5N-01

文の意味を考慮した音声の感情認識

加藤涼太　長名優子

東京工科大学 コンピュータサイエンス学部

1　はじめに

近年、画像や音声などを対象とした感情認識に関する研究が盛んに行われている。感情認識を実現することで、人とコンピュータのコミュニケーションが円滑に行えるようになる可能性がある。

本研究では、文の意味を考慮した音声の感情認識を提案する。提案手法で認識する感情は怒り・喜び・悲しみの3つと感情を表出していない状態である平静の4つとする。これらは、従来の音声を対象とした感情認識の多くで認識の対象とされている感情である。提案手法において、音声の感情認識は、一般化調和解析 [1] にてパワースペクトルと基本周波数を求め、それらを特徴量としてサポートベクトルマシン [1] に入力することで行う。文の感情認識は、ナイーブベイズ分類で行う [2]。音声を用いたサポートベクトルマシンによる認識結果と文を用いたナイーブベイズ分類による認識結果を比較し、判定したものを最終的な認識結果とする。

2　文の意味を考慮した音声の感情認識

2.1　音声からの情報取得

提案システムでは、ユーザが PC に向かって話している状況 を想定しているため、PC に向かって話している状況で声を録音し、感情認識を行う。音声は WAV 形式で録音し、音声データ (WAV 形式、サンプリング周波数：16000Hz) を取得する。音声をテキストデータに変換するのには Google Cloud Platform が提供している Google Cloud Speech API を用いる。音声とテキストデータの認識範囲を一致させるために pydub というライブラリを用いて音声を文単位に分割し、文単位で感情認識を実現する。

2.2　音声の感情認識

提案手法では日本語のみを対象としている。日本語での発話では、感情に韻律情報が大きく影響しているため、音声の特徴として基本周波数とパワースペクトルを用いる。ここで、基本周波数とパワースペクトルは、声の高さと声の大きさに当たる。音声特徴の推定には一般化調和解析を用いる。音声での感情認識は、図1のように、0.2秒単位で一般化調和解析により基本周波数とパワースペクトルを抽出し、それらを特徴量としてサポートベクトルマシンにより行う。0.2秒ごとの認識結果の中で最も多い感情を認識結果とする。また、**2.4** で述べる最終的な感情判断に用いるデータとして、2番目に多かった感情も候補として残す。

2.3　文の感情認識

提案手法では、ナイーブベイズ分類を用いた確率的な文書分類により、音声より変換されたテキストデータの感情認識を実現する。一般的にナイーブベイズ分類は学習データを必要とするため、学習データとして、感情文と感情文ごとのタグを格納したデータベースを感情コーパスとして作成する。

2.3.1　感情コーパス

感情コーパスに用いる文は、国語研日本語ウェブコーパスから抽出する。感情に関する文を感情ごとに1000文以上、総文数が5000文以上になるように抽出する。文字数は1文につき10から30文字程度とする。音声の認識で扱う怒り・喜び・悲しみと平静の感情のタグを1文ごとに付与し、表1のようなデータベースを作成する。タグの付与を1人で行うと主観による影響が出てしまうと考えられるため、1つの文に対して複数人でタグの付与を行うようにする。

2.3.2　ナイーブベイズ分類による感情分類

ナイーブベイズ分類では、過去の事例を元に未知の文書があらかじめ与えられているどのカテゴリに属す

Recognition of Emotion from Speech considering Meaning of Sentences
Ryota Kato and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

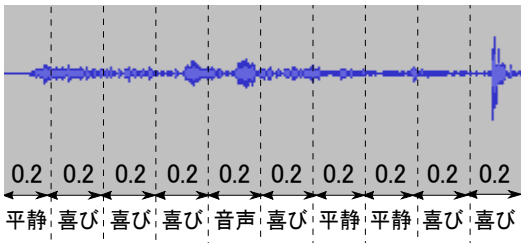

図 1: 0.2 秒ごとの認識のイメージ

表 1: 感情コーパス

	文	感情
1	なんかほのぼのしてて、好きだなあ	喜び
2	あー、腹立つ	怒り
3	よろしくお願いします	平静
4	いやあ、ほんと良かった	喜び
5	こんばんは	平静

