

ペンの持ち上げを考慮した神経力学モデルによる 手書き漢字の学習

野中 柚希[†][†]徳島大学 工学部西出 俊[‡][‡]徳島大学 大学院社会産業理工学研究部康 鑫[‡]任 福継[‡]

1. はじめに

手書き文字の認識は個人認証や入力 UI など、様々な分野において応用されている。大田らは、文字部品やストローク（画）を確率文脈自由文法によって学習し、少ない学習データしかない条件下でも高精度認識ができることを示した [1]。このモデルでは、画のダイナミクスを別々のものとして捉えており、漢字全体のダイナミクスは考慮していない。志久らは、傾斜を考慮した文字を畳み込みニューラルネットで学習し、手書き文字を高い精度で認識できることを示した [2]。このモデルでは漢字の構成に重点を置いており、筆跡などのダイナミクスは考慮していない。本研究では手書き文字における漢字全体の筆跡ダイナミクスを神経力学モデルで学習、認識することを目標としている。本稿では特に、ペンの持ち上げによって筆跡データが取得できない学習データに対してモデルの予測機能を用いて補完する手法を提案する。

2. 神経力学モデル MTRNN

本研究では、手書き漢字の学習モデルとして神経力学モデル Multiple Timescale Recurrent Neural Network (MTRNN) を用いる [3]。本章では MTRNN の構成、機能、提案する学習データ補完手法について述べる。

2.1 MTRNN の構成

MTRNN は現状態 $IO(t)$ から次の状態 $IO(t+1)$ を予測する予測器である。従来の RNN とは異なり、文脈層は Fast Context (C_f) 層と Slow Context (C_s) 層と階層的に構成されており、各層のニューロン群で異なる発火速度になるように設定することで階層的に情報を学習し、長い時系列を学習することが可能である。本研究では IO 層で漢字執筆時の座標系列を入力する。モデルの概略図を図 1 に示す。

2.2 MTRNN の機能

MTRNN には学習・認識・生成の三つの機能がある。本節では三つの機能を紹介する。

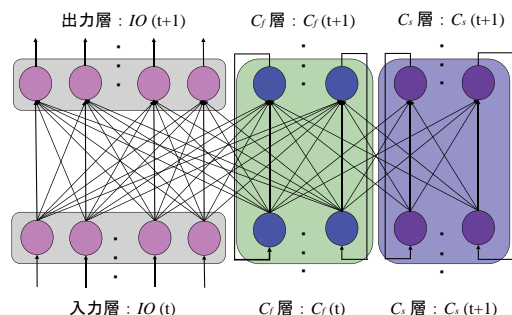


図 1: MTRNN の構成

Learning Handwritten Kanji Characters Using Neurodynamical Model with Consideration of Pen Lifting Yuzuki Nonaka (Tokushima Univ.), Shun Nishide (Tokushima Univ.), Xin Kang (Tokushima Univ.), and Fuji Ren (Tokushima Univ.)

MTRNN の学習：各学習データに対して前向き計算を行うことで各時刻の入力に対する出力値を計算し、最終時刻の出力誤差を初期時刻まで逆伝播する Back Propagation Through Time (BPTT) を用いる。各重みにかかる誤差値を積算し、初期時刻における誤差積算値によって重みと C_s の初期値 ($C_s(0)$) を更新していく。ここで $C_s(0)$ は各データ固有の値である。

MTRNN の認識：MTRNN の学習後、認識する時系列に対して学習と同様の計算を行う。ただし、重みは更新せず、 $C_s(0)$ のみを更新し、時系列を表現する $C_s(0)$ を求める。

MTRNN の生成：生成する時系列の $C_s(0)$ を MTRNN に入力し、前向き計算を行うことで各時刻における出力値を計算し、時系列全体を求める。

2.3 提案手法（欠損データの補完）

本研究で対象とするデータは手書き漢字であり、漢字の画と画の間にはペンの持ち上げによるデータの欠損が存在する。力学モデルは現状態から次状態を出力するため、データ欠損時には出力値を求めることができない。本研究では前時刻の出力予測値を入力することでデータ欠損時の補完を行う手法を提案する。

提案手法の概略図を図 2 に示す。本研究で用いる MTRNN は入力値として執筆時の座標値 (x, y) を用いる。時系列データにおいてデータが存在する場合は座標値 (x, y) を入力し、データが存在しない（ペンを持ち上げている）場合は前時刻の出力値を現時刻の入力値とする。また、提案手法では、ペンの状態（持ち上げか執筆中か）を判定するための state ニューロンも同時に予測する。state ニューロンはデータが存在する場合は発火せず、データが存在しない場合に発火するように学習する。state ニューロンは C_f 層と同じ階層に配置する。本稿では手書き漢字の認識にのみ重点を置いているため、state ニューロンは不要であるが、ロボットの描画応用などにおいて生成する場合には state ニューロンは必要になると考えられる。

3. 漢字データの取得と実験設定

提案手法を評価するために、Wacom 社の液晶タブレット Cintiq13HD (DTK-1301/KO) (図 3) を用いて取得した 13

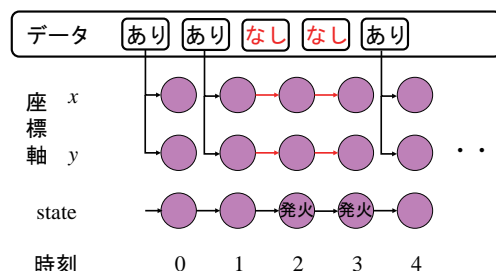


図 2: 提案手法の概略図

種類の漢字(土・丙・生・五・辺・冬・犬・石・仁・木・巧・民・四)を用いて評価実験を行った。実験では一人の被験者が各漢字を約9秒前後で10回ずつ書き、執筆中のタブレットから取得されるペン先座標データ (x, y) を10(frames/sec)で取得した。各漢字の座標系列例を図4に示す。各漢字について全10系列のうち5系列を学習用とし、残り5系列は評価用に未学習とした。

