

フラクタル×ジャズ

Fractal × jazz

1015238 吉田周平 Shuhei Yoshida

1 背景

音楽の自動生成の歴史は、古いものではジャスカンデブレ (1450/55 - 1521) の「サイコロのミサ」と呼ばれる「ミサ・ディ・ダディ」がある。また、モーツァルト (1756 - 1791) の「音楽のサイコロ遊び」などある。しかしながら、これらの音楽は、作曲者の知識や技術による編集が行われおり、自動生成とは言い難い。コンピューターの発達した現在では、人間が行っていた調整をも同時に行うことが可能になり、様々な音楽自動生成のシステムが開発されている。

音楽の研究をする際には「1/f ゆらぎ」が言及されてきた。「1/f ゆらぎ」の特性を持つ音楽の多くは心地が良い音楽であるとされている。しかし、その特性を持っていても心地よい音楽だとは限らないという事も事実である。本プロジェクトにある「フラクタル」は、「1/f ゆらぎ」のもつ自己相関性に基づいている。

ここ数年で「人工知能」や「ディープラーニング」などの言葉を、様々なメディアで見かけるようになった。「ディープラーニング」の基礎となる「ニューラルネットワーク」は、各方面で研究・活用されてきている。人の脳をモデルにした「ニューラルネットワーク」は、複雑にしていくことで人の脳に近付くと考えられている。しかしながら、現在の技術では人の脳と同等の人工知能は開発されていない。

2 課題の設定と到達目標

本プロジェクトの目標は「プロのジャズミュージシャンの演奏を模倣する自動生成システムの作成」である。ジャズのアドリブ演奏において、演奏者の癖というものが強く反映される。そこで、その癖を模倣してアドリブ演奏の自動生成を行うシステムを作成することにした。そこで、本プロジェクトでは、プロのジャズミュージ

シャンとして「マイルス・デイビス」と「チャーリー・パーカー」の二人の演奏を模倣させることにした。

そして、目標到達の基準として以下の条件を満たすこととする。

- 「マイルス・デイビス」の模倣をさせた場合、その曲と「チャーリー・パーカー」の曲に差異がある。
- 模倣して作成した曲と元のミュージシャンの曲に差異がない。

以上の条件をみたすことで、ある特定のジャズミュージシャンの演奏を模倣したものとする。また、差異があるか否かはコルモゴロフ・スミルノフ検定によって検定することとする。

3 課題解決のプロセス

本プロジェクトでは、前述した通り「乱数生成」と「ニューラルネットワーク」、「リカレントニューラルネットワーク」のそれぞれを研究することによって問題解決を試みた。

まずは、乱数による自動生成について述べる。「1/f ゆらぎ」の性質を持つ音楽の多くは心地よいと言われている。

そこで、本プロジェクトでは、「白色雑音」と「ブラウン雑音」そして「1/f ゆらぎ」の性質を持った音楽を乱数で自動生成した上で、それらを比較する実験を行った。また、乱数の自動生成のシステムは、カードナー [1] を参考し作成した。

カードナー [1] によると、「白色雑音」は音と音のつながりが無く、自己相関係数が0に近い。また、「ブラウン雑音」は音と音のつながりが強く、自己相関係数は1に近い。「1/f ゆらぎ」は双方の性質を折衷したものである。そして、自己相関が強すぎず弱すぎずであることが「1/f ゆらぎは心地が良い」とされる理由の一つである。

以上の三つの性質に沿った乱数生成を行い、その乱数に対応させて作曲し、その比較実験を行った。これによって、「1/f ゆらぎの性質を持つ音楽は心地よいのか」もしくは「1/f ゆらぎの性質を持つ音楽の心地良さは、他の音楽に対して優位性が見られるか」を確かめることが出来ると考えた。

実験では、スケールは「F メジャースケール」と「F ブルーノートスケール」の2種類、乱数生成方法は「白色」と「褐色 (反射壁)」、「褐色 (吸収壁)」、「1/f ゆらぎ」の4種類用意した。実験の方法については、4章で記述する。

