4K4-3

7感情の強度推定結果に基づく音響的特徴からの話者感情の推定手法

Speaker's mental state estimation method using intensities of seven emotions calculated from acoustic features

上村 譲史[*1] 新升 悠太[*1] 目良 和也[*2] 黒澤 義明[*2] 竹澤 寿幸[*2]
Joji UEMURA Yuta SHINMASU Kazuya MERA Yoshiaki KUROSAWA Toshiyuki TAKEZAWA

[*1] 広島市立大学情報科学部
School of Information Sciences, Hiroshima City University

[*2] 広島市立大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

Previous multi-class emotion classification method has two problems; it cannot deal with emotions except pre-defined emotions, and it cannot express situation when multiple emotions are arousing at the same time. In this paper, we apply hard and soft clustering methods into emotion classification. Hard clustering method gives information to re-classify emotion classes. Soft clustering method is valid to detect data which arouse multiple emotions. Seven types of emotion voice data (anger, anticipation, disgust, fear, joy, sadness, and surprise) are re-classified into eight clusters based on intensity values of the seven emotions calculated from acoustic features. The result of hard clustering suggested that some emotion data can be classified more finely. On the other hand, some data which have the features of multiple emotions could be found by soft clustering method.

1. はじめに

現在，感情を考慮した対応ができるインタラクションシステムについて研究が進められている．このようなシステムの実現に重要なのが，ユーザ感情の推定である．自然言語対話における話者感情推定に有効な情報源としては，バーバル情報（発話内容），音響的特徴（声の口調），視覚的特徴（表情，しぐさ）などが挙げられる．上村は，自然言語対話コミュニケーションにおいて感情推定に与える影響度についての調査を行い，音声＞表情＞言語の順で影響度が大きいことを示している[上村 2016]．

音響情報からの話者感情の推定についてはこれまで数多くの研究が行われているが，アウトプットとなる感情の形式については統一的に採用されているものは無い．感情のモデルとしては大きく多クラスモデルと次元モデルがある．多クラスモデルとしては，東アジアで古くから使われる喜怒哀楽の他に，表情分類に基づくモデル[Ekman1984]や，8 種類の基本感情をベースとして円錐状に感情を配置したモデル[Plutchik2011]がよく採用されている．一方次元モデルは 1 つまたは複数の次元の配置に基づき，情動を概念化するアプローチ方法である．多く用いられているのは快－不快の 1 軸の値で表すモデルであるが，これに興奮－冷静軸を追加したモデル[Russel11980]や緊張－弛緩，注目－拒否などの軸を加えたモデルもある．

しかし，実際の人間の感情を扱う上で，複数の感情クラスから 1 つを選択する手法には，分類に使用するクラスで過不足無く人間の感情の種類を網羅できるのかという問題と，複数の感情が混ざった心理状態をどう表すのかという問題がある．Plutchik や Russell の手法では類似した感情を近くに配置していると言われているが，類似していない感情が同時に生起している状況には対応できない．また次元モデルではアウトプットが感情クラスではなく次元軸の値であるが，やはり複数の感情が同時に生起している状態を表現できない．

そこで本研究では，7 種類の感情（喜び，恐れ，驚き，悲しみ，嫌悪，怒り，期待）の強度に基づくクラスタリングを用いた感情状態推定手法を提案する．本研究ではハードクラスタリングとソフトクラスタリングを用いることで，事前に想定した基本感情クラスの再分類のための情報の検出と，複数の感情的特徴を併せ持つ発話音声の検出について分析を行う．

2. 7 感情の強度推定結果に基づく感情推定手法

2.1. 機械学習による 7 感情の強度の算出

本手法では，7 種類の感情それぞれの強度値を算出するために，7 個の機械学習器を用いる．学習器としては，クラスの帰属度ではなく強度を学習し算出できる SMOreg[Smola1998]を用いる．

学習用音声データとしては，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)[有本 2013]に収録されている声優による感情演技音声を用いている．OGVC の演技音声は，8 種類の感情（喜び，恐れ，驚き，悲しみ，嫌悪，怒り，期待，受容）それぞれに対して，感情無し，弱感情，中感情，強感情と演じ分けられている．実験に用いた演技音声データの個数と学習のため付与した強度値を表 1 に示す．

表1 実験データの個数と学習強度値

		感情無し	弱感情	中感情	強感情
データの個数	怒り	80	80	80	80
	期待	80	80	80	80
	嫌悪	80	80	80	80
	恐れ	80	80	80	80
	喜び	84	84	84	84
	悲しみ	84	84	84	84
	驚き	96	96	96	96
学習用強度値		0.00	0.33	0.67	1.00

機械学習器に与える音響的特徴量は音響分析ツールopenSMILE[Eyben2010]を用いて計算する. 算出される素性値は, 音量, 12 次元のメル周波数ケプストラム係数, ゼロ交差率, 自己相関関数から算出した声である確率, 基本周波数とそれらの一次微分の波形(16×2=32 種類)に対して, 最大値, 平均値, レンジなど 12 種類の静的特徴量を掛け合わせた計 384 種類である. 本研究では 384 種類全ての素性を学習に用いた.