MTRNNの構成として、IOニューロン数は座標データ (x, y) であるので2、stateニューロン数は1、 C_f ニューロン数は30、 C_s ニューロン数は2とした。 C_f のニューロン数は複数の構成の中から最もよいものを実験的に設定し、 C_s のニューロン数は実験結果を平面上で解析しやすいように設定した。

評価実験はstateニューロンを付加した場合としなかった場合の2通りで行う。stateニューロンを付加しなかった実験は「土」以外の漢字データで行い、付加した実験は「辺」「石」「巧」以外の漢字データで行った。これはstateニューロンの付加によって入力次元が増え、筆跡ダイナミクスが類似している漢字が含まれると学習が困難になるためである。それぞれの実験において、各漢字データを認識した際の $C_s(0)$ の分布を解析した。

4. 実験結果と解析

本章では手書き文字の認識の実験結果について述べる。結果の解析は、各データ系列をMTRNNで認識した際に得られる $C_s(0)$ 値の分布 ($C_s(0)$ 空間) が形成するクラスタを評価して行った。 $C_s(0)$ 空間はデータのダイナミクスの類似度によって自己組織化され、 $C_s(0)$ 値のクラスタは類似した筆跡ダイナミクス(同一の漢字)を表す。stateニューロンなしの実験結果を図5に、stateニューロンありの実験結果を図6に示す。学習データは青色で、未学習データは赤色のマークでプロットした。

図5と図6より、双方の場合共に同一の漢字によって未学習・学習データ含めたクラスタが形成されていることが確認できる。一方で、データ間の重なりも存在しており、stateニューロンありの結果の方がstateニューロンなしの結果よりクラスタの重なりが大きい。これはstateニューロンを導入することで学習データの次元数が



図3: ペンタブレット



図4: 13種類の漢字

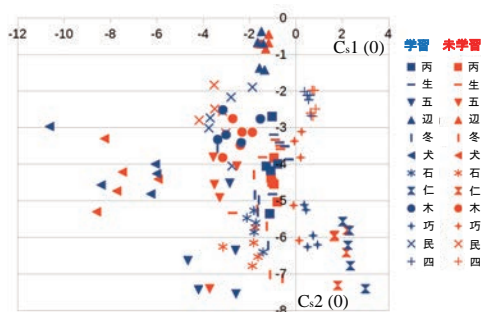


図5: stateニューロンなしの $C_s(0)$ 分布

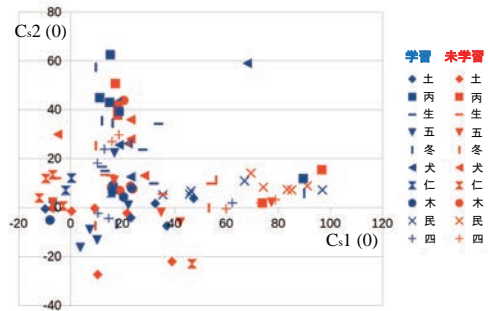


図6: stateニューロンありの $C_s(0)$ 分布

2から3に増えることに起因すると考えられる。また、解析の都合上、 C_s ニューロン数を2に設定したが、クラスタの重なりを防ぐためにはニューロン数をさらに増やすことが必要であると考えられる。多少のクラスタの重なりは存在するが、提案手法によって神経力学モデルによる手書き漢字の認識が可能であることを示唆する結果が得られた。

5. 考察

前章の実験の結果、stateニューロンを付加しないモデルの方がクラスタの重なりが少なくなる結果が得られた。本章ではstateニューロンの役割について議論する。

本稿で注目した手書き漢字の認識においてはstateニューロンは不要であると考えられる。一方で、ロボットによる描画動作生成などに応用する際、ロボットがペンを持ち上げるタイミングを把握する必要がある。本モデルで導入したstateニューロンを用いることでstateニューロンが発火した時にロボットはペンを持ち上げ、stateニューロンが発火していない時はペンを下ろし、描画動作を行うことが可能になる。双方の場合共、クラスタの重なりは存在するので、 C_s ニューロン数を増やした場合の検証は必要であると考えられる。

6. おわりに

本稿では、神経力学モデルMTRNNの予測機能を用いた学習データ欠損時の保管方法を提案し、ペンの持ち上げを考慮した手書き漢字の学習と認識について報告した。液晶タブレットを用いた実験の結果、多少の重なりはあるものの、各漢字系列ごとにクラスタが形成されることを確認した。今後は、モデルの改良を行い、ロボットへの応用なども検討していきたいと考えている。

謝辞

本研究は科学研究費補助金、若手研究(A)(課題番号16H05877)の支援を受けた。

参考文献

- [1] I. Ota, R. Yamamoto, S. Sako, and S. Sagayama, "On-Line Recognition of Handwritten Kanji Characters Based on Grammatical Description of Relative Positions between Strokes," The Institute of Electronics, Information and Communication Engineers, PRMU, PRMU106(606), pp.49-54, 2007.
- [2] O. Shiku, Y. Teshima, and S. Uchida, "Character Normalization Method for Recognizing Slanted Characters," The IEICE Transactions, Vol.J100-D, No.10, pp.902-906, 2017.
- [3] Y. Yamashita and J. Tani, "Emergence of Functional Hierarchy in a Multiple Timescale Recurrent Neural Network Model: a Humanoid Robot Experiment," PLoS Computational Biology, Vol.4, No.11, e1000220, 2008.