

連想記憶モデルによる顔の表情認識の研究

戴瑩† 中野康明† 柴田義孝† 橋本浩二†

† 岩手県立大学 ソフトウェア情報学部 ソフトウェア情報学科

† 信州大学工学部 情報工学科

本論文では、視覚の MRF モデルで形成した顔画像の連想型ホップフィールドネットと反学習との併用による顔画像の表情認識手法について、詳しく述べた。

まず、顔の標準パターンを連想型ホップフィールドネットに与えて学習させる。認識方法として、まず顔の未知サンプルの連想性想起パターンと各標準パターンによる連想性想起パターンとの類似度を計算し、最大類似度からある範囲に入る類似度を与える標準パターンを候補サンプルとして選ぶ。この候補サンプルにより新たに連想型ホップフィールドネットを構成して学習を行い、これを用いて類似度を求め、判別規則により認識結果を求める。

1 まえがき

計算機による顔画像の自動認識の研究はテレビ会議、human-computer interface による感情通信、その他で重要である。

これまでも顔識別について、多くの研究成果が得られている。Harmon[1] の先駆的な業績では、顔の輪郭を用いて認識している。阿部ら [2] は赤外線スキャナを用いて得た顔の 3 次元データを用い、3 次元空間内での回転を行なって、横顔を求め、この横顔輪郭を用いて認識している。

さて、画素レベルの濃淡情報を用い、統計的手法で顔画像を認識する方法などが提案されている。Kurita ら [3] は位置不変な特徴を持つ高次局所相関特徴を利用して、顔の認識を行なった。Dai ら [5] はホップフィールドモデルの連想特性を利用して、低画質顔画像の認識手法を提案した。

最近、Shibata ら [6][7] は感情の個人差のモデルを用いて表情分析および合成を行うことで、より正確な感情の伝達を可能とする感情通信の概念を提案し、実装してきた。

本論文では、ホップフィールドモデルの連想特性を利用して、顔画像に対する顔の表情認識手法を提案する。まず、連想性想起能力を有するホップフィールドモデルについて、反学習修正係数と記憶パターン数の影響を検討する。次に、顔画像をホップフィールドネットに入力して得られる連想性想起パターンと標準パターンの連想性想起の類似度を基にし、顔の表情を認識する。さらに、顔画像における表情認識のいくつかの実験結果を示す。

2 顔画像におけるホップフィールドネットの構成

2.1 ホップフィールドネット

ホップフィールド型ニューラルネット [8] は、相互結合型のネットワークであって、ニューロン間の結合における結合荷重が対称であり、自分自身への結合はないものと定義される。以下、簡単のため、これをホップフィールドネットと呼ぶ。

ニューロンの総数を N とすると、ホップフィールドネットにおける i 番目のニューロンの入力 U_i は次式で与えられる。

$$U_i = \sum_{j=1}^N W_{ij} V_j + I_i \quad (1)$$

また、ニューロンの出力 V_i は次式で与えられる。

$$V_i = f(U_i) = \frac{1}{1 + \exp(-2\lambda U_i)} \quad (2)$$

ここで、 W_{ij} はニューロン j からニューロン i への結合荷重、 I_i はニューロン i の入力バイアス、 λ は式 (2) で定義されるシグモイド関数のこう配を決定するパラメータである。なお、 $W_{ij} = W_{ji}$ 、 $W_{ii} \equiv 0$ 。

ホップフィールドネットは、次式で与えられるエネルギー関数を極小化する性質がある。

$$E = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N W_{ij} V_i V_j - \sum_{i=1}^N I_i V_i \quad (3)$$

