知能情報実験 III（データマイニング班）
ひらがなの文字分類

185761E, 185428D, 185767D

提出日：2020 年 8 月 13 日

目次

1　はじめに

1.1　実験の目的と達成目標

知能情報実験 III は、情報工学分野のより専門的な知識を理解・習得することを目的として、半年間でシステムの開発やデータ解析等に取り組む実施される。

その中の一つデータマイニング班においては機械学習外観ならびにその応用を通し、対象問題への理解、特徴量抽出等の前処理、バージョン管理やデバッグ・テスト等を含む仕様が定まっていない状況下における開発方法、コード解説や実験再現のためのドキュメント作成等の習得を目指す。

1.2　日本語文字認識とは

本グループでは画像認識を用いて手書き文字認識をすることを対象問題として設定した。また、今回は主にひらがな 75 種に関して文字認識を行なった。

画像認識とは [2] によると「画像から色や形といった特徴を読み取り、その特徴をさまざまな学習器に入れて新たな画像を認識できるようにしたパターン認識技術のひとつ」であり、手書き文字の画像に対して画像認識の技術を用いることで、その画像に書かれた文字の判別を行うことができる。

画像認識の技術は様々なところで利用されている。[2] によると、例えばテレビ業界での事例がある。映像内の人物と名前に間違いがあってはいけないので、それを防止するための確認作業があるのだが、顔認識技術を用いることで、その確認作業の効率化を図れたという事例がある。他にも [2] によると、農業分野においてはドローンと画像認識技術を用いることで、農作物の出荷量を予測するなどの事例もある。本実験を通して画像認識の知識を深めることで、様々な業種の発展に貢献するための知識を有することができるだろう。

2　実験方法

以下の様な手順を踏まえて実験を進めた。詳細はのちのセクションにて記載する。

1. 実験目的
2. 実験計画
3. データセット構築
4. モデルの選定
5. パラメータ調整

2.1　実験目的

MNIST とは、￥citeMNIST によると『「0」～「9」の手書き数字の画像データセット』のことである。また、「主に画像認識を目的としたディープラーニング／機械学習の初心者向けチュートリアルでよく使われており」とも￥citeMNIST で述べられている。

本実験では、まず MNIST を用いた手書き文字認識のプログラムを実装することで文字認識で行われている処理を理解した後に、手書きひらがな文字のデータに対して MNIST で行なった処理を再現、または独自の処理の追加 (例えば画像サイズの調整など) を行い、文字認識の精度をどこまで高められるかの検証を行った。

2.2　データセット構築

ETL 文字データベース
http://etlcdb.db.aist.go.jp/?lang=ja

2.3　モデル選定

LeNet と呼ばれる畳み込み層とプーリング層を 2 回繰り返すモデルを採用した。[3] によると LeNet モデルは MNIST データセットを 98% 近く正確に分類していることから、今回の課題に対しても良い結果を得られると考えたためこのモデルを実験に用いることにした。

2.4　パラメータ調整

入力画像サイズについて

実験当初は MNIST チュートリアルに従って進めていたために 28x28 の入力サイズだったが、おそらくこの画像サイズではひらがなの複雑な形状を捉えられないと判断し、最終的に 40x40 まで入力サイズを増加させた。

入力画像について

入力された画像を 2 値化する事でノイズを除去。ノイズ除去を行った後は、画像サイズをダウンスケールさせ、もう一度画像の 2 値化を行った。

勾配法について SGD から Adam に変更した

3　実験結果

実験の最終結果として、日本語のひらがな文字のテストデータを 90 %程度の割合で正しく認識する事ができた。

図 1 は入力画像の縦横のピクセル数を変化させた際の、エポック数とテストデータの正答率の関係性を示している。横軸はエポック数を表し、縦軸はテストデータの正答率を示す。

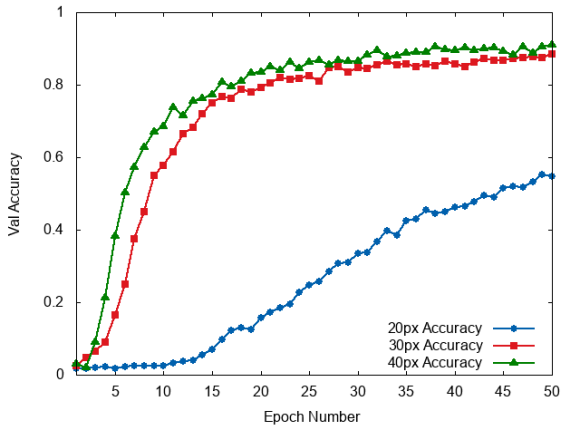

図 1　エポックごとの予測精度の変化

縦横 20px の画像データを入力として学習させると、学習の速度が非常に遅い事がわかる。縦横 30px の画像データの場合、早い段階から高い予測精度を示し、40px の場合はさらに良い予測精度を出している。