次に、ニューラルネットワークによる自動生成だが、まずはニューラルネットワークの概要について述べる。ニューラルネットワークは人の脳をモデルにした学習モデルである。ニューラルネットワークでは順番に、入力層、中間層、出力層の三種類の層で構成されている。また、それぞれの層は幾つかのニューロンで構成されている。

ある層から次の層へは、重み付きの入力信号の総和を活性化関数によって変換させる。例えば、中間層が一層のニューラルネットワークについて考えるとする。このとき、入力を x として重み行列を W, V 、バイアスペクトルを b, c 、シグモイド関数を $\sigma()$ とする。その場合、入力を式に表すと以下のようなになる。

$$h = \sigma(Wx + b) \quad (1)$$

$$y = \text{softmax}(Vh + c) \quad (2)$$

ニューラルネットワークの学習では、最適なパラメータを得る必要がある。そのために、損失関数を使用する。損失関数はニューラルネットワークの性能の悪さの指標である。本プロジェクトでは、主な損失関数として2乗和誤差や交差エントロピー誤差について調査したが、システムには交差エントロピー誤差を使用した。交差エントロピー誤差の式は、

$$E = - \sum_{k=0}^k t_k \log y_k$$

で表される。ここにおける t_k は訓練データ、 y_k はニューラルネットワークの出力である。上記の式を用いてパラメータを更新した。この手法を勾配法という。

学習の際は、元の楽譜を編曲したデータを使用した。編

曲は複数の種類がある音符を四分音符と四部休符にするもので、複雑なデータを簡易化するために行った。

最後に、リカレントニューラルネットワークによる課題解決を試みた。

リカレントニューラルネットワークは、時系列データを予測することに適したモデルであるとされている。そのため、音楽という時系列データを取り扱う本プロジェクトでは、このモデルを使用することが適当であると考えられる。

リカレントニューラルネットワークは現在の隠れ層に過去の隠れ層の情報を与えることによって、時系列データの予測に適応させている。 $h()$ を中間層、 $f()$ を活性化関数、 U, W を重みベクトル、 b をバイアスペクトルとして式に表すと以下のようなになる。

$$h(t) = f(Ux(t) + Wh(t-1) + b)$$

となる。これは、時刻 t の中間層の値に時刻 $(t-1)$ の値による影響を与えていることを示している。

音楽生成システムを作成する前段階として、sin 波を予測するシステムを作成し、そのシステムが適当であるかを検証した。検証の結果によると、学習は成功したと考えられる。

しかしながら、そのプログラムを音楽データように改変して自動生成させたところ、勾配が消失してしまい、学習に失敗してしまった。

原因として考えられるのは、勾配の計算の際に行う活性化関数の微分による勾配消失であると考えられる。そこで、LSTM(Long short-term memory) を使用することで解決可能であると予想したが、作成に時間がかかりすぎたため断念した。

そこで、データを音程だけのものにして、現在のシステムを使用した。コルモゴロフ・スミルノフ検定で目標到達の基準を満たすかを検証したが、良い結果を得る事は出来なかった。

4 実験

雑音の実験では、3章で示したように2種類のスケールと4種類の乱数生成方法の組み合わせで、全8種類の生成を行った。さらに、データに偶然を与えないように、それぞれの生成方法につき5曲生成させた。

さらに、以下の条件で行った。

- 評価はプロジェクトメンバー 8 名で行った。
- 作曲はアドリブ部分の 12 小節 48 音 (全て四分音符 または四部休符) とした。
- 作曲にはピアノの音を用いた。

以上の条件を満たした上で、次の手順で実験を行う。

1. スピーカーの音量を調整して、お手本のアドリブを流す。
2. 生成した曲の順番を入れ替えて流す。
3. 評価者は評価基準に基づいて評価する。

ここにおける評価は A,B,C として、「A：音楽になっている」「B：音楽として聴くことができる」「C：音楽になっていない」という評価基準を設定した。

実験による評価結果を A を 3 点、B を 2 点、C を 1 点として集計する。その結果を表 1、2 に示す。

表 1 評価点

	褐色 (反射)	褐色 (吸収)
F メジャー	61	67
F ブルーノート	71	66
合計	132	133

表 2 評価点

	白色	1/f ゆらぎ	合計
F メジャー	56	67	251
F ブルーノート	53	63	253
合計	109	130	504

この結果を「二元配置分散分析」を用いて分析した。二元配置分散分析は、1/f のブルーノートスケールの曲が優位であるかを統計的に検討するために使用した。表 3 は、分析したときの p-値と F-境界値の値である。結

