

# 時系列の音響特徴量を用いた音声感情分類についての一検討

田名網 那由多 林 実

明星大学大学院理工学研究科

## 1. はじめに

近年、コミュニケーションロボットや深層学習等の技術の飛躍的な発展に伴い、ユーザとコミュニケーションデバイス間の、音声対話によるコミュニケーションのニーズが急増している。同時に、対話における感情認識も益々必要とされている。そうした中、表情や言語情報のみならず非言語情報からの感情認識も重要である。

これまで、非言語情報による感情識別はサポートベクターマシンが多い。これは膨大な変数を学習させても過学習を起こしにくい特性を利用し、音声から音響特徴量を可能な限り取得し、それら全ての音響特徴量を使用して識別を行うというものである[1]。それらの音響特徴量の多くは、統計量のみを用いている[2]。

本研究では、統計量ではなく、時系列の音響特徴量を用いた音声感情分類の有効性を明らかにすることを目的として、時系列の音響特徴量を用いた LSTM[3]と統計的音響特徴量を用いた SVM との比較実験を行い、有効性を確かめる。

## 2. 時系列の音響特徴量と分類方法

実験では、感情音声に音声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)[4]の演技音声を用いた。その中から、4つの感情(喜び、悲しみ、怒り、平静)を演じさせた計960発話を分類の対象とした。

それらの感情音声に対して、25msの窓関数を10msずつシフトさせてメル周波数ケブストラム、ラウドネス、基本周波数の時系列の音響特徴量を取得した。図1はその例を示す。

取得した時系列の音響特徴量を使用して LSTM による音声感情分類を行った。併せて、比較のために統計的音響特徴量を用いて SVM による音声感情分類を行なった。



図1. 時系列の音響特徴量の取得例

## 3. 感情音声の分類結果

感情音声の分類結果を表1に示す。

表1に示した正解率の通り、時系列の音響特徴量を使用した LSTM による感情音声分類では、正解率は78.47%であり、統計的特徴量を用いた SVM による感情音声分類では、正解率は76.04%であった。

表1. 感情音声の分類結果

	時系列の音響特徴量を用いた LSTM	統計的音響特徴量を用いた SVM
正解率	78.47%	76.04%

## 4. 考察

表1の結果から時系列の音響特徴量を用いた LSTM が統計的音響特徴量を用いた SVM より正解率において2.43%上回ったことがわかる。従って、時系列の音響特徴量は音声感情分類においての有効性を確かめられたと考える。

LSTM は長期依存の学習が可能という点から考えて、感情音声に含まれる感情情報には時系列の音響特徴量における長期のパターンがあると考えられる。今後、発話全体に及ぶ長期のパターンを明らかにするとともに、感情ごとに正解率の確認を行う。

## 5. まとめ

本研究では、時系列の音響特徴量を用いた音声感情分類の有効性を明らかにすることを目的として、時系列の音響特徴量を用いた LSTM と統計的音響特徴量を用いた SVM の比較実験を行った。その結果、時系列の音響特徴量を特徴量として扱った LSTM の方が、統計量の音響特徴量を用いた SVM より正解率が2.43%高いことが確認でき、時系列の音響特徴量の有効性を確かめることができた。

## 謝辞

音声データの提供を頂きました国立情報学研究所音声資源コンソーシアムに対し心から感謝の意を表す。

## 参考文献

- [1] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Muller and S. Narayanan, "The Interspeech 2010 paralinguistic challenge," Proc. Interspeech 2010, pp. 2794-2797 (2010).
- [2] 有本泰子, 河津宏美, 大野澄雄, 飯田仁, "感情音声のコーパス構築と音響的特徴の分析" 情報処理学会研究報告. MUS, 74, 133-138, 2008-02-08
- [3] Felix A. Gers, Jürgen Schmidhuber, Fred A. Cummins, "Learning to Forget: Continual Prediction with LSTM" Published in Neural Computation 2000, pp. 2451-2471, (2000).
- [4] 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).