

音声によって表出する混合感情の主観評価と音響的特徴量からの推定

松林 一也[†] 政倉 祐子^{††} 大野 澄雄^{††}

[†]東京工科大学 ^{††}東京工科大学 ^{††}東京工科大学

1 はじめに

人は多彩な感情を有しており、それは日常的に行われているコミュニケーションにおいて多く表出される。そのため、感情認識という研究は多くの手法で行われているが、単一の感情の認識に限定したものが多く。しかし、実際に人が表出する感情は必ずしも1つとは限らず、複数の感情が混合した状態(混合感情)で表出される。[1] 本研究では、音声に含まれる混合感情が生起する事を確認し、推定を行うモデルの構築と推定精度の検証を行った。

2 音声資料

本研究では、自然に発話された音声を用いるため、Microsoft 者が提供しているリアルタイムストラテジーゲーム“Age of Empires II”を用いて、大学院生の男性4人を2チームに分け協力しながら対戦をしてもらった。対戦中の会話はSkypeを用いて行い、Tapurにより収録した。

録音した音声はPraatを用いて前後200msの無音区間を含むものを一発話として切り出し、咳やクリック音などを取り除いた評価用400発話とダミー用50発話の計450発話を選定した。

3 音響的特徴量

収録した音声から評価に用いる400発話の特徴量を参考にPraatを用いて抽出した。抽出した特徴量は[2]を参考にし、用いる特徴量を表1に示す。

4 評価実験

本研究ではP. Ekmanによって全人類に普遍的であると分類された“喜び(joy)”、“悲しみ(sad)”、“怒り(ang)”、“恐れ(fear)”、“驚き(sup)”、“嫌悪(dis)”の6感情を用いて評価実験を行った。

大学生9名(男性7名、女性2名)の評価者はダミーを含む音声をランダムに並び替えた440発話を6感情分で指定された感情がどの程度含まれているかを1:全然含まれない、3:やや含まれる、5:すごく含まれるの5段階で評価を行った。

表 1: 音響的特徴量

特徴量	説明
Pmax	短時間平パワーの発話内最大値
Pmed	短時間平パワーの発話内中央値
Pmin	短時間平パワーの発話内最小値
Pmean	短時間平パワーの発話内平均値
Pstdv	短時間平パワーの発話内標準偏差
Fmax	F_0 値の発話内最大値
Fmed	F_0 値の発話内 中央値
Fmin	F_0 値の発話内最小値
Fmean	F_0 値の発話内平均値
Fstdv	F_0 値の発話内標準偏差
Cmax	第1次ケプストラム係数の発話内最大値
Cmed	第1次ケプストラム係数の発話内中央値
Cmin	第1次ケプストラム係数の発話内最小値
Cmean	第1次ケプストラム係数の発話内平均値
Cstdv	第1次ケプストラム係数の発話内標準偏差

5 決定木によるモデル構築と推定結果

決定木により混合感情を推定モデルを構築するため、感情が同時に生起しているグループを見つめるため、クラスター分析による感情のグループ化を行い、クラス数より感情グループを推定するモデルの構築を行った。

5.1 クラスター分析による感情グループの作成

評価実験より得られた評価値から混合感情を見つけるため、感情のグループ化を行った。評価者の各感情の評価値よりユークリッド距離を求め、Ward法によるクラスター分析を行い、いくつかのクラスターに分類した結果を図1に示す。図1より、各クラスターに2つ以上の感情が生起している事を確認した。

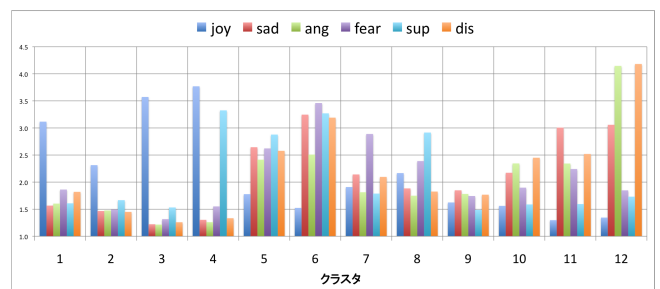


図 1: 各クラスターにおける感情の生起

Comparison of present and continuous feeling of acceptance for presentation in a single utterance of natural conversation

[†] Kazuya MATSUBAYASHI(Tokyo University of Technology)

^{††} Yuko MASAKURA(Tokyo University of Technology)

^{††} Sumio OHNO(Tokyo University of Technology)

5.2 クラスタ数の検証

クラスタ数の検証を行うため、クラスタ数を2から20までにクラスタリングを行い、決定木による、分岐数、RMSe、R2乗、誤認識率を求め、分岐数を図2に、RMSe、R2乗、誤認識率を図3に示す。

