

クラスタ分析の並列利用による音声からの感情推定

○天沼沙織[†] 樽松理樹[‡] 羽倉淳[‡] 藤田ハミド[‡]

岩手県立大学大学院ソフトウェア情報学研究科[†] 岩手県立大学ソフトウェア情報学部[‡]

1. はじめに

人間の音声には、言語情報と非言語情報が含まれている^[1]。言語情報に対しては音声認識として研究開発が進められ、近年では商用ツールも多くみられる。一方、非言語情報に対してもその重要性が見直され、近年、韻律情報に着目し感情を推定する研究が盛んになってきている^[2]。韻律情報から感情を推定するには、感情付けされた音声の音声特徴を韻律情報として抽出し、付与された感情との対応付けを構築する手法が主流である。この問題に対し、クラスタ分析を適用する場合、感情以外の特徴から、クラスタを生成することになる。しかし、感情毎の音声特徴の範囲が互いに重なっている、散在していることなどから十分な精度が得られない。そのため、データをクラスタに集約するクラスタ分析の能力をこの問題に活用するためには、何らかの工夫を追加する必要がある。

以上の背景から、本研究では、感情付けされた音声感情ごとにクラスタ分析を行い、その結果を利用する、いわば、クラスタ分析の並列利用することで、音声からの感情推定を行う手法を提案する。本手法の有用性が示せば、従来とは異なるクラスタ分析の活用方法の提案になることが期待できる。

2. 提案手法

本提案手法は、大きく複数のクラスタ集合（以後、クラスタ集合群と表記）を構築す

る学習過程とクラスタ集合群から感情を推定する推定過程に分類される。学習過程においては、感情が明示された音声データに対し感情ごとにクラスタ分析を行う。これにより各クラスタの精度向上を図る。推定段階では、新たな音声データを各クラスタ集合に適用し、その結果を統合することで推定精度の向上を図る。以下に各部分について説明する。

2.1 学習過程

学習過程においては、感情ごとに音声データを分割する。分割した各音声データから音声特徴を抽出し、これらに対してクラスタ分析を行う。クラスタ分析としては K-means 法^[3]を用いる。K-means 法ではデータの与え方により構成されるクラスタ間の距離が近くなる可能性があるため、本研究では、データの主成分値を求め、クラスタが分散するように調整する。また生成したクラスタ集合の適切さが推定精度に大きく影響を与える。しかし、集合の適切さを図る明確な基準がないのが現状である。そのため、本研究では次に示すクラスタ集合評価基準を導入する。これにより、そのクラスタ集合を評価し、精度が高い裏スタ集合の生成を試みる。

クラスタ集合評価基準としては、生成されたクラスタおよびクラスタ間の関係を指標として用いる。クラスタに関する指標としては、クラス内分散、クラスタ内の半径、クラスタ内の直径を用いる。クラスタ間に関する指標としては、クラス間分散、クラス内分散・クラス間分散比^[3]、エントロピーを用いる。それぞれの値について次に説明する。

クラス内分散は、各クラスタ内のデータの分散を足し合わせることで求める。クラスタ内の半径は、クラスタの中心から最遠点にあるクラスタ内のデータまでの距離になる。クラスタ間の直径は、そのクラスタに属するデータ間の最長距離になる。クラスタ間分散は、クラスタの中心の分散として求め、クラ

A Proposal for Estimating Emotion in Speech
Using Parallel Cluster Analysis

[†] Saori AMANUMA · Iwate Prefectural
University Graduate School, Graduate School of
Software and Information Science

[‡] M. Kurematsu, J. Hakura, H. Fujita · Iwate
Prefectural University, Software and Information
Science

ス内分散・クラス間分散比は、クラス間分散をクラス内分散で割ったものである。エントロピーとしては、クラス集合の情報の曖昧さ、不確かさを表すために用い、その計算式は、式(1)のとおりである。

$$H = - \sum_{i=1}^N \frac{C_i}{E} \log_2 \frac{C_i}{E} \quad \text{式(1)}$$

式(1)において、 H はエントロピー、 N はクラス数、 C_i はクラス i に属するデータ数、 E はある一つの感情に属するデータ数を示す。よって、クラス内のデータ数がすべて同じ場合に最大で、クラス内のデータ数の偏りが大きければ大きいほどエントロピーは小さくなる。

評価基準として利用するには、これらのクラス集合評価基準値とクラス集合による感情推定結果との関係を明らかにする必要がある。この点は明確になっていないため、本研究では予備実験を通して、値を調整する。

2.2 推定過程

感情推定方法においては、学習段階で生成した各クラス集合の推定結果を統合し、推定を行う。具体的には生成した全クラスタの中心を対象とした k NN 法によってクラスタを選択する。選択されたクラスタが持つ感情を、与えられた音声から推定した感情とする。

3. 評価実験

本手法の有用性を評価するために、次にあげる実験を実施する。

3.1 実験概要

実験データとしては、Ekman ら^[4]が提唱する「恐怖・怒り・嫌悪・幸福・悲しみ・驚き」に「平常」を加えた7感情が付与された音声データを用いる。音声データとしては、慶應義塾大学 研究用感情音声データベース (Keio-ESD)^[5]を用いる。本音声には 47 感情含まれることから、前述の 7 感情を訓練データ、残りをテストデータとして利用する。

音声特徴としては、有声区間におけるピッチ、パワーそれぞれの最大値・最小値・範囲・四分位値・四分位値範囲・平均・標準偏差を利用する。

感情推定手法の k NN 法では、 k の値を 1、3、5 の 3 パターンで評価する。

3.2 評価方法

3.2 で示した実験データに対し、提案手法で推定した結果を、①データに付与された感情を類義語辞典などから分類した結果、②決定木を用いて評価した結果、③同じ音声を人が聞いた場合の結果、とを比較することにより、本手法の有用性を評価する。②の決定木は、音声からの感情推定の分類器として多く用いられている手法^[2]である。この手法の精度を提案手法の精度が上回れば、本手法は有用であると評価できる。

また、 K -means 法の条件を変化させて精度を求めることで、2.2 で示したクラス集合評価基準と推定精度との関係を検証する。

実験結果、考察については当日発表する予定である。

4. おわりに

本稿では、音声から感情を推定するために、クラス集合を感情ごとに並列処理する手法を提案した。現在、その有用性を評価している段階である。今後の課題としては、実験に基づく本手法の評価とそれを踏まえた手法の精錬化があげられる。また、本手法は音声からの感情推定以外のタスクにも適用できると考えられることから、従来のクラスタ分析には適さない問題への適用を行い、手法の汎用性についても検証する。

参考文献

- [1] 広瀬、“韻律と音声言語情報処理—アクセント・イントネーション・リズムの科学”、丸善、2006
- [2] Pierre-Yves Oudeyer、“The production and recognition of emotions in speech: features and algorithms”、International Journal of Human Computer Interaction、Vol. 59(1-2)、pp. 157-183、2003
- [3] 石井ら、“わかりやすいパターン認識”、オーム社、1998
- [4] P. Ekman ら著、工藤力訳、“表情分析入門—表情に隠された意味をさぐる—”、誠信書房、1987
- [5] 慶應義塾大学研究用感情音声データベース (Keio-ESD)、<http://research.nii.ac.jp/src/list/detail.html#Keio-ESD>