

音楽の三要素表現からの生成モデルアプローチによる 音楽生成手法の提案

川村誠護[†] 寺田英雄[‡] 庄野逸[†]

電気通信大学 大学院情報理工学研究科[†] 株式会社オープンストリーム[‡]

1. はじめに

音楽情報処理の自動作曲の分野では、近年、Google の Magenta を初めとする自動作曲のためのフレームワークが多く公開されている。これらは、主に楽音を時系列データとして扱うモデルによって機械学習で新しい楽曲を生成しようとする試みである。

音楽として鑑賞に耐える楽曲を生成するには単なる時系列信号生成では足りず、様々な審美的な制約が必要となる。そうした制約は、古くから音楽理論という形で体系化されている。代表的な音楽形式である西洋音楽の理論では、楽曲表現の主要な要素はリズム、メロディ、ハーモニーという三つであると考えられている。これらの三要素は作曲者の個性によって表現に豊富な多様性があり、現代に至るまで多種多様な表現パターンを持つ楽曲が作られてきた。

現在公開されている自動作曲のフレームワークでも、このような音楽理論を制約として、表現パターンを学習によって獲得することを目指している。音楽の三要素表現を用いた自動作曲については既に研究が進められており、一定の成果を上げている[1]。一方、従来の研究では個々の作曲家の個性の抽出や比較、また、その個性を反映・強調した作曲や鑑賞者の好みも加味した作曲についての検討は不十分である。

本研究では、音楽の三要素から得られる表現パターンのばらつきを統計的に取り扱い、学習することが出来る生成系フレームを構築することで、楽曲の自動生成を試みる。また、このような生成系フレームを構築することによって、鑑賞者の好みによる自動作曲や伴奏などの生成や作曲者毎の個性を表現した楽曲の生成、音楽推奨システムなどに応用が可能か検討していく。

2. 手法

本研究では、図1のような枠組みで自動作曲

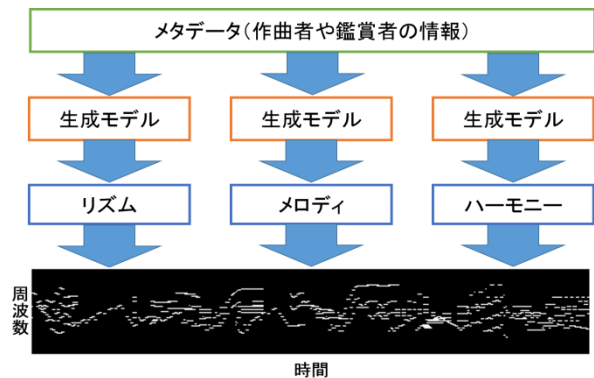


図1：フレームワーク

を試みる。音楽データをリズム、メロディ、ハーモニーの三要素に分解し、それぞれに対して分析および生成処理を行っていくことを考える。

リズムとは拍によって区別される音の長さの並びを示すもので、音符の種類によって表現することが出来る。本実験では、連続する二つのフレーム間に和音の変化があれば、それを拍の区切りとした。

メロディとは時系列に構成されている和音の中で最も高い音の遷移のことをいう。人間は連続した音の変化を周波数の比で認識するため、音の高さが異なっても周波数の比率が同じであれば同じメロディであると認識することが知られている。

ハーモニーとは二つ以上の音で構成された調和のとれた和音のことである。前述の通り、人間は周波数の比によって音の変化を認識するため、連続する和音同士の周波数の比が重要な枠組みを持つ。この和音同士の周波数の比を表現するには次元が大きいため、そのまま扱うことは困難である。そこで楽曲を、メロディを含む和音が連続したものであるとみなした場合、ハーモニーをメロディと独立してパターン化させることによって、より少ない次元でハーモニーを生成出来る。言い換えると、楽器を弾く際に、音の高さを決める腕の動きと、和音を決める指の動きを別々に考えるということである。

以上の音楽三要素を時系列に表現するため、それぞれのパターンを一つの事象であるとして、単純マルコフ連鎖モデルを用いて遷移を行うこ

“An application of music generation using triplet expression with generative model”

