1B1-OS-25a-4in1

表情・音響情報・テキスト情報からのリアルタイム感情推定システム

Construction of real-time emotion recognition system
from facial expression, voice, and text information

岡田 敦志 Atsushi OKADA　上村 譲史 Joji UEMURA　目良 和也 Kazuya MERA　黒澤 義明 Yoshiaki KUROSAWA　竹澤 寿幸 Toshiyuki TAKEZAWA

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

近年，表情，音響的特徴，発話文字列の情報を総合して発話者の感情推定を行う手法が提案されている．しかし，「半笑いで褒める」のように各情報源からの推定結果が食い違うことが意味を持つ状況も存在する．このような状況を捉えるため，本研究では表情，音響的特徴，発話文字列からの感情推定を並列かつリアルタイムに行うシステムを構築した．本稿では提案システムの構成および各感情推定処理の評価結果について述べる．

1. はじめに

近年，人間とコンピュータエージェントが自然言語による対話コミュニケーションをする機会が多くなった．しかし本当に自然な対話の実現のためには，相手の気持ちの変化を慮る「思いやり」の対話プランニングや，相手の気持ちに「共感する」態度を示す処理などが必要となる．[東中 16]では，理想の対話システムである「感情むき出しではないが，感情を持った対話システム」の利点として，ユーザの感情をくみ取ることができること，システムの状態をより直感的に示すことができること，機械との対話をより自然で豊かなものにできることの三点を挙げている．

しかし従来の対話システムにおける感情推定技術では，システムは人間の発話の字面しか考慮しないため，場合によっては不適切な返答をしてしまうことがある．例えば，人間が元気なく「元気だよ」と発話したときも，ノンバーバル情報に基づく感情推定を行っていなければ「それは良かったね」と表層的な返答をしてしまう．

そこで現在，ノンバーバル情報からの感情推定手法の研究が活発になっている．その情報源としては主に，表情や発話音声の音響的特徴などが挙げられる．ほとんどの手法は単独の情報源から感情推定を行っているが，実際の人間の心理状態としては，皮肉やツンデレのように，各情報源から推定される感情に食い違いがあることによって，複雑な心理状態が表現されていることもある．

そこで本研究では，このような表出感情の食い違いを対話しながら検出できるように，話者の表情，音響的特徴，発話文字列からそれぞれの感情をリアルタイムで推定し，出力するシステムを構築する．本システムは，発話者がカメラを見ながらマイクに向かって発話すると，端末上にリアルタイムで表情，音響情報，発話文字列それぞれからの感情推定結果が随時表示される．

2. リアルタイム感情推定システム

2.1 システムの処理の流れ

本研究で提案するリアルタイム感情認識システムの処理の流れを図 1 に示す．本システムでは発話者の表情を取得するためにカメラを，発話の音響情報と発話文字列を取得するためにマイクを用いている．

連絡先：目良和也，広島市立大学大学院情報科学研究科，
広島市安佐南区大塚東 3-4-1，mera@hiroshima-cu.ac.jp

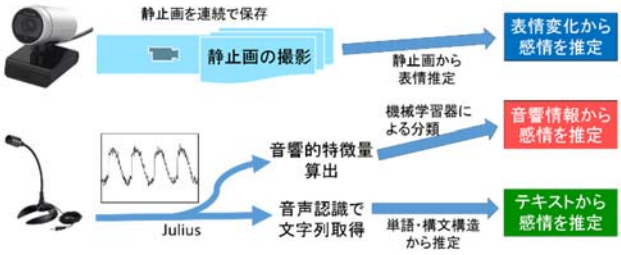

図 1: リアルタイム感情推定システムの処理の流れ

まず，発話者の顔をカメラで撮影し，連続で静止画像を取得する．次に取得した全静止画像に対し表情からの感情推定を行う．そして一発話の区切りを検出したら，発話区間に取得した静止画像の感情推定結果に基づいて発話区間全体の話者感情を推定する．

表情からの感情推定処理と並行して，マイクから取得した発話音声をもとに音響的特徴を抽出し，機械学習器によって音響的特徴からの感情推定処理も行う．

さらに，発話音声から音声認識装置を用いて取得した発話文字列に対して形態素解析を行い，解析された単語および構文構造から感情推定を行う．

これらの処理を並列かつリアルタイムに行うことで，一発話における 3 つの情報源からのそれぞれの推定感情を随時出力する．

2.2 表情変化からの感情推定

表情からの感情推定処理では，まず 30msec～50msec 間隔でカメラから取得した全静止画に対して OKAO Vision[オムロン]を用いて表情推定を行う．OKAO Vision は静止画から人間の顔を認識し，抽出した顔の特徴点を基に 5 感情（無，喜，驚，怒り，悲）の尤度を出力する．

