

ガウス混合モデルを用いた Human-Agent Interaction における ユーザの表情分析

Facial Expression Analysis using Gaussian Mixture Model in Human-Agent Interaction

坂戸 達陽[†] 曾 傑[‡] 中野 有紀子[†]
Tatsuya Sakato Jie Zeng Yukiko Nakano

1. はじめに

ユーザが自由に離脱できる情報検索、推薦対話システムの問題点として、ユーザがインタラクションの途中で離脱してしまうということが挙げられる。

ユーザはシステムとのインタラクション中、意識的、あるいは無意識的にユーザ自身の状態に基づく様々な非言語情報を発信している。システムがこれらの非言語情報からユーザの状態を推定し、振る舞いを適応的に変化させることであれば、ユーザとのインタラクションをより維持できるようになることが期待できる。

本研究ではユーザが発信する様々な情報の中から、ユーザの表情に着目し、インタラクション中のユーザの表情から、ユーザの状態を推定することを目指す。具体的には、観光案内を想定したヒューマン - エージェントインタラクション中に観測されるユーザの表情をパターン分類し、観測される表情パターンとユーザの状態との関係性を明らかにすることを目指す。

2. 手法

2.1 OpenFace[1]

OpenFace とは画像や動画を分析し、画像あるいは動画中の人物の視線の方向や、ランドマークと呼ばれる顔の特定部位の座標、顔の筋肉群の特定の動作を表す Action Unit (AU) [2]などの情報を抽出することのできるツールキットである。

本研究では OpenFace によって検出された AU の情報を表情の観測データとして用いた。

2.2 Wizard of oZ 法

Wizard of oZ (WoZ) 法とは、システムのふりをした人間 (wizard) がユーザとインタラクションを行う手法である。ユーザはシステムを相手にしていると思いながらインタラクションを行うため、システムの一部が未実装でも、実際のシステムに近い状況での対話データを収集することができる。本研究ではシステムとユーザのインタラクションデータを収集するために WoZ 法を用いた。

2.3 ガウス混合モデル

ガウス混合モデル (Gaussian mixture model; GMM[3]) とは、各カテゴリに正規分布を仮定した混合モデルである。

[†] 成蹊大学理工学部 Faculty of Science and Technology, Seikei University

[‡] 成蹊大学大学院理工学研究科 Graduate School of Science and Technology, Seikei University



図 1 エージェントシステム

観測されたデータを用いてパラメータ推定を行うことにより、観測されたデータからパターンを学習することができる。また、学習したモデルを用いて観測データのクラスタリングを行うことができる。

本研究では WoZ 法を用いたヒューマン - エージェントインタラクション中に観測されたユーザの顔表情から OpenFace によって抽出された AU の情報を観測データとしてガウス混合モデルのパラメータ推定を行い、モデルに観測された表情のパターンを学習させた。観測された表情データを学習されたモデルを用いたクラスタリングによって表情パターンとして分類し、分類された表情パターンを分析した。

3. 実験

3.1 インタラクションデータの収集

実験に用いるエージェントシステムを図 1 に示す。ユーザは PC の画面に正対し、画面内のエージェントとユーザは観光案内を想定したインタラクションを行う。エージェントはユーザにおすすめの観光スポットを提示し、ユーザは提示された観光スポットに関して質問を行う。エージェントは実験者が操作し、ユーザの質問を理解したうえで、質問に対する応答として、エージェントの言語、非言語行動を選択する。ユーザはエージェントを実験者が操作していることは知らされておらず、WoZ 法を用いてエージェントシステムとユーザのインタラクションデータの収集を行った。

実験の流れは以下の通りである。

1. エージェントがある観光スポットについてのコンテンツを 1 つ提示して説明する。
2. ユーザが質問し、エージェントが質問に回答する。
3. 質問が終了するまで 2. を繰り返す。
4. 3 つのコンテンツについて 1. から 3. を繰り返す。

