

# 動的パラメータ調整を用いた感情極性システム

山下 優衣<sup>†</sup> 小野寺 俊<sup>‡</sup> 田谷 昭仁<sup>†</sup> 戸辺 義人<sup>†</sup>

青山学院大学理工学部情報テクノロジー学科<sup>†</sup>

青山学院大学大学院理工学研究科理工学専攻知能情報コース<sup>‡</sup>

## 1. はじめに

近年、ストレスのある生活に苦しんでいる人が多くおり、その多くは自分の無意識のうちに精神的な健康状態が悪くなっている。さらに、結果としてうつ病などの病気を引き起こすこともあるため、身体面だけでなく、精神面の健康管理が懸念されている。また、健康管理を行う際、音声を用いた感情認識を利用することがあるが、この時クラウドソーシングを用いて不特定多数の人からの評価を得ることで、その評価を利用し感情認識の精度を高めることができると考える。これらの多数の評価を反映するために、パラメータの動的調整を行った。本稿では、設計、実装、評価結果について述べる。

## 2. 関連研究

クラウドソーシングを用いて顔の表情を認識する研究がある<sup>1)</sup>。少数の専門家の専門知識ではなく、大規模な一般集団の判断に基づいた社会的なイメージより、表情の自己認識を可能としている。また、うつ病患者へ向けた携帯電話のアプリケーションとサポートアーキテクチャを開発した研究がある<sup>2)</sup>。このアーキテクチャでは、センサ値に基づき患者の気分や感情、認知・動機付け状態、活動、環境コンテキスト、社会的コンテキストの予測を行った。

## 3. 設計

### 3.1 SONATA

我々は先の研究において、音響的特徴と言語的特徴を組み合わせた感情認識システム SONATA を開発した<sup>3)</sup>。SONATA では、短い発話文を録音し、その音声を2つの特徴量に基づき解析することでスコアを算出する。オープンソフトウェア openSMILE<sup>4)</sup> で取得できる音響的特徴量から算出したスコアを音響スコア $P_s$ 、単語感情極性<sup>5)</sup>から算出したスコアを構文スコア $P_t$ と定義し、重みを $w(0 \leq w \leq 1)$ としたとき、重み付け手法により総合的評価 $P$ は以下のように定義す

る。

$$P = wP_s + (1 - w)P_t \quad (1)$$

$P_s$ ,  $P_t$ ,  $P$ は-1に近いほどネガティブ寄り、+1に近いほどポジティブ寄りであるとする。本研究では、固定値であった重み $w$ を複数の評価により適切な重みへ変更する。次節からその方法について述べる。

### 3.2 評価値による重みの決定

録音された $n$ 種類の発話音声を評価者に評価をしてもらう。 $i$ 番目の評価者からの評価値を $P_{Ei}$ 、音響スコアおよび構文スコアを $P_{si}$ ,  $P_{ti}$ と表したとき、 $P_{Ei}$ から適切な重みを求める必要がある。評価値の2乗の和 $d$ を最小化する重み $w$ を $w_{eval}$ と表し、これを評価者からの重みとする。

$$d = \sum_{i=1}^n (wP_{si} + (1 - w)P_{ti} - P_{Ei})^2 \quad (2)$$

### 3.3 重みの更新

評価者ごとに算出した $w_{eval}$ を利用し、評価者 $m$ 人による最適な重み $W$ を求めるアルゴリズムを図1に示す。このアルゴリズムは、評価者が増える度に、今までに求めてきた重み $W$ に新しい評価 $w_{eval}$ を加える計算であり、どちらを重視するのかを重み係数 $\alpha(0 \leq \alpha \leq 1)$ によって決定する。

```
for  $i$  in range( $m$ )
```

```
if 最初の評価者
```

```
 $W_0 = w_{eval_0}$ 
```

```
else
```

```
 $W_i = (1 - \alpha)W_{i-1} + \alpha w_{eval_i}$ 
```

図1. 重み更新アルゴリズム

このアルゴリズムにより求めた $W$ を用い、式(1)から音声ごとに新たにシステムのスコア $P_{new}$ を算出する。 $P_s$ ,  $P_t$ は SONATA によって算出されたスコアを用いる。

$$P_{new} = WP_s + (1 - W)P_t \quad (3)$$

## 4. 実装

本研究の評価実験において、評価者に音声の

ラベル付けとして音声ごとに得点を付けてもらう。評価実験については次章で述べる。評価値を得るためのシステムとして Web アプリケーションを実装した。実行画面を図2に表す。本アプリケーションを用いることで、評価値がサーバ上に保存できるようにした。

