Vol.2020-MUS-128 No.6
2020/8/24

# 深層学習を用いたブルース進行のJAZZアドリブ自動生成の検討

小笠原 稜[1]　松川 晃大[1]　上瀧 剛[1]

**概要**：深層学習を用いて，ジャズ音楽の即興演奏を生成する機能を提案する．深層学習を用いた楽曲の自動生成はこれまでにも研究されているが，クラシック音楽やポピュラー音楽を対象としていた．今回ジャズ音楽，特にジャズブルースに着目した．提案手法では，Long short-term memory（LSTM）を用いてアルトサックス奏者である Charlie Parker の楽曲を学習させ，4 小節ごとに分けてフレーズ生成を行い，フレーズを繋ぎ合わせることで 12 小節のブルース進行のフレーズを生成する．その際フレーズの繋ぎ合わせの前後に関連性を持たせる必要があり，翻訳や対話生成のモデルである Sequence to Sequence を用いることで関連性を持たせた．実験の結果，生成されたフレーズはブルース進行を成しており，学習した楽曲を繋ぎ合わせたり，フレーズの一部を変更したりしたフレーズが生成された．このことから生成されたフレーズは単なるコピー&ペーストではなく，学習した楽曲の特徴を残したものであることが分かった．

## 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．この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．この二つの機構の内，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即興演奏や一定のコード進行を生成することが困難であり，セッションに対応した音楽の自動生成が少ない問題がある．提案手法ではジャズ音楽，特にジャズブルースに着目し，深層学習の LSTM を用いて，アルトサックス奏者である Charlie Parker の楽曲を学習させた．この学習によって機械に彼の楽曲の特徴を獲得させると同時に音楽理論の習得をさせる．また LSTM による学習と生成はブルース進行の 12 小節を分割し，4 小節ごとに行った．LSTM によって生成された 4 小節のフレーズを繋ぎ合わせることで，12 小節のブルース進行のフレーズを生成させ，即興演奏で見られる新しい演奏フレーズかつ音楽理論にあてはまる演奏フレーズの生成を再現する．この際のフレーズの繋ぎ合わせに Sequence to Sequence を用いることで関連性を持たせた．

[1] 熊本大学自然科学教育部情報電気工学専攻

## 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(1756-1791) 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，予め複数のフレーズが用意されており，それらは 1 小節毎に分割されている．新たにフレーズを生成する際はサイコロを振り，出た目に応じて小節を選択し，それらを組み合わせていく．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，適したジャンル，曲調の音楽生成に有効である一方で，「特定の作曲家，演奏家らしさ」を有する音楽生成は不向きであった．

### 2.2 Magenta

Google Brain チームは，2016 年から 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．これの利点は，人が音楽理論を 1 つずつニューラルネットワークに教える必要がないことである．ニューラルネットワークを適切に設計することで，楽曲データから音楽家や演奏上の特徴，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，ランダム性が低いことやブルース進行を中心とした特定のコード進行の音楽の生成をしているものが少ない問題が挙げられる．