すなわち、ニューロンをランダムに 1 個選択し、そのニューロンの出力を式 (2) に基き変化させることを繰り返すと、ホップフィールドネットでは

(1) 時間の経過とともにエネルギーが必ず減少する

方向に状態が遷移する。

(2) エネルギーの減少が止まるのは全てのニューロン状態の変化が止まった安定状態である。

という二つの性質がある。

このホップフィールドネットの結合荷重を適当に与えれば、連想性記憶の性質を示す神経回路網モデルになることも知られている。

画像をホップフィールドネットとして表すこともできる。すなわち、最も単純なモデルでは、一つの画素に相当する一つの処理モジュール $s_{i,j}$ の状態は、外部入力と前後左右にの画素に相当する処理モジュール $s_{i-1,j}$, $s_{i+1,j}$, $s_{i,j-1}$, $s_{i,j+1}$ から受ける相互作用に基いて定まる。したがって、これは簡単な相互結合型ネットワークである。

このようにホップフィールドネットとして表現したモデルは図1で表される。あるニューロン (i,j) と周囲4個のニューロンの間の結合荷重は $W_{(i,j)(i-1,j)}$, $W_{(i,j)(i+1,j)}$, $W_{(i,j)(i,j-1)}$, $W_{(i,j)(i,j+1)}$ であり、これら以外のニューロンとの結合荷重はすべて0になる。

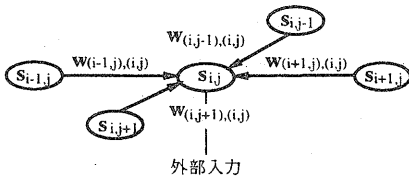


図1: ホップフィールドネット

連想性想起能力を持たせるため、結合荷重を次式のように定めることが試みられている [8]。

$$W_{(i,j)(i',j')} = \frac{1}{N} \sum_{m=1}^M s_{(i,j)}^m s_{(i',j')}^m \quad (4)$$

ここで、 (i,j) と (i',j') は隣接したニューロン、 N はニューロンの総数、 $s_{(i,j)}^m$ は記憶させたい m 番目のパターンの第 (i,j) 画素の値である。 M は記憶させたいパターンの数である。

式(4)からわかるように、結合荷重は対称で、自分自身への結合は0であるので、ホップフィールドネットの条件を満たす。このホップフィールドネットは自己想起型連想記憶モデルと呼ばれる。本論文では簡単のため連想型ホップフィールドネットと呼ぶことにする。

2.2 連想型ホップフィールドネットによる顔画像の記憶と想起

前節で述べた連想型ホップフィールドネットに学習用顔画像を与え、式(4)によって結合荷重 $W_{(i,j)(i',j')}$ を学習させる。そのとき、 M は記憶させる顔の総数である。

顔画像の認識のためには、顔画像の分解能の最低限界は 16×20 画素であるという予備的考察を得ている [4]。そのため、本論文では、学習用顔画像は 16×16 画素とした。1個のニューロンが1画素に対応するので、ニューロン数 N は256個である。

実際に用いた 16×16 画素の顔画像は比較的高い分解能の原画像からモザイク化、すなわちぼかしと再サンプリング、を行うことにより作成した。

学習の完了した連想型ホップフィールドネットにある顔画像を外部入力として与える。システムに非同期更新の状態遷移を行なわせると、2.1節で述べたように安定状態で止まる。安定状態におけるニューロンの各々がこの入力顔画像に対する連想性想起パターンを与える。

M が1であるとき、すなわち、連想型ホップフィールドネットの記憶パターンが入力顔画像の標準パターンのみであるとき、この標準パターンを与えたときの連想性想起結果は原画像の情報を完全に想起し、入力パターンと想起パターンの間の類似度は1である。

ここで、類似度は

$$u = \frac{\sum_x \sum_y (f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sqrt{\sum_x \sum_y (f(x,y) - \bar{f})^2 \sum_x \sum_y (t(x,y) - \bar{t})^2}}$$

で定義され、 $f(x,y)$ と $t(x,y)$ は二つのパターンを、 $\bar{f}$ と $\bar{t}$ はそれぞれの平均値を表す。



図2: 入力標準パターンとその連想性想起 ($M=1$)

図2は、顔の標準パターン画像と、 M が1の連想型ホップフィールドネットによる連想性想起パターンを示す。図2から、記憶した画像が濃淡レベル差

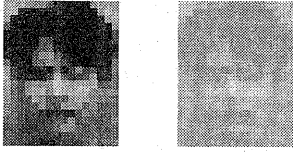


図 3: 入力試験パターンと
その連想性想起 (M=1)

