P-12

Deep-Learning による数字画像認識

Digit Character Recognition by A Deep-Learning Approach

柴　智彰[1]

*Tomoaki Shiba[1]

Abstract: Deep-learning is a recently well-established approach for machine learning using a multilayered neural network with multiple hidden layers. To date, various algorithms for deep neural networks with many layers have been proposed. In this paper, we use a convolutional neural network (CNN) which is often used for image recognition and report the simulation results of digit character recognition using CNN.

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には, 入力層, 隠れ層, 出力層と呼ばれる層がある．各層は，複数個のノードを持ち，その各ノードは値をもつ．また，ある層と次の層の間のノード同士はエッジで結ばれ，各エッジは重みを持つ[1]．

２．Convolutional Neural Network (CNN)

CNN は，畳み込みと入力データをより扱いやすい形に変換するために情報を圧縮するプーリングを繰り返し，際立った特徴を有する情報を得るような深層学習(deep learning)により構成されるニューラルネットワークモデルである．また，全結合層だけで構成される従来の階層型ニューラルネットワークとは異なり，前述の畳み込み層やプーリング層を有するのが特徴である．さらに，全結合層の代わりに，畳み込み層やプーリング層を使うことで，オブジェクト等の始点や終点，カーブなどの視覚的特徴を効果的に抽出できることも報告されている[2]．

３．畳み込み層

畳み込み層では, 様々なカーネルを使用して畳み込み処理を行う. ここにおけるカーネルとは, 一般的に n×n のデータ形式により重みパラメータが保持され，どのように畳み込むかが示されるようなフィルタを指す．また，畳み込み処理は，元の入力画像に対し，カーネルを左上から右下まで要素ごとに掛け合わせていくような処理のことである．このようにカーネルをスライドさせていくことから sliding window 処理とも呼ばれる．こうした処理によって得られたデータは特徴量マップと呼ばれ，次層に出力される．

４．プーリング層

CNN において，畳み込み層で出力された特徴量マップはプーリング処理されることが多い．プーリングとは，データサイズを減らし，対象領域の小さな情報の違いを認識し，その領域内の際立った特徴を的確に取得することである．

例えば，4×4の入力に対して，2×2の領域ごとにプーリング処理を行う．その結果，出力として2×2の特徴量マップを得る．この場合，データ量は4×4=16から2×2=4の4分の1に減ることがわかる． CNN では，このような特徴量マップから最大値を取る Max-pooling および平均値を取る Avg-pooling が良く利用されている．

５．全結合層

全結合層は一般的な多層ニューラルネットワーク同様，前層の全ノードとその層の各ノード同士が全て結合されている層である．今，i 番目の入力を xi，j 番目の出力を yj，xi と yj のノードの重みを wij，バイアスを bi とすると yj は

$$y_j = f\left(\sum_j w_{ij}\, x_i + b_i\right) \tag{1}$$

のように計算される．

６．出力層

出力層は, 全結合層で得た結果を元に活性化関数を用い，ニューラルネットワークの出力がカテゴリ毎の確率を表すように学習され，確率が最大値となるようなノードを識別結果として出力する層である．

７．活性化関数

CNN でよく使われる活性化関数には，主にシグモイド関数や ReLU 関数などがある. 本研究で用いた ReLU 関数は

$$\begin{aligned} y &= 0 \text{ if } x < 0 \\ y &= x \text{ if } x \geq 0 \end{aligned} \tag{2}$$

で表される．

８．重みの学習

１：日大理工・学部・数学

CNN を含めた階層型ニューラルネットワークの学習には，誤差逆伝播法[3]によりノード間の重みを更新し，モデルを最適化する手法が一般に広く用いられている．

誤差逆伝播法は，ニューラルネットワークに学習データを入力した際に，期待する出力の値と実際に得た出力の値から損失を求め，その値が小さくなるように各ノードの重みを繰返し計算により更新するような手法である．損失(Loss)とはモデルの精度の悪さを表し，損失関数から求めることができる．しかし，誤差逆伝搬法を用いたニューラルネットワークの学習においては損失関数が依存するパラメータ数は通常多い．それは損失関数のパラメータとして各ノードの重みなどがすべて含まれるからである．よって，損失関数の最小値を取るようなパラメータの組み合わせを一意に求めることは通常困難である．そこで，一般的には最急降下法を応用して損失が小さくなるように各ノードの重みを更新するような手法が用いられる[4]．

９．CNN による実験

本研究では，オープンソースである Caffe を使用して実験を行った．

実験ではトレーニング用の数字0~9の画像を各100枚，評価用の画像を各60枚の計1600枚用意して一種類ずつ実験し，どの数字が読み取りにくいのかについて調べた．その際，トレーニング用画像データ数を少しずつ増加しながら再度実験を試みた．

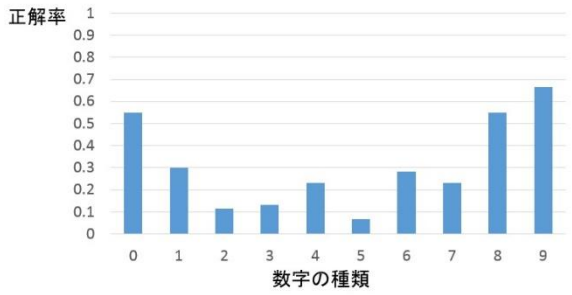

Figure 1. 数字毎の正解率

その結果，訓練用の画像が多いほど精度が良く，逆に少ない場合では過学習が起きやすいことが分かった．また，数字画像0，8，9以外は50%を下回り，5に関しては10%を下回った．

１０．　むすび

本研究では，近年急速に発展してきている Deep-Learning を使用した機械学習法を数字文字画像認識に適用した場合について考慮した．本研究では実験に用いた画像枚数が少なく正解率が低くなったものの，判別しにくい数字を正しく識別することができることがわかった．また，正解率の低かった数字については，被験者により書き方が異なり正しい画像と認識できなかったと考えられる．そこで，実験に用いる画像を収集する際，どのように手書きを行うのかあらかじめ被験者に対し指定する必要があるとも考えられる．

学術講演会当日では，画像枚数を増やし癖のある数字画像の正解率を上昇させること，また，正しく認識できなかった数字がどの数字と誤認識されているのかについて考察しその結果について報告する予定である．

１１．　参考文献

[1] 武井宏将：「初めてのディープラーニング」，リックテレコム，pp. 20-21，2016.

[2] 石橋崇司：「Caffe をはじめよう」，オライリー・ジャパン，pp. 45，2017.

[3] Rumelhart, David E，Hinton Geoffrey E，and Williams，Ronald J．“Learning representations by back-propagating errors”*Nature* **323** (6088)，pp．533-536，1986.

[4] 石橋崇司：「Caffe をはじめよう」，オライリー・ジャパン，pp. 51-56，2017.