Vol.2018-MUS-119 No.24
Vol.2018-SLP-122 No.24
2018/6/16

メロディの自動補間によるメロディ編集手法の検討

佐々木 真菜† 坂井 拓己† 平井 辰典†

概要：本稿では，ニューラルネットワークを用いて既存楽曲のメロディを部分的に自動編集する手法について検討する．画像処理分野において，ニューラルネットワークを用いた画像の自動補間手法（画像のインペインティング）が提案されており，欠損領域がある画像に対して，周辺のピクセル情報を用いることで自然な補間を実現している．この技術をメロディに応用し，部分的に欠損したメロディの自動補間（メロディのインペインティング）を実現できると考えた．本研究では，既存メロディの一部分を意図的に欠損させ，その欠損箇所をニューラルネットワークにより自動補間する手法について検討する．具体的には，欠損させたメロディの音高列を入力データとし，教師データとなる元のメロディを推定できるようにネットワークのモデルを学習する．これにより，既存楽曲のメロディの一部を元のメロディとは違うものに変えつつ，自然なメロディとなるような編集の実現を目指す．

1. はじめに

既存の音楽が常にリスナーにとって最高の音楽とは限らない．最高の音楽の形は創り手の数だけあり，また，リスナーの数だけある．現在，我々が好きな曲に関しても，もっと感動したり高揚したりするメロディの並びは他にもあるはずで，そういった意味では既存の音楽には拡張可能な可能性がリスナーの数だけ存在しているといえる．しかし，音楽経験や音楽的知識のないリスナーが音楽にアレンジを加えることは難しい．さらに，自身にとってよりよいメロディを生み出すことはより難しい．そのため，ほとんどのリスナーは既存楽曲に対してオリジナリティを発揮できず，ありのままに聴くことしかできずにいる．そこで我々は，既存楽曲をベースに，部分的なアレンジを加えることができる「メロディの自動補間（インペインティング）」の手法を検討する．本手法は，既存楽曲の一部分のメロディを意図的に欠損させ，教師データとなる元のメロディを推定できるようにニューラルネットワークのモデルを構築することで，任意の入力メロディの欠損箇所に自然に繋げることのできる別のメロディを生成する．これにより，既存楽曲ベースのメロディアレンジが可能となる．

本手法を用いることで，音楽経験や音楽的知識の有無に関係なくお気に入りの楽曲のアレンジをすることができる．つまり，既存の音楽をベースに個人のクリエイティビティを発揮できるという点において，誰もが新しい方法で音楽創作に関わることができるものである．本稿では，本提案手法の具体的な実装方法や結果について述べ，メロディの自動補間技術の可能性について検討するとともに，新しい音楽創作の形を提案する．

2. 関連研究

コンピュータを用いてメロディを生成することは，自動作曲と呼ばれる．これまでに自動作曲に関連した研究は多く提案されてきた．コンピュータを用いた自動作曲は，1957年のイリノイ大学のコンピュータ ILLIAC I による，イリアック組曲が，世界最古の作品であると言われている．このようにコンピュータが登場した初期の頃から，自動作曲を実現しようとする取り組みが，行われてきたことがわかる．深山らによると，自動作曲のこれまでのアプローチとして，音楽理論に基づきルールベースで自動作曲を行う手法，や既存の作曲家のスタイルを模倣した断片を繋ぎ集める手法，確率モデルによる自動作曲手法等が提案されてきた[1]．また昨今，自動作曲において，LSTM や RNN などといったニューラルネットワークに関連した技術が用いられることが多い．音楽コンテンツを生成するにあたり，単純なニューラルネットワークでは，時系列データを扱うことができなかったが，これらのモデルでは，時系列データを考慮することができる．一方で，CNN を用いて，GAN によりメロディを生成する MIDINET[2]等の手法も提案されている．このように自動作曲においても，ニューラルネットワークに関連した技術が成果を上げている．

これらの自動作曲手法はユーザに音楽経験や音楽的知識がなくても，比較的簡単に音楽を生成できる利点がある一方で，ユーザのクリエイティビティを反映することが難しいという欠点もある．このような問題を解決するために，Human-in-the-loop 型の自動作曲技術も提案されている[3]．しかし，自動作曲技術を利用している限りは，どうしてもユーザの思い通りにメロディを生成することができず，楽曲全体の品質が満足いくものとはなりづらい．そこで，既に品質の高い既存楽曲をアレンジするという形での音楽創作の支援について考える．既存楽曲について，ユーザのクリエイティビティに応じたアレンジを支援できれば，自動作曲とは違い，ユーザが関与できる形で簡単に質が高い音楽制作ができると考えた．

