機械学習による和太鼓音楽の生成

Generating Japanese Drum Music by Machine Learning

中央大学大学院　理工学研究科　電気電子情報通信工学専攻
Electrical Electronic and Communication Engineering Faculty of Science and Engineering, Chuo University
岡本　和巳
Kazumi OKAMOTO

1. はじめに

古来より日本では各地に民謡やお囃子など，固有の音楽が生まれてきた．しかし近年，地方都市などでは過疎化の流れが大きく，それに伴う少子高齢化や人口減少といった問題が発生している．この人口の減少に伴い，地方に存在する固有の音楽も規模が縮小している傾向がある．このまま規模の縮小が続けば音楽の担い手もいなくなり，各土地の音楽そのものが失われてしまうと考えられる．埼玉県の西部に位置する秩父市では、日本三大曳山祭の一つである秩父大祭が存在している.同祭で演奏されている囃子は秩父屋台囃子と呼ばれる.秩父屋台囃子は、一つの大太鼓と複数の小太鼓、鉦と笛で構成される[1].この内、大太鼓が全体の演奏をリードする役割を担い、小太鼓はリズムを刻む役割を担う.秩父屋台囃子では大太鼓と小太鼓は一体となって演奏されるため、通常の練習は一緒に行う必要がある.このため、個人で練習する機会は少ない.また、上述した過疎化などにより秩父屋台囃子の演奏者が減少すれば、練習機会を確保することが困難になる.そこで、演奏者が叩いた太鼓に呼応して他方の太鼓の音を自動的に生成するシステムがあれば、一人での練習が可能になり、練習機会の確保につながる.また、それにより秩父屋台囃子という文化の保存につながると考えられる.このようなシステムの実現には和太鼓の音楽を生成する必要がある.そこで本研究では、機械学習を利用して和太鼓音楽の生成を試みた.

当研究室では，これまで秩父屋台囃子を題材として，和楽器における機械学習を用いた演奏支援システムについて[2], 和太鼓のリズムの特徴量と機械学習を用いた評価[3], 機械学習による和太鼓の自動演奏システムについての研究[4]がされてきている.しかし，いずれも和太鼓音楽の生成について触れているものはない.また，機械学習を利用して音楽を生成する研究においては、コンピュータで扱いやすいメロディーやハーモニーを生成対象にしているものが多い.しかし、メロディーやハーモニーは直接音として生成することができない.本研究は，演奏支援に利用するための音楽生成であるので，直接音として生成できるオーディオ信号を生成する必要がある.オーディオ信号はCDであれば、1秒間に44100回の音圧を16bit(32768段階)でサンプリングすることで、音を表現している.忠実に音楽を表現できる一方で、計算コストが高すぎるため現状ではあまり広く使われていない[5].オーディオ信号を直接扱った数少ない例として、WaveNet[6]というモデルがある.このモデルでもオーディオ信号の生成には非常に時間がかかる.演奏支援を行うことを考えると、オーディオ信号の生成はできるだけ速いほうが良い.

そこで、本研究では秩父屋台囃子を題材として、オーディオ信号を間接的に扱う手法を提案している.提案手法により、計算コストを減らしても、元の曲と類似している曲を生成できるのかを確認する.この研究は演奏支援を行う際に役立つと考えられる.

2. 提案手法

本研究では、秩父屋台囃子の大太鼓のオーディオ信号に短時間フーリエ変換を適用し、スペクトログラムに変換する.その後、スペクトログラムのフレームを 3 つの手法で次元削減する.一つ目はスペクトログラムのフレームを主成分分析により次元削減する手法.二つ目はスペクトログラムにメルフィルタバンクをかけ、次元削減したメル周波数スペクトログラムを用いる手法.三つ目はメル周波数スペクトログラムに離散コサイン変換を適用し、さらに次元削減をした、メル周波数ケプストラム係数を用いる手法.これら三つの手法を行った後にそれぞれのフレームをベクトル量子化によりラベルデータ化し、同時にコードブックを得る.ラベルデータで機械学習による学習を行い、生成モデルを作成する．生成モデルに初期入力を入れることによってラベルデータを生成する.ラベルデータをコードブックと

逆変換でオーディオ信号に復元する.各節で詳細を述べる.

2.1. 短時間フーリエ変換

一定の大きさの窓関数を用いて信号を切り出し、その結果をフーリエ変換してスペクトルを計算する処理を短時間フーリエ変換[7]と呼ぶ.一つの窓に対して一組のスペクトルが求められるので、スペクトルの時間的変化を求めたことになる.大太鼓のオーディオ信号に短時間フーリエ変換を適用することで、スペクトログラムに変換する.大太鼓の 5 秒分のオーディオ信号をスペクトログラムに変換した例を図 2.1 に示す.

