

ニューラルネットワークを用いた表情認識

正員 宮嶋 照行[†] 正員 長谷川孝明[†]
 正員 羽倉 幸雄[†] 正員 羽石 操[†]

A Facial Expression Recognition Using a Neural Network

Teruyuki MIYAJIMA[†], Takaaki HASEGAWA[†], Yukio HAKURA[†]
 and Misao HANEISHI[†], Members

[†] 埼玉大学工学部電気工学科, 浦和市
 Faculty of Engineering, Saitama University, Urawa-shi, 338 Japan

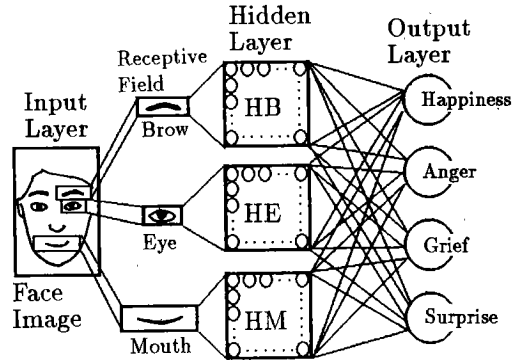


図1 表情認識方式
 Fig. 1 Facial expression recognition system.

あらまし 本論文では、ニューラルネットワークを用いた表情認識方式を提案する。特定の1人の顔の四つの表情の認識実験を行い、特にニューラルネットワークの構成法について検討し、提案する方式が有効であることを示す。

キーワード：ニューラルネットワーク、表情認識、パターン認識

1. まえがき

近年、表情認識の工学的研究が活発に行われている^{(1)~(7)}。密室で1人で作業をする人間の表情を監視し、その表情から精神状態や体調などを推定しそれに合わせて、警告したり、作業環境を調節するシステムなどにおいて、表情認識システムは有用なものであり、その構築が望まれる。

人間の脳に学んだ情報処理を行うニューラルネットワークは、その学習能力と並列計算性のために、パターン認識の分野で優れた能力をもつことが知られている^{(8),(9)}。表情認識にニューラルネットワークを用いることが有効であると考え、これまでにニューラルネットワークを用いた表情認識方式について筆者らは報告を行った⁽⁴⁾。筆者ら以外にも文献⁽⁵⁾、⁽⁶⁾などをはじめとし、盛んに注目されるようになってきた。ニューラルネットワークを用いて表情認識を行う場合、入力特徴量と、識別手法であるニューラルネットワークについての検討が必要である。しかしながらこれまでにニューラルネットワークの構成法については詳細な検討が行われていない。

そこで本論文ではニューラルネットワークを用いた表情認識方式⁽⁴⁾を提案し、ニューラルネットワークの構成法について検討を行う。提案するニューラルネットワークは中間層を三つに分割した3層ニューラルネットワークである。特定の1人の四つの表情の識別実験を行い本方式の性能を評価する。

2. 表情認識方式の提案

本章では、ニューラルネットワークを用いた表情認識方式について述べる。表情認識方式の構成を図1に示す。

2.1 入力特徴量

従来の表情認識の研究では、特徴量として、口の開き具合等のパラメータを用いる方法^{(1)~(3),(5)}、原画像データ、エッジデータ、2値データを用いる方法⁽⁶⁾、動画像のオプティカルフローを用いる方法⁽⁷⁾がある。しかしながら表情認識のための特徴量として決定的なものはまだ見つかっていない。

本論文では入力特徴量として2値画像について検討する。2値画像を用いることで特徴量抽出の手作業の軽減、メモリの節約、計算量の低減になる。

これまでの表情に関する研究^{(1)~(7)}より、おもに眉と目と口の情報から表情が特徴づけられると考えられる。従って本方式では、眉と目と口の部分の画像のみを用いる。