吉永らは，HMM を用いてピアノ用の楽譜から，運指を考慮したギター用のタブ譜を生成する手法を提案した[4]．これは，ギター運指の状態遷移確率を求めることで，できるだけ大きく運指が遷移せず，原曲の曲想を保つモデルを実現したものである．これはピアノ用の楽譜を，ギター用にアレンジ(編曲)したことに該当するものと考えられる．沼尾らは，ユーザの感性に合わせて，音楽を自動編曲/作曲する手法を提案した[5]．この研究では，被験者に既存の楽曲を聴取させ，6 種 12 語の形容詞対によって評価してもらう．そして，楽曲から受けた印象を用いて，楽曲構造と印象との間の関係を学習し，ユーザの感性に応じた編曲を実

† 駒澤大学

Vol.2018-MUS-119 No.24
Vol.2018-SLP-122 No.24
2018/6/16

現する．このように，元の楽曲をアレンジすることで，より弾きやすい楽曲，感性に応じた印象の変化をもたらす編曲等が可能である.同様に,アレンジというアプローチで，ユーザにとって，より好みにあったメロディとなるような楽曲の制作も実現できるのではないかと考える．

土屋らは，旋律の概形を曲線として描くことで，作曲の経験がないユーザでも，旋律の全体像を把握しながら，作曲ができるシステムを提案した．このシステムは，ユーザが完成した曲に対して不満があった場合も，満足するまで再編集することを可能としており，ユーザのクリエイティビティに基づき，曲を完成させることができる．また，音符の情報を始めとする音楽的な概念を直接取り扱わないため，初心者でも手軽に作曲ができる[6].

このように，音楽経験がないユーザに対する作曲支援技術は提案されているが，作曲未経験者が完全にゼロから作曲する事に対しては,依然としてハードルがあると考える．そこで我々は，既存の曲をベースとして，アレンジによってより自分好みの新たな楽曲にしていくというアプローチで，ユーザのクリエイティビティによる楽曲の拡張を目指す．本研究では，この目標を達成するための手法として，メロディの自動補間を提案する．

周辺情報を基に，情報を補間する研究は，様々な分野で提案されている．音声データにおいて，ノイズや音割れが発生している箇所を，周辺情報を基にした補間により自然な音になるように補間する手法が提案されている[7]．画像に関しても，欠損領域を補間するインペインティングに関する研究が行われている．Iizuka らは，CNN によって画像内の任意の欠損領域を，自然に補間する技術を提案している[8]．インペインティングは動画にも拡張されており，欠損したフレームの補間や不適切な物体の削除を可能としている[9]．このような補間技術では，ユーザが不要な領域や変更したい領域を指定することで，新たな画像や動画が生成できる．音楽のメロディにおいても同様に，補間による編集が可能なのではないかと考えられる．

本研究では，既存楽曲におけるメロディにおいて，ユーザが変更したい箇所を欠損情報とみなし，ニューラルネットワークで補間することでアレンジを加える手法を提案する．メロディの自動補間による置き換えが実現すれば，既存楽曲の特定部分の編曲が音楽経験のないユーザにも可能になる．

3. 前処理

本章では，ニューラルネットワークを用いて，メロディの自動補間を行うモデルを構築するために必要なデータセットの作成方法について記述する．

機械学習において，学習データの量と質がモデル精度の向上につながるため，適切なデータの準備と前処理が重要である．本研究では The Lakh MIDI Dataset v0.1 の Clean MIDI subset を使用した[10]．このデータセットは，176,581 曲分の MIDI ファイルを含んでおり，その多くがポピュラーな洋楽曲で構成されている．ここからメロディを抽出できれば，高精度なモデルの実現が期待できる．

