

音符・コード系列に注目した 神経力学モデルによる音楽情報の階層的学習

天羽 水穂[†] 西出 俊[‡] 康 鑫[‡] 任 福継[‡]
[†]徳島大学 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、ユーザの好みに合わせた音楽推薦システムや、専門知識のないユーザに対して作曲・演奏補助システムの構築に関する研究が進められている [1] [2]. 本研究でも同様に作曲補助システムの構築を目的としており、特に神経力学モデルによるモデルの構築を目指す. 本モデルは音符・コード系列情報を階層的に学習することでその力学的特性を自己組織化する. 本稿ではその基盤システムとして、少数の楽譜情報を学習した結果について報告する. 実験では自己組織化したパラメータに基づいて学習した楽譜情報を復元できることを示す.

2. 神経力学モデル MTRNN

本研究では、神経力学モデルとして階層的な情報の学習が可能な Multiple Timescale Recurrent Neural Network(MTRNN)[3] を用いる. MTRNN は現在の状態 $IO(t)$ を入力とし次の状態 $IO(t+1)$ を出力する予測器である. MTRNN のニューロンは入出力 IO 層, Fast Context(C_f) 層, Slow Context(C_s) 層に分かれており、各ニューロン群は時定数と呼ばれるパラメータによって発火速度を制御する. 発火速度は IO 層, C_f 層, C_s 層の順に遅くなる. ニューロンの発火速度の違いにより、MTRNN は異なるレベルの情報を階層的に学習することが可能である. 次節で本研究で用いる MTRNN の構成と機能について述べる.

2.1 本研究で用いる MTRNN の構成

本研究では音符情報とコード情報を MTRNN によって階層的に学習することを目標とする. 音符情報とコード情報の変化の速度に注目すると、コードの変化は小節単位のものが多く、音符より遅く設定されている. そこで、提案モデルで音符情報は IO 層, コード情報は C_f 層で学習する. また、楽譜全体の情報を C_s 層に配置した制御用 C_s ニューロンによって自己組織化する. 本モデルは、MTRNN による言語情報の学習に関する研究を参考にしている [4]. [4] では IO 層でアルファベット系列を入力することで文章を学習した結果、 C_f 層で単語情報が、制御用 C_s ニューロンで文法情報が自己組織的に学習されることが示されている. 本稿ではこれを楽譜の学習に応用することを目指している. モデルの構成を図 1 に示す.

2.2 MTRNN の機能

MTRNN の基本的な機能は、学習・認識・生成の 3 つである. 学習には順計算と Back Propagation Through Time(BPTT) を用い、結合部の重みと制御用 C_s の初期値、制御用 $C_s(0)$ を更新する. 学習により、学習データ系列のダイナミクス情報を制御用 $C_s(0)$ ニューロンの中に自己組織化する. 認識では、認識すべき時期列を入力し、BPTT によって制御用 $C_s(0)$ のみを更新する. 認識す

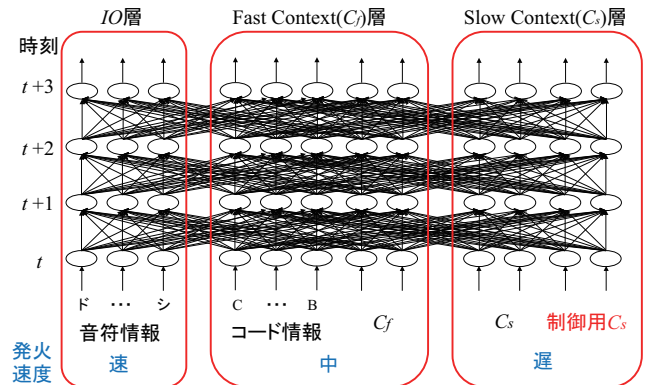


図 1: MTRNN の構成

べき時系列は (1) 音符系列, (2) コード系列, (3) 音符系列とコード系列のいずれかを用いる. 生成では、任意の制御用 $C_s(0)$ を MTRNN に入力することで、順計算によって各時刻における音符・コード値を逐次計算し、 $C_s(0)$ が表現する音符・コード系列データ (楽譜情報) を計算する.

