

構造に注目した神経力学モデルを用いた 漢字の筆跡ダイナミクスの学習

野本 楓[†]西出 俊[‡]康 鑫[‡]任 福[‡][†]徳島大学 理工学部[‡]徳島大学 大学院社会産業理工学研究部

1. はじめに

近年、文字認識に関する研究が盛んに行われているが、その多くは文字全体を1つの画像として認識しており、筆跡のダイナミクスを考慮したものはほとんどない。大田らは、文字部品やストローク(画)など漢字の階層構造に注目し、確率文脈自由文法を用い、少ない学習データでも十分な認識精度が得られることを示した[1]。梅田らは、手書き漢字を対象に正規化した特徴量の抽出を行いニューラルネットワークで学習を行うことにより、人間の学習、認知過程をモデル化した認識システムを提案し、そのシステムの有用性を示した[2]。本研究では、ペンの持ち上げを考慮した神経力学モデルを用い、漢字全体を執筆する動作を1つのダイナミクスとした学習手法を提案する。

2. 神経力学モデル MTRNN

本研究では漢字の筆跡ダイナミクス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 の構成と各機能について述べる。

2.1 MTRNN の構成

MTRNN は現時刻 (t) の状態を入力し、時刻 ($t+1$) の状態を出力する多層ニューラルネットワークであり、入出力層の他に複数の階層から成るコンテキスト層 (C_f 層, C_s 層, C_{ss} 層) で構成される。モデルの概略図を図1に示す。各層のニューロン群は時定数と呼ばれる値によって発火速度が制御される。本研究のモデルでは時定数の値を、入出力層, C_f 層, C_s 層, C_{ss} 層の順に大きくなるように設定することにより、時系列情報を階層的に学習する。各種計算式は紙面の都合上省略する。

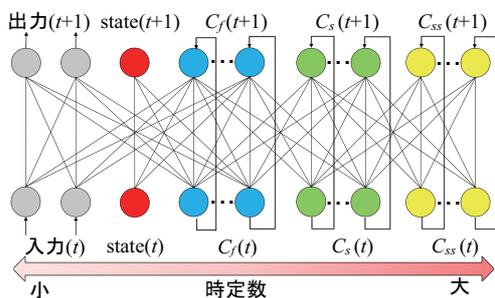


図1: MTRNN の構成

2.2 MTRNN の機能

MTRNN には学習、認識、生成の3つの機能がある。

MTRNN の学習: 教師データ(時系列)を入力し、出力誤差が収束するまで各ニューロン間の重みとコン

テキストの初期値 $C_f(0)$, $C_s(0)$, $C_{ss}(0)$ を更新する。得られたコンテキストの初期値は各時系列データの類似度を自己組織化する形で求まる。

MTRNN の認識: 認識するデータに対し、学習と同様の方法で $C_f(0)$, $C_s(0)$, $C_{ss}(0)$ を求める。得られた $C_f(0)$, $C_s(0)$, $C_{ss}(0)$ は認識時系列を表現する。

MTRNN の生成: $C_f(0)$, $C_s(0)$, $C_{ss}(0)$ を入力し、時系列データを生成する。本研究では認識で得られた $C_f(0)$, $C_s(0)$, $C_{ss}(0)$ を用いる。

3. 提案手法

本章では MTRNN を用いた漢字の筆跡ダイナミクスの学習法について述べる。

3.1 ペンの持ち上げと state ニューロン

本研究で対象とするデータは手書き漢字の筆跡ダイナミクスであるため、漢字の画と画の間にペンの持ち上げによるデータの欠損が存在する。力学モデルは現時刻の状態から次の時刻の状態を出力するため、データ欠損時は出力値を求めることができない。本研究では、前の時刻の出力を現時刻の予測値として入力することでデータ欠損に対応したモデルを提案する。

本研究では漢字の執筆時に取得するペンの座標値 (x, y) を MTRNN に入力する。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。提案モデルでは、ペンの状態(持ち上げ中か執筆中か)を判定するために state ニューロンを組み込んだ。state ニューロンはデータが存在する間(執筆中)は発火せず、データが欠損している場合(ペン持ち上げ中)に発火するように学習する。

3.2 漢字の階層的な学習

本研究では漢字の筆跡ダイナミクスを段階的に学習するために、モデルを学習データに合わせて階層的に変化させる手法を提案する。学習データを1. 偏の画、2. 偏または旁、3. 漢字全体と変化させ、それに伴って MTRNN のモデルのコンテキスト層も 1. C_f 層のみ、2. C_s 層を接続、3. C_{ss} 層を接続と変化させる。

1. 漢字の画のみを学習する。漢字の画執筆時の座標データを入出力、state ニューロン、 C_f 層のみを接続した MTRNN で学習する。これにより、 C_f 層に画のダイナミクス情報を学習する。
2. 1. で学習したモデルに C_s 層を接続したモデルを構築する。偏または旁を執筆した際の座標データを用いて学習する。これにより、 C_s 層に漢字の偏と旁のダイナミクス情報を学習する。
3. 2. で学習したモデルに C_{ss} 層を接続したモデルを構築する。漢字全体を執筆した際の座標データを用いて学習する。これにより、 C_{ss} 層に漢字全体のダイナミクス情報を学習する。

提案手法で各コンテキスト層が学習する情報を図2に示す。

Learning of Handwriting Dynamics of Kanji Characters Using Neurodynamical Model Focusing on Structure Fusei Nomoto (Tokushima Univ.), Shun Nishide (Tokushima Univ.), Xin Kang (Tokushima Univ.), and Fuji Ren (Tokushima Univ.)