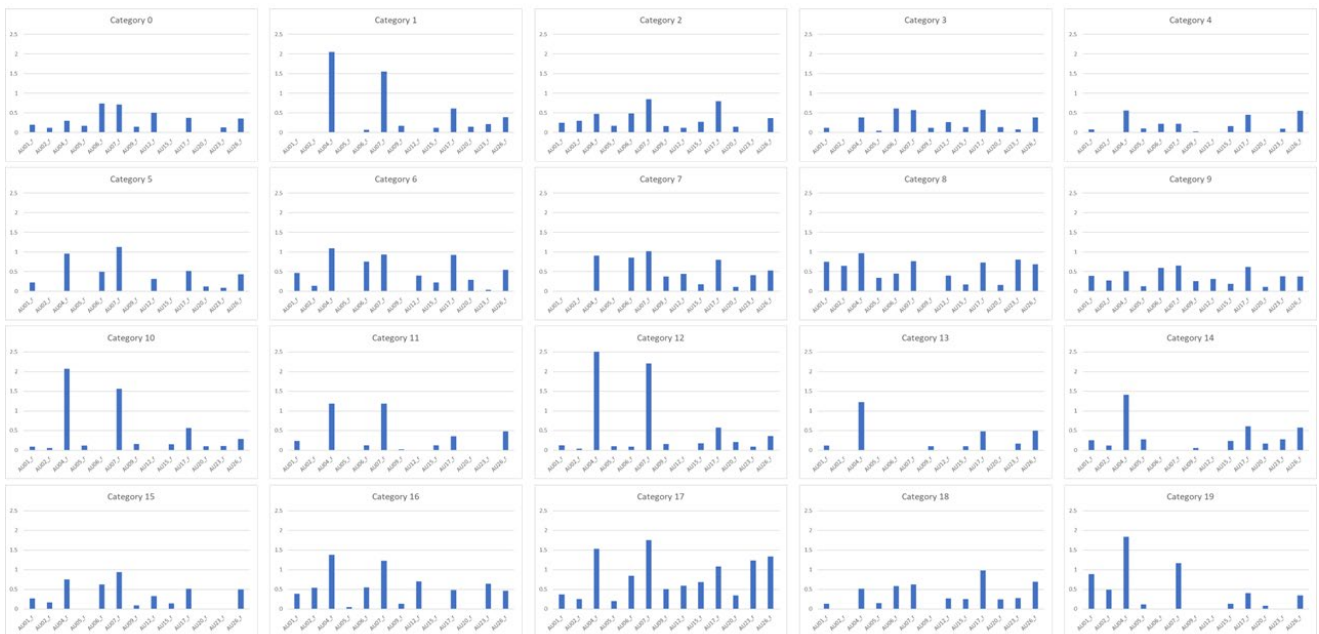


図 2 GMM の学習結果

5. 提示したコンテンツをまとめて表示して、ユーザーに1つ選択してもらう。
6. 3つの観光スポットについて1.から5.を繰り返す。
7. 選択された3つのコンテンツをまとめて表示し、ユーザーに最終選択をしてもらう。

実験参加者は21名で、20代から40代の男性10名、女性11名であった。

紹介する観光スポットは、大阪、仙台、博多の3つの中からランダムな順番で選択される。

各観光スポットを紹介する際には、以下の3種類の振る舞いが観光スポットごとにランダムな順番で選択され、実験者は選択された振る舞いに基づいてエージェントを操作した。

- エージェント A：共感反応なし
エージェントは無表情で、ユーザーの行動で振る舞いは変化しない。
- エージェント B：理解反応
エージェントは実験参加者からの質問に対し、考え込むしぐさや理解するしぐさ（うなずきなど）を行う。
- エージェント C：理解反応 + 感情同調
エージェントはエージェント B の理解反応に加え、ユーザーの感情に寄り添う感情を示す。