を除いて完全に想起されていることがわかる。このとき二つのパターン間の類似度は 1 である。

図 3 は図 2 に示した顔の試験パターンすなわち同一人の別画像と、 M が 1 のときの連想型ホップフィールドネットによる想起結果を示す。

このとき、同一人の標準パターンとこの試験パターンとの類似度は $u = 0.73$ であった。一方、標準パターンの連想性想起パターンと試験パターンの連想性想起パターンとの類似度は $S = 0.84$ であった。この二つの類似度の比較から、顔画像パターンを直接使用するより、連想性想起パターンを用いた方が、未知パターンと標準パターンとの類似性が高くなる傾向があることがわかる。

M が 1 より大きいとき、連想型ホップフィールドネットは複数の顔画像を記憶している。この場合、ある試験パターンを入力として与えると、正しくない想起を行うことがある。この現象を“にせの記憶”と名づける。にせの記憶を除去するために、Hopfield らの反学習 (unlearning) [8] という手法を用いた。すなわち、式 (4) により結合荷重 $W_{(i,j)(i',j')}$ の初期値を選び、以下のような修正を行なう。まずランダムに状態を選び、その状態を系のダイナミクスに従って遷移させる。その結果得られた状態 $s^r = (s_1^r, \dots, s_N^r)$ の値を用いて

$$\Delta W_{(i,j)(i',j')} = -k s_{(i,j)}^r s_{(i',j')}^r \quad (5)$$

を結合荷重 $W_{(i,j)(i',j')}$ に加えることにより、にせの記憶を忘却させるのである。

3 ホップフィールドモデルの認識性能

3.1 反学習修正係数 k について

前節で述べたように連想型ホップフィールドネットでは、標準パターンを記憶した後で、未知パターンを与えると、その人の標準パターンに良く類似したパターンを想起する。したがって、顔画像認識に

使用できる可能性がある。

前節では、にせの記憶を除去するために、式 (5) の反学習を採用した。しかし、この反学習はにせの記憶を除去すると同時に、入力顔画像の連想性想起の有効性を損なう可能性がある。これは顔の認識を行う場合に不利となる。にせの記憶を減じ、顔想起の有効性を最大限に保つために、反学習係数 k を最適に設定する必要がある。ここで、 k の値と顔の連想性想起による認識性能との関係について述べる。

記憶させた標準パターンと、この顔を入力したときの連想性想起パターンの間の類似度を u とする。連想型ホップフィールドネットに記憶された顔の総数 M が 4 のとき、 u の反学習係数 k による変化を図 4 に示した。●—● は 1 番目の顔画像、×—× は 2 番目の顔画像に対応する。

図 4 から、 u は、 k が増えるにつれて一旦上昇し、その後低下することがわかる。1 番目の顔では、 $k = 0.01$ のとき、類似度 u が最大になる。反学習係数 k は、本来連想型ホップフィールドネットのにせの記憶を除去するためのものであり、 k を大きくすると、除去の効果として類似度 u の値が大きくなる。しかし、 k が大きくなり過ぎると、反学習はにせの記憶を除去するとともに、顔の有効情報を消す副作用も大きくなる。その結果、類似度 u は小さくなると解釈される。

図 4 で、2 番目の顔に対して、類似度 u と反学習係数 k との関係は 1 番目の顔と同じ変化傾向を示す。ただし、 u を最大にする k は 0.02 である。すなわち、反学習係数 k は、顔が異なると、異った最適値を取る。

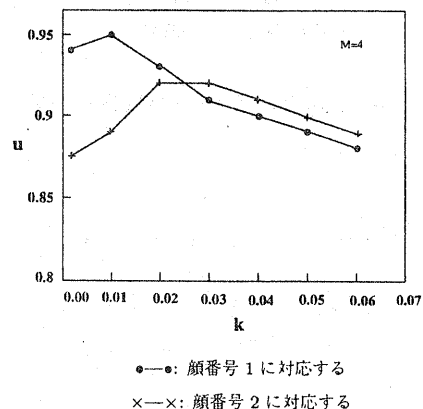


図 4: 類似度 u と k の関係 ($M=4$)

