

# SVM とニューラルネットワークを用いた 手書き数字の3段階認識法

山森 一人<sup>a)</sup>・井野 誠也<sup>b)</sup>・相川 勝<sup>c)</sup>

## Three-step Recognition Method for Handwritten Digits Using SVM and Neural Network

Kunihito YAMAMORI, Seiya INO, Masaru AIKAWA

### Abstract

Handwritten mathematical expression recognition is one of the research fields of handwritten character recognition. Recognition of handwritten mathematical expressions also attracts attention from the viewpoint of education, and applications that help students whose weak subject are mathematics have also been developed in recent years. In this paper, we focus on handwritten digits recognition. We propose three-step recognition method that consists of support vector machine, artificial neural network and the reevaluation of those results. Experimental results showed that our method achieved 93.4% accuracy for handwritten digit images.

**Keywords:** Support vector machine, Neural network, Image recognition, Handwritten digits

### 1. はじめに

手書き文字認識<sup>1)</sup>の研究分野の一つに手書き数字認識がある。手書き数式認識とは、紙やタブレット端末上に書かれた手書き数式に対して、数式中の文字や記号を、数式の意味も含めて認識を行う技術のことである。手書き数式認識は教育の観点からも注目を浴びており<sup>2)</sup>、数学の苦手な学生を手助けするアプリケーションも近年開発されている。

手書き文字認識を行う手法としてテンプレートマッチングや最近傍法がある。テンプレートマッチングは、ある文字を代表するテンプレート画像と入力画像の類似度によって認識を行うため、書き手によって形が変わってしまう手書き文字認識では膨大な数のテンプレートを準備する必要がある。最近傍法は、あらかじめクラス分類されているすべての参照データと入力画像とのユークリッド距離を算出し、ユークリッド距離が最も小さい参照データが属するクラスに入力画像を分類するため、参照データが多いと計算時間がかかるという欠点がある。そこで本研究では、比較的少ない計算量で認識精度も高いサポートベクターマシン<sup>3)</sup> (Support Vector Machine: SVM) とニューラルネットワーク<sup>4)</sup> (Neural Network) に注目する。

本研究では、手書き数式認識の中の手書き数字認識に着目する。本研究の目的は、SVMまたはニューラルネットワークのみを用いた認識より認識精度を向上させることである。提案手法では、SVMにより候補となる数字3種を選出する大分類と、大分類で決定した数字をニューラルネットワークにより一つに絞り込む詳細認識、大分類と詳細認識による認識結果の再評価という3段階により認識精度の向上を図る。

### 2. Support Vector Machine

SVM は2クラスのパターン識別器を構成する手法で、各クラスの学習データから識別面を構築し、未知のデータが識別面のどちら側にあるかでクラス分類を行う。識別面は学習データから識別面までの距離を最大化するように求める。学習データから識別面までの距離をマージンと呼ぶ。マージンを最大化することにより、未知のデータに対しても高い精度を得ることができる。

#### 2.1. 線形クラス分類

SVM は学習データである入力に対し、2値の出力を計算する。入力 $\mathbf{x}$ に対する出力 $y$ を式(1)に示す。

$$y = \text{sign}(\mathbf{x} \cdot \mathbf{w} - \theta). \quad (1)$$

式(1)中の重み $\mathbf{w}$ としきい値 $\theta$ はともにパラメータであ

a) 情報システム工学科教授

b) 情報システム工学科

c) 宮崎大学工学部教育研究支援技術センター技術職員

る。また、符号関数 $\text{sign}(u)$ を式(2)に示す。

$$\text{sign}(u) = \begin{cases} 1 & (u > 0), \\ -1 & (u \leq 0). \end{cases} \quad (2)$$

式(2)は、入力 $x$ と重み $w$ の内積がしきい値 $\theta$ より大きいとき1を出力し、しきい値 $\theta$ 以下のとき-1を出力する。

## 2.2. Multi-class SVM

SVM を多クラス分類が可能に拡張したものを多クラス SVM (Multi-class SVM) という。

SVM を多クラス分類可能にする方法の一つに1対1分類法 (one-versus-one) がある。1対1分類法は、全 $n$ クラスの中から2クラスずつのペアを順に選び、SVM を構築する。 $nC_2$ 通りのペアで構築された SVM を組み合わせることにより、すべてのクラス分類が可能になる。

## 3. ニューラルネットワーク

### 3.1. 階層型ニューラルネットワーク

階層型ニューラルネットワークとは、ニューロンと呼ばれるユニットを複数の層を構成するように並べたニューラルネットワークである。階層型ニューラルネットワークの例を図1に示す。