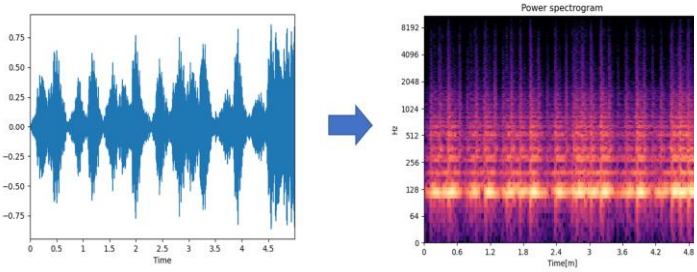

図 2.1 オーディオ信号からスペクトログラムへの変換

2.2. 次元削減

スペクトログラムは行列とみなすことができる. 行数は周波数、列数が時間、成分が音の大きさを表す.この行列の列ベクトルがフレームとなる. スペクトログラムは一つのフレームにつき高次元のデータを持つ. データの次元が多いと、高精度のモデルを作るために必要なデータが指数関数的に増えるため、後に述べるベクトル量子化をうまく行うことができない.そこで次元削減により、なるべく情報量を減らすことなく次元を削減する.次元削減には主成分分析、メル周波数スペクトログラム、メル周波数ケプストラム係数の三つの手法を用いた.それぞれ 2.2.1、2.2.2、2.2.3 で詳細について述べる.次元削減のイメージを図 2.2 に示す.

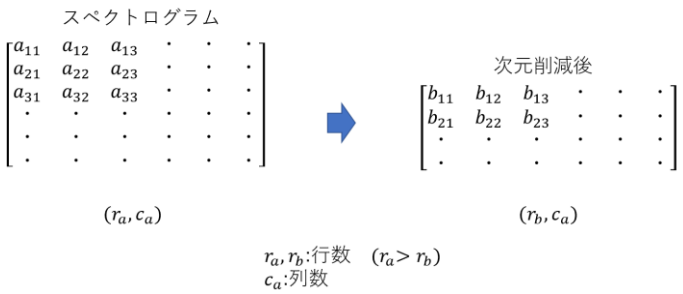

図 2.2 次元削減のイメージ

2.2.1. 主成分分析を用いた次元削減

主成分分析[8]により各フレームの次元を落とした.元のデータをどの程度表せているか示す累積寄与率は 86%であった.

2.2.2. メル周波数スペクトログラムを用いた次元削減

音の周波数は音の高さに対応する物理量であるが、人間の聴覚において音の高さを知覚する際の量としてメル尺度がある.人間の聴覚には周波数の低い音に対して敏感で、周波数の高い音に対しては鈍感であるという性質がある[9] 音の周波数をメル尺度に変換する際に、特徴量の次元を落とし、低周波成分ほど分解能を高く、高周波成分になるほど低くするためのフィルタバンクをメルフィルタバンクという. スペクトログラムにメルフィルタバンクをかけて、メル周波数スペクトログラムを求める.これにより、人間の聴覚に必要な情報を残しながら次元を落とすことができる.

2.2.3. メル周波数ケプストラム係数を用いた次元削減

メル周波数スペクトログラムに離散コサイン変換を適用して、メル周波数ケプストラム係数[10]を求める.離散コサイン変換を適用して、低次元成分のみを取り出すことでメル周波数スペクトログラムの余分な細かい成分を無視することができる.これにより、メル周波数スペクトログラムからさらに次元削減する.

2.3. ベクトル量子化

2.2 で求めた次元削減後の各フレームを k-means++法によりクラスタリングし、それぞれにラベルを割り当てる.クラスタ数は 128 としたので、各フレームは 0~127 のラベルのどれかを割り当てられる.同時に各クラスタのクラスタセントロイドも求めることができ、これをコードブックと呼ぶ.このような処理をベクトル量子化[11]と呼ぶ. 図 2.3 にベクトル量子化のイメージを示す.

