

# 個人特徴を用いた表情認識に関する研究

田口 雄太 指導教員：何 立風

## 1 はじめに

現在、分野や目的による多種多様なロボットの存在により、我々の生活は支えられている。近年では、家庭用向けのロボットも登場しており、ロボットがますます身近な存在になりつつある。このような中で、ロボットと人間が共同で作業する機会が増加していくことが予想され、ロボットと人間の円滑なコミュニケーション、特にロボットが人間の感情を理解することが必要不可欠な要素となってくる。

本研究では、人間の感情を理解するために表情に注目し、表情を認識するシステムを提案する。表情認識に関する研究は数多く行われているが、[1]では認識精度向上のために解決すべき課題の1つとして、表情の表出方法に関する個人差について指摘をしている。これは、不特定多数の人物のデータを使う従来の学習方法では、その人のクセや表情の程度といった個人特徴が反映されていないことを意味している。この問題を解決するためには、個人特徴を学習に反映する必要があるが、短期間で多くの個人の表情画像を収集するという、新たな問題が発生することになる。そこで本研究では、認識対象を1人に限定し、学習に個人特徴を反映しつつ、学習データを随時取得する形で、時間と共に学習性能を進化させるシステムを作ること目標とする。

## 2 提案手法

### 2.1 システム概要

提案システムの流れを図1に示す。処理の流れは、特徴点抽出部分、自己組織化マップを用いた学習部分、最近傍法を用いた認識部分の3つからなる。提案システムの最大の特徴は、ユーザと計算機が相互的なやり取りを行うことである。システムが間違っただけの認識をしても、ユーザ側が間違いを指摘、修正することができ、これにより、認識精度を向上させていく。

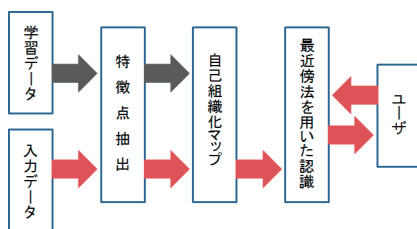


図1 提案システムの流れ

### 2.2 表情の特徴量

表情は、目や口などの顔パーツの形状や位置が変化することで、表出をする。そこで本研究では、顔パーツの形状や位置の変化を特徴量とし、学習に使用していく。目、口、眉に関する34個の特徴点を定義し、無表情とある表情の対応する特徴点の差を表情特徴量として計算していく。

### 2.3 特徴点抽出

本研究の特徴点抽出に関しては、Jason Saragihが開発したライブラリ[2]を用いている。これは、事前に特徴点情報を学習することで、リアルタイムに自動で特徴点の抽出を行うシステムである。このシステムでは、Constrained Local Models(CLM)

が利用されている。CLMは大きく分け、shapeモデルとpatchモデルの2つのモデルから構成されている。shapeモデルは、特徴点間の形状をモデル化したもので、 $n$ 個の特徴点の座標 $(x_i, y_i)_{i=1, \dots, n}$ から形状ベクトルが定義される。すべての学習データの形状ベクトルから平均形状を計算し、学習データに対して主成分分析を行うことでモデル化を行なっている。一方、patchモデルは、ある特徴点周りの輝度値などの情報をモデル化したものである。この2つのモデルを併用することで、入力データから特徴点を抽出している。

本研究ではモデル化にあたり、個人に限定した様々な顔の動きがあるデータを学習に用いる。個人のデータのみを学習に用いることで、抽出の性能を向上させている。

また、34個の特徴点の他に、鼻の下部分の特徴点も抽出を行う。これは、左右の目頭と鼻の下の特徴点を使い、正規化を行うためである。正規化の方法は、左右の目頭の距離と、目頭の中間点と鼻の下の特徴点との間の距離が一定になるように、また、顔の傾きが水平になるようにアフィン変換を行う。

### 2.4 自己組織化マップ

自己組織化マップ(Self-Organizing Map, 以下SOM)とは、Kohonenによって提案されたニューラルネットワークの一種であり、入力層と出力層の2層から成っている。SOMの特徴は、教師なし学習を行うことができる点である。さらに、高次元データ空間を低次元の空間に写像する働きもあり、クラスタリング問題によく用いられている。

出力層は $n$ 個のノードを持っており、ノードは1次元や2次元的に配置される。 $i$ 番目のノードには参照ベクトル $m_i$ と呼ばれるものが存在しており、 $m_i$ の次元は、入力層と同じ次元となる。図2にSOMの基本的な構造を示し、具体的なSOMの学習方法を以下に説明する。

1. すべての参照ベクトルをランダムな値で初期化
2. 学習データから入力 $x$ をランダムに選ぶ
3.  $\|m_c - x\| = \min_i \|m_i - x\|$ を計算し、勝者ノード $c$ を決める
4.  $m_i$ を $m_i + h_{ci}(x - m_i)$ で更新する  

$$h_{ci} = \alpha \exp\left\{-\frac{\|r_i - r_c\|^2}{2\sigma^2}\right\}$$
 $r_i$ :  $i$ 番目のノードの位置  
 $\alpha$ : 1より小さい正の整数  $\sigma$ : 近傍の広がり
5. 2に戻り、繰り返す