前処理として，この MIDI ファイルからメロディ情報の抽出を行った．MIDI ファイルは，トラックごとに音高や音符長などが記録されており，メロディに該当する要素を比較的容易に扱うことができるデータ形式である．ただし，ここで，どのトラックがメロディトラックであるかを明らかにしなければいけないという問題がある．そこで，以降は本研究における，MIDI ファイルからメロディトラックを抽出する方法について記述する．

メロディトラックの抽出を行うため，まず，176,581 曲分の MIDI ファイルをテキストデータに変換する．変換したテキストは，楽曲及び各トラックのメタデータ，音のオン/オフに関する note event で構成されている．さらに，この note event はトラックごとに分けることが可能であり，メロディに該当するトラックを見つけることができれば，メロディが抽出できる．

一部の MIDI ファイルでは，トラックのメタデータとして，楽器名等の情報を記述されている．そこで，各トラックにどの楽器が使用されているか記載されている MIDI ファイルに限り，“vocal”もしくは“melody”という単語をメタデータに含むトラックを判別することでメロディトラックを同定する．この手法により，メロディトラックの抽出を行った結果，176,5181 曲分の MIDI ファイルのうち，該当する単語を含んでいる MIDI ファイルは 10,853 曲であった．本稿ではこの手法により抽出できたファイルのみを処理の対象とする．

メロディを抽出したファイルを分析し，メロディを構成する「ノートナンバー」と，note event 間の音間隔を表す「デルタタイム」，デルタタイムがどの程度の長さかを示す「分解能」の情報を抽出する．これにより，ニューラルネットワークの学習に使用する音高列を取得する．

ニューラルネットワークに入力するメロディの長さを 1 曲分とすることもできるが,曲ごとの音符数は異なるため,入力の長さは固定長とする．本研究では，補間したい箇所の前後 1~2 小節程度が考慮できるような音符数として，入力層のユニットの数を 12 個に設定する．そこで,各楽曲のメロディデータを，休符も含めて 12 音ごとに切り分けることで，ニューラルネットワークに入力する学習データを用意する．実際には，ノートナンバーの列(12 次元)+各ノートのデルタタイム(12 次元)+分解能(1 次元)の情報を収めた計 25 次元のデータを学習用のひとまとまりのデータとする．10,853 曲分のメロディに対して上述の処理を行った結果，入力データ数は 326,721 個となった．これらのデータを用いてニューラルネットワークの学習を行う．

4. メロディの自動補間手法

本章では，ニューラルネットワークを用いたメロディの自動補間モデルの構築手法について記述する．学習データには，前章に記述した手法で抽出した既存楽曲のメロディから作成した 12 音分のデータを用いる．12 音のうちの中間 4 音分のノートナンバーをランダムな値に置き換えることでメロディの欠損を表現して入力データとする．なお，本稿においては，デルタタイムに関する情報は学習の対象とせず，各音符の長さは原曲のままで，ノートナンバーのみをアレンジするものとする.中間 4 音を欠損させた 12 音

Vol.2018-MUS-119 No.24
Vol.2018-SLP-122 No.24
2018/6/16

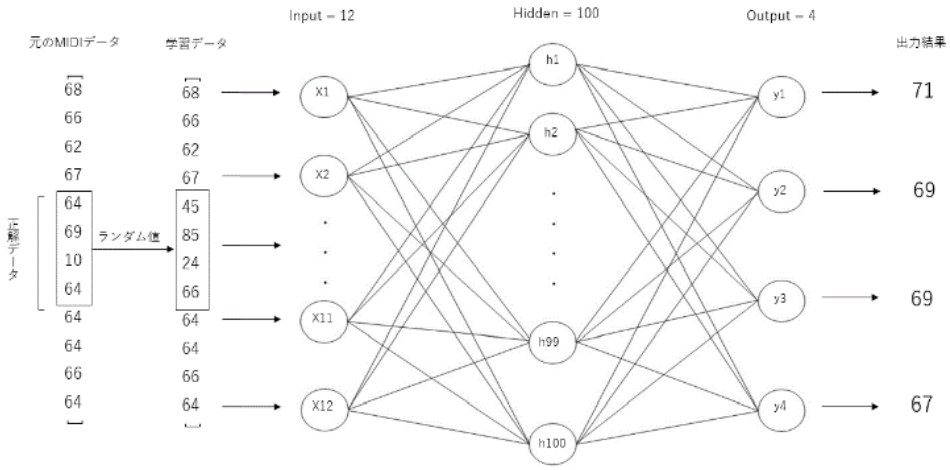

図 1：ニューラルネットワークによるメロディ補間モデルの構造

分のノートナンバーのデータをニューラルネットワークの入力層に与え，ランダムな値に置き換える前のノートナンバーを教師データとして学習させることで，中間のノートナンバーを予測するモデルを学習する．