分岐数はクラスタ数が16以上になると一定になった。またRMSeや誤認識率よりクラスタ数は12とし、決定木によるモデルの構築を行った。

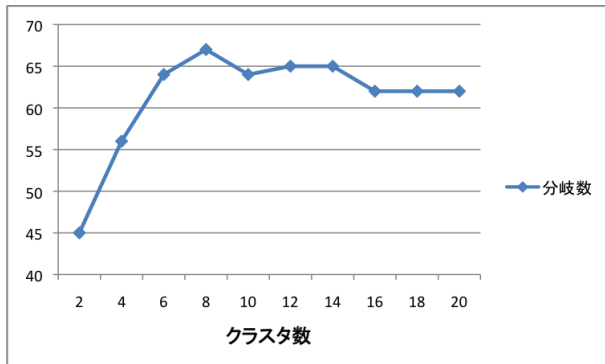


図 2: 決定木による分岐数

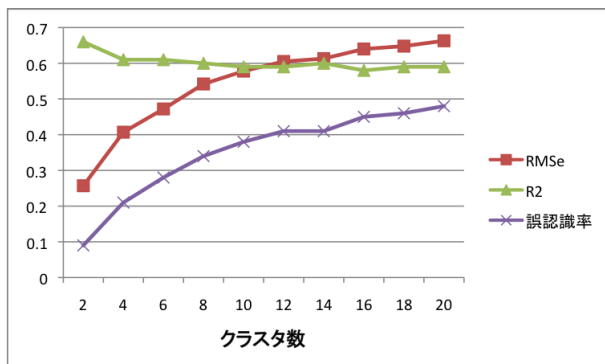


図 3: 決定木による RMSe、R2 乗、誤認識率

5.3 決定木によるモデル

決定木を用いて音響的特徴量から評価実験より求めたクラスタの推定を行った。分類したクラスタを推定するため、音響的特徴量を説明変数、クラスター分析によって得られたクラスタを目的変数として推定モデルの構築を行った。推定精度の誤差を回避するため n-fold 交差検証法 (n=10) を採用している。決定木による構築を行った結果、分岐数:65 分岐、RMSe:0.61、R2 乗:0.6 よりツリーが構築された。

5.4 モデルによる推定結果

決定木による推定結果を y 軸が実測値、x 軸が予測値とし表 2 に示す。推定結果はクラスタは 2、9 のクラスタが最も良かった。しかし、クラスタ 1 や 7 の推定結果が悪いという結果になった。

表 2: 決定木による推定結果

	1	2	3	4	5	6	7	8	9	10	11	12
1	6	3	0	0	5	0	0	2	1	1	3	0
2	0	39	2	2	0	0	0	1	5	4	2	1
3	1	0	24	1	1	1	0	0	3	1	0	1
4	0	0	1	18	0	0	0	4	0	1	1	1
5	0	3	2	2	23	1	0	3	2	1	2	1
6	0	0	6	1	0	5	0	1	0	0	0	1
7	0	6	1	2	4	0	0	4	4	1	0	0
8	2	2	2	1	4	0	0	26	1	4	0	0
9	0	8	1	0	0	0	0	0	38	1	6	1
10	0	4	0	0	0	0	0	3	1	26	4	0
11	0	1	1	1	2	0	0	0	0	3	25	1
12	2	1	1	0	2	0	0	1	1	1	2	8

6 考察・まとめ

1つの発話から混合感情が生起する事を確認し、推定モデルの構築を行った。推定モデルによる推定の正解率を表3に示す。全体的な認識率は約18%と低い結果となったが、クラスタ2から5、8から11においては60%以上の認識率となっている。認識率が低いクラスタは度数に起因していると考えられる。また、クラスタ7に関しては6感情が全体的に高く生起しているため、正確に推定できなかった可能性が高い。より細かな混合感情の推定を行うためにはクラスタ数と学習データ数を増やす必要が有ると考えられる。

表 3: 決定木による認識率

クラスタ	1	2	3	4	5	6
認識率	0.29	0.70	0.73	0.69	0.58	0.36
度数	21	56	33	26	40	14

クラスタ	7	8	9	10	11	12
認識率	0.00	0.62	0.69	0.68	0.74	0.42
度数	22	42	55	38	34	19

参考文献

- [1] 赤木正人、”音声に含まれる感情情報の認識”、日本音響学会誌 66 巻 8 号、 pp.393-398、 2010.
- [2] 有本泰子、大野澄雄、飯田仁、”怒りの程度の時間変化量を推定する音響的特徴量の検討”、日本音響学会講演論文集、 pp.324-325、 2009.