本手法ではコード情報のみを認識することで制御用 $C_s(0)$ 値を計算し、得られた制御用 $C_s(0)$ 値で音符情報を生成することで楽譜の自動生成を行う. 本稿では提案手法の学習能力を検証することが目的であり、制御用 $C_s(0)$ 値から学習した曲を復元することを目的とする.

3. 実験設定

提案手法を評価するために、一つの曲の楽譜を変調した楽譜情報を学習・認識・生成する. 使用した楽譜データの前提条件を図 3 に示す. 音符情報とコード情報については、各音符・コードに対応するニューロンを設定し、各時刻において音符・コードが存在する時に発火するように学習する. 本稿では単音のみを扱うため、発火関数として、音符・コードニューロン共に softmax 関数を用いてもよいが、和音への拡張を考慮し、音符ニューロンの発火関数は sigmoid 関数とし、コードニューロンの発火関数のみ softmax 関数とする.

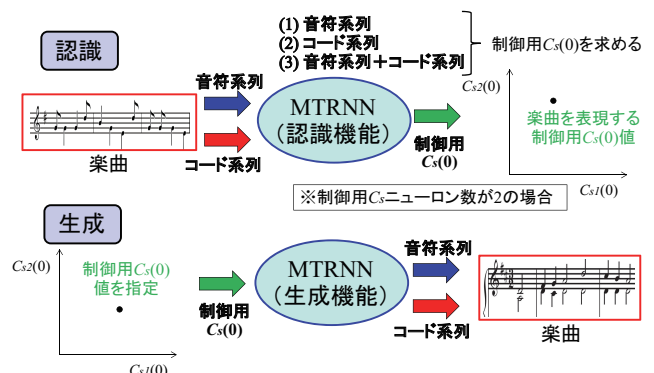


図 2: 提案手法の概略図

Hierarchical Training of Music Information Using Neurodynamical Model Focusing on Note and Code Sequences Mizuho Amo (Tokushima Univ.), Shun Nishide (Tokushima Univ.), Xin Kang (Tokushima Univ.), and Fuji Ren (Tokushima Univ.)

音符	<ul style="list-style-type: none"> ・単音のみ ・使用音域:3オクターブ ・最小単位:8分音符
コード	<ul style="list-style-type: none"> ・12種類のルート音 (C, C#, D,A#, B) ・3種類の和音(major, minor, 7)

図 3: 楽譜データの前提条件

実験で用いる MTRNN の各層のニューロン数を表 1 に示す。音符ニューロンは使用音域が 3 オクターブであるため $12 \times 3 = 36$ 個に休符の 1 個を加えた 37 個とする。コードニューロンについては 12 種類のルート音それぞれに 3 種類の和音が存在するため $12 \times 3 = 36$ 個とする。 C_f , C_s , 制御用 $C_s(0)$ のニューロン数については様々な構成で学習を行い、実験的に決定した。

本実験では同一の楽譜を変調して作成した 3 つの楽譜データについて実験を行った。3 つの楽譜データ全てを学習に用い、それぞれの楽譜データ (音符とコード系列) を認識して得られた制御用 $C_s(0)$ を用いて MTRNN で楽譜データを生成した。生成された楽譜データが元の楽譜データと一致するかどうか検証した。

4. 実験結果

本実験の結果の一例を図 4 に示す。図 4 の左に元の楽譜情報を、右に生成した楽譜情報を示しており、同一の音符・コードを同一色で塗っている。左上の数字は一つの楽譜におけるデータ数を表している。音符については各要素のアルファベット (do, re, mi, fa, so, ra, ci) の頭文字であり、数字は 3 オクターブ内の音符の高さを表している。元楽譜データにおけるハイフン (-) は一つ前の音

