

発話における言語的意図と音声的意図のギャップの検出

— ツンデレ発話を事例として —

森下康生 目良和也 黒澤義明 竹澤寿幸

広島市立大学大学院 情報科学研究科

morishita@ls.info.hiroshima-cu.ac.jp, {mera, kurosawa, takezawa}@hiroshima-cu.ac.jp

1. はじめに

昨今、しゃべってコンシェルや Siri などのコンシェルジュシステムや音声案内システム[1,2]など、音声対話によってコンピュータとコミュニケーションをとるシステムが開発されている。その主要な技術として音声認識が挙げられる。音声認識は、人間の音声発話を文字列に変換して出力する。しかし、その処理において、早口、沈んだ口調、強い調子などのノンバーバルな情報は欠落してしまう。一方、音響分析の分野では、声の高さや大きさの変化などの情報から、話者の感情を判別するための研究が進められている。しかし、発話の内容、音声の変化、いずれか片方の技術を使うだけでは十分な精度が得られないため、言語的（バーバル）特徴と音響的（ノンバーバル）特徴の両方を使った感情推定手法が提案されている[3]。音響的特徴には、単語のアクセントや文のイントネーションなどのパラ言語的特徴と、怒鳴ったり照れて早口になったりなどのノンバーバル的特徴があるが、本研究ではノンバーバル的な音響的特徴に注目する。

我々の研究グループでは、これまで、発話内容の格フレーム構造を用いたバーバルな感情判別手法を提案している[4]。また、ノンバーバル情報である発話の音響情報から、納得しているか否かや自信があるか否かといった心的状態を判別する手法も提案している[5]。

しかし実際の音声発話を分析したところ、バーバルとノンバーバルで異なる感情が表れている事例が発見された。人間同士のコミュニケーションにおいて、このような感情表出の食い違いが生じている場合は、話者が本心を隠している、あるいは、思ったことをうまく表現できていないという可能性がある。そうした場合においても聞き手が正しく話者の真意をくみ取れたならば、聞き手も話者の意図に応じた対話を行うことが可能であるが、聞き手が話者の意図を間違えて汲み取った場合、コミュニケーションが成り立たなくなってしまうといった事態も起こりうる。

そこで本研究では、バーバルな特徴とノンバーバルな特徴それぞれから話者の意図を推定し、それらの推定結果にギャップのある事例を収集し、分析する。さらに特定のギャップ状況から高次の話者の意図が推定できるか実験を行う。なお本研究では高次の話者の意図の例として、ツンデレ発話を対象に実験を行う。

2. ツンデレ発話の定義

本研究では、バーバルな感情表現とノンバーバルな感情表現のギャップがある典型例として、ツンデレ発話を扱う。萌え用語選定委員会[6]によると、ツンデレは以下のように定義されている。

- (1) いつもはツンツンしているのに、2人になるとデレデレする状態のこと
- (2) ツンデレとされる、強気で不器用なキャラクターが特定の相手に対して自分の心が動揺した時に、それをごまかすために使用する表現

本研究では(2)を対象として調査を進める。表1に感情属性およびツンデレを規定化したものを示す。ツンツンおよびデレデレの感情属性があり、それらが内面および外面で一致している場合、キャラクター属性としてはそれぞれ、ツンツンおよびデレデレとなる。しかし、両者が不一致である場合、キャラクター属性としてツンデレと規定できる[7]。

表1：ツンデレ状態における感情表出

外面	ツンツン	デレデレ	ツンツン	デレデレ
内面	ツンツン	デレデレ	デレデレ	ツンツン
属性	ツンツン	デレデレ	ギャップの存在	
			ツンデレ	

なお、ツンデレ発話における特徴的な言語的表現も存在する。富樫は以下のような言語的表現を挙げている[7]。

- (1) つっかえ
- (2) 接続助詞「から」による言いさし
- (3) 終助詞「ね」「よ」
- (4) 「～てあげる」「～でもいい」
- (5) とりたて詞「だけ」
- (6) 発話末の強調(文語体では、促音「っ」や記号「!」の付加)
- (7) 副詞類(「別に」「勘違いしないで」など)

