2T-03

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

関快斗　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

コンピュータを用いた自動作曲は古く、1957年のマルコフ過程を用いた自動作曲の研究以来、様々な手法を用いた自動作曲の研究が行われている。近年ではDeep Learningを用いた自動作曲 [1] に関する研究も盛んに行われている。

そのような手法の1つとして、LSTM (Long Short-Term Memory)[2] 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システム [3] が提案されている。このシステムでは、中間層にLSTMを用いたリカレントニューラルネットワークによって既存の楽曲の特徴を学習し、学習済みのリカレントニューラルネットワークを利用して曲の生成を行っている。学習データとしては、音形、音高、小節内の位置を用いている。音形とは和音を構成する音どうしの高さの関係を表したものである。しかし、この手法では音形を半音単位で表現しているため、同じ音形でも音高によっては調の音階以外の音が発生してしまうという問題がある。

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自動作曲における度数に基づく音形表現の導入を提案する。これは、LSTMを用いたリカレントニューラルネットワークによる自動作曲システム [3] における音形を度数に基づくものに変更することで、音階に属する音とそれ以外の音とを区別して扱えるようにしたものである。

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自動作曲は、従来手法 [3] に基づいたものである。この手法では、既存の曲の特徴をLSTMを用いたリカレントニューラルネットワークで学習し、それにランダムな初期値を与えて曲を生成する。リカレントニューラルネットワークには時刻 t の音形、音高、小節内の位置を入力し、次の時刻 $t+1$ の音形と音高を出力するように学習を行う。なお、従来手法では音形を半音単位で表現していたが、提案手法では度数単位に変更することで不自然な音、音階以外の音が生成されることを抑制する。

2.1　学習データ

学習データにはクラシック曲のMIDIのデータベース [4] のデータを用いる。データベースに含まれる曲には様々な調のものが含まれているが、すべて移調してハ長調に揃えたものを学習データとして用いる。音の高さや長さの情報はMIDIデータからMusic21[5] というライブラリを用いて抽出する。Music21はMITで開発されたPythonのライブラリで、音楽解析や出力された情報のMIDIデータやピアノロールへの変換を行うことができる。提案手法では、抽出した情報から16分音符の長さ単位で音形、音高、小節内の位置といった形でベクトル表現し、それを入出力として用いる。

2.2　曲のデータベクトル表現

曲のデータはベクトルとして表現する。提案手法では、16分音符の長さ単位 (以下、ブロックと呼ぶ) で考えるものとし、それぞれが音形、音高、小節内での位置の情報を持つ。

(1) 音形

音形は、同時に鳴っている音どうしの高さの関係と音の状態を表現したものである。音の存在する範囲のみを考慮し、音階の音とそれ以外の音を区別して表現する。

提案手法では音どうしの高さの関係は度数で表現する。度数とは音程を表す単位であり、同じ高さであれば1度、音階の隣りの音であれば2度のように数える。なお、度は音階の音が基準となっているため、図1の右の2つの和音における下の2つの音程はいずれも3度となる。ただし、厳密には D_4 と $F_4\sharp$ では半音

Introduction of Degree-based Sound Expression in Music Composition by Recurrent Neural Network with LSTM

Kaito Seki and Yuko Osana (Tokyo University of Technology, osana@stf.teu.ac.jp)

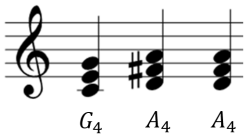

図 1: 和音の例

4 つ分だけ違うので長 3 度、D_4 と F_4 では半音 3 つ分だけ違うので短 3 度と区別される。提案手法では、臨時記号で表される場合はすべて ♯ を使った表現で表すように統一して扱うものとしている。

ブロックごとの音の状態は 0〜3 の数値で表現する。それぞれ、音が存在しない (0)、音の開始 (1)、前の音の継続 (2)、休符 (3) を表す。音形は同時に鳴っている一番上の音から音が存在するところに音の状態を表す 0〜3 の数値で割り当てる。図**??**の 5 つ目の和音と 6 つ目の和音はいずれも A_4, F_4, D_4 の 3 音から構成されているが、5 つ目の音形は [[1, 0, 1, 0, 1], [0, 0, 0, 0, 0]]、6 つ目の音形は [[2, 0, 2, 0, 2], [0, 0, 0, 0, 0]] と表すことになる。いずれも前半の 5 つの数字が音階の音、後半の 5 つの数字が音階以外の音を表している。

提案手法の音形の表現方法では、図 1 の左の和音と右の和音の開始位置のブロックはいずれも [[1, 0, 1, 0, 1], [0, 0, 0, 0, 0]] となる。それに対し、中央の和音の開始位置のブロックは [[1, 0, 0, 0, 1], [0, 0, 1, 0, 0]] のように表されることになる。**1** で述べたように従来手法では、左と中央の和音の音形が同じになっていたが、提案手法では、音階以外の音を含まない左と右の和音の音形が同じになる。

音形をベクトルデータとして表す際には、0〜3 で音の状態を表したのもではなく、学習データに含まれるすべての音形に対して番号を割り当てた音形番号を用いる。

(2) 音高

音高は同時に鳴っている音の中で最も高い音の音高を数値で表したもので表現する。数値は A_0〜C_8 の音に 0〜87 の番号を割り当てたものを用いる。最高音の情報を音高として用いているのは、最高音が主旋律を構成していることが多いと考えられるからである。

(3) 小節内での位置

小節内での位置は 1 小節を 16 分音符の長さ単位で区切ったブロックを 0〜15 の数値で表したもので表現する。

2.3 学習

学習では、その時刻の音形、音高、小節内の位置を入力し、次の時刻の音形と音高を出力するように学習を行う。出力された音形と音高と教師信号と誤差に基づいて重みを更新し、正しい出力が出せるように学習していく。

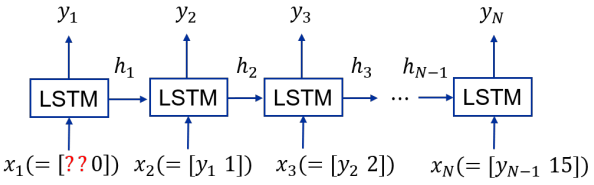

図 2: 曲の生成のイメージ

2.4 曲の生成

曲は、曲の特徴を学習させたリカレントニューラルネットワークに初期値を与えることで生成する。学習データの中からランダムに 1 音を取り出したものを初期値とし、それを入力として次の時刻の音形と音高を出力する。出力された次の時刻音形と音高に小節内の位置を加えたものが次の時刻のリカレントニューラルネットワークへの入力となる。図 2 に示すようにこれを繰り返すことで曲を生成する。

3 計算機実験

提案手法を用いて曲の生成を行い、従来の手 [3] に比べて不自然な音階以外の音の少ない曲が生成できることを確認した。

参考文献

[1] K. Choi, G. Fazekas, K. Cho and M. Sandler : “A tutorial on deep learning for music information retrieval,” http://arxiv.org/abs/1709.04396., 2017 (2022 年 1 月 5 日参照).

[2] S. Hochreiter and Jrgen Schmidhuber : “Long short-term memory,” Neural Computation, Vol.9, No.8, pp.1735–1780, 1997.

[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] MIDI クラシック音楽データ集 by Windy, https://windy-vis.com/art/, (2022 年 1 月 5 日参照).

[5] Music21: a Toolkit for Computer-Aided Musicology - MIT : http://web.mit.edu/music21/, (2022 年 1 月 5 日参照).