次に，音声認識装置による発話区切りの検出をトリガーとして，一発話区間全体における話者感情を推定する．発話区間におけるどのタイミングの表情を参照するかについて，本研究では事前に行った小規模実験の結果に基づいて【発話区間が終わった瞬間】の表情を採用している．

なお表情から推定される感情クラスは，OKAO Vision により最も尤度が高いとされた感情を，3 種類の感情クラス（positive（喜），neutral（無，驚），negative（怒，悲））に再分類して出力している．

2.3 音響情報からの感情推定

音響情報からの感情推定では，まずマイクを通して取得した発話音声からオープンソースソフトウェア openSMILE[Eyben 10]を用いて音響的特徴を抽出する．本研究では，特徴量セットの1つである IS09 セットに含まれる特徴量のうち，音のエネルギー，12 次元のメル周波数ケプストラム係数(MFCC)，基本周波数とそれらの一次微分（14×2=28 種類）に対して算出される最大値，平均値，レンジなど 12 種類の基本統計量計 336 種類を機械学習に用いる．

次に，openSMILE によって得られた 336 種類の音響的特徴量を機械学習器 Support Vector Machine（SVM）に学習させることで 11 種類の感情クラス（嬉しい，楽しい，リラックス，冷静，不安，怒り，恐れ，嫌悪，軽蔑，疲れた，悲しい）に分類する．機械学習に用いた学習データは，感情音声コーパス[目良 17]に含まれる音声データのうち，21～24 歳の大学生 11 名中 8 人以上が同じ感情と評定した演技音声 1,163 件である．学習データの内訳を表 1 に示す．

表 1:機械学習に用いた感情音声データ数

感情クラス	データ数
狂喜・楽しい	35
余裕・嬉しい	138
リラックス・気楽	126
眠い・疲れた	81
冷静	126
不安・緊張	180
憂鬱・悲しい	118
恐れ	40
怒り	175
嫌い	109
軽蔑	35
合計	1,163

2.4 テキスト情報からの感情推定

発話文字列からの感情推定処理では，まず発話音声を音声認識エンジン Julius[河原 05]を用いて文字列に変換する．次に，形態素解析エンジン MeCab[Kudo 04]を用いて発話文字列を形態素解析する．最後に，発話文字列に含まれる語に対して評価極性辞書を参照することで，発話文字列全体の表す感情を positive，negative，neutral の 3 クラスの感情に分類する．本研究では評価極性辞書として，現代形容詞用法辞典[飛田 91]と感情表現辞書[中村 93]を用いた．

提案手法では評価極性辞書との語句マッチングによって文字列の感情極性を判定する．まず，発話文字列中の単語一つ一つを評価極性辞書とマッチングさせ，positive が付与されている時に感情の度合を+1 し，negative が付与されている時に感情の度合を-1 する．感情極性を持つ語に「ない」という否定表現がかかっている場合，感情極性の符号を反転する．そして，最終的な感情の度合が+1 以上のとき positive，0 のとき neutral，-1 以下のとき negative と判断する．

3. 各感情推定手法の評価実験

3.1 表情変化からの感情推定実験

実験手順を以下に示す．

1. 複数の大学生による自由対話の様子を録画し，そこから一発話ずつ発話動画データを切り出す（計 400 件）
2. (1)の発話動画を音声無しで評定者（話者とは異なる大学生 5 名）に提示し，表情変化のみから推定される話者感情を 3 種類（positive，neutral，negative）から 1 つ選択させる
3. 5 人中 3 人以上の感情評定結果が一致したデータに対して提案手法を適用する
4. 人間による感情評定結果を正解として提案手法の出力を評価する

実験結果を表 2 に示す．positive および neutral の推定精度は良かったが，negative は neutral に多く誤推定されたため精度が下がった．また再現率については neutral が低かったが，これは表情検知の感度が良すぎたためと考えられる．なお全体の正解率は 0.41 となった．

