

動画からの表情変化特徴の抽出と 雰囲気・個性のモデリングのための表情の解析

小竹山 祐輝^{1,a)} 中村 理恵^{1,b)} 持橋 大地^{2,c)}

概要: 表情には個人の感情や雰囲気、印象など様々な情報が含まれており、近年、表情解析研究が盛んに行われ、社会実装され始めている。一方、これら表情解析の研究の多くは、感情推定タスクに代表されるような、教師あり学習によって表情を既存のある表情カテゴリに分類する研究であり、個人における表情特徴を明らかにするための解析手法はほとんど検討されていない。これに対し、本研究では教師なし学習による表情分類により、現実世界にどのような表情が存在するかを明らかにすることで、個人の表情パターン、雰囲気・個性の定量化を実現することを目指した。本表情解析手法の開発において、表情分類の際の個人の顔特徴に対する頑健性、顔ブレに対する頑健性、表情の文脈に重要な強度と表情パターン情報を保持した状態で高次元で高精度に分類する技術などが求められ、多くの技術的課題がある。本研究では上記を配慮して、現実データに応用性の高い解析フレームワークを提案する。具体的には、顔の造形に頑健な表情変化特徴量設計と正面化補正、表情強度による1段階目分類、主成分分析による特徴量の次元圧縮後、周辺化ギブスサンプリングによってガウス混合モデルを用いた2段階目分類を行う。コーセー化粧品販売員のオンラインカウンセリング研修動画を対象に実験した結果、販売員間の表情変化パターンの違いが示唆され、本解析手法の有効性が示されたため報告する。

キーワード: 表情解析, 表情変化特徴量, 教師なし学習, 混合ガウス分布, LFM

Modeling Facial expressions from Video for the Analysis of Atmosphere and Personality

YUKI KOTAKEYAMA^{1,a)} RIE NAKAMURA^{1,b)} DAICHI MOCHIHASHI^{2,c)}

Abstract: Facial expressions contain various information such as emotions and moods, and facial expression analysis studies have been conducted extensively. Most of these studies classify facial expressions into certain existing categories by supervised learning, such as emotion estimation, and few analysis methods have been investigated to reveal the characteristics of facial expressions of individuals. Therefore, in this study, we aimed at quantification of individual facial expression patterns and moods by using unsupervised learning for expression classification to identify real facial expressions. Expression classification requires robustness to individual facial features and blurring, and analysis that preserves context-important intensity and pattern information. In this study, we propose a framework suitable for real-world data analysis that takes the above into consideration. Specifically, we design and frontalize facial expression change features, perform first classification based on facial expression intensity, perform dimensionality compression of features using principal component analysis, and perform second classification using GMM with collapsed Gibbs sampling. Experimental results on online counseling training videos of KOSÉ cosmetics salespeople suggest differences in facial expressions among salespeople, and the effectiveness of this analysis method is reported.

Keywords: Facial Expression Analysis, Facial Expression Change Features, Unsupervised Machine Learning, Gaussian Mixture Model, LFM

1. はじめに

近年、人間とロボットの高度なインタラクション、ビジネス交渉の円滑化等を実現化するために表情・微表情を用いた解析研究が盛んに行われている [1][2][3]。特に表情・微表

情から喜怒哀楽といった相手の感情を予測する感情推定研究が盛んに行われている。一方で表情には話者の印象・雰囲気を形成する役割も存在し、美容や個性発見の文脈で表情に含まれる雰囲気をモデリングすることは、重要かつ社会的に幅広い応用が考えられる。しかしながら、表情から雰囲気や印象を研究した例は多くない [4]。また多くの研究は、表情を教師あり学習によって予めラベル付けされた既存の表情カテゴリに分類するものであり、教師なし学習で表情を分類するような研究例は少ない。動画中で現れる様々な

¹ 株式会社コーセー 研究所 東京都北区柴町 46-3

² 統計数理研究所 東京都立川市緑町 10-3

a) y-kotakeyama@kose.co.jp

b) r-itou@kose.co.jp

c) daichi@ism.ac.jp

表情を解析対象として教師なし分類する研究はさらに少ない。そこで、本研究では動画中から表情特徴量を抽出し、教師なし学習を用いて適切に分類することで人の持つ表情を明らかにし、個人の潜在的魅力の発見、雰囲気の定量化を実現することを目指した。