また、20px を入力画像のサイズとした場合のエポック数を 200 まで増加させて実験を続けたが、68% ほどの予測率で頭打ちとなり、30px 以上を入力サイズとした時より精度が良くなることはなかった。

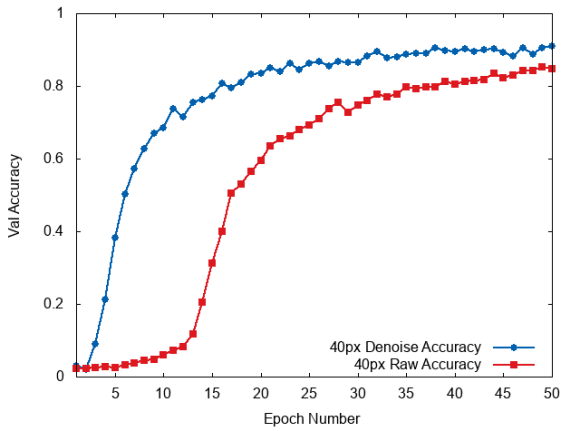

図 2　画像の前処理による予測精度の変化

図 2 はトレーニング及び予測に使用するデータにノイズ除去と 2 値化の前処理を行った画像と、サイズ変更のみを行った画像の 2 つを用意し、予測精度の変化を示したグラフである。

サイズ変更のみの処理を行った画像データによる学習では、前処理を行ったデータに比べ予測精度がある程度出るまでに必要なエポック数が増加し、同じエポック数では、前処理を行ったデータを入力したモデルの方が予測精度が高い。

さらに、前処理を行ったデータは 1 エポックを実行する時間は 2 秒程度だったのに対し、前処理

を行っていないデータでは 1 エポックに約 10 秒かかることもわかった。

4　考察

実験課題への取り組みを通し、実験の意義、実験からわかったこと、今後の展望などを述べる。失敗やつまづきがあれば、それらについての失敗分析を含めると良い。

図 1 の 20px と 30px の予測精度の間には大きな差があり、30px と 40px の間には大きな差がない事がわかる。これは、ひらがなの特徴を捉えるのに 20px では十分ではなく、30px 程度あれば十分にひらがな同士の区別ができるほど特徴の分離ができていると考えられる。

図 2 の結果を踏まえると、予測機のパフォーマンスの向上には、学習するモデルの設計だけでなく、データの前処理も大きな影響を持っていると考えられる。

5　意図していた実験計画との違い

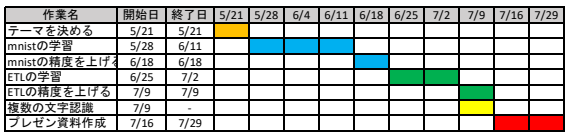

図 3　作業ガントチャート

まず元々の計画として、文章が載った画像から一文字ずつ文字を認識し、どのような文章が書いてあるかを判断するものを作成することを目的としていた。しかし、結果として今回は一文字ずつ認識する部分までの実装で作業を終えることとなった。

今回、実験の計画を立てる上で、段階的に進捗を作っていくことを念頭に入れ、以下のように実験を進めていくことにした。

1. 数字から読み取る
2. ひらがなから読み取る
3. 複数文字から読み取る

実際には二ヶ月間で図 3 のガントチャートの通りに進捗があり、1、2 までを実装することができたが、3 を終えることができなかった。理由としては、ひらがなの構造がアルファベットなどと違い、一文字のパーツが複雑な作りとなっており、ひらがな 1 文字を 1 文字として認識することが難しいことが問題として考えられた。しかし、データマイニングの実験として、”文字と文字の境目を認識する”部分の実装は間に合わなかったものの、”文字自体を認識する”という点に関しては成し遂げることができたと考えている。

6　まとめ

今回、本グループでは画像認識を用いて手書き文字認識を行った。結果として画像認識の段階では 1 文字ずつの認識を行うことができ、ETL を使った文字認識で 99% まで精度を上げることができた。このデータマイニング班の実験を通して、画像認識の仕組みや手書き文字認識で使用するデータセットについての知見を広げたり、畳み込み層を用いることで更に手書き文字認識についての精度を上げることができることを学んだ。

今後は機会があれば、複数のひらがなの文字の載った画像認識から、文字認識を実装できるように改善することや、またひらがなだけでなく漢字やアルファベットの文字認識まで広げていくことができると考えている。

参考文献

[1] MNIST とは, https://www.atmarkit.co.jp/ait/articles/2001/22/news012.html

[2] 画像認識とは, https://ledge.ai/image-recognition/

[3] LeNet – Convolutional Neural Network in Python, https://www.pyimagesearch.com/2016/08/01/lenet-convolutional-neural-network-in-python/