

手書き文字の正誤判定システムにおけるオートエンコーダと GAN の比較 Comparison of autoencoder and GAN for handwritten character accuracy judgment system

宮下 知也[†]
Tomoya Miyashita

縄手 雅彦[†]
Masahiko Nawate

伊藤 史人[†]
Fumihito Ito

多々納 俊治[‡]
Toshiharu Tadano

1. はじめに

文字の読み書きに困難のある児童を発見するために簡易的なスクリーニング検査が求められており、書字の検査においては「小学生のための読み書きスクリーニング検査 (STRAW)」[1] などがある。書字検査において必要な評価項目は流暢性と正確性であり[2]、流暢性の評価には書字にかかる時間、正確性の評価には文字が正しく書けたかどうかを測定する必要がある。現状ではビデオに録画したりストップウォッチで計測したりすることが行われており、PCを用いた検査の電子化が求められている。電子化を行うにあたって書字時間の測定は、問題提示から文字の書き始めまでの時間と文字の書き始めから書き終わりの時間を計測するのみであるため、1つの評価基準を設ければ自動で測定可能であるが、文字の正誤判定に関しては、個人によって字形が異なるため、1つの評価基準を設けて自動で測定するのは非常に難しい。

また、今までの文字認識は文字の一部が間違った文字を入力しても、機械の中で補完を行って正しい結果を返すというものであったため、文字の正誤判定が正しく行えないといった問題がある。そのような問題を解決し、文字の正誤判定手法を開発した研究として漢字検定の自動採点ツールに関する研究[3]もある。

我々は以前、スクリーニング検査の電子化に向けて、機械学習を用いて正解文字と不正解文字を学習させて文字の自動正誤判定の手法を開発しようと試みた。しかし、この手法では不正解文字のデータを大量に作成しなければならず、時間と手間がかかった[4]。そこで本研究では、正解文字のデータのみを学習させて文字の自動正誤判定を行う手法の開発を試みる。具体的には、機械学習の手法の1つであるオートエンコーダ[5]とDCGAN[6]を用いて2種類の手書き文字の自動正誤判定モデルを作成し、それぞれのモデルの精度を比較検討する。

2. モデルの作成

本研究では、文字画像を文字認識モデルに入力し、文字認識モデルの中間層のデータをオートエンコーダに入力して文字の正誤判定を行うモデル(以下、AEモデル)とDCGANのアルゴリズムを用いて訓練した画像生成器Gと画像判別器Dのうち画像判別器Dを用いて文字の正誤判定を行うモデル(以下、GANモデル)の2種類のモデルを作成し、それぞれの精度を比較した。

2.1 AEモデル

ここでは、文字認識モデルのデータをオートエンコーダに入力するタイプのモデルについて説明する。

[†] 島根大学 大学院総合理工学研究科, Graduate School of Science and Engineering, Shimane University

[‡] 出雲医療看護専門学校, Izumo Medical Care Nursing School

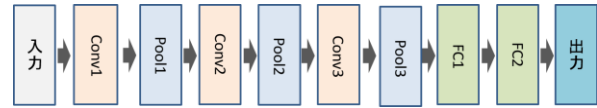


図1 文字認識モデルの構成

2.1.1 文字認識モデルの作成

今回作成した文字認識モデルは、入力した文字画像を881のクラスに分類するモデルである。モデルの構成を図1に示す。図1のConv1はサイズ 64×64 でチャンネル数8の畳み込み層、Conv2はサイズ 32×32 でチャンネル数16の畳み込み層、Conv3はサイズ 16×16 でチャンネル数32の畳み込み層、Pool1~Pool3はサイズ 2×2 のMaxプーリング、FC1はノード数1024の全結合層、FC2はノード881の全結合層であるモデルの学習には産総研のETL8B文字データベース[7]に収録されている881種類の漢字を用いている。また活性化関数にはReLU[8]を用いている。

2.1.2 オートエンコーダモデルの構成

オートエンコーダモデルは3層構造であり、入力層と出力層のノード数が1024、中間層のノード数が256である。入力層には、前節で作成した文字認識モデルのFC1層から出力されるデータを入力する。

2.1.3 オートエンコーダモデルの学習

モデルの学習にはETL8B文字データベース[0]の文字画像を文字認識モデルに入力し、その時に文字認識モデルのFC1層から出力されるデータを正規化したものをオートエンコーダモデルの学習データとした。1文字あたり160枚の画像があり、そのうち159枚の画像を学習用データとして使い、残りの1枚画像を正誤判定のしきい値を決めるためのデータとした。

2.2 GANモデル

モデルの概要を図2に示す。図の入力はランダムなノイズ、学習画像はETL8B文字データベースの文字画像である。この2つの入力データから生成器Gと判別器Dを学習させる。判定したい漢字1文字に対して1つのモデルを作成する。学習終了後は、判別器Dにテスト用の画像データを入力して、正誤判定を行う。

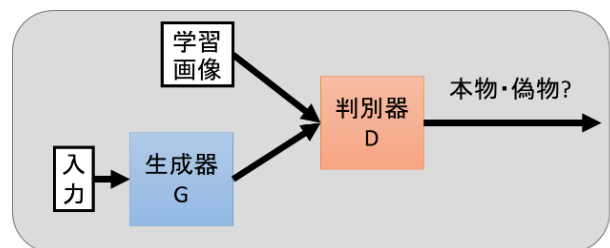


図2 GANモデルの概要

表1 間違いパターンの一例

間違い漢字の内容	漢字画像の一例
1・2画追加	間 芽 散
1・2画抜け	会 兎 孫
鏡文字	業 邨 省

3. モデルの評価方法

今回は、STRAW に収録されている漢字の一部(28文字)をモデルの精度評価の対象とした。対象漢字1文字あたり10枚の正解漢字画像と10枚の不正解漢字画像を学習用データとは別に用意し、テストを行った。評価に用いた間違い漢字の間違いパターンの例の一部を表1に示す。間違いパターンは、1・2画追加漢字、1・2画抜け漢字、鏡文字の3種類である。また、正誤判定を行うにあたってモデルの出力値に対してしきい値を設けた。AEモデルのしきい値は0.0128、GANモデルのしきい値は0.5である。AEモデルでは出力値がしきい値以下であれば「正解」、しきい値より大きければ「不正解」と判定し、GANモデルでは出力値がしきい値以下であれば「不正解」、しきい値より大きければ「正解」と判定する。