雰囲気の可視化・定量化に際し、我々は以下の研究仮説を設定した。印象・雰囲気とは表情に代表されるような人の発するノンバーバルな客観情報を第三者が捉えることで認識される。世の中に存在する表情特徴の1つ1つは、雰囲気を決定づける一つの要素である。一方で単一の表情特徴はあくまでも数ある表情の1つに過ぎず、個人の雰囲気は、個人に観測される様々な表情特徴の組み合わせと出現頻度により決定づけられるものだと考えた。本稿では以上の研究仮説を基に、動画における表情特徴量抽出と表情変化パターン分類による個性のモデリングのための解析手法の提案を行う。以下、表情から表情分類、雰囲気の定量化をする上での技術的な課題と本手法における解決手法について記述する。通常、表情特徴量を抽出する方法には、dlibによる顔特徴点取得や深層学習の一種のCNN(Convolutional Neural Network)特徴量などがあるが、動画の各フレームに両者を単純に用いて教師なし学習で分類を行うと、顔特徴(パーツや骨格)に依存するため、表情を分類することが難しい。本稿では、この問題を軽減するために表情特徴として、表情変化に着目し、Choiら[5]により提案された表情変化特徴量(LFM, Landmark Feature Map)を適用した。LFMは、dlib点の相対距離行列の時間差分特徴であり、Choiらの研究においては感情推定を目的に使用され、様々な人種、性別において高精度で感情予測できたことが報告されている。また本研究では、正面に顔を固定したような綺麗なデータセットではなく、カウンセリングなどの自然な環境の表情分類を目的としているため、顔のブレに頑健なモデルである必要もある。顔のブレに対する正面化補正には、Vassiliosら[6]の開発した正面化補正アルゴリズムを使用した。Vassiliosらのアルゴリズムはdlibランドマーク点を表情の情報保持して正面化する事に優れている。さらに表情分類において、高次元の表情変化特徴量データを表情変化の強度情報と表情変化の仕方(パターン)情報を保持しつつ、個性に該当する局所的な表情変化を上手く表現し、現実データを分類することが求められる。そのため、本研究では表情の強度とパターンで二段階分類を行った。初めに表情強度メトリクスを設定し、強度別に表情分類した後に、層別に主成分分析で次元圧縮を行い、ガウス混合モデル[7]を用いてクラスタリング分類した。モデルのパラメータの学習には高次元データで、局所解に陥りにくい手法の確率的サンプリング手法MCMCの一種である周辺化ギブスサンプリング[8]を適用した。さらに口を閉じる、開くなどの表情変化の対称運動を1つのクラスターとして分類するために、サンプリング過程で原点对称に仮想的に



図1: 提案する解析フレームワーク。動画から表情変化特徴量 LFM を作成し、LFM の強度とパターンによって二段階分類を行った後、個人における各クラスターの出現割合を算出する。

データを射影することによる GMM の拡張も行った。

本稿では本解析手法について、コーサー化粧品販売員のオンラインカウンセリング研修動画を対象に実験した結果、表情分類と表情に基づく個性的な雰囲気の定量化に対する本手法の有効性が示唆されたため報告する。

2. 関連研究

動画像における表情解析分野の研究の大半は、表情検出(Spotting)と感情推定タスク(Recognition)である。本節では表情解析の観点からの研究分野の概要として、表情の特徴量設計から感情推定タスクまでの流れを概説する。

2.1 マクロ表情とマイクロ表情と感情分類研究

表情に基づく感情推定タスクにおいて、表情はマクロ表情とマイクロ表情に分類される。マクロ表情に関するデータセットは、ニュートラルな表情から目標感情を代表する表情への顔変化を示す動画を収録した Extended Cohn-Kanade[CK+][1]が有名である。一方、一般的にマイクロ表情はマクロ表情に比べて表情強度が非常に弱く、その強さは1/9程度とも言われている[5]。また微表情は0.2秒から0.5秒の間に現れる表情であり、不随意的な反応であるため真の感情が含まれるとされている。そのため、マクロ表情解析研究に加え、微表情領域における感情推定研究も盛んであり、定期的に Facial Micro-Expression Grand Challenge(MEGC)も行われている[9]。MEGCでは、CASME II[2]やSAMM[10]という微表情に特化したデータセットに対する感情認識精度の評価がなされている。微表情に基づく感情推定手法には、LBP(local binary patterns)による推定やSVM, CNNによる分類などがある。

2.2 表情特徴量の抽出

表情解析研究における特徴量抽出方法には大きく3種類がある。1つ目はEkmanら[11]が、目に見えるすべての可能な顔面動作を解剖学的根拠に基づき区別するための包括的なシステムとして開発したAction Unitである。AU8:眉を下げる, AU10:上唇を上げるなど表情の動作にラベルが付いている。2つ目は深層学習の一種であるCNNによる画像からの特徴量抽出である[3]。ネットワークの構造としてはVGGやResnetなどが用いられる。CNNベースの表情推定は高精度である一方で画像全体を処理するための膨

大なメモリが必要である。3つ目の特徴量は、顔の点のランドマーク検出がある。回帰木のアンサンブル学習を用いた dlib[12] は、入力された顔画像に対して顔の検出と顔の目・鼻・口といった特徴点 68 点を検出する手法であり、高品質かつ高速での予測が可能である (図 2 (a))。

2.3 Landmark Feature Map (LFM)

LFM は、Choi ら [5] によって提案された表情変化特徴量である。特徴量には dlib のランドマーク点特徴量を使用する。ある時間 t における dlib により推定された顔の特徴点 68 点から各点間の距離行列 (68×68) を作成する。LFM は dlib 距離行列の時間変化パターンとして以下のように定義される。

$$LFM_t(i, j) = \|p(i, t) - p(j, t)\|_2 - \|p(i, t-T) - p(j, t-T)\|_2 \quad (1)$$