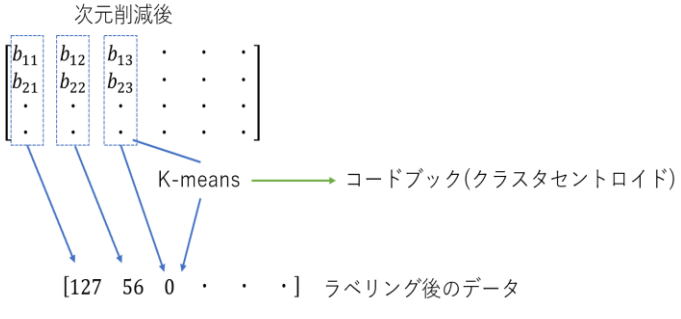

図 2.3 ベクトル量子化のイメージ

2.4. 機械学習モデル作成

機械学習であるディープニューラルネットワーク[12]を利用する.2.3 で作成したラベルデータを過去 32 個のデータから次のデータを予測するようなデータセットにする.このデータセットで学習させ, 作成したモデルは初期入力として 32 個のデータを与えると, 出力された次のデータを再び入力として利用することで, その後の時系列を生成する自己回帰型生成モデルである.

2.5. オーディオ信号合成

2.4 で作成した生成モデルから生成されたラベルデータを 2.3 で得たコードブックを用いて、ラベルに対応するフレームに置き換える.その後, 次元削減の逆変換を行うことでスペクトログラムに戻す.スペクトログラムからオーディオ信号に戻すために, Griffin-Lim 法[13]という位相復元法を利用してオーディオ信号に変換する.

3. 評価

提案手法で生成した曲が元の曲とどの程度似ているかを調べるために評価実験を行う. 評価指標は生成した曲と元の曲でそれぞれフレームごとに平均をとったメル周波数ケプストラム係数(MFCC)の相関係数とする. MFCC は楽器音に対しては音色に対応している.すなわち今回求める相関係数の値が大きければ、フレームごとの音色が似ているということがわかる.この評価指標を用いて提案手法でどの程度元の曲と似ているかを明らかにする. 作成したモデルは初期入力として、元データの曲を 0.74 秒(32 個のフレーム)入力するとその後の曲が生成される生成モデルである. ランダムに選んだ 0.74 秒分の曲を初期入力として与え、3 章の三つの手法でそれぞれ 10 秒の曲を 20 曲生成した.生成した曲と元の曲で評価指標を求める.結果を図 3.1 に示す.