本研究では、比較実験として上記の言語的表現のみを素性として用いた場合のツンデレ発話の検出実験を行う。

3. 言語的特徴と音響的特徴を用いた発話意図ギャップ検出手法

図1は本提案手法の概要である。本手法では、言語表現から推定できる感情を外面的な感情表現、音響的特徴から推定できる感情を内面的な感情表現とし、それらの推定結果が一致しない場合をギャップのある状態として検出する。表1より、ギャップ状態のうち外面的にネガティブで、内面的にポジティブな感情が表出されている時をツンデレ状態とする。

入力発話が与えられた時、音響分析による音響的特徴の抽出と、音声認識による発話文章の獲得および、発話文章からの言語的特徴の抽出が行われる。抽出された特徴量を素性として機械学習器により学習し、言語・音声の意図のギャップを検出する。本手法で使用する音響的特徴および言語的特徴については、次節で説明する。

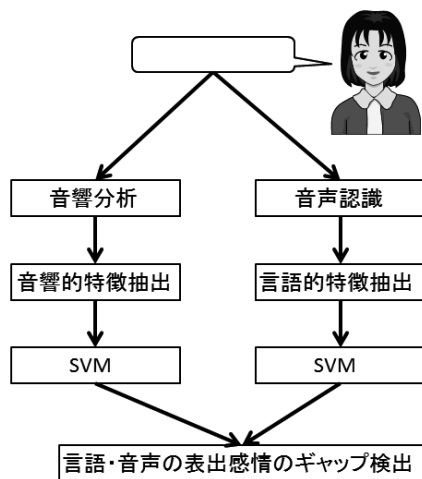


図1：提案手法の概要

4. 実験に用いる特徴量

4.1 音響的特徴量

本研究では OpenSMILE[8]によって抽出される音響的特徴量を用いる。まず、以下5つの特徴量グループを挙げる。

- ・エネルギーの二乗平均平方根
- ・1次-12次までのメル周波数ケプストラム係数
- ・基礎周波数
- ・エネルギーの二乗平均平方根の一次微分成分
- ・基礎周波数の一次微分成分

上記グループから、以下の特徴量を採用する。

- ・最大値
- ・最大-最小値の範囲
- ・最大の絶対値
- ・算術的平均値
- ・線形近似と実波形の差異として計算された平方誤差
- ・標準偏差

上記以外に、特徴量グループ“有声確率”から“算術的平均値”のみを採用する。これら96種類の特徴量をSVMの素性として使用する。

4.2 言語的特徴量

本研究における言語的特徴量として、高村の感情極性値[9]を使用した。単語の感情極性とは、ある単語が良い印象を持つ（ポジティブ）か悪い印象を持つ（ネガティブ）かを示す二値変数である。感情極性値は語彙ネットワークを利用して自動的に計算されたものであり、今回利用した公開資源中の単語感情極性対応表は-1から+1の実数値が割り当てられている。-1に近いほどネガティブ、+1に近いほどポジティブであることを示す。

本研究において、発話文中のポジティブな単語およびネガティブな単語の極性値をそれぞれ合計し、その2つの値を学習機械に入力する素性値とする。

5. 評価実験

本論文では、以下の4種類の実験を行う。

- I. 音響的に表出感情を分類する実験
- II. 言語的に表出感情を分類する実験
- III. IとIIの結果からギャップを検出する実験
- IV. 典型的ツンデレ表現のみを用い、ツンデレ発話を判別する実験（比較実験）

実験データとして、アニメDVD等[10,11,12,13,14]から同一の声優（釘宮理恵氏）の発話のみを収集し、作業員（20代男性1名）が発話かツンデレであるか否かをタグ付けした。その際、音響的にポジティブ・ネガティブ・ニュートラルのいずれであるか、および言語的にポジティブまたはネガティブであるかについてもタグ付けした。このようにして収集したツンデレ発話および非ツンデレ発話各263発話を実験に用いた。

なお学習機械は、LIBSVM[15]を使用した。また、検定手法として、5分割交差検定を用いた。

5.1 音響的特徴量を用いた感情分類実験

4.1節で述べた97種類の音響的特徴量を用いて、その発話がポジティブ（デレ）、ネガティブ（ツン）、ニュートラルのいずれであるかを判別する。表2に分類結果を示す。

表2：音響的特徴を用いた機械学習による感情分類結果

		Prediction			Recall
		Positive	Negative	Neutral	
Actual	Positive	61	26	29	0.52
	Negative	31	263	25	0.82
	Neutral	3	25	63	0.69
Precision		0.64	0.83	0.53	