[†]Seigo Kawamura [‡]Hideo Terada [†]Hayaru Shouno

[†]The University of Electro-Communications

[‡]Open Stream, Inc

とで楽曲を生成していく。単純マルコフ連鎖モデルとは確率過程の一種であり、現在の状態から次のフレームの状態を確率的に求めることが出来る。

3. 実験

本研究で使用する単純マルコフ連鎖モデルのパラメータを決めるための学習データとして、John らの MusicNet を使用した[2]。MusicNet は 10 人の作曲者によって作曲されたクラシック音楽 330 曲分の階層化されたデータである。このデータから音楽の三要素を抽出し、作曲者の個性を表現するような確率モデルによる楽曲生成を試みた。

メロディは音高の比率によって求まり、音高の比率はノート番号の差によって求められる。ノート番号の差を一つの状態とすると、全 153 状態を持つマルコフモデルとなり、その作曲者毎の遷移行列を図 2 に表す。

遷移行列は、現在の状態を縦軸に見た時の横軸の値が、次の状態に遷移する確率として表される。縦線が強く反応しているのは、どの状態にあっても遷移が発生しない確率が高いことを示し、斜め線が強く反応しているのは、遷移が発生した状態にあった場合、次にその遷移と逆方向に遷移する確率が高くなることを示す。図 2 を見ると、ベートーベンは激しい遷移が多く、モーツァルトは遷移が比較的緩やかであることが分かる。

ハーモニーの抽出には、作曲者毎に使われる和音パターンを全て抽出し、その中で良く使われるパターンのみを選択することで、確率モデルのパラメータの次元を削減した。選択されなかったパターンはコサイン類似度を用いて、選択されたパターンの中から一番似ているパターンに置換した。図 3 は原曲と、代表的なパターンで置換して得られた楽曲である。

図 3 上の原曲と比べると、置換後の楽曲は一部の音が消えてしまっており、表現能力が下がっていることが分かる。一方、二つの曲を聞き

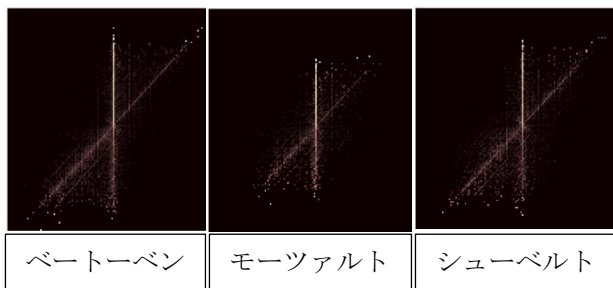


図 2 : 作曲者毎の遷移行列

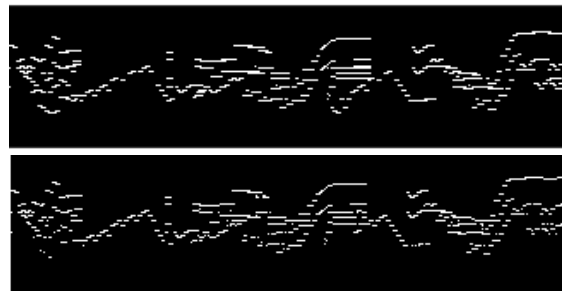


図 3 : 原曲 (上), 置換後 (下)



図 4 : 生成された楽曲

比べた結果、全体を通して楽曲に対する印象に変化は見られなかった。

4. 結果・考察

図 4 はベートーベンの楽曲を学習させた確率モデルから生成したものである。生成された楽曲は連続した和音として不自然な遷移は見られなかった。一方、トニック、サブドミナント、ドミナントといった順に進行するような基本的なコード進行が見られなかったため、音楽としては全体を通して不自然なものとなった。これは、単純マルコフ連鎖モデルでは、音楽の時系列情報を十分に表現出来ていないことが原因であると考えられる。

5. まとめ

本研究では、マルコフ連鎖モデルを生成モデルとして用いた楽曲生成システムを構築することにより、楽曲の自動生成を試みた。

今後の課題としては、生成モデルとして時系列データを扱えるようなモデルを適用することで、作曲者毎の特徴を持った楽曲生成の可能性を検討する。またそれと同時に、今回抽出したパターンから作曲者毎の楽曲に対する特徴表現を獲得する手法を考えていく予定である。

参考文献

- [1] 秋口俊輔. (2009). ソフトコンピューティング手法を用いた曲印象からの楽曲自動生成システムの構築. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 21(5), 782-791.]
- [2] John Thickstun, Zaid Harchaoui and Sham Kakade. Learning Features of Music from Scratch. arXiv preprint arXiv:1611.09827,2016