薫， 荒牧英治，“JIWC:クラウドソーシングによる日本語感情表現辞書の構築”，言語処理学会第 23 回年次大会発表論文集，pp. 771-774，2017

[4] Alex Krizhevsky et. al., Ilya Sutskever, and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25，2012

[5] 有本泰子， 河津宏美，”音声チャットを利用したオンラインゲーム感情音声コーパス”，日本音響学会 2013 年秋季研究発表会講演論文集，1-P-46a，pp.385-388，2013

[6] 目良和也，黒澤義明，竹澤寿幸，“演技感情音声における演技感情と他者評価の関係の分析”，日本音響学会 2023 年秋季研究発表会講演論文集，1-9-17，2023