

# コード進行と小節数の条件制約下での 深層学習を用いたジャズ音楽の自動生成の検討

小笠原 稜<sup>1</sup> 上瀧 剛<sup>1</sup>

**概要:** 深層学習を用いた音楽の自動生成はこれまでも研究されてきているが、生成結果を制御することはできていない。これは深層学習による音楽の自動生成がモデルの出力結果から確率的に選択されるためである。そのため、従来の深層学習による音楽の自動生成はコード進行や小節数等の条件に強い制約のあるジャズ音楽に適した手法ではない。これに対して、本研究ではジャズ音楽に特化した深層学習による自動音楽生成機構を提案する。具体的には学習した提案モデルに音符情報とコード進行情報を入力し、音符の候補を出力する。この候補の中から木探索によって条件に一致した音符を探索する。これを指定した小節数になるまで繰り返すことで音楽を生成することで、コード進行や小節数の条件に沿ったフレーズ生成が可能である。実験では、従来手法と比べて、よりコードに沿った自然なフレーズを生成可能であることを示す。

## 1. はじめに

本研究は人と機械間での音楽のセッションの実現を目指して研究を進めている。セッションとはコード進行に沿って演奏者が交互に即興演奏をしていくことを指す。セッションを人と機械間で行うためにはロボットによる演奏機構と音楽の自動生成機構の開発が必要である。本研究は2つの機構の内、音楽の自動生成機構について述べる。

これまで音楽の自動生成はマルコフ連鎖等の統計的手法 [1], [2] や深層学習を用いた手法 [3] が提案されている。しかし、これらの音楽の自動生成は確率的に音符を選択するために、生成結果を制御することができない。そのため、従来手法による音楽自動生成は、コード進行や小節数等の条件に強い制約のあるジャズ音楽に適した手法ではない問題がある。

本研究では従来の深層学習による生成にルールモデルを組み込むことで、コード進行や小節数の制約下でのジャズ音楽を生成する自動音楽生成機構を提案する。また、深層学習モデルに音楽的構造を学習させるために、音符情報とコード進行情報から音符を出力するモデルを採用した。これは音楽的構造を学習させることで、コード進行や小節数の制約下でも音楽的な生成を可能とするためである。評価実験では従来手法と比べて、指定した小節でよりコードに沿った自然なフレーズを生成可能であると確認した。

## 2. 関連研究

### 2.1 確率統計モデルによる音楽の自動生成

音楽の自動生成の先駆的な例として、Mozart の「サイコロ遊び」 [1] や、Lejaren らによる世界初のコンピュータが作曲した「イリアック組曲」 [2] がある。前者の場合、予め1小節毎に分割された複数のフレーズが用意されており、新たにフレーズを生成する際はサイコロの出た目に応じたフレーズを選択し、音楽を生成する。後者は元の楽曲の音高や音価に注目し、マルコフ連鎖に基づいて遷移確率を算出して楽曲を生成している。確率統計モデルによる音楽生成は、特定のジャンルや曲調の音楽生成に有効であるが、長期的な音楽の依存性関係を考慮できない問題を持つ。

### 2.2 Magenta

Google Brain チームはMagentaの開発を行っている [3]。これは膨大な量のクラシックやジャズ音楽のデータを基に楽曲の学習を行い、フレーズの生成を行っている。この利点は、人が音楽理論を1つずつコンピュータに教える必要がないことである。深層学習モデルを適切に設計することで、楽曲データから楽曲の特徴、人の感性情報を自動的に学習できる可能性がある。しかし、Magentaは音楽データを音符や小節などの音楽的な切り分けではなく、機械的に特定の秒数ごとに切り分けて学習、生成している。そのため、音楽的な構造を考慮しておらず、コード進行や音楽理論の制約の強いジャズ音楽に適していない問題を持つ。

