

LSTM を用いたリカレントニューラルネットワークによる コード進行と音高・リズムの2段階での自動作曲

島吟 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

自動作曲に関する様々な研究が行われているが、その中のひとつに LSTM (Long Short-Term Memory)[1] を用いたリカレントニューラルネットワークによる自動作曲 [?]-[?] が提案されている。しかしながら、コード進行を考慮した曲の生成ができない、曲の終わり方が不自然になることがある、曲の途中で曲の終わりのような音の並びが生成されてしまうことがあるなどの問題がある。また、人間が作曲を行う場合のようにリズムと音高を同時に考えた曲の生成を行うことはできない。さらに、曲を生成する際に、既存の曲の一部を入力として与え、それに続く形で曲を生成しているため、途中から曲が始まっているような感じになってしまうこともあるという問題もある。

本研究では、LSTM を用いたリカレントニューラルネットワークによるコード進行と音高・リズムの2段階での自動作曲を提案する。この手法では、まず既存の曲のコード進行の学習を行ったリカレントニューラルネットワークを用いてコード進行を生成する。次に、生成されたコード進行を考慮し、既存の曲の音高とリズムの学習を行ったリカレントニューラルネットワークを用いてリズムと音高を生成していく。これにより音高・リズムを同時に生成することを可能にする。

2 LSTM を用いたリカレントニューラルネットワークによるコード進行と音高・リズムの2段階での自動作曲

提案する LSTM を用いたリカレントニューラルネットワークによるコード進行とリズム・音高の2段階での自動作曲では、既存の曲の特徴について学習を行ったリカレントニューラルネットワークを用いて、コード進行を生成した後、音高・リズムの生成を行う。



図 1: 学習に用いる曲 (一部) の例

2.1 曲の特徴の表現

2.1.1 学習データ

学習に用いる曲には、様々な調が用いられている可能性がある。それをそのまま学習に用いてしまうと、様々な調の音階に含まれる音の高さが学習されてしまい、生成される曲に調の音階に含まれない音が含まれてしまう可能性が高くなる。そこで、提案システムではすべて八長調に移調したものを学習データとして用いる。

2.1.2 コード進行

コード進行は $1/2$ 小節単位で扱い、コードはコード名で表現する。図 1 の例では、 $Am7 \rightarrow Gsus4 \rightarrow Fadd9 \rightarrow C \rightarrow \dots$ となる。また、曲の始まる前の部分のデータに関しては、曲の始まりを表す S で表現することで曲の始まりを考慮できるようにする。

2.1.3 音高・リズム・小節内の位置・曲の中での位置

音高やリズムの特徴は 16 分音符の長さ単位で表現する。音高は C_5 や D_5 のようにオクターブの情報も含めた音名で表す。なお、休符は R で表す。リズムは休符の開始を 0、音符の開始を 1、前の状態の継続である状態を 2 で表す。音高とリズムの曲の始まる前の部分のデータに関しては、コードの場合と同様に、曲の始まりを表す S で表現する。また、小節内の位置は 4 分の 4 拍子であれば 0~15 の数値で表す。曲の中での位置は最後の小節は 1、それ以外は 0 で表す。

2.2 コード進行の学習・生成

コード進行を生成する LSTM を用いたリカレントニューラルネットワークでは、 M_c ブロック分のコード (ブロック $b - M_c + 1 \sim b$ のコード) と、ブロック

Automatic Composition in Two Stages: Chord Progression and Pitch/Rhythm using Recurrent Neural Network using LSTM
Gin Shima and Osana Yuko (Tokyo University of Technology, osana@stf.teu.ac.jp)

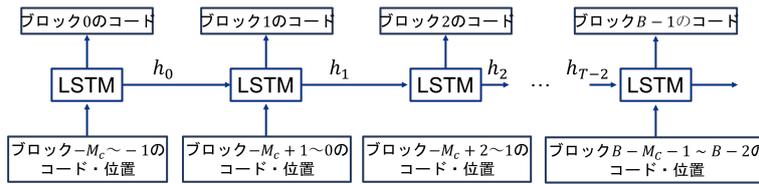


図 2: コード進行の生成を行う LSTM を用いたリカレントニューラルネットワーク

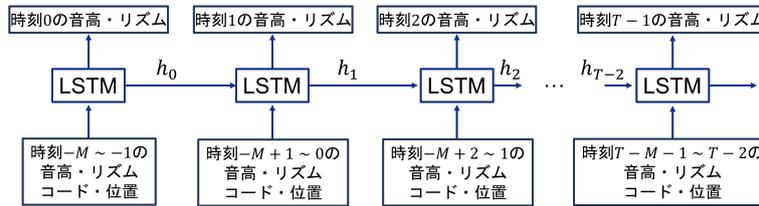


図 3: 音高・リズムの生成を行う LSTM を用いたリカレントニューラルネットワーク

$b+1$ の曲の中での位置を入力とし、次のブロック $b+1$ のコードを出力するように学習を行う。図 2 にコード進行の生成を行う LSTM を用いたリカレントニューラルネットワークの流れを示す。

$M = 3$ とし、図 1 の例に対して、学習を行う場合にはリカレントニューラルネットワークには

[ブロック b]	[入力]	[出力]
0	S S S 0	Am7
1	S S Am7 0	Gsus4
2	S Am7 Gsus4 0	Fadd9
3	Am7 Gsus4 Fadd9 0	C

のようなデータを用いて学習を行うことになる。

コード進行を生成する際には、 M_c 個の S と 0 を初期入力として与え、曲の先頭からコードを順に生成していく。次のブロック以降は、すでに生成されたブロックのコードも入力として利用する。なお、生成する曲の長さは学習した曲の中からランダムに選択することで決定するものとする。

2.3 音高・リズムの学習・生成

音高・リズムを生成する LSTM を用いたリカレントニューラルネットワークでは、 M 時刻分の音高・リズム (時刻 $t - M + 1 \sim t$ の音高・リズム) と、時刻 $t + 1$ のコード、小節内での位置、曲の中での位置を入力とし、次の時刻 $t + 1$ の音高・リズムを出力するように学習を行う。図 3 に音高・リズムの生成を行う LSTM を用いたリカレントニューラルネットワークの流れを示す。

音高とリズムを生成する際には、 $2M$ 個の S と位置を表す 2 つの 0 を初期入力として与え、曲の先頭から順に生成していく。次の時刻以降は、すでに生成さ



図 4: 生成された曲 (一部) の例

れた時刻の音高とリズムも入力として利用することになる。

3 計算機実験

図 4 に提案システムによって生成した曲の一部を示す。

参考文献

- [1] S. Hochreiter and J. Schmidhuber: "Long short-term memory," *Neural Computation*, No.9, Vol.8, pp.1735-1780, 1997.
- [2] 姫野雄大: "LSTM による自動作曲システムの構築," 東京大学大学院修士論文, 2016.
- [3] 小山凌平, 長名優子: "LSTM を用いたリカレントニューラルネットワークによる自動作曲," 情報処理学会第 83 回全国大会, 2021.
- [4] 関快斗, 長名優子: "LSTM を用いたリカレントニューラルネットワークによる自動作曲における度数に基づく音形表現の導入," 情報処理学会第 84 回全国大会, 2022.
- [5] 皆銭拓真, 長名優子: "LSTM を用いたリカレントニューラルネットワークによる弦楽四重奏の自動作曲," 情報処理学会第 85 回全国大会, 2023.
- [6] 加藤義隆, 長名優子: "LSTM を用いたリカレントニューラルネットワークによるコード・リズム・音高の 3 段階での自動作曲," 情報処理学会第 85 回全国大会, 2023.