るかを決定する。提案手法では、音声から得られたテキストデータを入力として、感情コーパスによって学習した結果を元に文に対してタグを付与する。文に対する感情の認識結果 c^w は以下のように決定される。

$$c^w = \underset{c_i}{\operatorname{argmax}} P^{c_i} \tag{1}$$

ここで、P^{c_i} はその文が感情 c_i を含んでいる可能性の高さを表しており、

$$P^{c_i} = P(c_i) \prod_{k=1} P(w_i|c_i) \tag{2}$$

で与えられる。ここで、$P(c_i)$ は感情コーパスにおいて感情 c_i のタグがつけられた文の含まれている確率、$P(w_i|c_i)$ は感情コーパスにおいて感情 c_i のタグがつけられた文に単語 w_k が含まれる確率を表している。式 (1) では、特定の感情と関連の深い単語が多く含まれていればその感情であると判断することができる。

2.4 音声と文に対する認識結果に基づく最終的な判断

2.2 と **2.3.2** で得られた認識結果に基づいて図 2 のような手順で最終的な判断を行う。音声の認識結果 c^s が平静であれば、最終的な認識結果 c を平静と判断する。また、音声の認識結果 c^s と文の認識結果 c^w が同じてあれば、その感情を認識結果とする。音声と文に対する認識結果が異なる場合には、音声による認識結果の 2 番目の候補 c_2^s と文に対する認識結果が一致しているかを調べ、一致していればその感情であるとする。一致していない場合には、文の認識における P^{c_i}

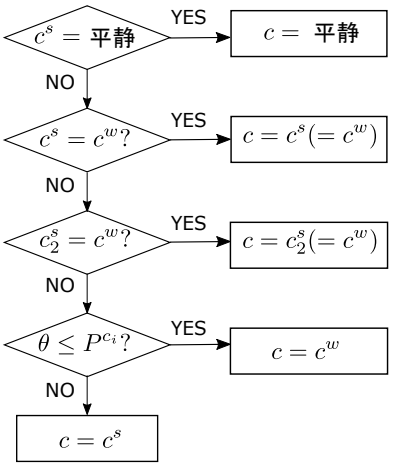

図 2: 最終的な感情判断

の値をしきい値 θ と比較し、しきい値以上ならば文に対する認識結果の感情であると判断する。しきい値以下であれば音声に対する認識結果の感情であると判断する。

提案手法では、文の意味を考慮した際に特定の感情と深く関わっている場合でも、音声が平静であると判断された場合には平静であると考えるのが自然であると考えたため、音声が平静であると判断された場合にはその認識結果を重視するものとしている。音声と文の認識結果が一致していればそれを認識結果として問題ないと考えられる。音声と文の認識結果が一致していない場合は、音声の認識結果の 2 番目の候補も含めて再検討することで、正しく感情を認識できる可能性が高まると考えている。それでも一致しない場合には、文の認識結果が信頼性が高いかどうかを調べ、信頼性が高い場合には文の認識結果を優先するものとしている。

3 計算機実験

計算機実験を行い、提案手法において感情認識が行えること、文に対する感情認識の結果を考慮することで認識精度が向上することを確認した。

参考文献

[1] 小谷野 信一, 長名 優子 : “表情と音声を用いたサポートベクトルマシンによる感情認識,” 第 76 回全国大会講演論文集, 2014.

[2] 山本 麻由, 土屋 誠司, 黒岩 眞吾, 任 福継 : “感情コーパス構築のための文中の語に基く感情分類手法,” 電子情報通信学会技術研究報告, NLC107-158, pp.31–35, 2007.