<sup>1</sup> 熊本大学大学院自然科学教育情報電気工学専攻

### 3. 提案手法

提案する自動音楽生成機構は既存の楽曲を用いた深層学習モデルの学習部分と、学習済みモデルによるジャズ音楽の生成部分に分かれている。

#### 3.1 深層学習モデルの学習

Charlie Parker の譜面 [4] を用いて、音楽を深層学習モデルに学習させた。

##### 3.1.1 データセット

音楽を学習させる際に Charlie Parker の 50 曲 [4] を用いた。また、データ数を増やすために楽曲をそれぞれの調子に転調させ、計 600 曲を学習させた。データセットは音符情報とコード進行情報の 2 つから構成される。音符情報は音階と音価、コードのルート、コードの構成音、小節内の位置から構成される。また、コード進行情報はコードのルート音と構成音から構成される。この際、コード進行情報の先頭には開始記号、末尾には終了記号を付与する。開始記号と終了記号を付与することで楽曲のコード進行だけでなく、楽曲の開始と終了のフレーズ感を学習させる。これらのデータセットに含まれる音階や音価等の情報はすべて onehot ベクトルを用いてエンコードされる。

**音階** 音階は MIDI データのノート番号を採用した。0 から 127 は MIDI 規格の音階情報を持った音符を指す。また、128 は休符を指す。

**音価** 各音符の継続時間を指す音価は TICK 表記を採用した。TICK 表記は 4 分音符を 480 Tick とする音価の表現方法である。

**コード** コードはルート音と構成音から構成される。ルート音は C から B の 12 の音階を用いて表記する。構成音はメジャーコードやマイナーコード、ディミニッシュコード等の表記を用いた。また、開始記号は S, 終了記号は En を使用する。

**小節内の位置** 小節内の位置は TICK 表記を用いて表記する。これによってモデルに音符の小節内の位置を認識させる。

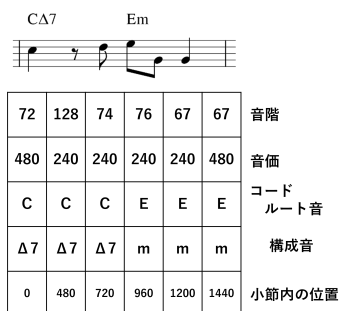


図 1 音符情報のデータ形式。音符情報は音階と音価、コードのルート音、コードの構成音、小節内の位置から構成される。

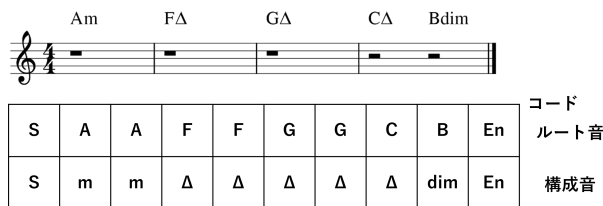


図 2 コード進行情報のデータ形式。コード進行情報はコードのルート音と構成音から構成される。

##### 3.1.2 モデルの構成

音楽学習モデルとして、2 層の Long short-term memory (LSTM) を用いたモデルを作成した (図 3)。作成したモデルは音階と音価を同時に生成する。音符情報には音符の音階と音価、コード、小節内の位置が含まれ、LSTM に入力することでフレーズの特徴を得る。また、コード進行情報には現在のコード、前後の小節のコードが含まれ、それぞれ LSTM に入力することでコード進行の特徴を得る。この際、それぞれの情報は学習された Embedding 層で表現される。次に、それぞれに得られた特徴を結合し、音階と音価のそれぞれの Embedding 層の次元と同じである 2 つの出力を得る。最後に、得られた出力からそれぞれ音階と音価を出力する。

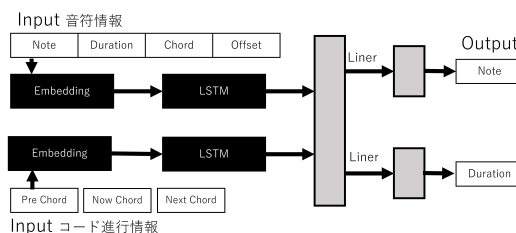


図 3 モデル構成。音符情報とコード進行情報を LSTM を用いてそれぞれ特徴を取得し、得られた特徴から音階と音価を出力する。

#### 3.2 学習済みモデルによる生成

提案する自動音楽生成機構は学習済みの深層学習モデルと木探索によるルールモデルを組み合わせてジャズ音楽を生成する (図 4)。

##### 3.2.1 木探索

提案する自動音楽生成機構は深層学習モデルに木探索によるルールモデルを組み込むことで、音楽理論やコードの制約を設ける。通常、LSTM から得られた確率ベクトルの中で最も値の高い音符を選択し、生成を行う。提案モデルは LSTM から得られた音符の候補の中でコードや拍数、音楽理論等の条件に合うものを選択し、その中で確率が最も高いものを選択する。条件に合う確率ベクトルがなくなる場合が生じても生成できるように木探索を用いる。この探索は次のステップで実行する。