2.2 ニューラルネットワークの構成

本方式では3層のニューラルネットワークを用いる。中間層はそれぞれ眉、目、口のみから入力を受ける三つのニューロン群 HB, HE, HM に分割されている。各ニューロン群 HB, HE, HM 内のニューロンは、それぞれ入力層の眉、目、口の部分に受容野をもちこの部分のみから入力を受ける。出力層には認識対象の各表情に対応するニューロンがあり、それぞれ中間層のすべてのニューロンから入力を受けており、中間層の結果を統合して表情を識別する。ネットワークの学習法として、パターン認識の分野で広く用いられているバックプロパゲーション⁽⁸⁾を用いる。

3. 表情認識実験

2.で提案した本方式の性能を評価するため表情認識実験を行う。

3.1 表情データ

顔画像は、撮影時の顔とカメラ間の距離および位置関係や、イメージスキャナによる画像の取込み時の位置等が必ずしも一定にならないために、位置ずれと大きさの正規化が必要である。正規化は次の手順で行う。

(1) 基準点として、左右の目頭の位置 R , L を求める。

(2) R と L を結ぶ線分 RL の長さが一定値となるように、画像を等方的に拡大・縮小し、また線分 RL が一定位置にくるように、平行および回転移動を行う。

なお目頭の位置 R , L は抽出の自動化が困難なため、目視によって求めた。目頭の位置の自動抽出は今後の課題である。

顔画像の正規化後、各顔面特徴(眉、目、口)の切出しを行う。特定の一人の顔画像を対象としているため、正規化後の顔画像のほぼ同じ場所に各顔面特徴が位置すると考えられるので、すべての顔画像に対して同一の座標位置から各顔面特徴を切り出した。

特定の一人の顔の幸福、怒り、悲しみ、驚きの四つの表情を対象にする。表情データは各表情について6個ずつ(合計24個)採取し、学習データはそのうちの3個(合計12個)、評価データは残りの3個(合計12個)を用いる。各顔面特徴データのサイズは眉が64×24画素、目が48×24画素、口が88×32画素とした。

3.2 表情認識実験

学習の係数は0.1、慣性の係数は0.9とした。結合の重みの初期値を10種類変えて実験を行い、その平均値を認識結果とした。学習の終了条件は、出力の2乗誤差の総和が12回続けて0.1以下になったとき終了するものとした。

3.2.1 データサイズの検討

ニューラルネットワークへの入力データの次元は、計算量とメモリを少なくするために、できるだけ小さいことが望ましい。そこで3.1で述べたオリジナルのデータを S_1 、オリジナルのデータを縦、横それぞれ1/2に縮小したものを S_2 (眉: 32×12, 目: 24×12, 口: 44×16)、1/4に縮小したものを S_3 (眉: 16×6, 目: 12×6, 口: 22×8) とし、これら3種のデータについて適切なデータサイズについて検討する。

中間層の各ニューロン群のニューロン数を2個から20個まで変えて実験を行った。認識率を図2に示す。

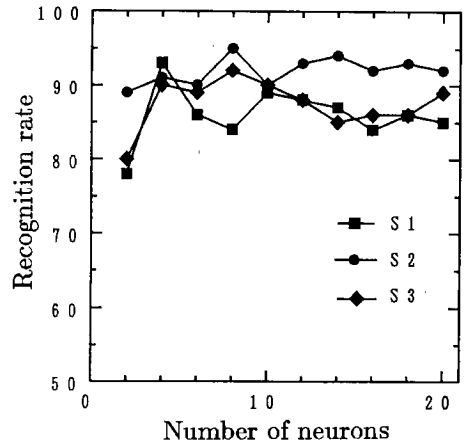


図2 ニューロン数と認識率の関係

Fig. 2 Relation between number of neurons and recognition rate.

ほとんどのニューロン数において、 S_2 は S_1 , S_3 と比較して認識率が高く、90%以上の十分な高い認識率が得られた。これは S_1 については、正規化時の若干のズレや、雑音などの影響で認識率が低下し、また S_3 については、過度の縮小により表情識別に必要な情報が失われたために認識率が低下したためと思われる。一方 S_2 については、 S_1 に含まれる悪影響成分を吸収し、かつ必要な情報を失わない程度に縮小が行われたため最も高い認識率が得られたものと思われる。

