

## サポートベクター回帰による自発音声の感情次元推定\*

永岡篤, 森大毅 (宇都宮大)

## 1 はじめに

音声伝達するパラ言語情報の研究には音声対話コーパスの利用が必要不可欠である。大部分の研究は単一のコーパスを対象としている。多数のコーパスを併用することで、様々なシチュエーションでのパラ言語的現象を包括的に研究の対象とすることができる。しかしながら、パラ言語情報のラベルを有するコーパスは数少ない。また、パラ言語情報の評価方法はコーパスによって異なりこのままでは併用することが不可能である。

コストが大きい人手ラベリングを避けつつパラ言語情報ラベルを共通化するためには、パラ言語情報ラベルを有するコーパスを基にした別のコーパスに対する自動ラベリングを相互に行うことが有効であると考えられる。本論文では、複数コーパスに対する相互ラベリングを目的とした、機械学習によるパラ言語情報の自動推定の有効性を検証した。

## 2 パラ言語情報自動ラベリング

まず本論文では、宇都宮大学パラ言語情報研究向け音声対話データベース(UUDB)[1]に付与されているパラ言語情報と同様のラベルを他の音声対話コーパスに付与することを考える。

UUDBには、パラ言語情報ラベルとして以下に示す6軸の感情次元による評価が付与されている。評価値は1から7まで(4が中立)の数値である。

- 1) 快-不快
- 2) 覚醒-睡眠
- 3) 支配-服従
- 4) 信頼-不信
- 5) 関心-無関心
- 6) 肯定的-否定的

この評価方法に基づくパラ言語情報の自動ラベリングは、音声の音響特徴量から各感情次元の連続的な評価値を推定する回帰問題となる。この自動ラベリングの枠組みを Fig. 1 に示す。他コーパスへの適用に先立ち、今回の実験では正

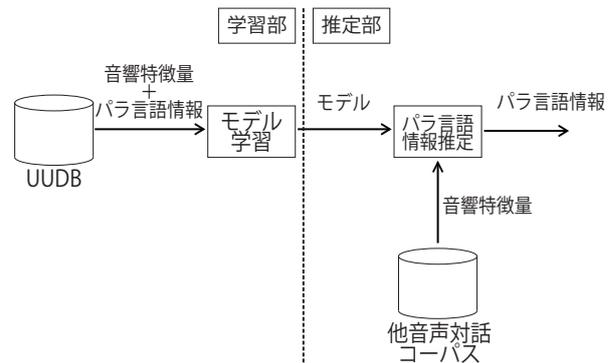


Fig. 1 自動ラベリング方法

解が分かっている UUDB のデータを用いて感情次元の推定を行った。

## 3 感情次元推定実験

## 3.1 実験条件

感情次元を推定するために使用したデータは UUDB に収録されている女性話者 12 名の全発話、計 3706 発話である。各発話に対して Interspeech 2010 Paralinguistic Challenge ベースラインシステム [2] と同様の 1581 次元の音響特徴量を抽出した。そして、UUDB に付与されている評価者 3 名の平均評価値を目的変数、抽出した音響特徴量を説明変数とする回帰モデルをサポートベクター回帰 [3] によって学習した。カーネル関数にはガウシアンカーネルを用いた。使用した統計ソフトは R の kernlab パッケージである。得られた不特定話者のモデルは leave-one-speaker-out 交差検証によって評価した。

## 3.2 感情次元推定結果および考察

各感情での平均評価値と推定値の相関係数を Table 1 に示し、その相関図を Fig. 2 に示す。また、平均評価値と推定値の RMS 誤差を Table 2 に示す。加えて、Table 1, Table 2 には [1] で報告されている相関係数と RMS 誤差も示している。なお [1] で使用された特徴量は  $f_0$ , 強度, 非周期性指標などの 6 次元であり、推定には重回帰モデルが用いられている。

Table 1, Table 2 を見ると、本研究の手法を用いることで全体的に [1] より良い精度が得られ

\* Emotion dimension estimation for spontaneous speech by support vector regression, by NAGAOKA, Atsushi, MORI, Hiroki (Utsunomiya University).

