

## RealSense カメラを用いた議論の場の空気の推定手法の検討

熊崎 滉大<sup>†</sup> 白松 俊<sup>‡</sup>名古屋工業大学 工学部 情報工学科<sup>†</sup> 名古屋工業大学 大学院工学研究科 情報工学専攻<sup>‡</sup>

## 1. はじめに

近年スマートフォンやタブレット端末の急速な普及に伴い Web 上で意見を述べるのが一般的になっている。政治や趣味など様々なことで議論がなされている。しかし、Web 上の議論では参加者が各々の意見を主張し続けるのみで合意形成には至らない。対面の議論では合意形成を助ける目的でファシリテーターが付く場合が多い。Web 上では特に参加者を増やしようので対面の議論よりもファシリテーターが重要になる。しかし、参加者が増えるにつれファシリテーターの認知的負荷が大きくなる。そこでファシリテーターを支援する技術が必要であると考えた。本研究では、ファシリテーターエージェントの実現を目指す。

では、ファシリテーターにはどんな要素が必要であるか。柳原[1]によると、中立な立場に立つこと。状況への感受性が豊かであること。先走らないこと。失敗を恐れないこと。援助的であること。以上の要素が必要であると言われている。ファシリテーターエージェント作成においてこれらの要素全てを満たす必要がある。この中でも状況への感受性というところに注目し、場の空気を読む機構の実現を考えていく。先行研究により、Web 議論のテキストデータのみからその議論の「場の空気」を把握することが困難であると分かった。Web 議論での空気推定が目的ではあるが本研究では、対面の議論にも目を向けて、議論における場の空気の推定を検討する。人間は無意識に場の空気を感じ取っている。それを自動化するために視線や表情などの顔から得られる情報が重要であると考えた。センサーカメラを用い、対面の議論参加者の行動をセン

シングし、マルチモーダルコーパスを作成する。対面の議論で場の空気が推定できるようになれば、言語表現と場の空気の関係が明らかになり、言語表現のみで場の空気を推定する手法を確立できると期待する。

## 2. 場の空気のモデル化

場の空気の推定には場の空気の定義が必要になる。そこで場の空気を表す数理モデル化する。先行研究として白松ら[2]がモデル化を試みている。いくつかモデル式が存在するが本稿では、

- 時刻  $t$  時点での参加者の  $a$  の感情極性:  $\text{spol}(t) \in [-1, 1]$ .
- $t$  時点で  $a_1$  の  $a_2$  に対する感情極性:  $\text{spol}(a_1, a_2, t) \in [-1, 1]$ .
- 発言  $u$  に対する  $a$  の傾聴の度合:  $\text{lisn}(a, u) \in [-1, 1]$ .

これらのモデル式を取り扱う。議論中に誰が誰を向いているのかが分かれば  $\text{spol}(a_1, a_2, t)$ ,  $\text{lisn}(a, u)$  の推定に役立つと考える。議論参加者の表情から感情を推定できれば、 $\text{spol}(t)$ ,  $\text{spol}(a_1, a_2, t)$  の推定に役立つと考えている。

## 3. RealSense カメラを用いた注視対象者の推定

対面の議論の行動センシングに Intel 社の RealSense カメラという 3D モーションセンサーカメラを用いる。顔の検出機能が搭載されており、顔の姿勢情報 (pose, pitch, yaw), 顔の表出情報 (目の動き, 口の動き等) を取得できる。これらの機能から被験者の顔の向きを自動判定し、議論のターンテイクも合わせることで話者の方向を向いて話を聞いていたか知ることができ、その議論への参加度合いを推定できると考える。

まず、RealSense カメラの利用可能性を図るため、予備実験として Misrepresentation Game[3] と呼ばれる騙し合いゲームを被験者 4 名で行った。被験者 1 名の前にカメラを置きセンシングを行った。本予備実験では、姿勢情報、表出情報 (目の動き) の取得の検証を目的とした。騙

Considering Method for Estimating Atmosphere of Debate using RealSense camera

Koudai Kumazaki, Shun Shiramatsu

<sup>†</sup>Department of Computer Science,  
Nagoya Institute of Technology

<sup>‡</sup>Department of Computer Science, Graduate  
School of Engineering, Nagoya Institute of  
Technology