中間層と出力層の各ユニットは、入力の出力への影響の大きさを表すパラメータである重み $w$ とバイアス $b$ を持つ。入力 $x$ と重み $w$ の内積を求め、バイアス $b$ を減じた値に活性化関数を作用させた値をユニットの出力 $y$ とする。ユニットの出力 $y$ は式(3)で導出できる。

$$y = f(x \cdot w - b). \quad (3)$$

ここで、式(3)の $f(u)$ は活性化関数と呼ばれ、よく用いられる活性化関数の例としてシグモイド関数を式(4)に示す。

$$f(u) = \frac{1}{1 + e^{-u}}. \quad (4)$$

### 3.2. 誤差逆伝播法

誤差逆伝播法<sup>5)</sup> (Backpropagation) は、ニューラルネットワークを学習させる際に用いられるアルゴリズムであり、出力層から入力層に向かって重みとバイアスを更新する。更新前のバイアスを $b_{\text{old}}$ 、更新後のバイアスを $b_{\text{new}}$ 、更新前の重みを $w_{\text{old}}$ 、更新後の重みを $w_{\text{new}}$ としたときのバイアスと重みの更新式を式(5)に示す。

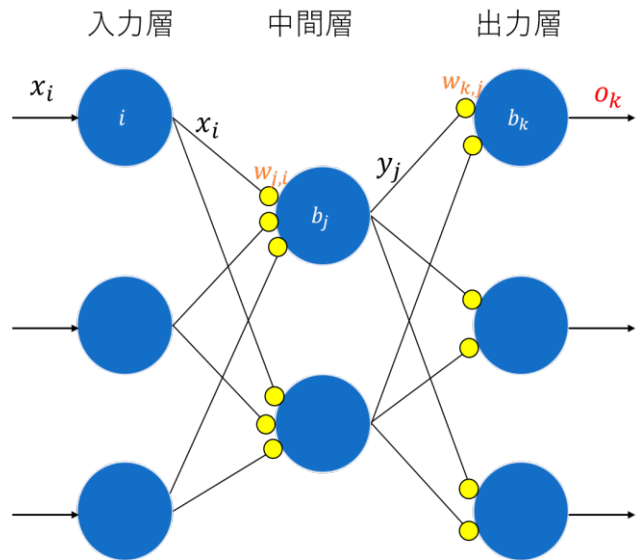


図 1. 階層型ニューラルネットワークの例。

$$\begin{cases} b_{\text{new}} = b_{\text{old}} + \eta \frac{\partial E}{\partial b_{\text{old}}} \\ w_{\text{new}} = w_{\text{old}} + \eta \frac{\partial E}{\partial w_{\text{old}}} \end{cases} \quad (5)$$

式(5)の $\eta$ は学習率と呼ばれ、学習速度の調整を行うパラメータであり、 $E$ はニューラルネットワークの出力と教師信号との誤差を表す。

## 4. 提案手法

提案手法での処理の流れを図2に示し、SVMによる大分類とニューラルネットワークによる詳細認識での処理を図3に示す。図2に示した通り、入力画像に前処理として平滑化、二値化、細線化を行う。次に、各数字に対して間違われやすい数字の上位2種を、SVMを用いてあらかじめ決定しておく。以降、間違われやすい数字を候補文字、本来の数字を代表文字と呼ぶ。図3に示す「0、2、3」のブロックでは、「0」が代表文字、「2、3」が候補文字であることを表す。さらに、代表文字と候補文字からなる手書き数字画像を学習させたニューラルネットワークで詳細認識を行う。SVMの分類結果と、ニューラルネットワークの認識結果を再評価することで、入力画像の数字を最終的に決定する。

### 4.1. Multi-class SVM を用いた大分類

前処理を加えた数字画像を入力とし、Multi-class SVM で候補文字を決定する。候補文字の決定法を以下に示す。

- (1) 各数字に対して、代表文字とは異なる数字として分類された数 (誤分類数) を調べる。

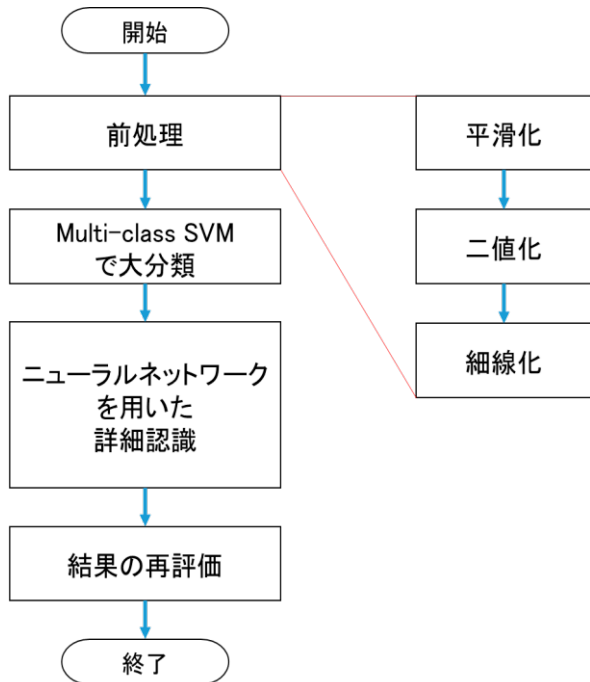


図 2. 提案手法の処理の流れ.

