

方向線素特徴とノイズ重畳を用いた ニューラルネットワークによる手書き文字認識

大山輝光[†] 黒田英夫^{††} 宮原末治^{††}
志久^{†††} 修^{†††} 高比良^{†††} 秀彰^{†††}

本論文では、手書き文字認識などの実問題へ階層型ニューラルネットの汎化能力を応用するために、方向線素特徴量とノイズ重畳を併用する手法の解析的・実験的検証を行う。ノイズ重畳の効果が学習初期の探索的な振舞いにあるとの解析的考察に基づき、ノイズ量を変えながら手書き文字認識実験を行った結果、特に学習データが少ないときの有効性を確認した。また、学習後の出力素子の平均的な振舞いからノイズ量と構造化との関係を調べた結果、より大規模な実験に本手法を適用するためのノイズ量決定法へ応用可能であることを確認した。

Handwritten Character Recognition Using Directional Element Feature and Noise Injection into Neural Network

TERUMITSU OHYAMA,[†] HIDEO KURODA,^{††} SUEHARU MIYAHARA,^{††}
OSAMU SHIKU^{†††} and HIDEAKI TAKAHIRA^{†††}

To apply the generalization ability of multi-layered neural networks to handwritten character recognition, we clarify the efficacy of directional element feature learning and noise injection into inputs by a theoretical analysis and simulation. As a result of the recognition experiment of handwritten character database ETL9B, the efficacy of the noise injection was confirmed. Then, we discuss an experimental consideration about outputs of output units with noise injection.

1. はじめに

階層型ニューラルネットを用いたパターン認識では、未知パターンに対して適切な出力を与える汎化能力の改善が課題となっている。階層型ニューラルネットの汎化能力向上に関しては、学習時の入力にノイズを重畳する方法^{1),2)} や、位置ずれ画像を学習に用いる方法³⁾ などがある。また実問題にニューラルネットを適用する場合、限られた訓練データから有効な特徴量を抽出し、より高い汎化能力を獲得する必要がある。本論文では手書き文字認識などへ階層型ニューラルネットの汎化能力を応用するために、ノイズ重畳と方向線素特徴量とを併用する手法の解析的・実験的検証を行う。そして、方向線素特徴とノイズ重畳を併用することで、

特に訓練データが少ない場合において有効であることを示す。最後に、素子の平均的な出力を調べニューラルネットの構造化に対する考察を行う。

2. ETL9B 認識実験

階層型ニューラルネットの学習時にノイズを重畳すると、学習初期にはノイズの影響を大きく受けて学習がゆっくりと探索的に進み、最終的にはノイズの影響が無視できることが解析的に示されている(付録 A.1 参照)。そこで実問題への応用に向け、方向線素特徴の学習時に付加するノイズ量と学習パターン数、特徴のオーバーラップとの関係を定量的に調べるため、これらを変化させながら手書き文字認識実験を行う。実験システムはノイズ除去と正規化、輪郭線抽出を行う前処理部と、水平、垂直、 ± 45 度の4種類の方向線素特徴を算出する特徴抽出部、そしてニューラルネット認識部の3つによって構成される。方向線素特徴抽出部では、 64×64 [pixel] の文字データを 8×8 [pixel] の格子状領域(以下、受容野)に分割し、受容野ごとに4方向の特徴量を求めることで1文字につき256次

[†] 和歌山信愛女子短期大学
Wakayama Shin-Ai Women's Junior College

^{††} 長崎大学工学部
Faculty of Engineering, Nagasaki University

^{†††} 佐世保工業高等専門学校
Sasebo National College of Technology

元のベクトルで構成される方向線素特徴量が求められる。さらに、有効な特徴を増やすため受容野を互いに4[pixel]ずつオーバーラップさせ特徴量を算出する。この場合、受容野数は 15×15 個となり方向線素特徴量は1文字につき900次元となる。認識部では、入力-中間-出力層の素子数が256-256-100と、900-256-100の2種類の3層ニューラルネットを用い、学習方法には一般的な誤差逆伝播学習法を用いる。