この結果より3.2.2では S_2 を用いて実験を行った。

3.2.2 中間層を分割しないニューラルネットワークとの比較

3.2.1の結果より本方式が十分な高い認識性能を有することがわかったが、次に本方式と、中間層を分割せずすべての中間層のニューロンが入力層から同じ入力を受けるネットワークとの性能を比較する。

計算量とメモリを最小限に抑えるために、中間層のニューロン数はできるだけ少ないことが望ましい。図2より、中間層の各ニューロン群のニューロン数が4個で認識率はほぼ飽和しているため、ここでは中間層の各ニューロン群のニューロンを4個とする。

比較するネットワークは、中間層のニューロンが4個で、調節可能な重みの数が本ネットワークとほぼ同数のネットワークとする。ネットワーク内の結合の重みの総数は(バイアスも含める)、本ネットワークの方が(中間層のニューロン数)×10だけ多く、厳密には二つのネットワークの規模は同じではない。しかし、その数は結合の重みの総数の0.7%程度であり、二つの

表1 中間層を分割しないネットワークとの性能比較

| ネットワーク | 認識率[%] | 学習回数 |
|---------|--------|------|
| 中間層3分割 | 91 | 488 |
| 中間層分割なし | 89 | 1203 |

ネットワークは同規模のネットワークと考えると差し支えない。

評価データに対する認識率と学習回数を表1に示す。認識率は両ネットワークともほぼ同じであるが、学習回数は本ネットワークが1/3程度少なくて済み、本ネットワークは各ニューロン群に四つのニューロンをもつため中間層全体で12個のニューロンをもつことになり、等価的にネットワークの規模が大きくなったために少ない回数で学習が済んだものと思われる。

また別の被験者の顔の表情データを用いて同様の実験を行った結果、ほぼ同様な結果が得られ本方式の被験者によらない有効性が確認された。

4. むすび

本論文では、ニューラルネットワークを用いた表情認識方式を提案し、その性能評価を行った。識別方法として中間層を三つに分割した3層のネットワークを用いた。表情認識実験を行った結果、評価データに対して90%以上の十分高い認識率が得られ本方式の有効性が確認された。中間層を三つに分割する本方式は、ネットワークの規模が同程度の中間層を分割しない方式よりも学習回数が少なくて済み、工学的に有効であることが確認された。

2値画像以外の入力特徴量に関する検討が今後必要である。また本論文で行った実験は、本方式の基礎的的性能を評価するため小規模であるが、更に複雑な表情の認識実験、および不特定の顔の表情認識方式への拡張も今後の課題として挙げられる。

文 献

- (1) Suwa M., Sugie N. and Fujimura K.: "A Preliminary Note On Pattern Recognition of Human Emotional Expression", Proc. of The 4th International Joint Conference on Pattern Recognition, pp. 408-410 (1978-11).
- (2) 石井威望, 岩田洋夫: "コンピュータ画像処理を利用した顔の表情の自動認識", 日本機械学会論文集(C), 52, 483, pp. 2989-2992 (1986-11).
- (3) 石井威望, 広瀬通孝, 岩田洋夫: "画像理解による運転者の表情の自動認識", 自動車技術, 41, 12, pp. 1398-1403 (1987-12).
- (4) 宮嶋照行, 長谷川孝明, 羽倉幸雄: "ニューラルネットを用いた顔の表情認識", 1990 信学秋季全大, D-408.
- (5) 小林 宏, 原文雄: "ニューラルネットによる人の基本表情認識", 信学技報, HC90-29 (1991).
- (6) 大橋洋史, 徐 剛, 辻 三郎: "ニューラルネットによる表情認識—学習データに関する考察", 信学技報, PRU91-4 (1991).
- (7) 間瀬健二, 前田英作, 末永康仁: "表情動画像からの感情の認識の1手法", 信学技報, PRU91-24 (1991).
- (8) Rumelhart D. E., Hinton G. E. and Williams R. J.: "Learning representation by back-propagating errors", Nature, 323, pp. 533-536 (1986).
- (9) Lippmann R. P.: "Pattern Classification Using Neural Networks", IEEE Commun. Mag., 27, 10, pp. 47-64 (1989).

(平成3年9月11日受付)