4. 評価結果

テストデータを作成したモデルに適用し、モデルの精度検証を行った。AEモデルとGANモデルの判定精度の結果を表2、AEモデルとGANモデルの間違いパターンごとの判定精度を表3に示す。表2より、全体の判定精度はならびに正解漢字の判定精度はGANモデルよりAEモデルの方が高かった。一方で、不正解漢字の判定精度はAEモデルよりGANモデルの方が高かった。また表3より、AEモデルとGANモデル共に鏡文字の正誤判定精度が高いのに対して1・2画追加漢字と1・2画抜け漢字の判定精度は低かった。さらに、図3には文字の画数と判定精度の関係を示す。図3のデータに対して相関分析を行うと、GANモデルの相関係数は-0.25で弱い負の相関が認められ、文字の画数が増えると若干精度が低くなる傾向が認められた。さらに、AEモデルの相関係数は-0.60で負の相関が認められ、文字の画数が増えると判定精度が低くなる傾向が見られた。

また、文字ごとの判定精度においてAEモデルでは「史」の判定精度が90%で一番高く、「業」の判定精度が50%で一番低かった。GANモデルでは「完」の判定精度が85%で一番高く、「散」の判定精度が45%で一番低かった。

表2 モデルの精度

	AEモデル	GANモデル
全体の精度	68.8%	61.8%
正解漢字の精度	94.3%	67.9%
不正解漢字の精度	43.2%	55.7%

表3 間違いパターンごとの精度

	AEモデル	GANモデル
1・2画追加漢字の精度	30.2%	57.1%
1・2画抜け漢字の精度	39.1%	39.1%
鏡文字の精度	93.2%	93.5%

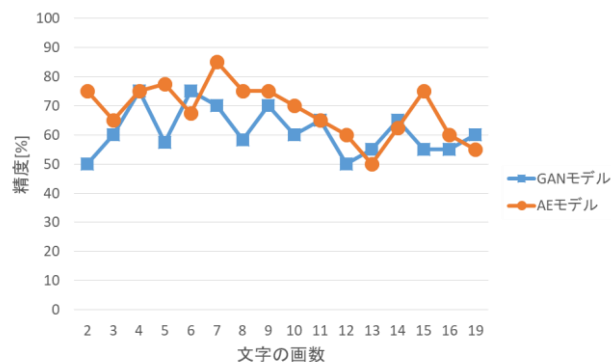


図3 画数ごとの精度

5. 考察

今回の検証では、GANモデルよりAEモデルの方が全体の判定精度が高かった。これは、AEモデルでは、文字の概念を習得した文字認識モデルに判定する画像を入力し、そこから得られたデータをオートエンコーダに入力していたためであると考えられる。また、GANは元々画像生成アルゴリズムとして提案されたものであり、今回のような文字の正誤判定にはあまり向いていないのではないかと考えられる。さらに間違いパターンごとの正誤判定精度より、AEモデルとGANモデル共に1・2画追加漢字、1・2画抜け漢字の判定精度が低かった。これは、テスト時に入力した文字画像が学習時の文字画像とほとんど見た目の変化がないため、判定が難しかったのではないかと考えられる。

6. まとめと今後の課題

検証の結果より、全体の判定精度が高かったAEモデルの方が文字の自動正誤判定システムに適していると考えられる。今後、実際の現場で用いるためには、正誤判定の精度を向上させていく必要がある。判定精度を向上させていくにあたって、文字ごとに評価のしきい値を自動で設定できる仕組みや評価の方法、アルゴリズムの見直しを行っていく必要がある。また、1・2画抜け漢字や1・2画追加漢字の判定方法についても見直しを行う必要がある。

参考文献

- [1] 宇野彰, 春原則子, 金子真人, Taeko N Wydel, “小学生の読み書きスクリーニング検査-発達性読み書き障害(発達性 dyslexia) 検出のために-”, インテルナ出版, (2006).
- [2] 平林ルミ, 河野俊寛, 中村賢龍, “小学生の視写における書字行動プロセスの時間分析”, 特殊教育学研究会, 48, pp.275-284, (2010).
- [3] 井戸伸彦, “画と辺とを同定した後の手書き漢字入力のプロット評価方法”, 信学技報, PRMU2015-25, (2015)
- [4] 宮下知也, 縄手雅彦, 伊藤史人, 多々納俊治, “機械学習を用いた手書き文字の自動正誤判定システムの開発”, FIT情報科学技術フォーラム講演論文集DVD, K-007, (2017).
- [5] G.E.Hinton, R.R.Salakhutdinov, “Reducing the dimensionality of data with neural networks”, Science, Vol.313, No.5786, pp.504-7, 2006.
- [6] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks.” arXiv preprint arXiv:1511.06434, (2015).
- [7] 産業技術総合研究所, ETL 文字データベース, 1973-1984.
- [8] Xavier Glorot, Antoine Bordes, Yoshua Bengio, “Deep Sparse Rectifier Neural Networks”, In Aistats, Vol.15, pp.275, (2011).