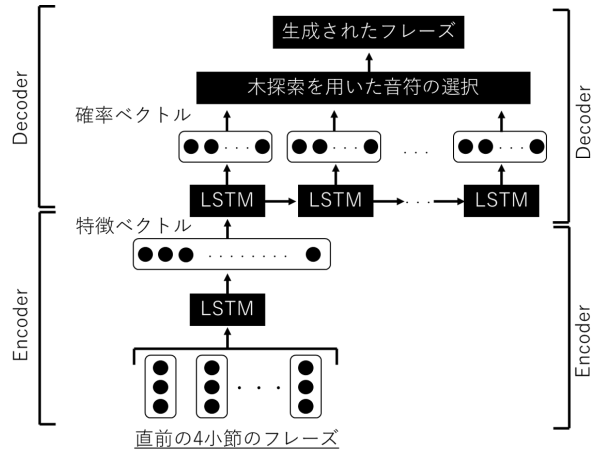



**図 1** 提案モデルの構造．提案モデルは直前の 4 小節のフレーズを特徴ベクトルに変換する Encoder と，音符を予測する Encoder 及び予測の確率ベクトルから探索し，4 小節のフレーズを生成する木探索から構成される．

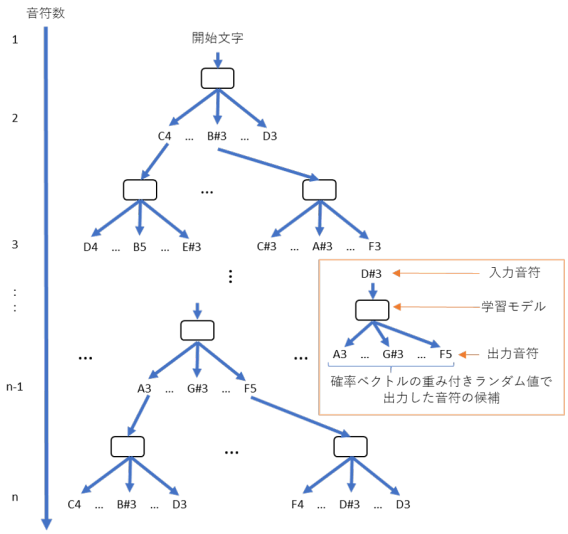



**図 2** フレーズ生成に用いる木探索．学習モデルに音符を入力し，条件に合う確率ベクトルの重み付きランダム値で複数の音符を選択する．その後次のモデルに選択した音符を入力し，選択した音符の組み合わせに対して探索する．

## 3. 提案手法

提案手法では直前の 4 小節の演奏フレーズを入力し，次の 4 小節の演奏フレーズを出力するモデル提案する．

提案モデルの構造を図 1 に示す．提案したモデルは Sequence to Sequence に木探索を組み合わせたモデルである．Encoder で演奏フレーズを特徴ベクトルに変換し，開始文字と合わせて Decoder 内の LSTM に入力することで確率ベクトルを出力する．確率ベクトルから次入力の音符を選択し，LSTM に入力する．これを繰り返すことで 4 小節の演奏フレーズを生成する．

入力する音符を選択する際はコード，拍数の条件に合う確率ベクトルを抜き出し，重み付きランダム値を用いて選択する．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．図 2 にフレーズ生成に用いる木探索を示す．始めに開始文字をモデルに入力し，条件に合う確率ベクトルの重み付きランダム値を用いて音符を複数選択し，次のモデルに入力する．このとき複数選択した音符の組み合わせに対して探索を行い，探索範囲が 5 以上で楽曲フレーズ全体のスコアが低い枝の経路を枝刈りによって探索を省くことで，探索回数を減らす．楽曲フレーズのスコアは音符の出現確率の総乗したものである．

## 4. 実験

実験ではデータセットの作成，提案モデルの学習と生成，生成されたフレーズの評価を行う．

### 4.1 データセット

使用する楽曲の MIDI データからテキストデータに変換し，データセットを作成する．今回提案したモデルには Charlie Parker の譜面 [4] の中で，ブルース形式の楽曲（表 1）を学習させる．

**表 1** データセットに使用した Charlie Parker の楽曲

| No. | 曲名 | （小節数，音数） |
|---|---|---|
| 1 | Another Hair Do | （60，372） |
| 2 | Au Private 1 | （60，343） |
| 3 | Back Home Blues | （60，398） |
| 4 | Barbadas | （48，313） |
| 5 | Billie's Bounce | （72，421） |
| 6 | Bloomdido | （72，473） |
| 7 | Blue Bird | （36，287） |
| 8 | Blues for Alice | （48，344） |
| 9 | Buzzy | （48，266） |
| 10 | Cheryl | （48，262） |
| 11 | Chi Chi | （96，591） |
| 12 | Cosmic Rays | （36，291） |
| 13 | KC Blues | （36，262） |
| 14 | Laird Baird | （48，354） |
| 15 | Mohawk 1 | （72，463） |
| 16 | Mohawk 2 | （48，464） |
| 17 | Now's The Time 2 | （48，305） |
| 18 | Perhaps | （48，293） |
| 19 | Si Si | （48，298） |
| 20 | Visa | （60，392） |

まず MIDI データから音階，音価，コードの情報を取得し，音階__音価__コードの形式で表現したテキストデータに変換する．テキストデータの音階には，音符の場合は MIDI データのノート番号である 0 から 127，休符の場合は-1 が代入される．音価には 4 分音符を 480Tick とした音の長さを表す単位であるティック表記を用いて代入し，コードにはブルース進行で使われる 3 種類のコード（I：1，VI：4，V：5）が代入される．

| 行 | | |
|---|---|---|
| 1 | 60_480_1 60_480_1 60_240_1 60_240_1 -1_480_1 -300_0_0 ¥t | 65_480_4 65_240_4 65_240_4 65_480_1 -1_480_1 -300_0_0 |
| 2 | 65_480_4 65_240_4 65_240_4 65_480_1 -1_480_1 -300_0_0 ¥t | 67_480_5 67_240_4 67_240_4 67_480_1 -1_480_1 -300_0_0 |
| 3 | 67_480_5 67_240_4 67_240_4 67_480_1 -1_480_1 -300_0_0 ¥t | 60_480_1 60_480_1 60_240_1 60_240_1 -1_480_1 -300_0_0 |
| 4 | 60_480_1 60_480_1 60_240_1 60_240_1 -1_480_1 -300_0_0 ¥t | 65_480_4 65_240_4 65_240_4 65_480_1 -1_480_1 -300_0_0 |
| 5 | 65_480_4 65_240_4 65_240_4 65_480_1 -1_480_1 -300_0_0 ¥t | 67_480_5 67_240_4 67_240_4 67_480_1 -1_480_1 -300_0_0 |