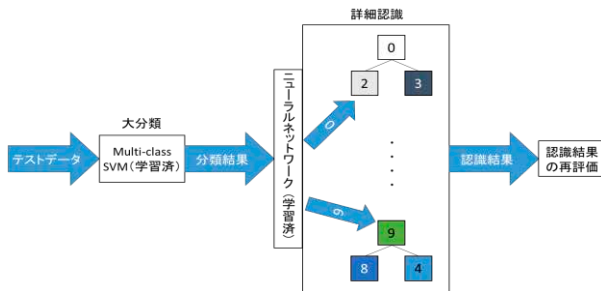


図 3. 大分類と詳細認識の処理.

- (2) 各数字において、誤分類数の多いものから順に 2 種を候補文字として選択する。

また、あらかじめ学習しておいた Multi-class SVM に未知のテストデータを入力した際の分類結果を代表文字候補とする。

#### 4.2. ニューラルネットワークを用いた詳細認識

代表文字と 4.1 節で決定した候補文字を用いて、ニューラルネットワークを学習させる。教師信号は、入力された数字に対応する出力層のユニットに「1」を、それ以外に「0」を与える。認識結果は、ニューラルネットワークの出力値が最も大きいユニットに対応するクラスとする。

#### 4.3. 認識結果の再評価

ニューラルネットワークの出力層の各ユニットの中で、最も出力の大きいユニットに対応する数字（第 1 出力）と、2 番目に出力の大きいユニットに対応する数字（第 2

出力）、さらに第 1 出力と第 2 出力の出力差を求める。SVM での分類結果と、第 1 出力、第 2 出力、出力差をもとに再評価を行うことで入力画像の数字を最終的に決定する。最終的な認識結果の決定方法を以下に示す。

- (1) 大分類における代表文字候補と第 1 出力が一致した場合  
両者の一致した認識結果を最終的な認識結果とする。
- (2) 大分類における代表文字候補と第 1 出力が異なる場合  
以下の 2 パターンに従い、最終的な認識結果を決定する。
  - i) 大分類における代表文字候補と第 2 出力が一致した場合  
出力差がしきい値より大きければ第 1 出力を最終的な認識結果とし、小さければ第 2 出力を最終的な認識結果とする。
  - ii) 大分類における代表文字候補と第 2 出力が異なる場合  
第 1 出力を最終的な認識結果とする。

## 5. 実験と考察

### 5.1. 実験環境

本研究で用いた手書き数字データは、NIST（アメリカ国立標準技術研究所）の「Special Database 19 2nd Edition」を用いる。SVM とニューラルネットワークの学習データに 60,000 枚を使用する。テストデータには、学習データに使用されていない 10,000 枚を使用する。実験結果は 7 分割交差検証により評価する。

SVM は、7 つのデータセットを用いた「0~9」の 10 クラスでモデルを構築する。詳細認識で使用するニューラルネットワークのユニット数は、入力層を画像サイズの 784、中間層を予備実験により 50、出力層を代表文字と候補文字 2 種による 3 に設定する。また、再評価におけるしきい値は予備実験の結果から 0.1 に設定する。

### 5.2. 実験結果

数字の認識精度は、式(6)に示す認識率で評価する。

$$\text{認識率 (\%)} = \frac{\text{正答数}}{\text{テストデータ総数}} \times 100. \quad (6)$$

各データセットのテストデータに対して、SVM のみ、ニューラルネットワークのみを用いた認識率と、提案手法を用いたときの認識率を図 4 に示す。図 4 より、提案手法の認識率は SVM のみの認識率より約 1%、ニューラルネットワークのみの認識率より約 2% 向上したことが