ここで $p(i, t)$ は t 秒目の dlib 距離行列における i 番目の座標を表す。LFM _{t} は t 秒目の dlib 距離行列における LM の点 i と点 j の距離と $t-T$ 秒目の dlib 距離行列における点 i と点 j の距離の差分であり、つまり各顔座標の距離行列の時間差分である。LFM の各座標は各顔の変化領域に対応している。Choi らの研究 [5] においては、マクロ表情とミクロ表情において LFM のパターンが類似することを見出し、特に感情推定が難しいミクロ表情において高精度で感情予測を達成していることが報告されている。

2.4 顔向き補正 (frontalization)

動画データ解析において、顔の向きを正面補正することは重要である。正面補正の研究は盛んに行われており、一般的に 3D 顔モデルに点を推定して Appearance-based frontalization と顔のランドマーク点を正面化した際の推定 Landmark-based approaches の二種類に分類される。Appearance-based frontalization はあらかじめ定義された線形形状モデルを、対象物を含む画像に位置合わせする優れた方法である。GAN などでも使用され、表現力が高い一方で計算コストが高い課題がある。Landmark-based approaches について、一部情報損失は存在するものの計算量が少なく実応用に向いている。表情解析におけるランドマーク点データを用いた顔向き補正について、表情情報を保存したまま正面化推定する方法を Vassilios ら [6] が提案している。

3. 手法

本節では提案手法について詳細に説明する。表情解析フレームワークを図 1 で示す。

3.1 動画画像からの表情変化特徴量抽出 (LFM の作成)

表情変化特徴を細かく表現するために、今回は dlib ランドマーク点の相対距離行列の時間差分特徴量 LFM を使用す

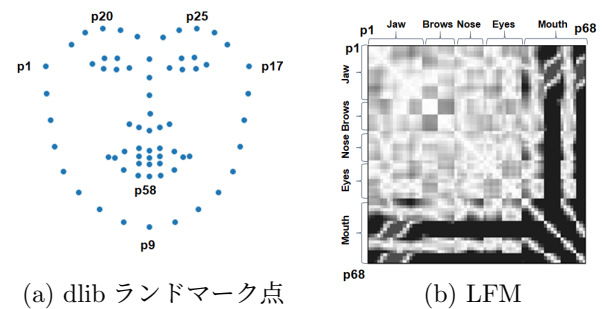


図 2: 表情変化特徴量の設計。(a) は顔特徴 68 点の dlib ランドマーク点である。(b) は dlib ランドマーク点の 68×68 の時間差分特徴量として作成される LFM である。

る。LFM は表情変化量であり、ランドマーク点のみに比べて、被験者の顔の造形への影響を軽減し、頑健性を確保できることが期待できるからである。

3.1.1 顔の検出とランドマーク点の取得

60fps の動画から 1 フレーム毎に dlib により顔検出とランドマーク特徴点 68 点の取得を行う。

3.1.2 表情特徴量の正面化補正

Vassilios ら [6] により提案された表情特徴の保存を目的としたランドマーク特徴点の正面化アルゴリズムを用いて、3.1.1 で取得したランドマーク点を正面化補正する。

3.1.3 表情変化特徴量 (LFM) の作成

正面化補正済み顔特徴点データから各 dlib 点間の距離行列を作成し、20 フレーム (1/3 秒) 毎に dlib 距離行列の差分行列 LFM を作成する。

3.2 表情変化特徴量の 2 段階分類

先行研究に比した際の本研究の特徴の一つは、喜怒哀楽のような予め決められた数個の感情ラベルを教師あり学習によって予測するものではなく、動画中に現れる様々な表情変化を教師なし学習を用いて分類することである。感情認識タスクにおいては、表情変化特徴を表情変化の強さで正規化し、表情変化パターンを主な情報として感情予測することがあるが、雰囲気・印象の文脈においては強度もパターンと同様に重要な情報を有する。そこで本研究では、正規化を行わず、3.1 で作成した表情変化特徴量 LFM の強度と表情変化パターンの 2 つの情報を保持したまま適切に分類することを目指した。

3.2.1 表情変化の強度による LFM 分類 (1 段階目分類)

いくつかの感情認識の表情研究において簡易的な表情強度メトリクスが作成されており、表情強度に応じて推定アルゴリズムを使い分けることが有効であることが報告されている [5][13]。また雰囲気、印象の文脈において、表情変化パターンと別軸として表情変化の強度で分類をすることは自然な仮定である。さらに、最初に表情強度でデータセットの分離を行うことで、今回使用する周辺化ギブスサンプリングによる GMM の学習効率 (サンプリング効率) を上げ

ることもできる。そこで本研究では、先行研究を参考に以下の表情強度メトリクス (FEM: Facial Expression Metric) を定義した。FEM を算出し、箱ひげ図と実際の動画の確認を合わせて行うことで表情を 4 段階に分類した。上位 25% の表情を強表情と中間表情とした。強表情は、箱ひげ図のうち外れ値に該当し、動画中の中で著しく大きな表情変化である (2.2%)。上位 25% のうち、強表情以外の表情変化が中間表情である (22.8%)。残りの 75% について、動画により表情変化を確認して、表情変化が認識でき、表情変化をパターン別に綺麗に分離できる閾値を 5% 単位で検討し、最終的に弱表情 I を 35%、ほぼ微表情か無表情である弱表情 II を 40% とした (計 100%)。表情強度 FEM は以下の式で算出される。ここで FEM は全 FEM の最大値 - 最小値 (今回のデータセットにおいては 8.199) で正規化した値を使用した。

$$FEM_t = \sqrt{\sum_{i=1}^{68} \sum_{j=1}^{68} LFM_t(i, j)^2} \quad (2)$$

3.2.2 各強度層別の LFM データセットの主成分分析

各 LFM データは 68 × 68 次元を持つ特徴量であるが、対称行列であり、かつ表情変化の中で互いに連動して動く箇所が存在するため、主成分分析による次元圧縮を行った。累積寄与率 95% までの主成分スコアを LFM の主成分特徴量として使用した。