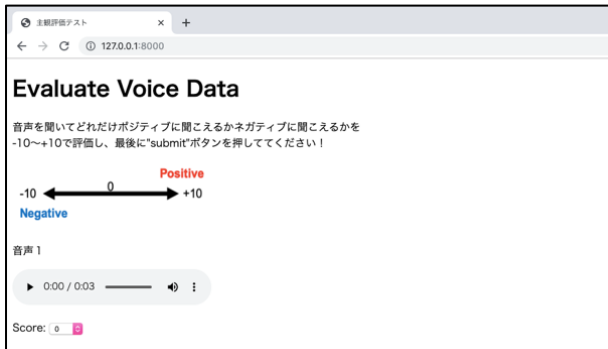


図2. Webアプリケーション実行画面

## 5. 評価

### 5.1 予備実験

評価者に評価を行ってもらう音声データを4種類の発話内容、男女計3名の発話者から8種類を用意した。発話内容と発話者について表1にまとめる。

表1. 各音声データの内容

	発話内容	発話者
T <sub>1_1</sub>	仕事で失敗してしまった	女性①
T <sub>1_2</sub>		男性
T <sub>2_1</sub>	成果を褒められた	女性①
T <sub>2_2</sub>		男性
T <sub>3_1</sub>	週末は趣味を楽しめた	女性①
T <sub>3_2</sub>		女性②
T <sub>4_1</sub>	人前での発表で緊張した	女性①
T <sub>4_2</sub>		女性②

### 5.2 評価実験

表1に表した8種類の音声データを7名の評価者に-10から+10の範囲でスコアを付けてもらい評価を行った。この評価値 $P_E$ はシステムから算出したスコア $P_s$ ,  $P_t$ ,  $P$ と範囲を統一するために $1/10$ にし、 $-1 \leq P_E \leq 1$ とする。

### 5.3 実験結果

評価値 $P_E$ および式(3)より求めた $P_{new}$ を用いて平均二乗誤差を求めることで本研究の評価を行う。図1のアルゴリズムによる計算において $\alpha$ を0.1刻みとし、それぞれの $\alpha$ に対し $W$ ,  $P_{new}$ の計算をする。 $\alpha$ ごとに式(4)に示す平均二乗誤差MSEを算出し、その結果を図3のグラフにまとめる。本実験では音声データ数 $n = 8$ , 評価者数

$m = 7$ とする。

$$MSE = \sum_{i=1}^n \sum_{j=1}^m (P_{new_i} - P_{E_{ij}})^2 \quad (4)$$

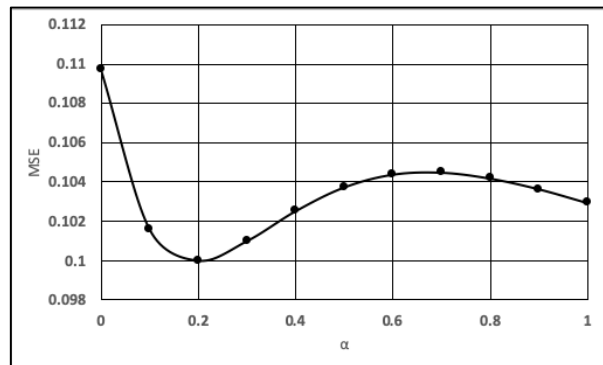


図3. 評価者とシステムの比較

計算結果から、 $\alpha = 0.2$ のとき評価者による評価とシステムによる計算結果との平均二乗誤差の最小値が0.100になるという結果が得られた。

## 6. むすび

本研究では、音声から感情認識をする際の音響的特徴量および言語的特徴量に対する最適な重みを複数の評価に基づいて動的に変更することを試みた。本研究の結果より元の重みに新しい評価値を少しずつの割合で加えていくことで、より人の認識にシステムの評価が近付くことがわかった。クラウドソーシングを用い評価を得ることで、新しい評価を過去の評価に反映させることができ、感情認識の精度を向上させることができると考えられる。

## 参考文献

- 1) S. Li, and W. Deng, “Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition”, IEEE, 2019.
- 2) M. N. Burns, M. Begale, J. Duffecy, D. Gergle, C. Karr, E. Giagrande, and D.C. Mohr, “Harnessing Context Sensing to Develop a Mobile Intervention for Depression”, J. of Medical Internet Research, 2011
- 3) 岡崎晴, 米岡良, 高橋淳二, 戸辺義人, “SONATA: 音声解析と構文解析を組み合わせた感情認識システム”, 電子情報通信学会東京支部学生研究会, 2017.
- 4) F. Eyben, M. Wöllmer, B. Schuller, “openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor”, ACM Multimedia Conference – MM, pp.1459-1462, 2010.
- 5) 高村大也, 乾孝司, 奥村学, “スピンモデルによる単語の感情極性抽出”, 情報処理学会, vol.47, No.2, 2006.