2.1 実験データ

入力データとしてETL9B手書き文字データの中から100字種 \times 200パターン(200筆跡)、合計20,000個を選んで使用する。学習データ数の変化に対する未知データの認識率とノイズ量との関係を明らかにするため、200パターンの各文字について前半20~100パターンを学習用、後半100パターンを汎化能力評価用の未知データとして使用する。これらに前処理を施し(学習パターン, 未知パターン)=(前半20パターン, 後半100パターン), 以下(40, 100)(60, 100), (80, 100)(100, 100)の5種類のデータセットを作成する。未知データは全データセットで共通である。したがってデータセット(20, 100)を用いた実験の場合、2,000個(前半20パターン \times 100字種)の学習を行い、共通の10,000個(後半100パターン \times 100字種)の未知データを提示して認識率を調べる。そして5,000回の学習を行いながら、50回学習ごとに未知データに対する認識率を記録する。また重量するノイズは乱数によって作成し、その最大値を0(ノイズなし)から6の範囲で変化させながら中間素子の入力へ付加する。

2.2 方向線素特徴の学習におけるノイズの効果

中間素子の入力に重量するノイズ量を変えながら学習を行った結果を図1に示す。受容野のオーバーラップは0、受容野数は 8×8 個である。

データセット(20, 100)に対する認識率はノイズなしの場合93.97%、ノイズ2のときに最も高い95.08%であった。データセット(100, 100)に対する認識率はノイズなしの場合97.24%、ノイズ2のときに最も高い97.44%であった。また、図のようにノイズ4の場合、ノイズ2のときに比べ認識率が低下している。このように、ノイズによって特に学習データが少ないときの認識率が改善されている。解析的な考察では、学習初期にはノイズの影響を大きく受け、学習の収束にともなってノイズの影響が無視できるようになることが示されたが、実験の結果からさらに、学習データが少ないときには変動パターンを入れて認識率を向上させる効果があることが推測できる。

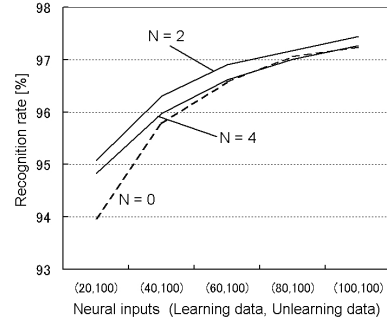


図1 中間素子の入力へのノイズ重量による認識率の変化

Fig. 1 The relation between the directional element feature learning and the recognition rate by using noise injection into inputs of hidden units.

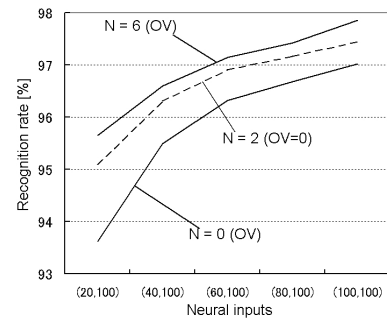


図2 方向特徴のオーバーラップによる認識率の変化

Fig. 2 The relation between the overlap of the directional element feature learning and the recognition rate.

2.3 方向線素特徴のオーバーラップ効果

同じ大きさの受容野を4[pixel]ずつ重ねて方向線素特徴量を算出し、ノイズ量を0から8の範囲で変化させながら学習を行った結果を図2に示す。図中、 $N=8$ のグラフは $N=2$ のものと同様の部分が多いため省略している。データセット(20, 100)に対する認識率はノイズなしの場合93.62%、ノイズ6のときに最も高い95.65%であった。また、ノイズ量が8の場合、認識率は95.45%とノイズ6のときに比べ低下した。データセット(100, 100)に対する認識率はノイズなしの場合97.02%、ノイズ6のときに最も高い97.86%であった。このように、特徴のオーバーラップによる認識率向上に加え、ノイズによって学習データが少ないときの性能が向上している。この実験からも、ノイズ重量には学習データが少ないときに変動パターンを入れて認識率を向上させる効果があることが示された。