3.2.3 周辺化ギブスサンプリングを用いたガウス混合モデル推定 (2 段階目分類)

LFM を分類する手法として、教師なし学習の一種であるガウス混合モデル (GMM) を用いる。GMM は与えられたデータセットが複数のガウス分布の重ね合わせで与えられると仮定して、各ガウス分布の平均や分散、ガウス分布ごとの重みを学習し、各データの潜在変数を確率的に割り当てる手法である。一般的に用いられるクラスタリング手法の kmeans は、GMM の EM アルゴリズムの $\sigma^2 \rightarrow 0$ の特殊な場合であり、GMM はクラスタの分布の形に応じてソフトにクラスタを割り当てることができ、実データの表現に適している。また局所的なクラスタ構造を抽出することも可能であり、非線形クラスタリングに比べると分類時の解釈性も高い。本研究では GMM のパラメータ推定方法として、EM アルゴリズムによる推定ではなく、確率的サンプリングの MCMC 法 (マルコフ連鎖モンテカルロ法) の一種の周辺化ギブスサンプリングを用いることにより、高次元空間において局所解に落ちることなく推定することを目指した。また口を開く・口を閉じるなどの原点对称な運動を数学的に明示して同じクラスタとするように、パラメータ推定時に各クラスタの原点对称なクラスタを仮想的に設定してサンプリングする原点对称なアルゴリズムを設計した。原点对称な本 GMM アルゴリズムを以下 Shadow clustering と呼ぶ。各表情強度層において、算出し

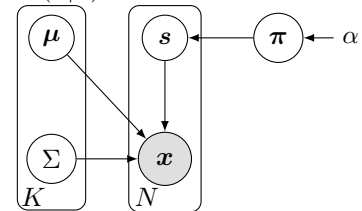
た主成分スコアデータセットに対して、それぞれで上記の GMM 分類を行った。ディリクレ分布のハイパーパラメータ α 、クラスタ数 K 、反復数 iter は各表情強度において、 $\alpha = 0.1$, $K = 20$, iter=1000 で実施した。原点对称 GMM の周辺化ギブスサンプリングアルゴリズムを Algorithm1 に示す。

本研究の原点对称性を考慮した多次元ガウス混合モデル (Shadow clustering) と周辺化ギブスサンプリングアルゴリズムについて以下で説明する。以下には一般的な GMM のグラフィカルモデルを示す。混合モデルの各成分はガウス密度を持ち、i 番目の観測値を x_i 、クラスタ k の平均ベクトルを μ_k と共分散行列を Σ_k 、混合比率パラメータを $\pi = (\pi_1, \pi_2, \dots, \pi_K)$ とする。混合比率パラメータはハイパーパラメータ α で条件づけられるディリクレ分布を使用する。また、 s_n は潜在変数を表す K 次元のベクトルであり、ある k に対して $s_{n,k} = 1$ が成り立つときに k 番目のクラスタが指定されたことを表す。上記条件の元、ガウス分布と混合ガウス分布は以下のように表現される。ここで

$$\mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_k)^T \Sigma_k^{-1} (\mathbf{x} - \mu_k) \right\} \quad (3)$$

$$p(x_i | \mu_k, \Sigma_k) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k), \sum_{k=1}^K \pi_k = 1 \quad (4)$$

GMM のグラフィカルモデルは以下のように表現される。ここで $p(\pi | \alpha) = \text{Dir}(\pi | \alpha)$ である。



GMM の周辺化ギブスサンプリングに関して記載する。π, μ, Σ を周辺化除去したモデルを考える。サンプルしたい変数 s_n 以外の全ての潜在変数の集合を $\mathbf{S}_{\setminus n} = \{s_1, \dots, s_{n-1}, s_{n+1}, \dots, s_N\}$ とすると s_n に関する事後分布は以下の 2 つの項の形で表せる。

$$p(s_n | \mathbf{X}, \mathbf{S}_{\setminus n}) \propto p(\mathbf{x}_n | \mathbf{X}_{\setminus n}, \mathbf{S}_{\setminus n}, s_n) p(s_n | \mathbf{S}_{\setminus n}) \quad (5)$$

各項について計算すると最終的にはカテゴリ分布と学生 t 分布の形で書き表せる。

$$p(s_n | \mathbf{S}_{\setminus n}) = \text{Cat}(s_n | \eta_{\setminus n}) \quad ; \quad \eta_{\setminus n, k} \propto \hat{\alpha}_{\setminus n, k} \quad (6)$$

$$p(\mathbf{x}_n | \mathbf{X}_{\setminus n}, \mathbf{S}_{\setminus n}, s_n, k = 1) = \text{St}(\mathbf{x}_n | \hat{\mathbf{m}}_{\setminus n, k}, \frac{(1 - D + \hat{\nu}_{\setminus n, k}) \hat{\beta}_{\setminus n, k}}{1 + \hat{\beta}_{\setminus n, k}} \hat{\mathbf{W}}_{\setminus n, k}, 1 - D + \hat{\nu}_{\setminus n, k}) \quad (7)$$