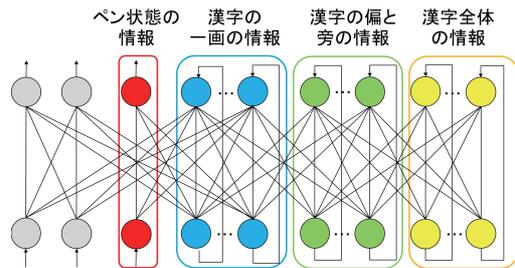


図 2: MTRNN における階層的な学習

4. 漢字データの取得と実験設定

本研究では, Wacom 社の液晶タブレット Cintiq13HD (DTK-1301/KO)(図 3) を用いて漢字執筆時の座標データを取得する. 学習する文字として偏 3 種 (てへん, きへん, さんずい) と旁 5 種 (反, 由, 白, 分, 主) の組み合わせから成る計 15 種の漢字 (扳, 抽, 拍, 扮, 挂, 板, 柚, 柏, 粉, 柱, 浜, 油, 泊, 汾, 注) を対象とする. 画, 偏・旁, 漢字をそれぞれ 5 回ずつ書き, 液晶タブレットでペン先座標データ (x, y) を 10(frames/sec) で取得した. 各文字の座標系列例を図 4 に示す.

MTRNN の構成として, 入出力ニューロン数は座標データ (x, y) の 2, state ニューロン数は 1, C_f ニューロン数は 45, C_s ニューロン数は 70, C_{ss} ニューロン数は 15 と設定した. 各コンテキスト層のニューロン数は複数の設定で実験を行い, 最も良い結果が得られたものを採用した. 実験では画, 偏・旁, 漢字全体の順に学習データを用いて MTRNN を学習し, 得られた MTRNN モデルを用いて漢字を生成した.



図 3: 液晶タブレット



図 4: 15 種類の漢字

5. 実験結果

複数の MTRNN の学習結果のうち, 最も学習性能が良かったもの (学習誤差が小さくなったもの) を用い, 漢字の生成を行った. 主観的に評価し, 比較的漢字の生成に成功したものを図 5 に, 失敗したものを図 6 に示す. 漢字の生成に失敗したものもあるが, 図 5 のように本手法の有効性を示唆する結果が得られた.

6. 考察

実験結果より, 提案手法によって漢字の筆跡ダイナミクスを段階的に学習し, 生成することが可能であることが示されたが, 一方で生成に失敗した系列も見られた. 本章では生成に失敗した原因について考察する.

本実験では, 第 2 段階の偏と旁の学習の際に学習誤差が十分に小さくなるものが得られなかった. その誤差が第 3 段階の漢字の学習に影響したものと考えられる. また, 本実験結果では偏よりも旁の生成結果の方が乱れていることがわかる. これは力学モデルにおける生成誤差

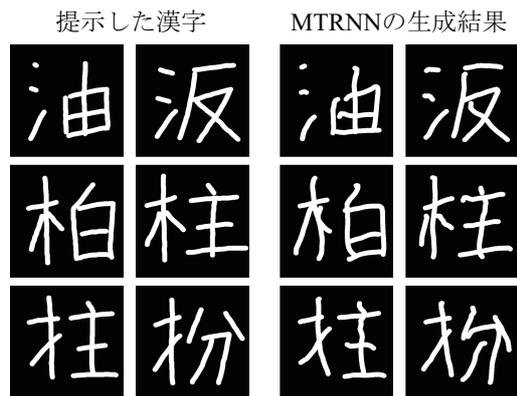


図 5: 成功した生成結果 (左が提示漢字, 右が生成結果)



図 6: 失敗した生成結果 (左が提示漢字, 右が生成結果)

の累積が原因であると考えられる. モデルの学習性能を向上させることが今後の課題である.

7. おわりに

本稿では, 神経力学モデル MTRNN を用いた漢字の筆跡ダイナミクスの学習システムの構築について報告した. 提案手法では, 漢字の構造に注目し, 画, 偏・旁, 漢字全体と段階的な学習を行う. 実験結果から, 提案手法によって漢字の生成が可能であることを確認した.

今後は, 学習性能を向上し, より複雑な漢字についても評価する予定である. さらに, 従来の画像ベースの漢字認識手法も導入し, 形状と筆跡を統合した漢字認識手法へと展開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] 梅田 三千雄, 横田 宏, "ニューラルネットワークの 2 段階構成による手書き漢字認識", 情報処理学会論文誌 42(11), 2724-2732, 2001-11-15, 一般社団法人情報処理学会.
- [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.