し合いゲームを選択した理由として、疑い、対立などが顕著に現れ、顔を覗き込むような行動が起きると考えたからだ。予備実験を実施し、姿勢情報は実用的な精度で得ることができた。

しかし、表出情報（目の動き）に関しては満足な結果が得られなかった。そのため顔の向きの自動判定には姿勢情報のみを用いることとする。更に、カメラから被験者までの距離によって得られるパラメータの大きさの幅に変化があることを確認した。離れるほど被験者は顔をあまり動かさなくとも周りを見渡せてしまうことが原因であると考えられる。もう一つパラメータに影響を与える要因として机の大きさも関係してくると思う。実際に実験を行い、データを収集する際は机の大きさ、カメラからの距離を最初に設定することで正確に誰が誰を向いているのか判定する必要がある。初期設定画面の図を以下に示す。手前の人物のどちらが自分であるか選択し、机の長さを入力、どのあたりに他の人がいるかは角度を示す紙を自分の前に置きある程度の角度を決めてもらい入力する。カメラからの距離は RealSense の機能により得られるためそれを用いる。

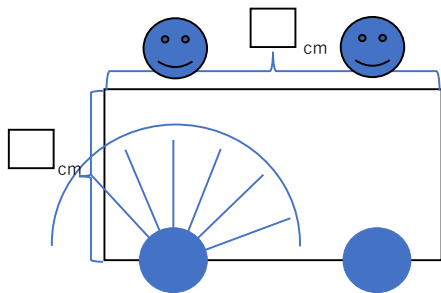


図 1：初期設定画面のモックアップ

#### 4. 表情による感情の推定

場の空気を読む時、表情は大事な要素になっていると思う。表情からその人の感情を考えていると思う。そこで Microsoft 社の EmotionAPI を用いて議論参加者の表情から感情推定を行い、その情報もコーパスに組み込むことを考えた。EmotionAPI とは、Microsoft 社が提供する WebAPI で、画像を与えると感情の極性が数値として返ってくる仕組みになっている。得られる感情として、怒り、軽蔑、嫌悪感、恐怖、喜び、中立、悲しみ、驚きの 8 つがある。

RealSense カメラでセンシングを行う際、数クロックに一度画像を保存しておき、その画像に対して感情推定を行うことで時間推移に対する感情の変化を得ることができる。実際に

EmotionAPI を使用した際の結果を以下に示す。



図 2：EmotionAPI の使用結果

画像を与えるとそれぞれの感情の信頼度を返す。上の画像は驚いた時、下の画像は無心であることを想定して撮影した。結果として上の画像では surprise の信頼度が一番高く、下の画像では neutral の信頼度が一番高いことが確認できた。

この感情推定と前節で述べた注視対象推定とを組み合わせることで  $spol(t)$ ,  $spol(a_1, a_2, t)$ ,  $lispn(u)$  の推定に利用できる可能性がある。

#### 5. まとめ

本稿では、顔から得られる情報から議論の場の空気の推定手法の提案を行った。現段階では空気の推定までは至っていないが、実験を重ねることでデータを集めていく。それらのデータから数理モデルを正確な物にしていきたい。また、同研究室で音声、脳波の方面からも空気推定について考えているので今後それらと共同して手法の確立を目指していく。

#### 謝辞

本研究は、JST CREST (No. JPMJCR15E1), JSPS 科研費 (No. 17K00461) の支援を受けた。

#### 参考文献

- [1]柳原:Creative0.D.～人間のための組織開発シリーズ～, Vol. 1, プレスタイム, 2003
- [2]白松ほか: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「みらいらぼ なごや」予稿集, pp. 30-33, SIG-CCI-003-06, 2017
- [3]Jonathan Gratch, Zahara Nazari, and Emmanuel Johnson. The misrepresentation game: How to win at negotiation while seeming like a nice guy. In proceedings of th 2016 International Conference on Auto-nomous Agents & Multiagent Systems, pp. 729-737, 2016