

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

加藤義隆 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

近年、Google の Magenta[1] などに代表されるようにニューラルネットワークを用いて自動作曲を行おうとする試みも多く行われている。そのような手法の1つとしてリカレントニューラルネットワークを用いた自動作曲が提案されている。この手法では、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を用いて、既存の曲の特徴を学習し、それを利用して曲の生成を行っている [3]。このニューラルネットワークは、現在の音形と音高、小節内の位置を入力し、次の時刻の音形と音高が出力されるようなものとなっている。しかしながら、この手法では、コード進行を考慮して学習を行っていないため、コード進行を考慮して曲を生成することはできない。また、曲を生成する際に、曲の終わりかどうかを考慮していないため、生成した曲の最後が違和感のある終わり方になっ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コード・リズム・音高の3段階での自動作曲を提案する。従来手法 [3] のコード進行を考慮した作曲が行えないという問題に対しては、コード進行をまず生成し、それを考慮して音高の生成を行うようにすることで解決する。また、曲の最後が違和感のある終わり方になるという問題に対しては、音高を生成する際に曲の中での位置 (最後の小節かどうか) も入力として与えることで解決を目指す。

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・リズム・音高の3段階での自動作曲

ここでは、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・リズム・音高の3段階での自動作曲について説明する。提案手法では、

LSTM を用いたリカレントニューラルネットワークを用いて既存の曲のコード進行、リズム、音高を学習し、学習したリカレントニューラルネットワークを用いて曲の生成を行う。

2.1 学習データとその表現

2.1.1 学習に用いる曲

学習データとしてはコード進行を考慮して作曲されていると考えられる J-pop などの既存の楽曲を想定している。また、学習曲はすべて八長調に移調したものを学習データとして利用する。

2.1.2 コード進行

コード進行は $1/2$ 小節単位で考える。リカレントニューラルネットワークへの入出力として扱う場合には、すべての学習曲に含まれるコードを考慮したワンホットベクトルで表す。

2.1.3 リズム

提案システムでは、コードと同様に $1/2$ 小節単位のブロックでのリズムパターンで表現する。ブロック内でのリズムのパターンは 16 分音符の長さ単位で、それぞれに対して 0 (休符の開始)、1 (音符の開始)、2 (前の状態の継続) のいずれかを割り当てることで表現する。さらにタイで表現される場合のように次のブロックにリズムが継続していることもあるため、次のブロックに音符が継続しているかも 0 (継続なし)、1 (継続あり) の状態を与えることで表現する。リカレントニューラルネットワークへの入出力として扱う場合には、すべての学習曲に含まれるリズムパターンを考慮したワンホットベクトルで表す。

2.1.4 音高

音高は、コードやリズムとは異なり 1 音単位で表現する。リカレントニューラルネットワークへの入出力

Automatic Composition in Three Steps of Chord, Rhythm, and Pitch by Recurrent Neural Network with LSTM
Kato Yoshitaka and Osana Yuko (Tokyo University of Technology, osana@stf.teu.ac.jp)

として扱う場合には、すべての学習曲に含まれる音符の高さを考慮したワンホットベクトルで表す。

2.1.5 音の長さ・小節内で位置・曲の中での位置

音高を考える際には音の長さや小節内での位置も考慮する。音の長さは16分音符の長さ単位で考え、すべての学習曲に含まれる音符の長さを考慮したワンホットベクトルで表す。小節内での位置は、小節を16分音符単位の長さで区切ったときの位置であり、4分の4拍子であれば0~15で表現する。リカレントニューラルネットワークへの入力として扱う場合にはこれをワンホットベクトルで表現したものをを用いる。

また、コード、リズム、音高を考える際に、曲の最後であるかを考慮できるように曲の中での位置も0と1で表す。最後の小節は1で、それ以外は0で表される。

2.2 コード進行の学習・生成

コード進行は、現在のブロックのコードと曲の中での位置を入力し、次のブロックのコードを出力するように学習を行ったリカレントニューラルネットワークを使って生成する。コード進行は、学習曲の先頭のコードからランダムに選択して決定した先頭のコードを初期の入力として、次のコードを次々と出力させることで生成する。

2.3 リズムの学習・生成

リズム系列は、現在のブロックのリズムパターンと次のブロックへの継続の有無、曲の中での位置を入力し、次のブロックのリズムパターンと継続の有無を出力するように学習を行ったリカレントニューラルネットワークを使って生成する。

リズム系列は、学習曲の先頭のリズムからランダムに選択して決定したものを初期の入力として、次のリズムを次々と出力させることで生成する。

2.4 音高の学習・生成

音高は、現在の音の音高、次の音のコード、長さ、小節内での位置、曲の中での位置を入力し、次の音の音高を出力するように学習を行ったリカレントニューラルネットワークを使って生成する。

音高の系列は、学習曲の先頭の音の音高からランダムに選択して決定したものを初期の入力として、次の音高を次々と出力させることで生成する。



図 1: 生成された曲の例

3 計算機実験

提案システムを用いて生成した曲の例を図1に示す。

参考文献

- [1] Google Brain Magenta, <https://magenta.tensorflow.org/>, (2022年12月21日参照).
- [2] S. Hochreiter and J. Schmidhuber : “Long short-term memory,” *Neural Computation*. No.9, Vol.8, pp.1735-1780, 1997.
- [3] 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.
- [4] 松村昂輝, 大谷紀子, 木村司, 福井健一, 沼尾正行 : “リズム・コード・メロディの三段階深層学習による音楽生成,” 研究報告音楽情報科学, Vol.2019-MUS-125, No.5, pp.1-4, 2019.