- step1 学習済みモデルに開始文字を入力。
- step2 学習済みモデルの出力の内、現状の条件（コード）に適した音階を取得。
- step3 取得した音階の内、最も確率の高い音階を次入力とする。
- step4 step2の際に現状の条件（コード）に適した音階が存在しない際は親木に戻り、取得した音階の中でstep3で選択した音符の次に確率の高い音階を次入力とする。
- step5 step2からstep4を特定の小節数になるまで繰り返す行い、音楽を生成する。

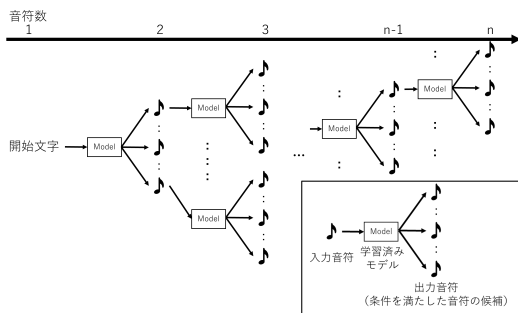


図4 生成時に用いる木探索。学習済みモデルから得られた出力の中で条件に適した音符を確率の高い順に選択し、次のモデルを入力することで条件に適した音楽を探索する。

## 4. 評価実験

提案する自動音楽生成機構を用いて、コード進行と小節数の制約下で音楽生成が可能か検証するために提案モデルを用いて音楽を生成した。本節は提案モデルによる音楽生成とコード条件による音楽生成の制御、オリジナル性の検証で構成されている。

### 4.1 提案モデルによる音楽生成

提案する自動音楽生成機構を用いて生成した音楽を図5に示す。生成音楽1は4分休符を入力とし、12小節かつBbのブルース進行の条件を組み込み生成を行った。また、生成音楽2は4分休符を入力とし、8小節かつCm/F7/Bb7/Bb7/Cm/F7/Bb7/Bb7のコード進行の条件を組み込み生成を行った。生成音楽はいずれも特定の小節で完結したフレーズかつ、指定したコード進行に沿ったフレーズが生成されていることが確認できた。

### 4.2 コード条件による音楽生成の制御

提案モデルと音階情報のみで生成するモデル(NOTE LSTM)による各キーのブルース進行の楽曲と各コードの音階使用率を調査した(図6, 図7)。

提案モデルが調子に合った音楽を生成可能か確認するために、各キーのブルース進行の楽曲の音階使用率を求める。



(a) 生成音楽 1. 12小節かつBbのブルース進行の楽曲



(b) 生成音楽 2. 8小節かつ

Cm/F7/Bb7/Bb7/Cm/F7/Bb7/Bb7のコード進行の楽曲。

図5 提案モデルを用いた音楽生成結果の譜面。4分休符を入力とし、特定の条件を組み込んで音楽生成を行った。

ブルース進行はI7, IV7, V7のコードから構成されるコード進行である。今回生成したブルース進行と類似した既存のブルース進行の楽曲(Another Hairdo)をC, Bb, Fの調子に転調させたときの音階の使用率を図6に示す。図6より、既存のブルース進行の楽曲ではブルース進行に含まれるI7, IV7, V7のコードのルート音を多用している。提案モデルは既存のブルース進行の楽曲と同様に各調子の1度と4度, 5度の音階を多用した音楽であることが確認できる。また、提案モデルはいずれの調子と比較してもNOTE LSTMより、既存のブルース進行の楽曲と類似した音階の使用率である。よって、提案モデルはいずれの調子でもそれぞれの調子に合った音楽を生成可能であることが確認できる。