3. ニューラルネットの構造化に対する考察

学習後のニューラルネットがどのように構造化されているかを調べるため、素子の平均的な出力を次のような評価基準で算出し、ニューラルネットの構造化

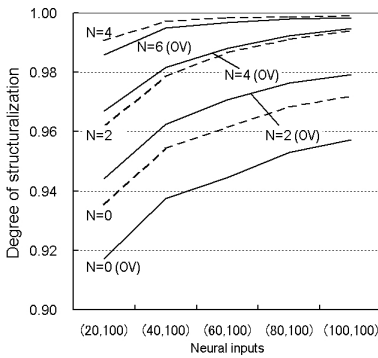


図3 ノイズ重畳による構造化度の変化

Fig. 3 The relation between the degree of structuralization and noise injection.

に対する考察を行う(以下、この評価値を構造化度と呼ぶ)。

$$W = \frac{\sum_k g\left(\gamma \left| \sum_j b_{kj} \right| \right)}{K} \quad (1)$$

b_{kj} は中間素子 j ($j = 1, 2, \dots, J$) と出力素子 k ($k = 1, 2, \dots, K$) 間の結合荷重, g は出力素子の入出力を規定するロジスティック関数である. $\left| \sum_j b_{kj} \right|$ は, 出力素子 k と全中間素子との結合強度(出力素子の受容野結合)の絶対値である. また, γ は微小定数で今回は $\gamma = 0.2$ とした. 実験で用いるニューラルネットは, 学習開始時に平均が 0 の同じ乱数値で初期化されるため構造化度は最小値の 0.5 である. そして学習の進展によって出力素子の平均的な出力が 0 か 1 になるように構造化されるほど W は 1 に近づく. 2.2 節および 2.3 節での実験後の構造化度を図 3 に示す. 図中, 点線は 2.2 節の結果, 実線はオーバーラップした特徴を用いた 2.3 節の結果である. 図のようにノイズを重畳することで構造化度 W が 1 に近づいている. これは出力素子の平均的な出力が 0 か 1 に確定的になるように構造化されていることを示している. 未知データの認識率が最も高いのは 2.2 節の実験ではノイズ 2 付近, 2.3 節の実験ではノイズ 6 付近であった. またノイズ 0 の場合, オーバラップした場合よりオーバラップなしの構造化度が高く, 認識率も良好であった. 入力素子数が異なるためオーバラップありとなしの構造化度を直接比較することはできないが, どちらの場合もデータセット (20, 100) の構造化度が 0.98 ~ 0.99 付近で認識率が最も高く, 構造化度が大きすぎると認識率が低下している. ニューラルネットによる文字認識システムの性能を比較する場合に多く用いられるのは, 学習パターンと未知パターンを 50% ずつとしたデータ

セット (100, 100) の認識率である. したがって, 今回の研究結果に基づき, 構造化度 0.98 ~ 0.99 を基準にデータセット (20, 100) の認識率が最大化されるようにノイズ量を決定し, データセット (100, 100) の認識に用いることで, より大規模な実験にノイズ重畳法を適用する場合の計算量低減が期待できる.

4. おわりに

本論文では実問題へ階層型ニューラルネットの汎化能力を応用するために, 方向線素特徴量とノイズ重畳を併用する手法の解析的・実験的検証を行った. また, 学習後のニューラルネットにおける出力素子の平均的な出力を調べ, ノイズ量と構造化に対する考察を行った. まずノイズ量を変えながら手書き文字認識実験を行い, 特に学習データが少ないときの性能向上に対する有効性を確かめた. 次に, 方向線素特徴量をオーバーラップさせて有効な特徴数を増やし機能の向上を確認した. 最後に, 学習後の出力素子の平均的な振舞いを調べてノイズ量と構造化との関係を考察し, より大規模な実験に本手法を適用する場合のノイズ量決定法への応用が期待できることを示した.

今後, 位置ずれや方向特徴のメッシュサイズ変更によって有効な特徴を増やすなど, さらなる認識率向上を検討したい. また, 構造化度を利用した適切なノイズ量決定法を確立し, より大規模な認識実験を行う予定である.