**図 3** データセットの構造．音符の音階，音価，コードの情報を音階__音価__コードの形式で表現し，入力データと正解データをタブ文字で対応させている．

次に Sequence to Sequence のモデルでは入力のデータに対応する正解データが必要である．よって楽譜上で入力した 4 小節のフレーズの次に来る 4 小節のフレーズを正解データとし，図 3 のように入力のテキストデータと正解のテキストデータをタブ文字で対応させたデータセットを作成する．

### 4.2 演奏フレーズ評価

生成されたフレーズの評価として主観評価を用いると評価者ごとに評価が変わってしまい，生成されたフレーズを定量的に評価ができない．以上のことから生成されたフレーズを定量的に評価するために正規化相互相関を用いる．

#### 4.2.1 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 3）．

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），$x' = 0 \ldots w-1$，$y' = 0 \ldots h-1$ としたとき，座標（x，y）での相関係数 $R_{ccoeff}$（x，y）は式（1）で表される．

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

ここで，式（1）の T' と I' は以下で定義される．

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

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

式（1）を正規化係数（式（2））を用いて正規化する．これにより，入力画像とテンプレート画像の間にある照明の差の影響を減らすことができる．最終的に類似度計算に用いる正規化相互相関を式（3）に示す．

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

$$R_{ccoeff\_normed}(x,y) = \frac{R_{ccoeff}(x,y)}{Z(x,y)} \quad (3)$$

#### 4.2.2 正規化相互相関による類似度調査

生成されたフレーズ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，類似度を求める．正規化相互相関を行うためには MIDI 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生成されたフレーズの MIDI データをピアノロール画像に変換する．生成されたフレーズのうち 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の抜き出す 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．

### 4.3 実験手順

Sequence to Sequence のモデルに学習させるときの各パラメータを表 2 に示す．実験手順は Sequence to Sequence のモデルに作成したデータセットを学習させる．次に提案モデルに対して楽曲「Another Hair Do」の冒頭 4 小節を入力し，その後のフレーズが生成されるかを確認する．また生成されたフレーズの主観評価に加え，生成されたフレーズがデータセットの楽曲のコピー＆ペーストになっていないか確認するためにデータセットの楽曲と生成されたフレーズの類似度を正規化相互相関を用いて求め，生成されたフレーズの定量的な評価を行う．

**表 2** 学習モデルのパラメータ

| | |
|---|---|
| 入出力層ノード数 (enc, dec) | (590, 603) |
| 中間層 | LSTM2 層 |
| 中間層ユニット数 | 100 |
| 損失関数 | 交差エントロピー誤差 |
| Optimizer | Adam |
| Learning rate | 0.001 |
| バッチサイズ | 10 |
| epoch 数 | 300 |

## 5. 考察

提案モデルによって同条件で 3 つのフレーズを生成した．生成されたフレーズのピアノロール画像を図 4 に示す．図 4 は 4 小節目までは入力したフレーズであり，残りの 8 小節が生成されたフレーズである．また生成されたフレーズ

[a] 生成されたフレーズ 1：16 分音符だけでなく，3 連符や 8 分音符など様々な音符で曲が構成されている．

[b] 生成されたフレーズ 2：全体的に 8 分音符で構成されており，ゆっくりとしたフレーズである．

[c] 生成されたフレーズ 3：全体的に 16 分音符で曲が構成されており，速いフレーズである．

**図 4** 生成されたフレーズのピアノロール画像．4 小節目までは入力したフレーズであり，5 小節目から 12 小節目が生成されたフレーズである．

の MIDI データをオンラインストレージ上にアップロードする．[*1]