Algorithm 1 原点对称の周辺化ギブスサンプリング

- 1: クラスター数 K に対して, 原点对称の潜在変数を以下のように設定 $K' = (K_+, K_-) = (1, \dots, K, \dots, 2K)$
その他パラメータも $\hat{\mu} = (\hat{\mu}_+, \hat{\mu}_-)$ のように設定
潜在変数のサンプル s_1, \dots, s_N に初期値を設定
 $\hat{\mu}_+, \hat{W}_+, \hat{\beta}_+, \hat{\nu}_+$ を計算
(k_- に属する値は k_+ に属するとして計算, $\hat{\mu}_-$ は $-\hat{\mu}_+$ で算出)
- 2: **for** $i = 1, \dots, MAXITER$ **do**
- 3: **for** $n = 1, \dots, N$ **do**
- 4: 式 (8) から式 (11) を用いて x_n に関する統計量を除去
- 5: **for** $k = 1, \dots, 2K$ **do**
- 6: 式 (7) を用いて $p(x_n | X_{-n}, s_{n,k} = 1, S_{-n})$ を計算
- 7: **end for**
- 8: 式 (6) と $p(x_n | X_{-n}, s_n, S_{-n})$ から s_n をサンプル
- 9: 式 (8) と式 (11) から x_n に関する統計量を追加
- 10: **end for**
- 11: **end for**

式中ハイパーパラメータについては以下に示す. $\mathbf{m} \in \mathbb{R}^D$, $\beta \in \mathbb{R}^+$, $\mathbf{W} \in \mathbb{R}^{D \times D}$ である. D はデータの次元を表す. \mathbb{R}^+ は正の実数を表す.

$$\hat{\beta}_{\setminus n,k} = \sum_{n' \neq n} s_{n',k} + \beta \quad (8)$$

$$\hat{\mathbf{m}}_{\setminus n,k} = \frac{\sum_{n' \neq n} s_{n',k} \mathbf{x}_{n'} + \beta \mathbf{m}}{\hat{\beta}_{\setminus n,k}} \quad (9)$$

$$\hat{\nu}_{\setminus n,k} = \sum_{n' \neq n} s_{n',k} + \nu, \quad \hat{\alpha}_{\setminus n,k} = \sum_{n' \neq n} s_{n',k} + \alpha_k \quad (10)$$

$$\hat{\mathbf{W}}_{\setminus n,k}^{-1} = \sum_{n' \neq n} s_{n',k} \mathbf{x}_{n'} \mathbf{x}_{n'}^\top + \beta \mathbf{m} \mathbf{m}^\top - \hat{\beta}_{\setminus n,k} \hat{\mathbf{m}}_{\setminus n,k} \hat{\mathbf{m}}_{\setminus n,k}^\top + \mathbf{W}^{-1} \quad (11)$$

(6), (7) について $s_{n,k} = 1$ となる場合の確率をそれぞれ求めて正規化することにより, s_n をサンプリングするカテゴリ分布のパラメータを得ることができる. また以下に, 原点对称を同一クラスターとみなす Shadow clustering のアルゴリズムを説明する. 初めに, クラスター数 K に対して原点对称な $2K$ の潜在変数を準備する. ここで $K_+ = (1, \dots, K)$, $K_- = (K+1, \dots, 2K)$ とする. 次に潜在変数 $\mathbf{S} = \{s_1, \dots, s_N\}$ の初期値に 1 から $2K$ までの値をランダムに割り当てる. 他のパラメータ $\hat{\mu}_+, \hat{W}_+, \hat{\beta}_+, \hat{\nu}_+$ の初期値については, k_- に属するサンプルを k_+ に属するとみなして算出する. この時, $\hat{\beta}_+, \hat{\nu}_+$ は k_- を k_+ としてカウントすれば良いが, $\hat{\mu}_+, \hat{W}_+$ についてはデータ点の \mathbf{x} にマイ

表 1: 本研究で用いた動画データセット.

動画 ID	販売員 ID	時間	使用部分
1	BC1	38m38s	6m19s
2	BC2	25m46s	7m16s
3	BC2	33m40s	8m54s
4	BC3	29m04s	8m09s
5	BC4	33m13s	3m37s
6	BC5	22m48s	7m34s
7	BC5	31m24s	3m15s
8	BC6	25m27s	1m18s
9	BC7	22m55s	12m53s
10	BC8	30m54s	3m32s
11	BC8	26m09s	3m07s

ナスをかけることに注意する. 周辺化ギブスサンプリングによって更新する際も同様である. また (6), (7) を用いて潜在変数をサンプリングする際には, 上記で求めた $\hat{\mu}_+$ にマイナスをかけたものを $\hat{\mu}_-$ とし, $\hat{W}_+, \hat{\beta}_+, \hat{\nu}_+$ を $\hat{W}_-, \hat{\beta}_-, \hat{\nu}_-$ として射影することで 1 から $2K$ までの範囲で潜在変数をサンプリングする. 最終的には, k_- に属した潜在変数を k_+ とみなして, $\mathbf{S} = \{s_1, \dots, s_N\}$ が求まる. 以上のアルゴリズムにより原点对称なクラスターを同一のクラスターとして分類ができる.