参考文献

- 1) Matsuoka, K.: Noise Injection into Inputs in Back-Propagation Learning, *IEEE Trans. Systems, Man and Cybernetics*, Vol.22, No.3, pp.436-440 (1992).
- 2) 栗田多喜夫, 麻生英樹, 梅山伸二, 赤穂昭太郎, 細美章隆: 多層パーセプトロンの学習における中間層に付加したノイズの影響とネットワークの構造化, *電子情報通信学会論文誌*, Vol.J79-D-2, No.2, pp.257-266 (1996).
- 3) 易 建強, 黒木秀一, 松岡清利: 誤差逆伝搬法による位置ずれパターンの学習, *電子情報通信学会論文誌*, Vol.J74-D-2, No.1, pp.27-35 (1991).

付 録

A.1 中間素子の入力に対するノイズ重畳の解析
入力素子 I 個, 中間素子 J 個, 出力素子 K 個から構成される 3 層ニューラルネットについて考える. 本論文では従来の議論をより一般化し, 入力素子の入出力を線形関数, 中間素子と出力素子の入出力を微分可

能な任意の関数 f と g によって次のように規定する .

$$y_j = f(\xi_j), \xi_j = \sum_{i=1}^I a_{ji}x_i + a_{0j} \quad (2)$$

$$z_k = g(\eta_k), \eta_k = \sum_{j=1}^J b_{kj}y_j + b_{0k} \quad (3)$$

ξ_j と η_k は中間素子 j と出力素子 k への入力, a_{ji} と b_{kj} は入力 ~ 中間および中間 ~ 出力層間の結合荷重, a_{0j} と b_{0k} はバイアスである . 中間素子 j の入力に微小ノイズ m_j を付加した場合の出力 \tilde{y}_j は $\tilde{y}_j = f(\xi_j + m_j)$ となるから, ノイズがない場合と比較するために f を ξ_j の回りに Taylor 展開すると次のようになる .

$$\tilde{y}_j \cong y_j + \left\{ f'(\xi_j) + \frac{1}{2} f''(\xi_j) m_j \right\} m_j \quad (4)$$

出力素子 k への入力は $\tilde{\eta}_k = \sum_{j=1}^J b_{kj} \tilde{y}_j + b_{0k}$ となるので, 式 (4) を用いて 2 次の微小量までとれば,

$$\tilde{\eta}_k \cong \eta_k + \tilde{l}_k + \frac{1}{2} \sum_{j=1}^J b_{kj} f''(\xi_j) m_j^2 \quad (5)$$

となる .

なお \tilde{l}_k は 1 次の微小量 $\tilde{l}_k = \sum_{j=1}^J b_{kj} f'(\xi_j) m_j$ である . ノイズ付加後の出力素子 k の出力は式 (3) より $\tilde{z}_k = g(\tilde{\eta}_k)$ となるから, 式 (5) を用いて g を η_k の回りに Taylor 展開すると次のようになる .

$$\begin{aligned} \tilde{z}_k \cong & z_k + g'(\eta_k) \tilde{l}_k + \frac{1}{2} g''(\eta_k) \tilde{l}_k^2 \\ & + \frac{1}{2} g'(\eta_k) \sum_{j=1}^J b_{kj} f''(\xi_j) m_j^2 \end{aligned} \quad (6)$$

よって,

$$\begin{aligned} \tilde{\epsilon}^2 \cong & \frac{1}{2} \sum_{k=1}^K \left[(t_k - z_k) - \left\{ g'(\eta_k) + \frac{1}{2} g''(\eta_k) \tilde{l}_k \right\} \tilde{l}_k \right. \\ & \left. - \frac{1}{2} g'(\eta_k) \sum_{j=1}^J b_{kj} f''(\xi_j) m_j^2 \right]^2 \end{aligned} \quad (7)$$