本研究では，中間層が 1 層であるニューラルネットワークのモデルを構築した．メロディの自動補間モデルの概要を図 1 に示す．入力層を 12 ユニット，中間層を 100 ユニット，出力層を 4 ユニットで構成し，中間層，出力層の活性化関数にはシグモイド関数を利用した．正解データに対する出力の誤差を，出力層と中間層を結ぶ各重みの割合で分配して，逆伝播させる誤差逆伝播法を利用して学習を行った．出力層の 4 ユニットは補間対象となる 4 音を表すものであり，ランダム値の前後のメロディ情報から，欠損箇所に当てはまるメロディのノートナンバーを予測するようなモデルとなっている．学習は，10,853 曲文のメロディから得られた 300,257 個のデータに対して 20 エポック行い，残りのデータはテストデータとした．

5. 結果と考察

学習したモデルに対して，テストデータを入力し，出力された 4 音分のメロディを欠損箇所に当てはめることで，新たな MIDI ファイルを生成した．これらの結果として，既存楽曲の一部のメロディと提案モデルで新たに補間したメロディとを比較する．生成されたメロディの例を図 2 と図 3 に示す．図 2，図 3 は，既存楽曲のメロディと本手法によるメロディ補間後の MIDI ファイルのメロディを楽譜に起こしたものである．図 2 について，既存のメロディのノートナンバーの並びは[64, 69, 10, 64]（E4, A4, R, E4）であったのに対し，出力は[71, 69, 69, 67]であった．また，図 3 については，既存のメロディのノートナンバーは[71, 72, 67, 67]であったのに対し，出力のノートナンバーは[72, 70, 69, 67]であった．MIDI 形式には休符を表すノートナンバーがないため，便宜的に 10 を休符のノートナンバーと置いており，出力で 10～20 の出力を得た場合には休符に変換した．上記の例を見ると，図 2，図 3 ともに一致したのは 1 つのノートナンバーだけであるが，実際にそれぞれの既存のメロディと補間後のメロディとを聴き比べてみると，どちらが既存楽曲のメロディかわからないほど自然に補間

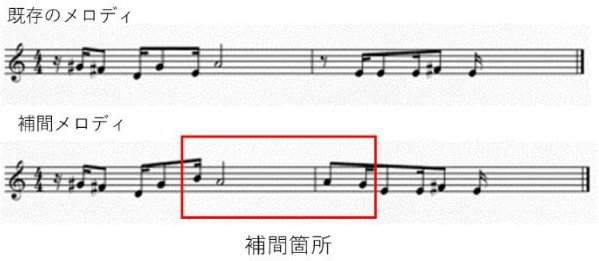

図 2.既存楽曲のメロディと本手法による
補間後のメロディとの比較（パターン 1）

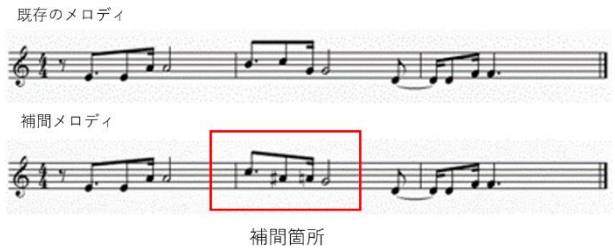

図 3.既存楽曲のメロディと本手法による
補間後のメロディとの比較（パターン 2）

することができていた．それ以上に，メロディ補間モデルから出力されたノートナンバーの並びの方がよりよいと思える出力を得ることができた．しかし，全体的な評価としては，欠損箇所 4 音をすべて正しく推測することができたテストデータは少なく，まだまだ工夫や改善が必要であると感じた．解決案としては，まず一つ目に，今回は中間層が 1 層であるニューラルネットワークを構築したが，中間層を複層にした深層学習とすることで，より精度の高いモデルを構築できると考えている．また，学習データの形式について，今回は欠損箇所を表現するのにランダム値を当てはめたが，欠損させたい箇所を 0 で表現したり，マスク処理を施してみたりと形式を変えて実験することも検討している．他にも，入力層のユニット数を減らし，欠損箇所を入力しない方法なども試してみたい．これらは今後の課題である．

