

## 議論の場の空気推定のための韻律情報の分析手法

幸浦 弘昂 白松 俊

名古屋工業大学 名古屋工業大学

## 1. はじめに

近年スマートフォンなどの普及に伴い、多くの人が Web 上で議論に参加できる時代になった。多数の議論参加者による意見や発言が増えることで、合意形成、相互理解の促進がこれまで以上に困難になる。このとき、議論の流れを整理したり、発言や参加を促すファシリテータが重要になると考えられる。ただし、人間のファシリテータにとって Web 上の多数の発言を扱うことは認知的負荷が高い。そこで本研究では、自律的なファシリテーション機構の実現によりファシリテータの負担を軽減できると考えた。

まず、ファシリテータは参加者のノンバーバルな振る舞いやパラ言語から場の空気を読んでいると考えられる。しかし、Web 上の議論ではこれらを観測できないため、まずは対面の議論で分析を行い、得られた知見を将来的に Web 議論にも応用する。本稿では、特に参加者発言の音声(特に韻律情報)に着目する。具体的には、議論のマルチモーダルコーパスを構築し、そのコーパス中の音声データを分析し、特に韻律情報を用いることで、場の空気を読むような機構を実現する。

## 2. 先行研究

石井ら[1]は、韻律パラメータとして F0move と発話の持続時間を用い、韻律パラメータによる発話行為の分布を示した。その結果、肯定・同意・相槌の際は「短下降型」、聞き返しは「短上昇調」、疑いは「動きの幅が広い短上昇調」、考え中・戸惑いは「平坦、長下降調」、否定的または曖昧な表現は「長上昇調、長平坦調」であり、これらの発話行為の識別には有効的であることを示した。

平野の研究[2]によると、音声入力→pitch, power 抽出→主要 pitch の抽出→特徴量の分析→感情の分類というように感情認識システムを構築した。特徴量分析において、pitch, power の値から平均値・最大値・最小値を求め、感情による変化傾向「喜び」、「怒り」、「悲しみ」の発話がそれぞれ「普通」と比較してどの程度変化しているのかを分析した。その結果、「喜び」では特徴量が高く、「怒り」では平均 pitch は低く平均

power は高くなり、「悲しみ」では最小 pitch は高く最大 power は低くなるということを示した。

白松ら[3]は、いくつかのパラメータの分布により、議論の場の空気のモデル化を試みている。その中のパラメータの一部に

- 発言  $u$  の発話行為の種別  $\text{spact}(u)$
- $u$  に現れている感情の種別  $\text{senti}(u)$
- 発話の遷移確率  $p(a_2 | a_1)$

があり、本稿では特にこの 3 つのパラメータに着目し、場の空気のモデル化を試みる。

## 3. 分析手法

図 1: システム構成図

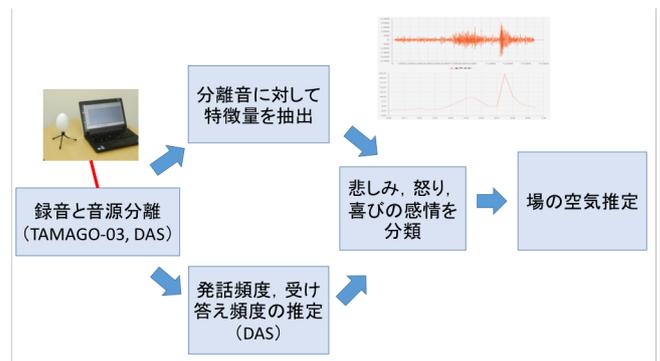


図 1 に本研究の大まかなシステム構成を示す。音声の録音は TAMAGO-03 という 8ch マイクを用いて行う。会議の場では参加者の発言が重なるバージョンが起りうるため、音源分離を行う必要があった。そのため録音した音声に対して HYLABLE 社の議論評価サービス (DAS) を用いて実験を行う。DAS では音源分離、発話頻度、発話のターンテイクの推定などを行うことができる。分離された音声に対し、Pitch, Power などの特徴量の抽出を行う。Pitch の推定は、REAPER を用いて得られた F0 を用いて推定し、Power の推定には短時間フーリエ変換を行い、推定する。

下の図 2, 図 3 はそれぞれ DAS を用いて得られたターンテイク図と、短時間フーリエ変換によって得られた時間変化ごとのパワーの推移グラフである。図 2 は発話の遷移確率  $p(a_2 | a_1)$  に対応する。図 3 で得られた Power や F0 推定値を特徴量として用い、SVM (SupportVectorMachine) を用いた機械学習により場の空気

Analysis method of prosodic information for atmosphere estimation in discussion space

Koura Hiroaki · Nagoya Institute of Technology

Shun Shiramatsu · Nagoya Institute of Technology

の推定を行う。本稿ではデータマイニングツール Weka を用いて、SVM を使用する。学習データ作成の際に感情推定、場の空気推定の指標として、先行研究で紹介したものを参考にする。