4. 実験

4.1 実験に用いたデータ

客役・化粧品販売員役 2 人 1 組で構成される, コーサー化粧品販売員のオンラインカウンセリング研修動画 (60fps, 解像度 1440×1920) から化粧品販売員役の映っている画面を Davinci resolve (<https://www.blackmagicdesign.com/jp/products/davinciresolve/>) により切り出し抽出した. 抽出した動画の中から販売員役が真横を向く, 大きく頷く, 画面から外れる等, ノイズの大きい箇所については除外し, 解析対象のデータセットとした (表 1).

4.2 実験方法

本実験では, 上記データセットを用いて表情特徴量から実際に個人間に現れる類似した表情変化を分類することが可能か, 個人に出現する表情変化クラスターをプロファイリングすることで雰囲気の数量化が可能か, 提案する解析手法の有効性について検証した. 以下の手順で解析した.

4.3 提案手法による解析

4.3.1 ランドマーク点検出・正面化と LFM 作成

抽出した 60fps の動画から 1 フレーム毎に dlib により顔検出とランドマーク特徴点 68 点を取得した. Vassilios らの正面化アルゴリズムにより dlib 点を正面化後, 時間差分 1/3 秒 (20 フレーム) ごとに LFM を作成した. 最終的に $n=9628$ の LFM データが取得された. 以下では本データを使用する.

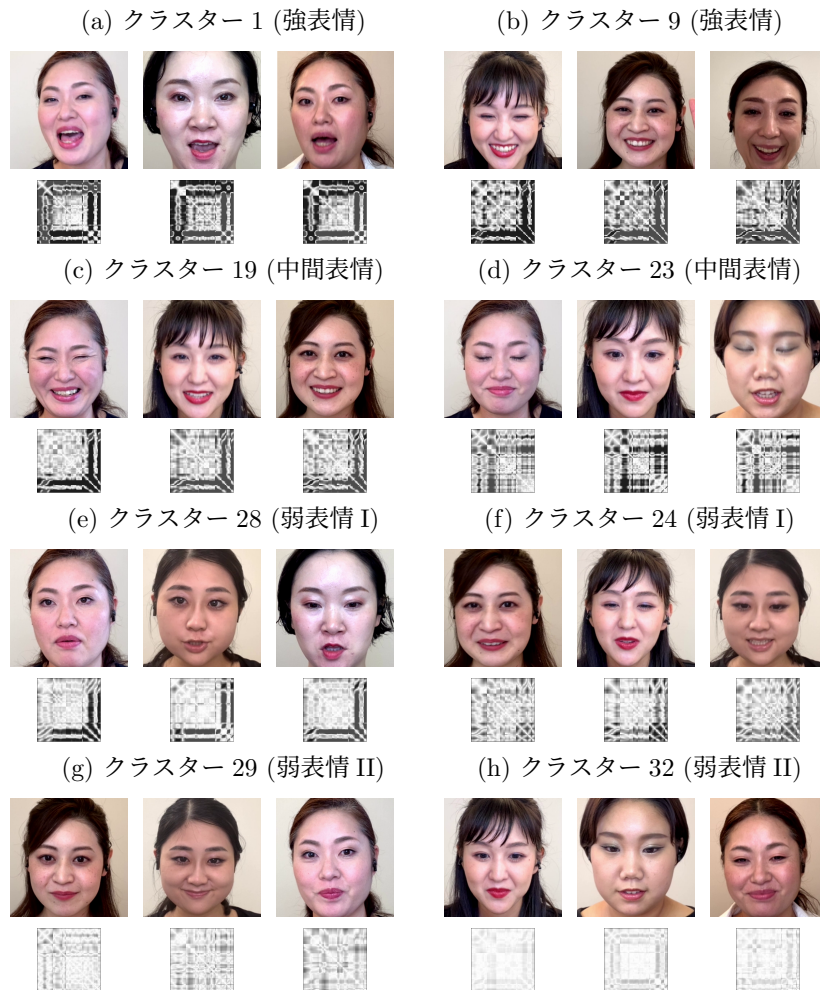


図 3: 提案フレームワークにより分類された LFM の各クラスターと表情の一例. (a), (b) は強表情, (c), (d) は中間表情, (e), (f) は弱表情 I, (g), (h) は弱表情 II のクラスターの例である. 表情変化を一枚の写真で表しているため, 対象となる運動については口を開けるなど意味のある方を載せている.

4.3.2 表情変化強度による LFM の 1 段階目分類と各強度層別の主成分分析

表情評価マトリクスに基づき, 表情を分類した. また各表情の層別に主成分分析を実施し, 次元圧縮を行った. 主成分特徴と寄与率の高い各主成分を代表する実際のフレームを抽出し, 主成分分析によりどのような表情特徴が次元圧縮されているかを確認した.