表 2:表情変化からの感情推定結果

		人間による評定			精度
		positive	neutral	negative	
提案手法	positive	83	38	4	0.66
	neutral	33	58	7	0.59
	negative	37	93	12	0.08
	推定失敗	3	3	1	
再現率		0.53	0.30	0.50	

3.2 音響情報からの感情推定実験

表 1 の感情音声データに対して Leave-one-out 交差検証実験を行った．正解率を表 3 に示す．精度については，余裕，不安，怒りが高かった．再現率については恐れと軽蔑が高かった．全体の正解率は 0.43 となった．

表 3:音響情報からの感情推定結果

	精度	再現率	F 値
狂喜・楽しい	0.36	0.57	0.44
余裕・嬉しい	0.54	0.42	0.47
リラックス・気楽	0.45	0.39	0.42
眠い・疲れた	0.40	0.53	0.46
冷静	0.37	0.51	0.43
不安・緊張	0.54	0.34	0.42
憂鬱・悲しい	0.46	0.39	0.42
恐れ	0.33	0.68	0.44
怒り	0.57	0.31	0.40
嫌い	0.44	0.47	0.46
軽蔑	0.25	0.71	0.36

3.3 テキスト情報からの感情推定実験

実験手順を以下に示す．

1. 複数の大学生による自由対話から 400 発話を選択し，人手で文字に書き起こす
2. (1)の発話文字列を評定者（話者とは異なる大学生 5 名）に提示し，推定される話者感情を 3 種類（positive，neutral，negative）から 1 つ選択させる
3. 5 人中 3 人以上の感情評定結果が一致したデータに対して提案手法を適用する
4. 人間による感情評定結果を正解として提案手法の出力を評価する

結果を表 4 に示す. 推定精度は, positive が 0.21, neutral が 0.96, negative が 0.05, 全体の正解率は 0.59 となった. 400 発話中 373 が neutral と推定されていた. これは, 評価極性辞書の単語が少なく, 発話文字列とマッチしなかったためだと考えられる. また, 評価極性辞書が話し言葉に対応していなかったことも一因である.

表 4:テキスト情報からの感情推定結果

		人間による評定			精度
		positive	neutral	negative	
提案手法	positive	3	11	0	0.21
	neutral	5	189	4	0.96
	negative	5	133	7	0.05
再現率		0.23	0.56	0.54	

4. おわりに

本研究では, 表情, 音響情報, テキスト情報それぞれから並列かつリアルタイムに話者感情を推定し出力する手法を提案した. 表情からの感情推定は, 発話終了時の静止顔画像の表情を参照した. 音響情報からの感情推定は, 336 種類の音響特徴量に基づき機械学習により推定した. テキスト情報からの感情推定は, 評価極性辞書とのワードマッチングをベースとした手法を用いた. 各感情の推定結果は, 表情からの推定正解率が 0.41, 音響情報からの推定正解率が 0.43, テキスト情報からの推定正解率が 0.59 であった.

今後はそれぞれの感情推定精度の向上だけでなく, 本システムを用いて 3 つの感情推定結果の一致あるいは不一致状況を分析し, “皮肉”や“ツンデレ”のように, より複雑な心理状態を検出できるようにする予定である.

謝辞

本研究を行うにあたり, オムロン(株)から画像センシング技術 OKAO(R) Vison をご提供いただいております. また本研究は, 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

[東中 16] 東中竜一郎, 岡田将吾, 藤江真也, 森大毅: 対話システムと感情, 人工知能学会誌, Vol.31, No.5, pp.664-670, 2016.

[オムロン] OKAO Vision | オムロン人画像センシングサイト, http://plus-sensing.omron.co.jp/technology/, (2017/03/07 アクセス).

[Eyben 10] Florian Eyben, Martin Wöllmer, Björn Schuller: “openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor,” ACM Multimedia Conference – MM, pp.1459-1462, 2010.

[目良 17] 目良和也, 谷有希, 村田唯, 黒澤義明, 竹澤寿幸: 演技感情と推定感情のタグを付与した感情音声コーパスの構築, 日本音響学会 2017 年春季研究発表会講演要旨集, p.142, 2017.

[河原 05] 河原達也, 李晃伸: 連続音声認識ソフトウエア Julius, 人工知能学会誌, Vol.20, No.1, pp.41-49, 2005.

[Kudo 04] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.

[飛田 91] 飛田良文, 浅田秀子: 現代形容詞用法辞典, 東京堂出版, 1991.

[中村 93] 中村明, 感情表現辞書, 東京堂出版, 1993.