表 1: 実験におけるニューロン数

音符情報	コード情報	C_f	C_s	制御用 C_s
37	36	100	30	6

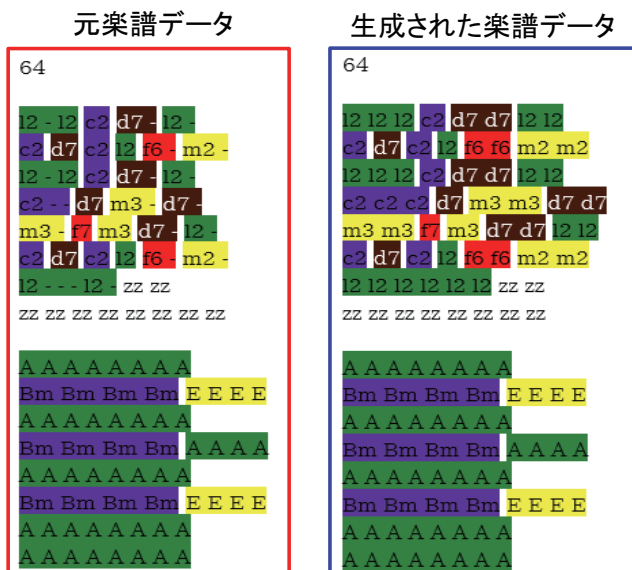


図 4: 実験結果例

符と同じ音符が続いていることを表し、zz は休符を表している。

元楽譜データと生成された楽譜データを比較すると、音符系列とコード系列共に全て再現できていることが確認できた。以上から、MTRNN の認識と生成の機能により、学習した楽譜については音符とコード系列情報を 1 つのパラメータ (制御用 $C_s(0)$) で表現可能であることが示された。

5. 考察

本稿では神経力学モデル MTRNN を用いた楽譜の自動生成を行うための基盤となるモデルの構築について報告した。本実験では学習データについてのみ評価を行っており、目指す楽譜の自動生成においては未知の楽譜情報も扱う必要がある。MTRNN はロボットの実験などにおいて未学習のダイナミクスも汎化能力によって扱うことができることが示されており、本実験結果は未学習の楽譜情報についても適用できることを示唆する結果になっていると考えている。今後はモデルの汎化能力と未知楽譜生成の関係性について検証することが必要である。

本実験では単音のみを扱ったが、モデルでは発火関数を sigmoid 関数としているため、和音にも適用可能である。一方で、入力データ量が多くなることで MTRNN の学習が困難になることも考えられ、入力音を単音から和音にした場合や、学習する楽譜数を増やす場合はニューロン数も増やすことが必要になると考えられる。学習時間はニューロン数に大きく影響されるため、今後はより効率のよい学習法を開発することも必要になる。

6. おわりに

本稿では神経力学モデル MTRNN を用いた音符系列とコード系列の階層的な学習について報告した。提案手法では MTRNN の異なる階層で音符情報とコード情報を学習し、認識と生成の機能を用いて、楽譜情報と制御用 $C_s(0)$ を相互に計算する。評価実験では 1 つの楽譜を変調した 3 つの楽譜データを用いてモデルの学習を行い、学習した楽譜を認識・生成した結果、元の楽譜が再現できることを確認した。

今後は、より多くの楽曲データにおいて評価し、モデルの汎化能力について検証し、未知楽譜の生成可能性についても検証していきたい。

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] 須見 康平, 糸山 克寿, 吉井 和佳, 駒谷 和範, 尾形 哲也, 奥乃 博, “ベース音高と和音特徴の統合に基づく和音系列認識,” 情報処理学会論文誌, Vol. 52, No. 4, pp. 1803-1812, 2011.
- [2] 白井 亨, 谷口 忠大, “可変長階層 Pitman-Yor 言語モデルを用いたメロディー生成手法の提案,” 知能と情報 (日本知能情報ファジィ学会誌兼論文誌), Vol. 25, No. 6, pp. 901-913, 2013.
- [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.
- [4] W. Hinoshita, H. Arie, J. Tani, H. G. Okuno, and T. Ogata, “Emergence of Hierarchical Structure Mirroring Linguistic Composition in a Recurrent Neural Network,” Neural Networks, Elsevier, Vol. 24, Issue 4, pp. 311-320, 2011.