4.3.3 GMM による各表情強度の LFM 分類と動画評価

表情強度毎に算出した主成分スコア (累積寄与率 95%) を使用し, GMM による LFM の分類を行った. 分類した LFM に対応する実際のフレームを抽出し, 類似した表情変化を抽出できているかを確認した. また分類された表情変化がある個人にのみ出現する特有の表情変化であるか, 複数人に共通する一般的な表情変化であるか評価するためにエントロピー指標を算出した. さらに今回表情特徴量として LFM 使用した場合と dlib 単体を使用した場合の各クラスターのエントロピーを評価することで LFM の本アルゴリズムにおける顔特徴への頑健性への有効性も確認した.

エントロピー指標は以下の式に従い, 算出した. 今回, 通常のエントロピー H に対して, エントロピーが最大の時, つまりある表情変化が 8 人から等確率で出現する場合に, 1 となるように, 11 人中同一人物を 1 人とカウントした際の合計 8 人 ($K = 8$) の対数 $\log K$ で正規化を行ったものをエントロピー指標とした.

$$H' = H / \log K = (-\sum_{i=1}^K p_i \log p_i) / \log K \quad (12)$$

個人における各クラスターの出現割合を算出し, LFM パターンの個人プロファイリングを行った.

5. 結果

5.1 強度による LFM 分類と各強度の主成分分析

LFM を表情強度マトリクスを基に 4 つのグループに分類した. 強表情は, 感情や表情癖に由来すると考えられるような大きな表情, 中間表情は自然に会話する程度の表情, 弱表情 I は微笑む程度の少しだけ変化する表情, 弱表情 II

は無表情か微表情レベルの微量な変化をする表情であった。主成分分析の結果、累積寄与率 95%までの主成分として、各グループはそれぞれ PC7, PC9, PC13, PC17 まで圧縮されることが分かった。各主成分として、例えば強表情の主成分 1 には口の動き、2 には目と頷くような動き、3 には目の動き、と表情変化特徴が各主成分に次元圧縮されている様子が確認された。

5.2 GMM による各表情強度の LFM 分類と動画評価

各表情強度において GMM により LFM の分類を行った。強表情は最終的に 15 クラスター、中間表情は 8 クラスター、弱表情 I では 5 クラスター、弱表情 II では 10 クラスターに分離された。クラスター数の α への依存性について、 α の値に応じてクラスター数が 2, 3 個のぶれは見られるものの依存性は低かった。図 3 には GMM 分類されたクラスターの一例を示す。各表情強度別に 2 クラスターずつを示している ((a), (b): 強表情, (c), (d): 中間表情, (e), (f): 弱表情 I, (g), (h): 弱表情 II)。初めに個人の顔に依存することなく、複数人に対して類似する表情変化を分類できていることが分かる。対称運動についても同一クラスターとして分類できている (図中の写真は変化前後で代表的写真を使用している)。また口の開閉という動作の中でも、縦に口を開ける動作 (図 3 (a), (f)) と斜め上 (横方向) に口を開ける動作 (図 3 (b), (c), (d)) の微妙なニュアンスの違いを区別して抽出できていることも確認される。さらにこの動作に由来し、一般的に smile とラベルとつけられる笑顔に関して、笑顔形成パターンを見ると口を縦に開ける笑顔と口角を上げる笑顔のタイプに分かれることが示唆された。また 1 段階目で行った表情強度分類によって、同じパターンの表情変化でも満面の笑顔から微笑みまでニュアンスを捉えられている事が示唆される ((b), (c), (e))。また弱表情 II に関しては、GMM 分類において、高次元空間上で強度情報の影響を受け、共分散の大きいジャンクなクラスターが形成されている。これは弱表情 II において、主成分空間上でデータ点が過度に離れており、GMM 分類の際、局所的なクラスターを検出できない代わりに、共分散の大きいジャンクなクラスターを形成し、強度の大きなクラスターと小さなクラスターに分離したのだと考えられる。この結果は、一

表 2: LFM と dlib を特徴量した際の分類結果のエントロピーの比較。

特徴量	平均	中央値
LFM 全体	0.59	0.70
LFM 強表情	0.35	0.34
LFM 中間表情	0.74	0.82
LFM 弱表情 I	0.74	0.84
LFM 弱表情 II	0.76	0.83
dlib	0.17	0.12



図 4: dlib 特徴量に基づく各クラスターの表情一覧。表情よりも個人の骨格に依存していることが示唆される。

段階目に表情強度で分類したことの有効性も示唆している。また LFM と dlib に基づいて作成した各クラスターについて、エントロピーを算出した結果を表 2 に示す。LFM を基に作成されるクラスターは dlib に比べてエントロピーが高く、顔特徴への頑健性が示される。表情特徴量として dlib のみを使用する場合、図 4 に示すように個人の顔特徴に依存し、表情特徴を上手に検出できないことが示唆される。