5.2 言語的特徴を用いた感情分類実験

4.2 節で述べた感情極性値を用いて、対象発話がポジティブであるかネガティブであるかを判別する。

まず対象の発話を形態素解析し、単語の原形を求める。次に、単語感情極性対応表を用い、各単語の感情極性値を求める。そして、ポジティブな単語同士、およびネガティブな単語同士で極性値の和を求める。これら 2 種類の値を機械学習の素性値として使用する。表 3 に分類結果を示す。

表 3：言語的特徴を用いた機械学習による感情分類結果

		Prediction		Recall
		Positive	Negative	
Actual	Positive	224	10	0.95
	Negative	29	263	0.88
Precision		0.88	0.96	

5.3 表出感情のギャップを検出する実験

5.1, 5.2 節で述べた言語的・音響的特徴量による分類結果を組み合わせて調査を行う。具体的には、言語的特徴量における分類結果と音響的特徴量による分類結果で食い違い（言語的にツン、音響的にデレ）があった発話をツンデレとして判断する。表 4 に結果を示す。

この結果、提案手法のツンデレ検出精度は 0.70、再現率は 0.27、F 値は 0.39 となった。また、SVM 出力ではなく、人手で付与したタグをもとに言語および音声の表出にギャップのある発話を検出した結果を表 5 に示す。

表 4：表出感情のギャップを用いたツンデレ検出結果

		Prediction		Recall
		言語=ツン 音声=デレ	左記以外	
Actual	ツンデレ	73	190	0.27
	非ツンデレ	31	232	0.88
Precision		0.70	0.54	

表 5：人手で付与したタグを使ったツンデレ検出結果

		Prediction		Recall
		言語=ツン 音声=デレ	左記以外	
Actual	ツンデレ	51	212	0.19
	非ツンデレ	20	243	0.92
Precision		0.71	0.53	

5.4 比較実験

5.3 節の実験に対する比較として、2 章で述べた特徴的なツンデレ表現を用いた実験を行った。2 章の 7 つの特

徴を素性とし、文中にいずれかが出現した場合、その素性値を "1" として、SVM に学習させた。例として、「べ、別にあなたのことが好きなんじゃないんだからねっ！」という発話の場合、素性番号 1, 2, 3, 6, 7 の素性値が 1、それ以外の素性値は 0 となる。

分類結果を表 6 に示す。ツンデレ検出精度は 0.63、再現率は 0.77、F 値は 0.69 であった。

表 6：言語的表現のみを用いたツンデレ検出結果

		Prediction		Recall
		ツンデレ	非ツンデレ	
Actual	ツンデレ	203	60	0.77
	非ツンデレ	119	144	0.54
Precision		0.63	0.70	

5.5 考察

本実験では提案手法の再現率が低かった。その理由としては表 5 からわかるように、ツンデレ発話といえなくても言語・音声の表出にギャップがあるとは限らないことが挙げられる。より具体的には、ツンデレ発話であっても、例えば恥ずかしい思いをしたことに怒りながら（ネガティブ）否定／罵倒（ネガティブ）をする場合もあり、音声・言語のそれぞれの表出感情にギャップのないツンデレ発話が一定数存在することも再現率が低い一因と考えられる。

また今回、言語的特徴=外面表現、音響的特徴=内面表現としてとらえ、言語=ツン、音声=デレの際にツンデレとして判断したが、言語=デレ、音声=ツンのパターンもツンデレとして判断すると精度 0.81、再現率 0.67 となった。例として、咄めるような口調（ネガティブ）で相手を労わるような（ポジティブ）発話があり、前述したような言語=ツン、音声=デレのパターンを扱うのみでは十分なツンデレ発話を検出できないと考えられる。この場合、表 1 で示されるような富樫が提唱する一般的なツンデレ表現と乖離している可能性もありうるが、上述のツンデレ発話のパターンが存在することを考えると、ある程度留意する必要がある。