インタラクシオン中のユーザーの振る舞いを正面から映像と音声で記録した。映像は30フレーム毎秒で記録した。実験者の操作もタイムスタンプとともにログを記録した。

インタラクシオンデータの分析のため、以下の3つの分析区間を設定した。

- 質問中：ユーザーの質問開始時から終了時まで

- 応答待ち：ユーザーの質問終了時からエージェントの応答開始時まで
- 応答中：エージェントの応答開始時から終了時まで

ユーザーの質問開始時からエージェントの応答終了時までを1インタラクシオンとし、各観光スポットについてのユーザーの最初のインタラクシオンの質問開始時からエージェントの最後のインタラクシオンの応答終了時までを1セッションとした。エージェントが質問に適切に応答できなかったインタラクシオンを成功、分からないなど、エージェントが質問に適切に応答できなかったインタラクシオンを失敗とした。

3.2 ガウス混合モデルのパラメータ推定

収録したインタラクシオンデータのうち、8名分のデータについて、インタラクシオン区間および、インタラクシオンの成功、失敗についてのアノテーションを行った。

アノテーションしたインタラクシオンデータの全セッションの映像データから、OpenFace を用いてユーザーの AU を抽出し、AU の時系列データを作成した。抽出した AU のうち、AU 1, AU 2, AU 4, AU 5, AU 6, AU 7, AU 9, AU 12, AU 15, AU 17, AU 20, AU 23, AU 26 の強度を用いて AU ベクトルを作成した。あるフレーム以前 30 フレームの AU ベクトルの平均を、そのフレームの観測データとした。アノテーションしたインタラクシオンデータ、全セッション分の観測データを入力としてガウス混合モデルのパラメータ推定を行い、ユーザーの表情パターンをモデルに学習させた。ガウス混合モデルのクラス数数は 20 とした。

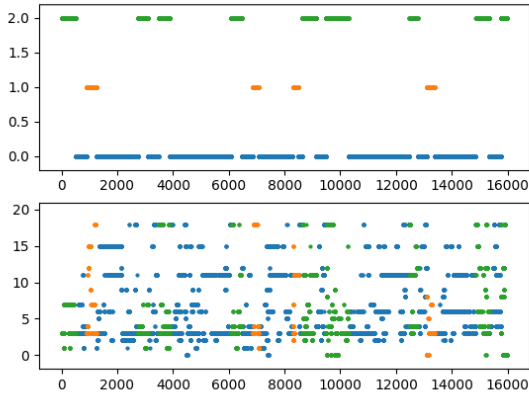


図 3 分類結果 (参加者 2, セッション 1)

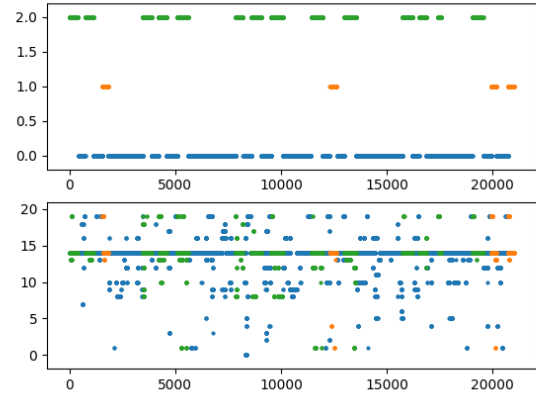


図 4 分類結果 (参加者 6, セッション 1)

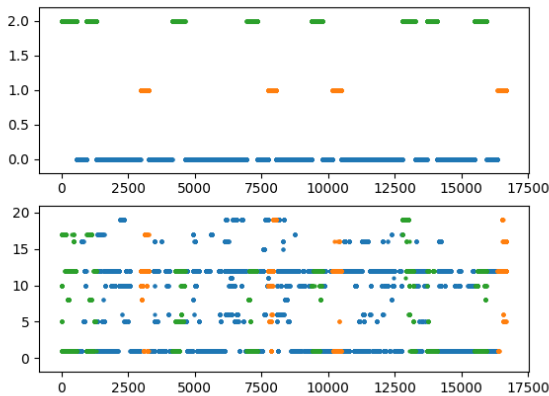


図 5 分類結果 (参加者 7, セッション 2)