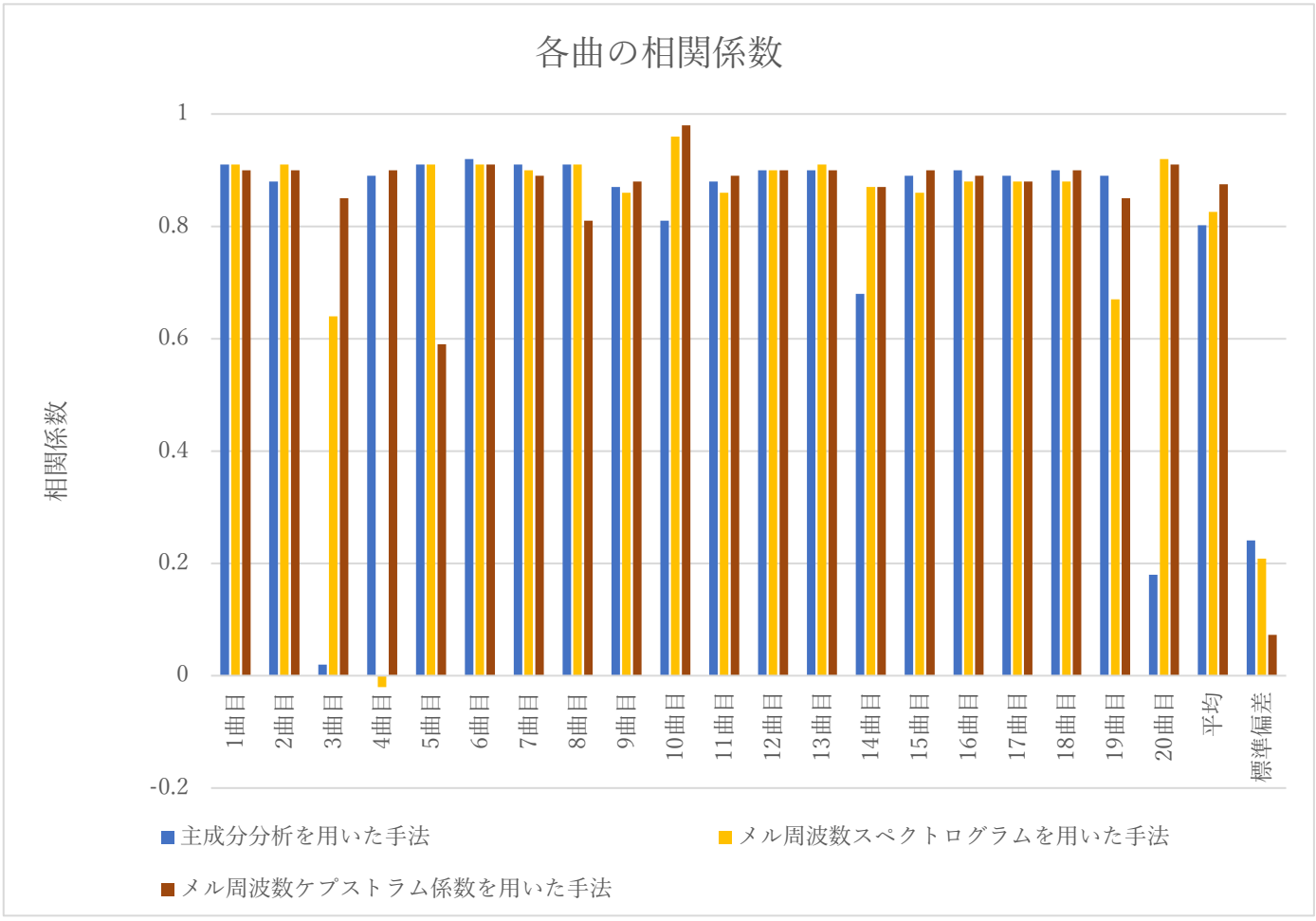

図 3.1 各曲の相関係数

4. 結論

本研究では、秩父屋台囃子の大太鼓を題材として提案手法でその曲を生成することを試みた.提案手法では三つの手法を提案し、それぞれの手法で 20 曲作成し、MFCC の相関係数を用いて評価を行った.その結果､すべての手法で 20 曲分の相関係数の平均は 0.8 を超える値となった.よって､この評価指標では元の曲と類似している曲を生成できていると結論付けることができる.したがって､提案手法により打楽器の音楽は生成できることが分かった.

また、主成分分析を用いた手法、メル周波数スペクトログラムを用いた手法､MFCC を用いた手法はそれぞれ 20 曲分の相関係数の平均が 0.80、0.83、0.88 となっている.この結果より、MFCC を用いた手法が最も評価指標が高くなることが分かった.

さらに、モデルの安定性を示す標準偏差の値も MFCC を用いた手法が最も低い値となった.したがって､MFCC を用いた手法が最も安定感のあるモデルを作成できることが分かる.

参考文献

[1] 浅賀 ひろみ, “秩父の祭りと秩父屋台囃子の歴史に関する研究, ” 白鴎大学論集 vol.23, no.2, pp.399-421, 2009

[2] 佐藤 信太郎, “和太鼓における機械学習を用いた演奏支援システムについて, ” 2018 年度 中央大学修士論文.

[3] 海老 原俊輔, “和太鼓のリズムの特徴量と機械学習を用いた評価, ” 2019 年度中 央大学修士論文.

[4] 趙 涌, “機械学習による和太鼓の自動演奏システムについての研究, ” 2020 年度 中央大学修士論文.

[5] Jean-Pierre Briot, Gaëtan Hadjeres and François-David Pachet, “Deep Learning Techniques for Music Generation – A Survey, ” arXiv:1709.01620, pp.17-19, Sept.2017.

[6] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu, “WAVENET: A GENERATIVE MODEL FOR RAW AUDIO, ” arXiv:1609.03499v2, Sept.2016

[7] 小野 順貴, “短時間フーリエ変換の基礎と応用, ” 日本音響学会誌, 72 巻, 12 号, pp.764-769, 2016.

[8] Sebastian Raschka, Vahid Mirjalili, “達人データサイエンティストによる理論と 実装 Python 機械学習プログラミング, ” 高橋 隆志 (編), pp.129-142, (社)株式 会社インプレス, 東京, 2020.

[9] S.S.Stevens, J.Volkmann and E.B.Newman, “A Scale for the Measurement of the Psychological Magnitude Pitch, ” J.A.S.A. vol.8, Jan.1937

[10] Beth Logan, “Mel Frequency Cepstral Coefficients for Music Modeling, ” In International Symposium on Music Information Retrieval, 2000.

[11] Dr. H.B. Kekre, Ms. Tanuja K. Sarode, “Vector Quantized Codebook Optimization using K-Means, ” International Journal on Computer Science and Engineering Vol.1(3), pp.283-290, 2009.

[12] Ian Goodfellow, Yoshua Bengio and Aaron Courville, “Deep Learning, ” MIT Press, pp.166-227, 2016.

[13] ANIEL W. GRIFFIN AND JAE S. LIM, “Signal Estimation from Modified Short-Time Fourier Transform, ” IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL.ASSP-32, NO.2, Apr.1984.