非ツンデレ発話に関しても、声は笑っているが（ポジティブ）相手を脅す・けなす（ネガティブ）発話があり、言語と音声の意図にギャップがある発話でも、必ずしもツンデレ発話であるとは限らない事例もいくつか確認された。

比較手法として、特徴的ツンデレ表現を使用したツンデレ発話の検出実験を行ったが、こちらでは精度は 0.63、再現率が 0.77 と一定の高さを示した。逆に非ツンデレ発話の場合、精度は 0.70 と高めといえるが再現率は 0.54 と、低めの結果となった。これらに関し、例えば典型的ツンデレ表現の「～から」はツンデレ発話において「言いさし」として使用されている場合もあったが、非ツンデレ発話における「理由・原因」を述べるための語として使用されている場合もあった。また、「ね」「よ」などの語尾としての表現も、これらはいわゆる「女性ことば」

であるため、ツンデレ以外の発話でも頻出していた。これらから、典型的ツンデレ表現が使用されている発話でも、その表現がツンデレ以外の意図で使用されている発話が存在するため、典型的ツンデレ表現を使用したツンデレ発話検出手法でも、必ずしも高い正解率を得られるわけではないといえる。

6. まとめ

本研究では、音響的特徴および言語的特徴量のそれぞれにおける表出感情を SVM により学習・分類し、それぞれの表出感情のギャップを検出することを目的とした。音響的特徴としては音の高さや大きさなどから算出される 96 種類の特徴量を用いた。言語的特徴量としては単語の感情極性値から求めた正負値それぞれの和を特徴量として用いた。実験の結果、ツンデレ発話 263 発話中 73 発話を提案手法により検出、0.70 の精度と 0.27 の再現率を得た。

今後は収集したギャップ事例をより詳細に分析し、音声・言語の意図の一致・不一致についてより細かいパターンで考察する予定である。さらに、きつい口調で「バカッ！」と言っているが顔を見ると照れているというような、表情やしぐさなど音響的特徴以外のノンバーバル情報も考慮したギャップ検出手法についても検討する予定である。

参考文献

- [1] 双方向音声案内デジタルサイネージ メイちゃん (<http://mei.web.nitech.ac.jp/>)
- [2] 西村竜一ほか, 実環境研究プラットフォームとしての音声情報案内システムの運用, 電子情報通信学会論文誌 D-II, J87-D-II(3), pp.789-798 (2004).
- [3] Björn Schuller, et al., "Speech Emotion Recognition Combining Acoustic Features and Linguistic Information in a Hybrid Support Vector Machine – Belief Network Architecture," Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing, Vol.1, pp.577-580 (2004).
- [4] 目良和也ほか, 語の好感度に基づく自然言語発話からの情緒生起手法, 人工知能学会論文誌, Vol.17, No.3, pp.186-195 (2002)
- [5] Michihisa Kurisu, Kazuya Mera, et al., "A Method Using Acoustic Features to Detect Inadequate Utterances in Medical Communication," IEEE Int'l Conf. Systems, Man, Cybernetics, pp.116-119 (2012).
- [6] 萌え用語選定委員会, 萌え萌え用語の萌え知識, イーグルパブリッシング, 2005
- [7] 富樫純一, ツンデレ属性と言語表現—役割語的アプローチとケーススタディー, 第 6 回現代日本語文法研究会, pp.1-18 (2008)
- [8] Björn Schuller, Stefan Steidl, Anton Batliner, The INTERSPEECH 2009 Emotion Challenge, Proceedings of INTERSPEECH 2009, pp.312-315 (2009)
- [9] 高村大也ほか, スピンモデルによる単語の感情極性

抽出, 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp.627-637, 2006

- [10] 釘宮理恵, ツンデレカルタ, インディーズメーカー, 2007
- [11] 釘宮理恵, ツンデレカルタ 2008, インディーズメーカー, 2008
- [12] メディアファクトリー, ゼロの使い魔 DVD-BOX[1-3]
- [13] ジュネオンユニバーサル, ハヤテのごとく! DVD-SET1
- [14] 釘宮理恵, ツンデレ百人一首, インディーズメーカー, 2008
- [15] C.C.Chang and C.J.Lin: LIBSVM: A Library for Support Vector Machines. (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>)