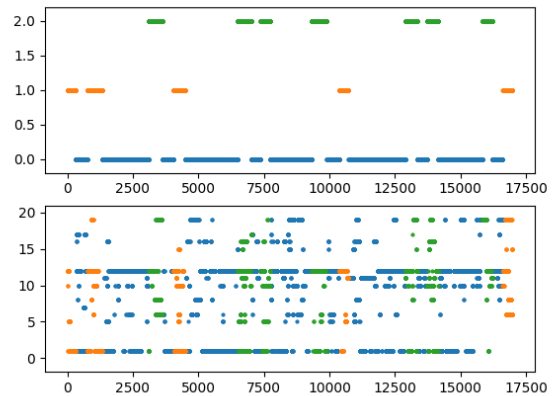


図 6 分類結果 (参加者 7, セッション 3)

3.3 表情パターンの分類

アノテーションしたインタラクションデータ、全セッション分の観測データに対して、観測データに対して最も尤度が高いクラスタをその観測データの表情パターンとする。

4. 結果

4.1 表情パターンの学習結果

図 2 に学習したモデルの各クラスタが生成する AU の平均値を示す。図 2 より各クラスタで異なる表情パターンが学習されていることが見て取れる。

4.2 表情パターンの分類結果

図 3 から図 6 に表情パターンの分類結果の例を示す。横軸はフレームである。各図において、上図はインタラクションの状態である。2.0 は成功のインタラクションのフレーム、1.0 は失敗のインタラクションのフレーム、0.0 はそれ以外のフレームをそれぞれ表している。下図は各フレームの表情パターンである。時系列全体に渡って様々な表情パターンが観測された。図 3 から図 6 が示すように、特定

の表情パターンとインタラクションの状態との間に直接的な対応付けを見出すことは難しい。他の特徴としては、実験参加者ごとに表出されやすい表情パターンに違いがあるように見える。例えば、図 3 と図 4 は異なる実験参加者の同じセッションの表情パターンの分類結果であるが、表出された表情パターンは大きく異なるように見える。一方、図 5 と図 6 は同じ実験参加者の異なるセッションで表出された表情パターンの分類結果であるが、表出された表情パターンが同じような傾向を示しているように見える。

5. 考察

実験の結果、特定の表情パターンとインタラクションの状態との間に直接的な対応付けを見出すことはできなかった。インタラクション中に表出されるユーザの表情パターンからユーザの状態を推定することはまだ難しいと言える。

表情パターンの分類結果では、特定の表情パターンとインタラクションの状態との間に直接的な対応付けを見出すことはできなかったが、実験参加者間で表出しやすい表情パターンに違いがあるように見えた。表出しやすい表情に個人差があるということは十分に考えられる。