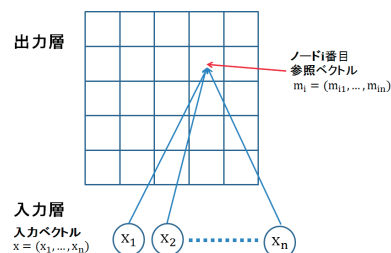


図2 SOMの基本構造

本研究では、出力層の大きさが $30 \times 30$ であるSOMを使い、

様々な顔の動きをデータとして学習を行う。学習データは、データベース [3] を使用する。このデータベースは、様々な表情を無表情から表情が表出するまでの連続データとしてまとめたものであり、この中からランダムに 1000 枚を選び、学習を行う。さらに、同じデータベースから表情の変化が大きい 309 枚の表情画像 (怒り, 嫌悪, 恐れ, 喜び, 悲しみ, 驚き) を用意し、先ほど学習した SOM に当てはめ、各ノードがどんな表情情報を持つのかを学習する (図 3)。

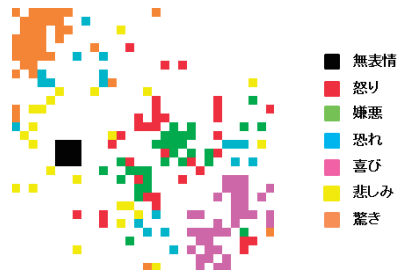


図 3 各表情の SOM 上の分布

### 2.5 最近傍法を用いた認識

最近傍法とは、テストデータに最も近い学習データを選び、その学習データが属するクラスにテストデータを分類する手法である。本研究では、近傍の  $k$  個の学習データを選び、 $k$  最近傍法を使用する。この部分では、認識した結果に関してユーザとの相互的なやり取りを行なっていく。入力データが与えられた場合の動作を以下に説明する。この部分は、似たデータを持つノードが近くに固まりやすくなる SOM 学習の特性を利用している。

1. 2.4 節で学習した SOM を使い、入力データがどのノードに近いかを計算
2. 求めたノードを中心に  $k$  最近傍法を行い、認識した結果をユーザに示す
3. ユーザは入力データと認識結果が正しいかどうかを判断
4. 結果をノードに反映

ユーザ側に結果の確認を求めた時点で、そのノードがどんな表情の情報を持つかは確定される。そこで、この情報を別に保持しておくことで、今後の判断に活用していく。さらに、計算機側からユーザ側に質問する形で、以前学習したデータが本当に正しいかどうかの再確認や、SOM のノード上でより曖昧な部分を確認するといった動作を追加する。これにより、認識効率や認識精度向上を行なっていく。

## 3 評価実験

### 3.1 実験内容

ユーザ 2 名の怒り, 嫌悪, 喜びの各認識率を実験し、評価する。例えば、喜びの認識率を求める場合、1 回の実験で計 8 枚 (喜びの表情を 3 から 5 枚, その他の表情を 3 から 5 枚) を入力データとし、実験を 4 回繰り返す。4 回の実験では、それぞれが前の実験の結果を保持した状態となる。すべての入力データは、微笑み, 喜び, 大喜びのように 3 種類の表情の程度が存在している。

また、比較実験として不特定多数のデータを用いたニューラルネットワーク (NN) を行う。この時、学習データとしてデータベース [3] から 309 枚の表情データを利用する。

評価方法は 1 つの表情に対して、3 つの程度を各 3 枚、計 9 枚を評価データとし、それぞれの認識率を計算する。

### 3.2 結果・考察

図 4, 図 5 に認識率を示す。実験回数が 0 回目のデータは、個人特徴を反映する前の状態に対する認識率である。結果を見ると、NN 法や個人特徴を反映する前のデータと比較して、ほとんどの場合で高い認識率を得ている。また、実験回数が増えるごとに認識率が高くなる、進化的な学習が行われているデータがある一方で、極端に認識率が高くなる場合や低くなる場合も確認できる。前者の原因としては、評価データの不足が考えられる。また後者は、違う意味を持つ表情でも似た特徴量になってしまうデータが存在することが原因だと考えられる。現在の特徴点の変化量を特徴量にする方法では、不十分であると言え、例えば、シワの有無といった別の情報を追加する必要がある。

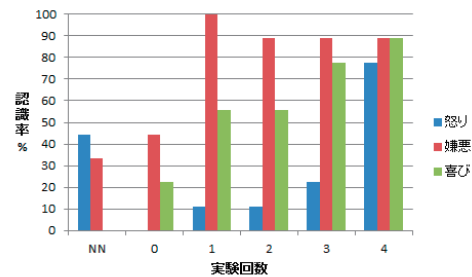


図 4 ユーザ A の認識率

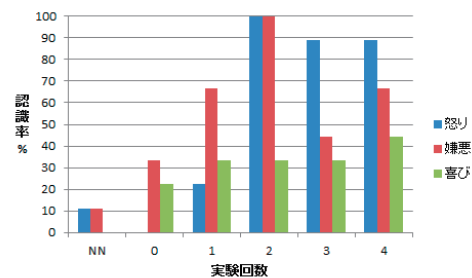


図 5 ユーザ B の認識率

## 4 おわりに

本研究では、個人特徴を反映する学習方法を提案し、評価を行った。今後は、シワ情報や時系列データといった新たな情報を追加した表情特徴量を検討する必要がある。また、複数の表情データが両立した確率的アプローチを検討することにより、さらなる認識率向上が期待される。

## 参考文献

- [1] 赤松茂: "人間とコンピュータによる顔表情の認識 [II] —コンピュータによる顔表情認識技術 (1): 表情による感情の認識—", 電子情報通信学会誌 85(10), 766-771, 2002-10-01
- [2] Jason Saragih: "6 Non-rigid Face Tracking", Mastering OpenCV with Practical Computer Vision Projects, PACKT, 189-233, 2012
- [3] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, "Extended Cohn-Kanade (CK+) database", <http://www.pitt.edu/~emotion/ck-spread.htm>