5.3 LFM パターンの個人プロファイリング

分類した LFM の個人における出現割合を図 5 に示す。各色はクラスターを表し、黒線は各強度の境目を表している。一連の動画の中で表情強度により強表情、中間表情を出しやすい人に差があることが分かる。また同一人物内で類似した表情変化パターンと出現頻度を示すことが読み取れる。特に中間表情内においてその傾向が顕著に読み取れる (図 6)。中間表情で個人内の類似性が高いのは、中間表情が通常使用される表情であり、表情癖として身につけているからだと考えられる。また個人によって希少性の高い表情変化を高い割合で持つ人と一般的な表情変化のみを持つ人が存在する (図 5, 図 6)。例えば美容部員 7 は個性的な表情特徴の出現割合が高い。これら表情強度と表情変化パ

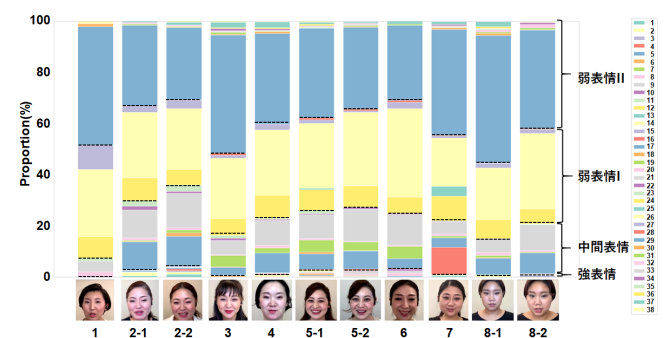


図 5: 全強度表情の各クラスターの個人における出現割合プロット。各色はクラスターを表している。黒の点線は表情強度の境目を示している。

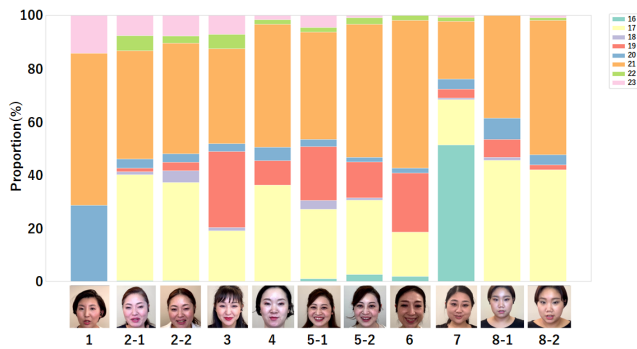


図 6: 中間表情の各クラスターの個人における出現割合プロット. 各色はクラスターを表している. 同一人物内での表情割合の類似性が見て取れる.

ターンの個人内での出現頻度は個性や個人の雰囲気を表していると考えられる.

6. 結論

本稿では、表情特徴量を基にした個人の雰囲気モデリングのための表情解析フレームワークについて提案した。提案手法は、表情変化特徴量 LFM を作成し、1 段階目で表情強度で分類後、主成分分析による次元圧縮を行い、2 段階目で周辺化ギブスサンプリング GMM により表情を分類する手法である。我々は、コーセー化粧品販売員のオンラインカウンセリング研修動画データセットにおいて、LFM 抽出と 2 段階分類を行うことで個人間に共通する表情変化、個人に特徴的な表情変化を検出できることを確認した。さらに、各 LFM の個人における出現割合が個人の雰囲気・印象を形作っている可能性も示唆され、今回の手法が現実データに対して、表情分類と表情に基づく雰囲気を定量化する手法として有効であることが示された。

参考文献

[1] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 94–101, 2010.

[2] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. Casme ii: An improved spontaneous micro-expression database and the baseline evaluation. *PLOS ONE*, Vol. 9, pp. 1–8, 01 2014.

[3] Van Quang Nguyen, Chun Jinhee, and Tokuyama Takeshi. CapsuleNet for micro-expression recognition. In *2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019)*, pp. 1–7. IEEE Computer Society, 2019.

[4] Won-Chan Lee and DaeKyun Yoon. A study on facial expression and first impression through machine learning. In *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*,

pp. 298–301, 2019.

[5] Dong Yoon Choi and Byung Cheol Song. Facial micro-expression recognition using two-dimensional landmark feature maps. *IEEE Access*, Vol. 8, pp. 121549–121563, 2020.

[6] Vonikakis Vassilios and Stefan Winkler. Identity-invariant facial landmark frontalization for facial expression analysis. In *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 2281–2285, 2020.

[7] J. McLachlan Geoffrey and Peel David. *Finite Mixture Models*. Wiley, 2000.

[8] Jun S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, Vol. 89, No. 427, pp. 958–966, 1994.

[9] Jingting LI, Su-Jing Wang, Moi Hoon Yap, John See, Xiaopeng Hong, and Xiaobai Li. Megc2020 - the third facial micro-expression grand challenge. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 777–780, 2020.

[10] Chuin Hong Yap, Connah Kendrick, and Moi Hoon Yap. Sann long videos: A spontaneous facial micro- and macro-expressions dataset, 2019.

[11] P. Ekman and W. Friesen. *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, 1978.

[12] Xuanyi Dong, Shoou-I Yu, Kinshuo Weng, Shih-En Wei, Yi Yang, and Yaser Sheikh. Supervision-by-Registration: An unsupervised approach to improve the precision of facial landmark detectors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 360–368, 2018.

[13] Rui Zhao, Quan Gan, Shangfei Wang, and Qiang Ji. Facial expression intensity estimation using ordinal information. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3466–3474, 2016.