6. まとめ

本研究では，既存楽曲の前後のメロディ情報から中間のメロディを推定するようなニューラルネットワークのモデルを構築することで，該当箇所に自然に当てはまる別のメロディを自動補間するメロディ編集手法を検討した．本研究の特徴は，メロディを意図的に欠損させることで欠損箇所に当てはまるメロディを推定できるモデルを学習するという手法にある．モデルの精度は決して望ましいものとはならなかったが，生成された新しいメロディの自然さに関しては，主観的な評価ではあるが，満足のいく結果となった．このメロディについての検証，評価については今後の課題である．

本稿では，音高列のみについてのメロディ補間しか考慮しておらず，デルタタイムを考慮したメロディの補間はできていない．そのため現状では，音符の長さは元のメロディとまったく同じになっている．今後，デルタタイムと音高列の両方を考慮することで，より柔軟なメロディの自動補間の実現を目指す．また，学習データの形式やモデルのパラメータに関する実験と改善を繰り返すことで，メロディ補間の精度を上げることを今後の目標にしたいと考えている． その後，より自然なメロディ補間を目指し，リスナーの好みに合った補間メロディの出力も目指す．

本手法によって実現できるのはメロディの部分的なアレンジにとどまらない．本手法を応用することで，曲と曲とを自然に繋ぐメロディを自動補間することによるメドレーの自動生成や，曲を短縮するための編集作業の自動化なども可能になると考えている．また，本稿では，メロディの自動補間についてのみ検討したが，同じ構造のネットワークと学習データの形式を用いることで，歌詞などの別のドメインにおける自動補間も可能であると考えている.これについても，メロディ補間の精度が上がり次第実現を目指す．

最終的には，本手法を組み込んだ楽曲編集を実現するインタフェースを実装することを考えている．それにより，音楽経験のないユーザでも個人のクリエイティビティを発揮して満足のいく品質の楽曲を創作できるようなシステムの実現を目指したい．

参考文献

[1] 深山覚，後藤真考：音楽コンテンツと生成 ，情報処理，Vol.57，No,6，pp.516--518 (2016).
[2] Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang： MIDINET:A Convolutional Generative Adversarial Network for Symbolic-Domain Music Generation，Proceedings of the 18th ISMIR Conference，pp.324−331 (2017).
[3] 北原鉄朗，深山覚，片寄晴弘，嵯峨山茂樹，長田典子 :OrpheusBB:Human-in-the-loop 型の自動作曲システム，pp.1--8，情報処理学会インタラクション (2011).
[4] 吉永悠馬，堀玄，深山覚，嵯峨山茂樹：隠れマルコフモデルによるギターのための運指決定および自動編曲，pp.1--4，日本音響学会 (2009).
[5] 沼尾正行，高木将一，中村啓佑：ユーザの感性に合わせた音楽自動編曲及び作曲，音楽情報科学，Vol.41，No.9，pp.44−54 (2001).
[6] 土屋裕一，北原鉄：音符を単位としない旋律編集のための旋律概形抽出手法，Vol.54，No.4，pp.1302−1307 (2012).
[7] Amir Adler, Valentin Emiya, Maria G. Jafa-ri, Michael Elad, Remi Gribonval, and Mark D.Plumbley： Audio Inpainting，IEEE Transactions on Audio, Speech, and 1Language Processing, Vol.20，No.03，pp.922−932 (2011).
[8] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa："Globally and Locally Consistent Image Completion"，ACM Transaction on Graphics (Proc. of SIGGRAPH)，Vol.36，No.4，pp.107:1--107:14 (2017).
[9] Alasdair Newson, Andrés Almansa, Matthieu Fradet, Yann Gousseau, and Patrick Pérez:Video Inpainting of Complex Scenes:SIAM Journal on Imaging Sciences，Vol.07，No.4，pp.1993−2019 (2014).
[10] Colin Raffel: Learning-based Methods for Comparing Sequences, with Applications to Audio-to-midi Alignment and Matching, Columbia University (2016).