図 5 は M が 11 のとき、1 番目及び 2 番目の顔の類似度 u と反学習係数 k の間の関係を示す。 M が

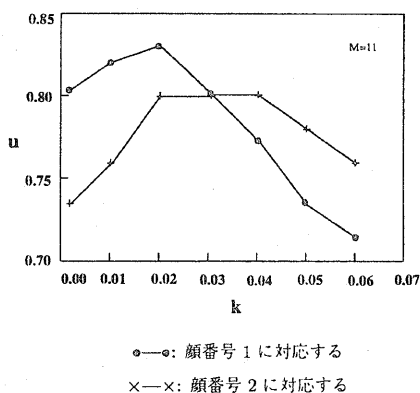


図 5: 類似度 u と k の関係 ($M=11$)

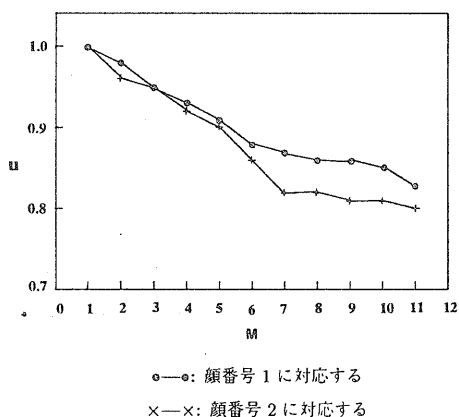


図 6: 類似度 u と記憶パターン数 M の関係

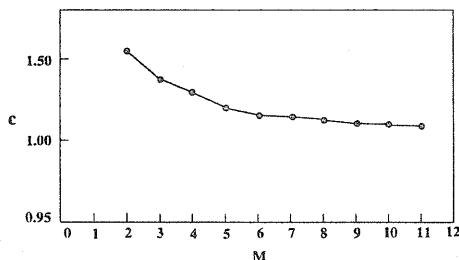


図 7: d と M の関係

11 の場合も、 u と k の間の関係は、 M が 4 のとき同様であるが、 u を最大にする k の値は M が 4 の場合と少し異なり、また、その最大値自体も減少していることがわかる。

以上の知見から、連想型ホップフィールドネットには顔の最適連想性想起を与える反学習係数 k があり、それぞれの顔に対して異った最適値があることがわかる。

この実験的考察により、できるだけ、にせの記憶を除去し、顔の有効情報を保持するために、反学習係数 k の設定は次のようにする。

1. 記憶させた顔の各々に対して、この顔と連想型ホップフィールドネットによる連想性想起パターンの間の類似度を最大にする k を求める。
2. これらの k のなかで、値が一番大きい k を反学習係数として採用する。

このようにすれば、連想型ホップフィールドネットが記憶した顔について、にせの記憶の除去と顔パターンの有効情報の保持の最適なトレードオフが得られる。

3.2 記憶パターンの数 M について

3.1 節で述べたように、連想型ホップフィールドネットに記憶された顔の総数 M が大きくなると、想起された顔と元々の顔との類似度は低下する。この節では M と類似度 u の関係をさらに検討する。

図 6 は、 M を変化させた場合の u の変化を示す。 $\circ$ — $\circ$ 及び $\times$ — $\times$ は、それぞれ 1 番目と 2 番目の顔に対応する。

図 6 でわかるように、 M が大きくなるに伴って、 u が小さくなる。すなわち、記憶パターンの数が大きいほど、連想性想起パターンは別の記憶情報から妨害を大きく受けて、自分の顔との類似度が小さくなる。記憶パターンの数が小さいほど、顔の連想性想起パターンは標準パターンとの差が小さく、顔の本来の様子をよく保持している。