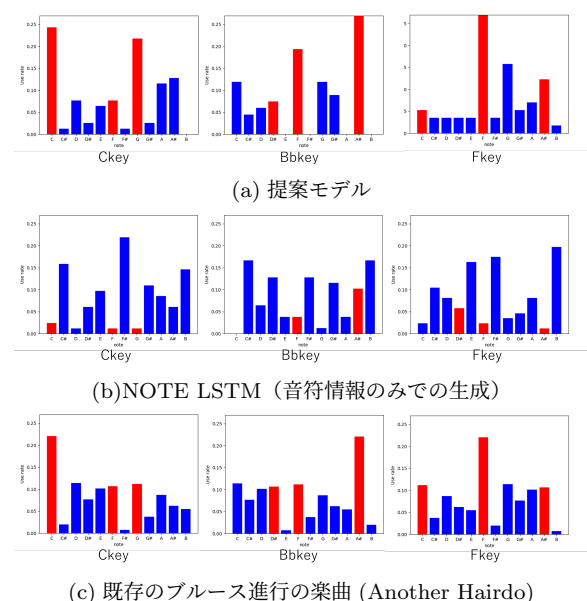
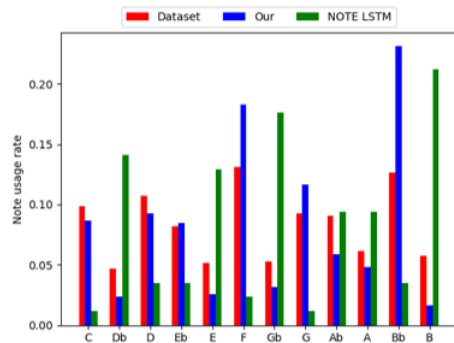
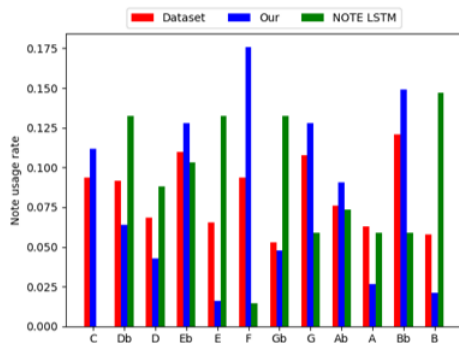


図6 各キーのブルース進行の楽曲を生成した際の音階の使用率。各キーの1度と4度, 5度の音階を赤色, その他の音階を青色で示す。

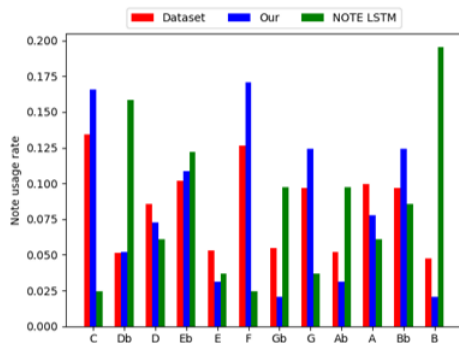
提案モデルがコードに沿った音楽を生成可能か確認するために、各コードの音階使用率を求める。図7より、提案モデルはNote LSTMより、Charlie Parkerの楽曲のコードの中で使われている音階使用率と近いことが確認できる。よって、提案モデルはコード条件に沿った音楽生成が可能であることが確認できる。



(a) コード：Bb7



(b) コード：Eb7



(c) コード：F7

図7 コードごとの音階の使用率。データセットと提案モデル、音符情報のみで生成のコードごとの音階の使用率の比較。いずれのコードも提案モデルが最もデータセットに近い結果が得られた。

### 4.3 オリジナル性の検証

提案モデルによる生成音楽がデータセットの楽曲のコピー&ペーストでないか確認するために正規化相互相関を用いて、楽曲の類似度を求める。

#### 4.3.1 正規化相互相関

正規化相互相関とは入力画像とテンプレート画像をマッチングして入力画像中からテンプレート画像の位置を発見する手法であり、入力画像とテンプレート画像が完全一致の時に最大値1を返し、単に相関がない時に0を返す(式5)。