となるので, 2 次の微小量まで評価すると,

$$\begin{aligned} \tilde{\epsilon}^2 \cong & \epsilon^2 - \sum_{k=1}^K (t_k - z_k) g'(\eta_k) \tilde{l}_k \\ & + \frac{1}{2} \sum_{k=1}^K \left\{ g'(\eta_k)^2 - (t_k - z_k) g''(\eta_k) \right\} \tilde{l}_k^2 \\ & - \frac{1}{2} \sum_{k=1}^K (t_k - z_k) g'(\eta_k) \sum_{j=1}^J b_{kj} f''(\xi_j) m_j^2 \end{aligned}$$

となる . よってノイズの平均的な影響は

$$\tilde{\epsilon}^2 - \epsilon^2 \cong - \sum_{j=1}^J \tilde{S}_j m_j + \frac{1}{2} \sum_{p=1}^J \sum_{q=1}^J \tilde{T}_{pq} m_p m_q \quad (8)$$

となる . ただし,

$$\begin{aligned} S_j &= \sum_{k=1}^K (t_k - z_k) g'(\eta_k) b_{kj} \\ \tilde{S}_j &= f'(\xi_j) S_j \\ T_{pq} &= \sum_{k=1}^K \left\{ g'(\eta_k)^2 - (t_k - g(\eta_k)) g''(\eta_k) \right\} b_{kp} b_{kq} \\ \tilde{T}_{pq} &= f'(\xi_p) f'(\xi_q) T_{pq} - f''(\xi_p) S_p \delta_{pq} \end{aligned}$$

である . なお, δ_{pq} はクロネッカー記号である .

ノイズ m_j は独立であり確率密度関数が,

$$\rho(m_j) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left(-\frac{m_j^2}{2\sigma_j^2}\right)$$

で与えられるとき, ノイズ m_j の期待値と分散は, $E[m_j] = 0, V[m_j] = \sigma_j^2$ となるからノイズの平均的な影響は

$$E[\tilde{\epsilon}^2 - \epsilon^2] \cong \frac{1}{2} \sum_{p=1}^J \sum_{q=1}^J \tilde{T}_{pq} \sigma_p^2 \delta_{pq} \quad (9)$$

$$\begin{aligned} \tilde{\epsilon}^2 - \epsilon^2 \cong & \frac{1}{2} \sum_{j=1}^J \left(\sum_{k=1}^K [f'(\xi_j)]^2 \left\{ g'(\eta_k)^2 \right. \right. \\ & \left. \left. - (t_k - z_k) g''(\eta_k) \right\} b_{kj}^2 \right. \\ & \left. - (t_k - z_k) g'(\eta_k) f''(\xi_j) b_{kj} \right) \sigma_j^2 \end{aligned} \quad (10)$$

となる . 出力素子の入出力をロジスティック関数 $z_k = g(u) = 1/\{1 + \exp(-u)\}$ で規定すると, $g'(\eta_k) = g(\eta_k)(1 - g(\eta_k))$ となる . $g''(\eta_k)$ にも同様に $g(\eta_k)(1 - g(\eta_k))$ の項が含まれる . したがって, 学習の最終局面において出力素子の出力が 0 または 1 に収束していくと, $g'(\eta_k) \cong 0, g''(\eta_k) \cong 0$ となるから, 式 (10) は $\tilde{\epsilon}^2 - \epsilon^2 \cong 0$ となり, ノイズの平均的影響が無視できるようになることが分かる . さらに, $f'(\xi_j)$ と $g'(\eta_k)$ は, それぞれ $f(\xi_j) = 0.5$ および $g(\eta_k) = 0.5$ の場合に最大となる . これは, 学習初期においてニューラルネットの出力が 0.5 のように不確定な場合, 式 (10) よりノイズの影響 $\tilde{\epsilon}^2 - \epsilon^2$ の値も大きくなることを示している . 以上の解析により, 学習初期にはノイズの影響を大きく受けて学習がゆっくりと探索的に進み, 最終的にはノイズの影響が無視できることが分かる .

(平成 14 年 11 月 25 日受付)

(平成 15 年 3 月 4 日採録)