想起能力の信頼性を評価するため、接近係数 d を次のように定義する。

$$d = \frac{u_s - u}{u_s} \quad (6)$$

ある顔の試験パターンの想起において、その想起パターンと、この標準パターンによる想起パターンとの類似度を u_s とし、ほかの標準パターンによる想起パターンとの類似度を u とする。そうすると、 d が大きいほど、ネットが第一候補に自信を持って想起していることを示す。

図7は M を変えたとき、1番目の顔についての d の変化を示す。当然のことではあるが、 M の増加に伴って、 d が小さくなり、想起の信頼性は低下する。すなわち、連想型ホップフィールドネットは顔の記憶数が少ないほど、未知サンプルに対する分類性能は高い。興味深いことは、 $M > 6$ のとき、 d の値はほとんど変わらない。すなわち、 M がある程度大きくなると、多少 M を減らしたところで分類性能は改良されない。

いろいろな顔のサンプルに対する結果も同様であった。そのような M の値、すなわち d が漸近値に到達するような M の値を、連想型ホップフィールドネットの M の臨界値と呼ぶ。

4 ホップフィールドモデルによる顔の表情の認識

前節では、連想型ホップフィールドネットに記憶された顔の数が少ないほど、未知サンプルの認識の信頼性が高くなることを示した。さらに、 $M > 6$ のとき、多少 M を減らしたところで分類性能は改良されない。本節では、この性質を利用した顔の表情の認識手法を提案する。

感情が主に6つの表情(喜び、悲しみ、恐れ、怒り、嫌悪、驚き)とそれぞれの感情の度合で現れる(図8のように示す)。この6つの標準表情を学習セットとして連想形 Hopfield Net に記憶させれば、前述の性質が利用できると考える。この原理に基く、顔の表情手法の流れ図を図9に示す。

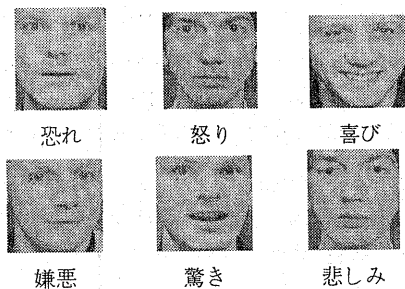


図8: 顔の表情パターン

まず、表情を認識したい入力顔画像を試験パターンとして、表情の標準パターン6つとの類似度をそれぞれ計算し、最大類似度を求める。この最大類似度に対応するある表情の標準パターンを第一候補 (C_1) として選び、次の類似度に対応するもう一つの標準パターンを第二候補 (C_2) として選ぶ。

表1: 表情認識の判別手法

M	combination of 1st and 2nd candidates at each step				
	C_1, C_2 including other(s)	C_1, C_2 C_1, C_2	C_1, C_2 C_1, C_2	C_1, C_2 C_2, C_1	C_1, C_2 C_2, C_1
1		C_1, C_2	C_1, C_2	C_1, C_2	C_1, C_2
2		C_1, C_2	C_2, C_1	C_2, C_1	C_1, C_2
decision	reject	$ d_2 < d_6 \Rightarrow C_2$ $ d_2 > d_6 \Rightarrow C_1$	C_2	C_2	reject