感情次元	相関係数 (本研究)	相関係数 [1]
快-不快	0.65	0.57
覚醒-睡眠	0.85	0.85
支配-服従	0.85	0.80
信頼-不信	0.53	0.51
関心-無関心	0.74	0.70
肯定的-否定的	0.50	0.49

感情次元	RMS誤差 (本研究)	RMS誤差 [1]
快-不快	0.58	0.63
覚醒-睡眠	0.53	0.57
支配-服従	0.62	0.74
信頼-不信	0.63	0.67
関心-無関心	0.57	0.61
肯定的-否定的	0.64	0.67

ている。特に快-不快では精度が顕著に上昇した。覚醒-睡眠の相関係数は変化がなかったが、RMS誤差は減少しているためモデルの精度は向上している。よって、本論文の手法は有効であると考えられる。Fig. 2より快-不快と信頼-不信、肯定的-否定的の3つの次元は推定値にバイアスがかかっていることが分かる。また、支配-服従と関心-無関心の2つの次元では、平均評価値が小さい発話は推定値のレンジが大きく推定が不安定であることが分かる。しかし、覚醒-睡眠では平均評価値と推定値のほとんどがグラフの対角線近傍に分布していることから平均評価値が小さい発話も比較的安定した推定ができていると言える。

ところが、覚醒-睡眠の推定においても平均評価値と推定値の差が大きい発話がある。それらには、以下に示すような傾向があることが分かった。

- 1) 相づちの発話が多い
- 2) 発話長が極端に短い発話が多い
- 3) クリーキーな発話や氣息性の発話が多い

これらの発話は、適切に音響特徴量抽出ができず、そのために感情推定が不安定になっている。または、本実験で抽出した音響特徴量にはこれらの発話が伝達するパラ言語情報を精度良く推定できるような特徴が含まれていないことが考えられる。よって、他のコーパスに対して推定を行う場合、高精度な自動推定が困難であることが予想できる発話は人手でラベリングする方法が有効だと考えられる。

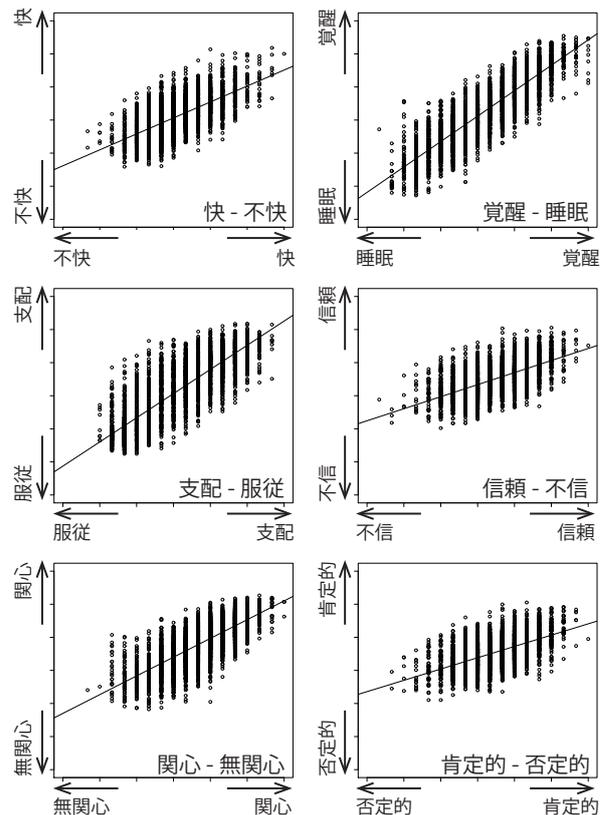


Fig. 2 各感情次元推定の相関図 (横軸: UUDBに付与された平均評価値, 縦軸: 推定値)

#### 4 おわりに

本論文ではパラ言語情報ラベルを有するコーパスを基にした別のコーパスへの相互自動ラベリングを目的として、サポートベクター回帰を用いたUUDBに対する感情次元の推定を行い、先行研究よりも高い精度を得た。

今後は、UUDB以外の既存コーパスの一部に対して人手ラベリングを実施し、自動ラベリングの有効性を検証する。また、既存コーパスが独自に持つパラ言語情報ラベルの有効利用も興味深い課題である。

謝辞 本研究の一部は立石科学技術振興財団研究助成 (課題名: 言語情報と音響情報の統合的利用による感情音声コーパスの大規模化) およびJSPS 科研費 (26280100) の支援を受けた。

#### 参考文献

- [1] Mori *et al.*, Speech Communication, Vol. 53, pp. 36–50, 2011.
- [2] Schuller *et al.*, in Proc. Interspeech 2010, pp. 2794–2797, 2010.
- [3] C. M. Bishop, 元田他監訳, “パターン認識と機械学習 下,” 丸善出版, 2006.