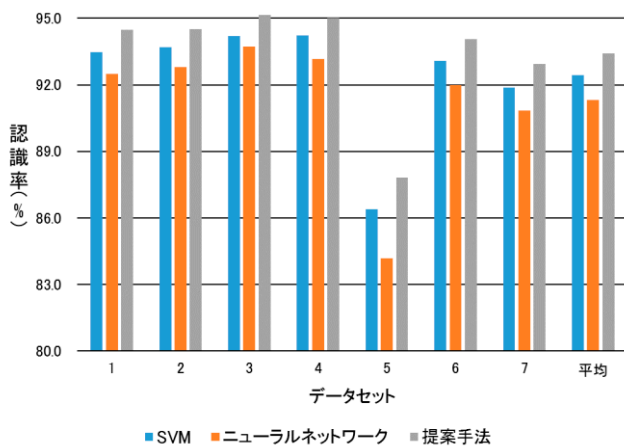


図 4. 各手法における認識率。

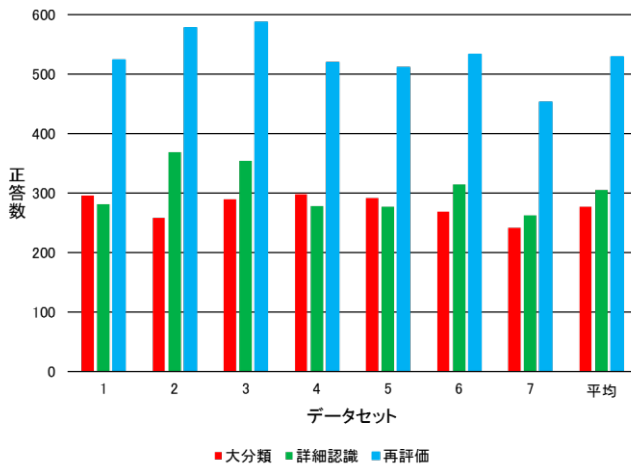


図 5. 各段階における正答数の比較。

分かる。図 5 は、提案手法における各データセットの、テストデータに対する各段階での正答数を示す。図 5 での正答数とは、大分類と詳細認識の認識結果が不一致の場合での正答数である。図 5 より、大分類と詳細認識での結果のみを採用するより、再評価を行うことで正答数が増加していることが分かる。

### 5.3. 考察

図 5 の詳細認識の正答数は、再評価結果に関わらず第 1 出力を最終的な認識結果としたときの正答数を示している。図 5 より、再評価を行うことで正答数は、SVM のみ、ニューラルネットワークのみの場合の約 1.8 倍に増加していることが分かる。このことから、大分類における代表文字候補と第 2 出力が一致し、かつ第 1、第 2 出力の差が小さい場合に、第 2 出力を最終的な認識結果とすることで、第 1 出力を最終的な認識結果とするより正答数が増加したということがいえる。

## 6. おわりに

本研究では、パターン認識の手段として一般的な SVM とニューラルネットワークより認識率が高い手法の開発を目的とした。そのため、SVM を用いて候補文字を決定する大分類と、代表文字と候補文字のみを学習させたニューラルネットワークによる詳細認識、大分類と詳細認識の認識結果の再評価という 3 段階による手法を提案した。

提案手法について、SVM のみとニューラルネットワークのみを用いたときの手書き数字認識での認識率の比較を行った。実験結果から、提案手法は SVM やニューラルネットワークのみを用いた場合より約 1% 高い結果を得ることができた。また、提案手法において大分類、詳細認識、再評価の各段階における正答数の比較を行った。実験結果から、大分類と詳細認識において異なる認識結果となった場合の正答数を、再評価を行うことで増加させることができた。

今後の課題としては、候補文字に選択する数字を増やし、候補文字が多い場合には複数のニューラルネットワークに学習させ、各ニューラルネットワーク間での認識結果の統合を行うことが挙げられる。

## 参考文献

- 1) 篠沢佳久, 大駒誠一, “テンプレートマッチングによるオフライン手書き文字認識ニューラルネットワークの作成”, 情報処理学会論文誌, Vol. 42, No. 1, pp. 16-25 (2001).
- 2) 鈴木雅人, 松本章代, 吉本定伸, 大塚友彦, “手書き数式解析に基づく基礎数学学習支援システムの開発 (障害者教育・特別支援教育/一般)”, 電子情報通信学会技術研究報告. ET, 教育工学, Vol. 109, No. 453, pp. 147-152 (2010).
- 3) 糟谷勇児, 山名早人, “二種類の SVM を用いたオンライン類似数式文字識別(テーマセッション, 文字認識・文書理解)”, 電子情報通信学会技術研究報告.PRMU, パターン認識・メディア理解, Vol. 105, No. 614, pp. 55-60 (2006).
- 4) 岩田彰, 高橋秀和, 堤田敏夫, “ニューラルネットワークによる郵便番号手書き数字認識”, 電子情報通信学会技術研究報告.PRU, パターン認識・理解, Vol. 95, No. 43, pp. 9-16 (1995).
- 5) D. Rumelhart, G. Hintont and R. Williams, “Learning representations by backpropagating errors”, Nature, Vol. 323, pp. 533-536 (1986).