なお, 実験は Leave-one-out 方式によって行われるため, 表 1 の各実験データについて 7 種類の感情の強度が算出されている.

2.2. クラスタリングによる感情音声の特徴分析

表 1 の実験データのうち感情表出度が高い強感情 584 件に対しクラスタリングを実施することで分析を試みた. 本研究では, k-means 法によるハードクラスタリング手法と, fuzzy c-means 法によるソフトクラスタリング手法について比較を行った. なお, 事前にギャップ統計量によるクラスタ数推定[Tibshirani2001]を行い, クラスタ数を 8 と設定した.

2.2.1. ハードクラスタリングによる分析

k-means 法によるハードクラスタリング結果を表 2 に示す. クラスタ h1 には期待や喜びといったポジティブ感情音声が多く分類されており, クラスタ h4 には恐れや悲しみといったネガティブ感情音声が多く分類されている傾向が見られた. またクラスタ h4 にはポジティブ感情音声がほぼ分類されず, クラスタ h1 にはネガティブ感情音声がほぼ分類されていないことから, ポジティブ感情の代表となるクラスタとしてクラスタ h1 が, ネガティブ感情の代表となるクラスタとしてクラスタ h4 が形成されたと想定される.

表 2　ハードクラスタリングによる感情音声の分類結果

	クラスタ番号								計
	h1	h2	h3	h4	h5	h6	h7	h8	
怒り強	**15**	2	2	**16**	5	**22**	7	11	80
期待強	**15**	5	11	0	9	**12**	10	**18**	80
嫌悪強	3	**26**	0	**16**	0	**18**	4	13	80
恐れ強	3	7	**24**	**13**	8	9	6	**10**	80
喜び強	**25**	3	6	5	**22**	1	**14**	8	84
悲しみ強	1	**18**	**16**	**16**	10	0	6	**17**	84
驚き強	**16**	4	13	10	**17**	**15**	12	9	96

次に, 同じ感情でも違うクラスタに属している発話音声データを実際に聞き比べることで, クラスタ毎の違いや特徴を分析した.

- **怒りの音声に対する調査(クラスタ h1, h4, h6)**
  クラスタ h1 の音声を基準とすると, クラスタ h4 は嫌悪が混ざったような音声, クラスタ h6 は声のトーンが低く落ち着いた音声を多く含んでいた.
- **期待の音声に対する調査(クラスタ h1, h6, h8)**
  クラスタ毎で違いや特徴を感じることはできなかった.
- **嫌悪の音声に対する調査(クラスタ h2, h4, h6)**
  クラスタ h2 の音声を基準とすると, クラスタ h4 は嫌悪の度合が薄い音声, クラスタ h6 は声のトーンが高く怒りが混ざったような音声を多く含んでいた.
- **恐れの音声に対する調査(クラスタ h3, h4, h8)**
  クラスタ h3 は掠れた声で細々と喋っている音声, クラスタ h4 はクラスタ h3 と比較すると説明口調であり, 演技っぽさが表れている音声を多く含んでいた. クラスタ h8 については違いを感じることはできなかった.
- **喜びの音声に対する調査(クラスタ h1, h5)**
  クラスタ h1 は純粋に喜んでいる音声だったが, クラスタ h5 は喜びの中に嘲笑が混ざっているような音声を多く含んでいた.
- **悲しみの音声に対する調査(クラスタ h2, h3, h4, h8)**
  クラスタ h2, クラスタ h3 は掠れた声で細々と喋っている音声を多く含んでいたが, クラスタ h4, クラスタ h8 については特徴を感じることはできなかった.
- **驚きの音声に対する調査(クラスタ h1, h5, h6)**
  クラスタ h1, クラスタ h5 は驚きの中に嫌気を含んだような印象を受けたが, クラスタ h6 は純粋に驚いている音声を多く含んでいた.

以上より, ハードクラスタリングを用いることで, 事前に想定した 7 種類の感情クラスにとらわれず耳で聴いたときの印象に近い形で再分類するための特徴が得られることが示された.

2.2.2. ソフトクラスタリングによる分析

クラスタ数=8, ファジィ度合=1.4 で fuzzy c-means 法を実施した. ソフトクラスタリングはハードクラスタリングのように 1 データに対して 1 つのクラスタを割り当てるのではなく各クラスタに対する帰属度を表現できる手法であるため, 各クラスタについて帰属度が高いデータ上位 50 件を対象に, どの感情が多く含まれているかを調査した. 各クラスタの内訳を表 3 に示す. クラスタ s1 には悲しみや嫌悪が多く分類される, クラスタ s5 には喜びや驚きのように非ネガティブな感情が多く分類される, クラスタ s6 には驚きや怒りといった口調の強い感情が多く分類される, 期待は各クラスタに偏り無く分類されているなどの傾向が確認できた.