次に、実験参加者間だけでなく、実験参加者内でも、実験参加者の状態によって表出しやすい表情に違いがある

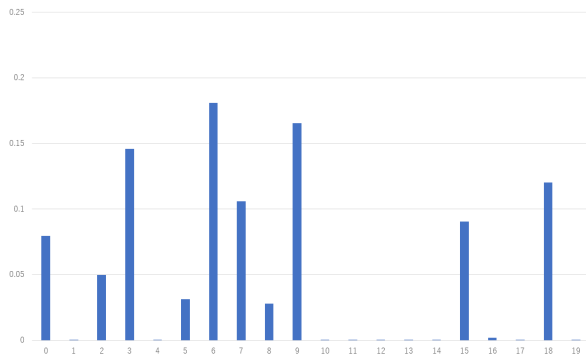


図 7 成功したインタラクションの表情パターン確率分布

のではないかと考えた. そこで, 成功したインタラクション中にはポジティブな表情が, 失敗したインタラクション中にはネガティブな表情が表出されると仮定し, 実験参加者内で, 成功したインタラクション中に表出された表情パターンの確率分布と, 失敗したインタラクション中に表出された表情パターンの確率分布間の Kullback-Leibler 情報量を計算し, 分布間に違いがあるかどうかを調べた. 計算の結果, 最大値は実験参加者 4 の 0.39, 平均は 0.21 であった. 実験参加者 4 の成功したインタラクション中に表出される表情パターンの確率分布を図 7 に, 失敗したインタラクション中に表出される表情パターンの確率分布を図 8 に示す. 実験参加者 4 の成功したインタラクション中に表出された表情パターンの確率分布と, 失敗したインタラクション中に表出された表情パターンの確率分布では, 特に表情パターン 7 と 9 の出現確率に顕著な差があるように見える. 確率分布に差がみられたことから, インタラクション中に表出される表情パターンの確率分布の変化からインタラクション中のユーザの状態の変化を推定できる可能性が考えられる.

6. おわりに

本稿ではヒューマン - エージェントインタラクション中のユーザの表情に着目し, WoZ 法によって実施されたヒューマン - エージェントインタラクション中のユーザの表情を分析した.

OpenFace によって検出された AU の情報を観測データとしてガウス混合モデルのパラメータ推定を行い, インタラクション中のユーザの表情パターンを分類した.

収録されたインタラクションデータの表情パターンを分析した結果, インタラクションの状態と表出する表情パターンとの間に直接的な対応付けを見出すことはできなかった.

実験結果では実験参加者間で表出しやすい表情パターンに違いが見られた. 実験参加者内でも, 成功したインタラクション中に表出された表情パターンの確率分布と, 失敗したインタラクション中に表出された表情パターンの確率分布間の Kullback-Leibler 情報量を計算した結果, 最大値は 0.39, 平均は 0.21 であり, 成功したインタラクション中に表出された確率分布と失敗したインタラクション中に表出された確率分布間に違いがあることが分かった. このことから, インタラクション中に表出されるユーザの表情パターンの確率分布の変化から, インタラクション中のユーザ

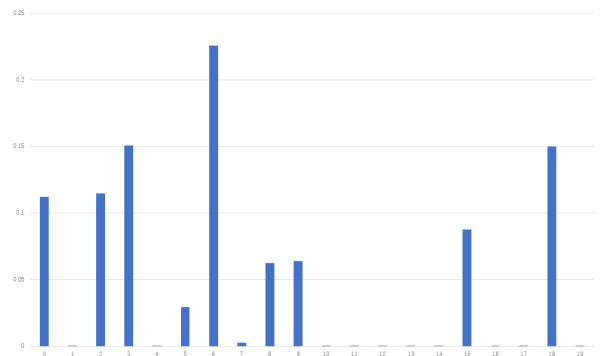


図 8 失敗したインタラクションの表情パターン確率分布

の状態の変化を推定できる可能性がある. しかし, 表出される表情の確率分布のどのような変化が, ユーザの状態のどのような変化に対応するかということについては議論の余地がある.

今後の課題としては, 表出される表情の確率分布とユーザの状態あるいはその変化のより具体的関係を明らかにすることが挙げられる. また, 今回用いなかったインタラクションデータについての追加の分析も今後の課題として挙げられる.

謝辞

本研究は沖電気工業株式会社との共同研究によって実施されたものである. 沖電気工業株式会社 深澤 伸一様, 赤津 裕子様, 鈴木 雄介様には, 本研究に関して大変有意義な議論をしていただきました. ここに感謝の意を表します.

参考文献

- [1] Baltrušaitis, T., Zadeh A., Lim, Y.C., Morency, L.P., "OpenFace 2.0: Facial Behavior Analysis Toolkit," IEEE International Conference on Automatic Face and Gesture Recognition (2018).
- [2] Du, S., Tao, Y., Martinez, A.M., "Compound facial expressions of emotion," Proceedings of the National Academy of Science, Vol. 111, No. 15 (2014).
- [3] Reynolds, D., "Gaussian mixture models," Encyclopedia of biometrics (2015).