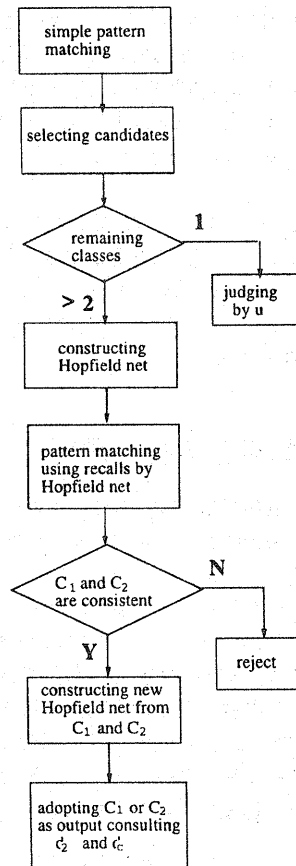


図9: 顔の認識手順

次に、顔表情の標準パターンを連想型ホップフィールドネットに与え、式(4)、(5)によって学習させる。この後、認識したい入力顔画像に対して、これによる連想性想起パターンと、学習された M 個の標準パターンによる連想性想起パターンとの類似度をそれぞれ計算し、新しい第一候補 (C'_1) と第二候補 (C'_2) を決める。 C'_1 , C'_2 が C_1 , C_2 と異なる場合、この試験パターンを reject する。もしそうではなければ、式(6)で定義した d の値を計算する。次に、 C'_1 , C'_2 に対応する二つの標準パターンを新しい Hopfield Net に与え、式(4)、(5)によって学習させる。その後、先の試験パターンに対して、これによる連想性想起パターンと、現在の二つの標準パターンの連想性想起パターンとの類似度をそれぞれ計算し、新しい第一候補と第二候補を決めて、 d の値を求める。

表1に未知顔の表情を判別するアルゴリズムを示す。記述を簡単にするために、 $M=2$ の場合の d の値は d_2 と表し、 $M=6$ の場合の d の値は d_6 と表す。 $M=2$ と $M=6$ のときに、 C_2 がどちらも第二候補の場合には、もし $|d_2|$ が $|d_6|$ より小さければ、 C_2 を認識結果とする。そうでなければ、 C_1 が認識結果とする。

$M=2$ と $M=6$ のとき、 C_2 をどちらも第一候補になる場合には、 C_2 が認識結果とする。そうでなければ、試験パターンを reject する。

$M=2$ のときは、 C_2 が第一候補になる場合には、 C_2 を認識結果とする。

5 実験結果と今後の課題

今回行った表情認識実験は複数の人間の顔を対象とした。顔の表情における学習セットは図8に示したものである。

図10に、学習サンプルに含まれない試験顔画像(恐れ)と $M=2$ と $M=6$ のときのこの顔画像の連想性想起を示す。表2には各stepの計算結果と認識結果を示す。この結果から顔の表情を判別するアルゴリズムの有効性が示された。このサンプルの他に、5つの顔画像を試験パターンとして、顔表情の認識実験を行った。結果は全部正解である。

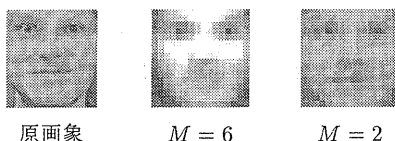


図 10: 顔画像とこの連想性想起

表 2: 認識結果

候補の状態			$ d_6 $	$ d_2 $	結果
$M=1$	$M=6$	$M=2$			
C_1	C_1	C_1	0.03	0.15	恐れ

いままで試験した顔の sample 数はまだ不十分である。今後の課題として、もっとたくさんの顔画像を用いて、表情認識の実験を行い、結果を分析することが挙げられる。

謝辞

実験データの一部として Olivetti Research Laboratory から公開されている ORL データベースを使用した。

参考文献

- [1] L.D.Harmon et al. "Automatic Recognition of Human Face Profiles", Computer Graphics and Image Processing 6, pp. 135-156 (1977).
- [2] 阿部他:" 3次元データより得られる横顔輪郭線を用いた人間の顔の自動識別", 信学論 D-II, J73-D-II, pp.1477-1484(1990)
- [3] T.Kurita, et al. "A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis", The Proceedings of 11th International Conference on Pattern Recognition, pp.213-216, (1992).
- [4] Y. Dai, et al. "A Study of Face Recognition with Low Quality Images", The Proceedings of ICARCV'94, pp.1442-1446(1994), Singapore.
- [5] Y. Dai, et al. "Recognition of Facial Images with Low Resolution Using a Hopfield Model", Pattern Recognition, pp159-167, (1998).
- [6] Y. Shibata, et al. "A User Model for Feeling Communication System", 東洋大学工学部研究報告 第31号, pp75-84, (1995).
- [7] 坂本他:" 顔画像を用いた人間・エージェント共生空間実現のためのインタフェース", マルチメディア通信と分散処理 87-27, pp155-160, (1998).
- [8] 乾 敏郎:" 認知と学習", 丸善株式会社 (1993).