入力画像の各画素の画素値を  $I(x, y)$ 、テンプレート画像の各画素の画素値を  $T(x, y)$ 、 $x' = 0 \dots w-1$ 、 $y' = 0 \dots h-1$  としたとき、座標  $(x, y)$  での相関係数  $R_{CC}(x, y)$  は式(1)で表される。

$$R_{CC}(x, y) = \sum_{x', y'} [T'(x', y') \cdot I'(x + x', y + y')] \quad (1)$$

ここで、式(1)の  $T'$  と  $I'$  は以下で式(2)、式(3)定義される。

$$T'(x', y') = T(x', y') - \frac{\sum_{x'', y''} T(x'', y'')}{(w \cdot h)} \quad (2)$$

$$I'(x + x', y + y') = I(x + x', y + y') - \frac{\sum_{x'', y''} I(x + x'', y + y'')}{(w \cdot h)} \quad (3)$$

式(1)を正規化係数(式(4))を用いて正規化する。正規化することで  $[-1, +1]$  の範囲に収まるので異なる楽曲でも統一した見方ができる。

$$Z(x, y) = \sqrt{\sum_{x', y'} T'(x', y')^2 \cdot \sum_{x', y'} I'(x + x', y + y')^2} \quad (4)$$

最終的に類似度計算に用いる正規化相互相関を式(5)に示す。

$$R_{NCC}(x, y) = \frac{R_{CC}(x, y)}{Z(x, y)} \quad (5)$$

#### 4.3.2 正規化相互相関による類似度調査

正規化相互相関を用いて楽曲の類似度を求める。正規化相互相関を行うためには音楽データを画像化しなければならない。よって、データセットの楽曲と提案モデルによる生成音楽をピアノロール画像に変換する。提案モデルによる生成音楽のうち4小節(ピアノロール画像)をテンプレートとして保存し、データセットの楽曲のピアノロール画像との正規化相互相関を行い、類似度を算出する。その後、生成されたフレーズから抜き出す4小節のピアノロール画像を1pxごとスライドさせ、生成されたフレーズ全体の類似度を求める。また、同様にデータセット同士の類似度も求める。

正規化相互相関を用いて求めたデータセット同士と、データセットと生成音楽の楽曲の類似度の平均と最大値、最小値を表1に示す。表1より、生成音楽の中にはデータセットの楽曲と類似した楽曲が生成されていることが確認された。しかし、全体としてはデータセットと同様な類似度であることが確認できる。よって、提案モデルによる生成音楽は単なるコピー&ペーストではないことが確認できる。

表1 正規化相互相関による楽曲の類似度

音楽データ	平均値	最大値	最小値
データセット	0.23	0.42	0.10
生成音楽	0.21	0.65	0.04

## 5. 結論

提案手法では、人と機械間でセッションを行うための機能として、機械が自動で音楽を生成する機能を提案した。提案モデルは改良した深層学習モデルにルールモデルを組み込むことで、小節数やコード進行の条件を組み込んだ音楽生成を可能なモデルを構築した。実際に提案モデルに Charlie Parker の楽曲を学習させ、提案モデルで指定した小節とコード進行のフレーズが生成できることを確認した。また、正規化相互相関による楽曲の類似度はデータセット同士の楽曲と同程度の類似度であり、オリジナル性のある音楽が生成可能であることが確認できた。しかし、実際の即興演奏では状況に応じて学習した楽曲以外の特徴を持つフレーズが求められるが、提案モデルでは状況に応じてフレーズの特徴を調整できない。よって、今後の展望として、ユーザーが選択したジャンルや曲調のフレーズを生成可能なモデルを提案することが挙げられる。

## 参考文献

- [1] 中島さち子: 音楽から聴こえる数学, 株式会社講談社 (2018).
- [2] 小寺未知留: <イリアック組曲>と『実験音楽』コンピュータ音楽の創作を対象とした研究の一事例として, 先端芸術音楽創作学会会報, Vol. 6, No. 4, pp. 5-11 (2014).
- [3] GoogleAI: Magenta, GoogleAI (online), available from (<https://magenta.tensorflow.org/>) (accessed 2020-01-14).
- [4] H.Goldsen, M.: *Charlie Parker Omnibook: For C Instruments. (Treble Clef Version)*, Criterion Music Corp (1982).