表 3 二元配置分散分析

変動要因	p-値	F-境界値
スケール	0.88842365	4.149097744
乱数	0.6921029	2.90111958
交互作用	0.48548861	2.90111958

果として、p-値が F-境界値の範囲内に収まったために帰無仮説は棄却されず、ブルーノートスケールの 1/f が優位性を示すことは出来なかった。

表 4 個人報告書の文字サイズ・文字数・行数

文字サイズ	1 行あたりの文字数
9 ポイント	24-28 文字

5 結果と考察

結果として、乱数によって生成された音楽はどれも良いものではなかった。1/f ゆらぎの特徴を有した乱数によって生成した音楽も、他の雑音との有意差は見られず、1/f ゆらぎの特徴を有していれば必ず心地よい音楽になるとは限らないことがわかった。

本プロジェクトでは、ニューラルネットワークによる学習に目を付けた。ニューラルネットワークは人の脳をモデルとした学習方法であり、今回の「ジャズミュージシャンの模倣」という目的には最適であると推測されていた。しかしながら、簡易的な構造のニューラルネットワークでは、上手く学習することが出来なかった。

そのため、時系列データを予測することに適したモデルであるリカレントニューラルネットワークを使用することにした。しかしながら、リカレントニューラルネットワークにおける学習の際に勾配消失が起こってしまった。

勾配消失の原因として考えられたのは、再帰的に学習することで勾配の計算の際に行う活性化関数の微分の値がより多くのかけてられてしまったことにあると考えられる。そこで、LSTM(Long short-term memory) を使用することで解決可能であると予想したが、作成に時間がかかりすぎたため断念した。また、活性化関数の変更によって回避することも解決策の一つとして考えられる。

6 今後の課題

今後の課題の一つとして勾配消失させないことがあげられる。そのために、LSTM のモデルを使用したり、活性化関数を変更することによって改善出来るかを検証することが必要であり、そのプログラムを作成することが課題となる。

もし、本プロジェクトを継続するのであれば、作成したプログラムによって生成された曲の分析の際に、1/f ゆらぎの特徴を持つかを検定して「1/f ゆらぎの特徴を持つ音楽は心地よい」という説を再検討するのも良いだろ

う。

本プロジェクトでは、模倣出来たかの判断をコルモゴロフ・スミルノフ検定のほかにアルペジオを使用した回数など、音楽理論的に分析する予定であったが、この方法についても正当性を確認する必要があった可能性がある。また、他にも適当な分析方法があった可能性もあり、その調査など課題は山積みである。

参考文献

- [1] ガードナー マーチン GardnerMartin. フラクタル音楽 ガードナー数学マジック (一松信 訳). W.h.freeman & Co Ltd, 1996.
- [2] かどかつひろ 加度克紘. ドレミで覚えるジャズ。アドリブの法則. リットーミュージック, 2015.
- [3] さいとうゆうき 斎藤康毅. ゼロから作る DeepLearning - Python で学ぶディープラーニングの理論と実装. オライリー・ジャパン, 2016.
- [4] すごもりゆうすけ 巢籠悠輔, 詳解ディープラーニング TensorFlow・Keras による時系列データ処理. マイナビ出版, 2017.
- [5] とうじょうさとし ひらたけいじ 東条 敏, 平田圭二. 音楽・数学・言語 情報科学が拓く音楽の地平. 近代科学社, 2017.
- [6] よこおか ゆかり 横岡ゆかり. 大好きな曲で楽しくレッスンジャズピアノはじめよう!. 日刊工業新聞社, 1989.