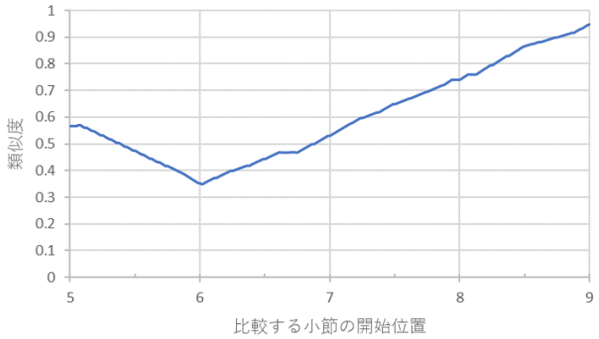
実際に生成されたフレーズは 12 小節のブルース進行のフレーズであることが確認でき，原曲と比べると複数の原曲フレーズを組み合わせるようなフレーズが生成されていることが分かった．また Charlie Parker の特徴を持ったフレーズであり，音楽理論に外れた音や音の並びがないことから提案モデルでは Charlie Parker の特徴だけでなく，音楽理論も機械が学習できていると言える．小節ごとに比較していくと，生成されたフレーズの 2 小節までは Encoder から受け取った特徴ベクトルの影響から入力した曲の音符の確率ベクトルの値が大きく，6 小節まではいずれの生成されたフレーズも同じ音の並びとなった．しかし後半部分は特徴ベクトルの影響度が小さくなり，それぞれ違うフレーズが生成された．

次に 12 小節のフレーズの内，生成された 5 小節目から 12 小節目のフレーズを正規化相互相関を用いて原曲との類似度を比較した結果を図 5 に示す．いずれも最初と最後の類似率が高く，比較する小節の開始位置が 6 小節のところで最小値となる結果が得られた．前半部は先ほど述べたように Encoder から受け取った特徴ベクトルの影響から類似度が高い．また後半部では拍数の条件が厳しくなり，条件に当てはまる音符が少なくなることから類似率が上がっていると考えられる．しかし，平均で類似率は生成されたフレーズ 1 から順に 0.60，0.61，0.57 であり，生成されたフレーズは単なるコピー＆ペーストではないことが確認できる．

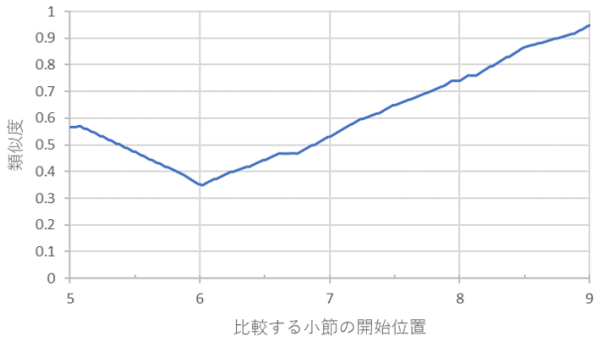
## 6. むすび

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．Charlie Parker の楽曲を深層学習である Sequence to Sequence を用いて学習させ，提案モデルで 12 小節のブルース進行のフレーズが生成できることを確認した．また，今回のモデルでは学習した楽曲に対して，忠実性の高いフレーズが生成された．しかし，実際の即興演奏では，状況によって忠実性の高いフレーズだけでなく，創造性，新規性の高いフレーズも要求される．そのため状況に応じてフレーズの特性を調整できるモデルが必要となる．よって今後の展望として，忠実性と創造性のバランスをとることができるフレーズ生成のモデルを提案することが挙げられる．

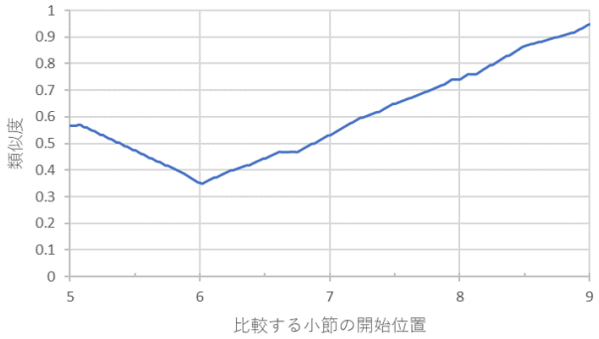
[*1] https://drive.google.com/drive/folders/1jEFA0IZoJKnWZuXbt0c86hLrgomkMO-t




[a] 生成されたフレーズ 1 の類似度




[b] 生成されたフレーズ 2 の類似度




[c] 生成されたフレーズ 3 の類似度

**図 5** 生成されたフレーズの類似度．12 小節のフレーズの内，生成された 5 小節目から 12 小節目までのフレーズを正規化相互相関を用いて原曲との類似度を求める．

## 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，2014，先端芸術音楽創作学会会報 Vol.6 No.4 pp.5-11.
[3] Magenta，https://magenta.tensorflow.org/，（参照日：2020/01/14).
[4] Charlie Parker Omnibook: For C Instruments.(Treble Clef Version)，Criterion Music Corp，1982.