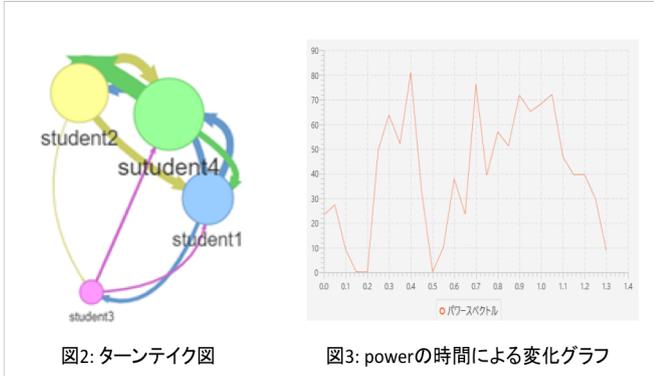


図2: ターンテイク図

図3: powerの時間による変化グラフ

#### 4. 評価実験の設定

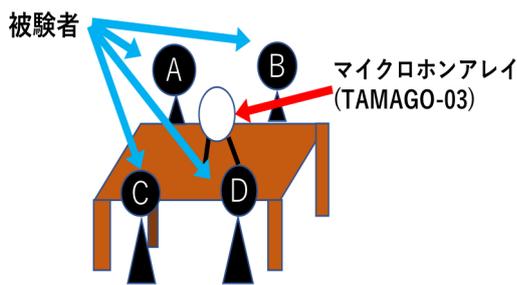


図4: 評価実験における対面の議論の録音

まず議論の話題について検討する。空気の悪い議論の設定として、時間内に他の人より良い意見を発言すると罰ゲームを免れ、その他の人は罰ゲームを受ける。つまり、それぞれが、他人の意見を潰しつつ自分がより良い意見を発言しようとする事で場の空気は悪くなると考えた。空気の悪い議論の設定として、時間内に全体として良い意見が出ればご褒美がでる。つまり、全体で協力し合い、良い意見を出す事で場の空気は良くなると考えた。

次に場の空気の推定実験の手順として、まず図4のように四人で円状に座りその真ん中にマイクを置き四人の会話を録音し、その音声を分離する。分離した音声に対して韻律情報などの分析を行い、その際に抽出される特徴量を元に一人一人の感情を推測し、場の空気を推測する。ここで感情推定、場の空気推定の指標として、先行研究で紹介したものを参考にしたい。具体的には、平野らの研究[2]から場の空気はが良いときは、それぞれの特徴量が高くなり「喜び」の感情が多く検出されると予測される。逆に場の空気が悪い場合には、平均 pitch は低く

平均 power は高くなり、最小 pitch は高く最大 power は低くなる。つまり、「怒り」、「悲しみ」の感情がそれぞれ多く検出されると考えられる。

#### 5. 予備実験結果

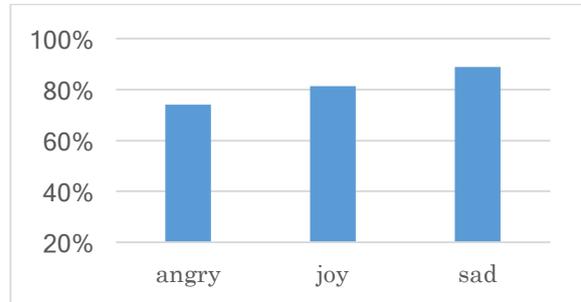


図5: SVM による感情分類の精度

取得した 27 個の分離音に対して、Weka の SVM で 10 交差検証を用いて教師あり学習を行い、「怒り」、「悲しみ」、「喜び」の感情を分類した結果が図4である。正解率として、「怒り」74%、「喜び」81.5%、「悲しみ」88.8%という精度になった。ただしサンプル数が少ないため今後はもう少し多くのサンプルを取得し、検証する。

#### 5. まとめと今後の展望

今回学習に用いたデータが 27 と非常に少なく、正確な精度が得られていない可能性があるため、評価実験を重ね、もっと多くのデータを取得する必要がある。また、上記の特徴量のうちどれが場の空気推定に有効であるのかを明らかにしていく予定である。新たな特徴量を加えていくことも検討している。

また、ファシリテーションの専門家からの意見を参考に、場の空気を読むプロセスの一環としてグループワーク中に各自の持ち時間をカウントして教えてくれたり、発言回数の少ない人を教えてくれる、いわばグループワークの時間を調整してくれるようなファシリテータ支援エージェントを開発していく予定である。

**謝辞** 本研究は、JST CREST (No. JPMJCR15E1) , JSPS 科研費 (No. 17K00461) の支援を受けたものです。DAS について助言を頂いた株式会社ハイラブル 水本武志氏、ファシリテーション支援に関する助言を頂いた株式会社ソーシャル・アクティ林加代子氏に深謝します。

#### 文 献

- [1] 石井他: 対話音声における韻律と音声の特徴を利用したパラ言語情報の抽出の検討. 第22回AIチャレンジ研究会, 2005
- [2] 平野: 音声からの感情認識を目的とする特徴量の検討. 岩手県立大学 感情情報学講座卒業論文, 2007
- [3] 白松俊, 北川昂, 幸浦弘昂, 熊崎滉大. “議論参加者の行動センシングに基づく場の空気の自動推定に向けた検討”, 人工知能学会 第3回市民共創知研究会「みらいらば なごや」予稿集, pp. 30-33, SIG-CCI-003-06, 2017.