表 3　ソフトクラスタリングによる感情音声の分類結果

	クラスタ番号							
	s1	s2	s3	s4	s5	s6	s7	s8
怒り強	2	5	6	**15**	4	**16**	0	**14**
期待強	4	7	**9**	9	6	5	5	5
嫌悪強	**17**	2	7	**17**	0	5	0	**11**
恐れ強	4	0	8	6	7	1	**18**	7
喜び強	2	**17**	5	0	**14**	6	4	3
悲しみ強	**16**	5	**10**	0	8	0	**16**	10
驚き強	5	**14**	5	3	**11**	**17**	7	0
計	50	50	50	50	50	50	50	50

次に, 2 クラスタの帰属度が突出して高い発話音声データを実際に耳で聞いてみることで, 2 つのクラスタの特徴を含んでいるか調査した.

- **嫌悪強(クラスタ s1 の帰属度:0.37, クラスタ s4 の帰属度:0.50)**
  クラスタ s1 は嫌悪と悲しみの音声を, クラスタ s4 は嫌悪と怒りの音声を多く含んでいることから, クラスタ s1 とクラスタ s4 への帰属度がどちらも高い音声は嫌悪, 悲しみ, 怒りの 3 要素を有していると想定される. この音声を聞いてみたところ, ため息と突き放すような口調を含む音声であった.
- **怒り強(クラスタ s4 の帰属度:0.40, クラスタ s6 の帰属度:0.48)**
  クラスタ s4 は怒りと嫌悪の音声を, クラスタ s6 は怒りと驚きの音声を多く含んでいることから, クラスタ s4 とクラスタ s6 への帰属度がどちらも高い音声は怒り, 嫌悪, 驚きの 3 要素を有

していると想定される. この音声を聞いてみたところ, 怒りと嫌悪が感じられる音声であった. 驚きについては強くて明確な発声ではあるが, 発話内容の影響もあり驚きが生起しているとは感じられなかった.

- **驚き強(クラスタ s2 の帰属度:0.36, クラスタ s6 の帰属度:0.52)**
  クラスタ s2 は驚きと喜びの音声を, クラスタ s6 は驚きと怒りの音声を多く含んでいることから, クラスタ s2 とクラスタ s6 への帰属度がどちらも高い音声は驚き, 喜び, 怒りの 3 要素を有していると想定される. しかしクラスタ s2 とクラスタ s6 の帰属音声にほとんど差が感じられなかったため, クラスタ s2 とクラスタ s6 への帰属度がどちらも高い音声についても複数の要素を有しているとはいえなかった.

以上より, ソフトクラスタリングを用いることで, 複数の感情の特徴を併せ持つ音声の特徴を検出できる可能性があることが分かった.

3. おわりに

本論文では, 事前に想定した基本感情クラスの再分類のための情報の検出と, 複数の感情的特徴を併せ持つ発話音声の検出を行うため, 7 種類の感情(喜び, 恐れ, 驚き, 悲しみ, 嫌悪, 怒り, 期待)の強度に基づいてハードクラスタリングとソフトクラスタリングを行う感情状態推定手法を提案した. ハードクラスタリングによる分析を行った結果, 事前に想定した 7 種類の基本感情クラスにとらわれず耳で聴いたときの印象に近い形で再分類できる可能性が示された. またソフトクラスタリングによる分析を行った結果, 複数の感情の特徴を併せ持つ音声の特徴を検出できる可能性が示された

今後は, Arousal(活性)値の低い感情の音声データも分析に用いることで, 本研究で分析に用いた基本感情の偏りを軽減する予定である.

謝辞

本研究は国立研究開発法人科学技術振興機構(JST)の研究成果展開事業「センター・オブ・イノベーション(COI)プログラム」及び JSPS 科研費 26330313 の助成を受けたものです.

参考文献

[有本 2013] 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 年秋季研究発表講演論文集, (2013)

[Ekman1984] Ekman, P., Expression and the Nature of Emotion. In Scherer, K. & Ekman, P. (Eds.), Approaches to Emotion (pp. 319-343). Hillsdale, NJ: Lawrence Erlbaum (1984)

[Eyben2010] Eyben, F., Wöllmer, M., Schller, B., openSMILE – The Munich Versatileand Fast Open-Source Audio Feature Extractor, Proc. ACM Multimedia (MM), ACM, pp.1459-1462 (2010)

[Plutchik2011] Plutchik, R. "The Nature of Emotions", American Scientist, (2011)

[Russell1980] Russell, J., "A circumplex model of affect". Journal of Personality and Social Psychology 39: 1161–1178 (1980)

[Smola1998] Smola, A.J., Schoelkopf, B., A Tutorial on Support Vector Regression. In NeuroCOLT2 Technical Report Series (1998)

[Tibshirani2001] Tibshirani, R., G. Walther, and T. Hastie. "Estimating the number of clusters in a data set via the gap statistic." Journal of the Royal Statistical Society: Series B. Vol. 63, Part 2, 2001, pp. 411–423 (2001)

[上村 2016] 上村譲史, 目良和也, 黒澤義明, 竹澤寿幸, 字句情報, 音響情報, 表情から推定した話者の感情の食い違い状況の分析と食い違い自動検出手法の提案, 第 78 回情報処理学会全国大会発表論文集, pp.321-322 (2016)