4V-04

Transformerによる二値分類を用いたメロディ生成

石田 裕太郎† 杉本 徹‡
芝浦工業大学大学院 理工学研究科† 芝浦工業大学 工学部‡

１． はじめに

計算機によるメロディ生成の研究は古くから行われており，これまでに遺伝的アルゴリズムやリカレントニューラルネットワーク(RNN)を用いた手法，近年では，特に自然言語処理分野において大きな成功を収めている Transformer[1]を用いたものなど，様々な手法が提案されている．

従来研究において，Transformer のデコーダ部分を用いたメロディ生成手法の研究が行われている[2]．この研究は長期的な構造を持つ自然なメロディの生成を可能にし，学習用データには既存の楽曲を用いた．しかし，メロディは言語と違い，正確さはそれほど重要ではないと考え，悪いメロディの特徴を避けることでも良いメロディが生成できるのではないかと考えた．本研究では，Transformer のエンコーダ部分を用いて良いメロディ悪いメロディの二値分類を行うモデルを作成し，メロディを生成する手法を提案する．

２． 研究の概要

本研究では，以下の手順で研究を進めた．

(1) 良いメロディ悪いメロディの二値分類モデルの作成
(2) メロディの生成
(3) 悪いメロディデータの増量・改良
(4) メロディ生成の結果に対する評価実験

３． データの仕様

本研究では，REMI[3]を基にメロディを機械学習モデルに入力可能な時系列データに変換する．音符の高さ(Pitch)は 128 種のトークンで表現し，音符の長さ(Duration)は 64 種のトークンで表現する．小節の始まりに Bar_None トークンを置き，小節の各ポジションは 32 種のトークンで表現される．例えば，音符列は図１のようにトークン列に変換される．

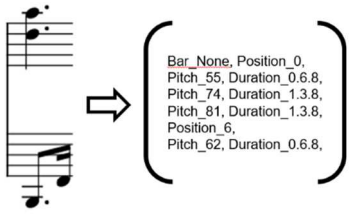

図１：音符列とトークン列の対応例

４． 二値分類モデルの作成

４．１ 良い・悪いメロディデータの作成

良いメロディデータは，邦楽，洋楽，ボーカロイド，アニソン，BGM など様々なジャンルの楽曲 210 曲分の MIDI データから作成する．また，4 分の 4 拍子の楽曲に限定する．楽曲に対して，手作業でメロディが一区切りする印象を受ける箇所で区切り，シーケンス長が最大 256 である学習データの 1 個分を作成する．この結果，3,300 個のメロディデータを得た．

悪いメロディデータを作成するために，まず良いメロディデータにおけるそれぞれの Pitch, Duration, Position の値の出現回数をカウントし，出現確率を求める．そして求めたそれぞれの出現確率に基づいて，最大シーケンス長 256 のランダムな悪いメロディデータを 3,300 個作成する．

４．２ 二値分類モデルの作成

本研究では，自然言語処理分野の二値分類タスクにおいて高い精度が得られることが知られている Transformer のエンコーダ部分を用いて，良いメロディ悪いメロディの二値分類を行うモデルを作成する．図 2 に提案モデルの概要を示す．

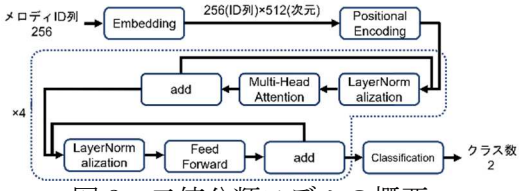

図２：二値分類モデルの概要

提案モデルの入力は，最大シーケンス長 256，バッチサイズ 32 の 2 次元データとし，Embedding 処理により追加される埋め込みベクトルの次元数は 512 に設定する．Multi-Head Attention のヘッド数は 8 に設定し，出力は要素数 2 の配列とする．出力の一つ目の要素は，良いメロディであるかどうかの評価値を表し，二つ目の要素は，悪いメロディであるかどうかの評価値を表す．

５． メロディの生成

５．１ 初期メロディの生成

初期メロディを完全にランダムに生成すると

Melody generation using binary classification with Transformer
† Yutaro Ishida, Graduate School of Engineering and Science, Shibaura Institute of Technology
‡Toru Sugimoto, Faculty of Engineering, Shibaura Institute of Technology

収束が遅くなるため，Duration，Position は悪いメロディデータを作成した際と同様の出現確率とし，Pitch は，3,300 個の良いメロディデータの中から 1 個を選択し，その 1 個から Pitch の出現確率を求める．そしてそれぞれの出現確率に基づいて初期メロディを作成する．

５．２　二値分類モデルを用いたメロディの生成

初期メロディから開始して以下の手順でトークンの変更を繰り返すことによりメロディを生成する．

(1) 次に実行するトークン変更操作を，置換，追加，削除，ポジション追加の 4 種類の操作の中からランダムに 1 つ選択する．図 3 にトークン変更操作の具体例を示す．

図 3：トークン変更操作の具体例

(2) 変更操作における Pitch, Duration, Position の出現確率は初期メロディの生成の際と同様とし，(1) で選択した操作を実行する．

(3) 変更後，良いメロディと判断される評価値と悪いメロディと判断される評価値の差が大きくなった場合，この変更を確定する．評価値の差が小さくなった場合，変更を取り消す．

(4) (1)～(3) を 500 回繰り返す．

現段階での二値分類モデルは，評価値が高く出やすい Pitch の値や，低く出やすい Pitch の値があるなど，学習にムラがある．そのため，繰り返しの回数を増やした場合に評価値が高く出やすい Pitch の値が多く出現するような傾向が見られた．そこで今回が繰り返し回数を 500 回に設定した．また，収束に必要な繰り返し回数を抑えるために Pitch の出現確率を限定した．

６．　悪いメロディデータの増量・改良

二値分類モデルを用いてメロディの生成を行い，生成後のメロディが悪いメロディに感じる場合，悪いメロディデータに追加する．この際のメロディ生成では，Pitch, Duration, Position の出現確率は悪いメロディデータ作成時と同様の値にする．また，繰り返す回数を 100 回に変更する．

７．　評価実験

提案手法の効果を確認するため，学生 6 名を被験者として，初期メロディと二値分類モデルを用いて生成したメロディを比較する評価実験を実施した．評価対象は任意に生成した初期メロディとその初期メロディから提案手法を用いて生成したメロディのペア 20 組である．各ペアをどちらが初期メロディか二値分類モデルを用いて生成したメロディか分からない形で被験者に聞いてもらい，4 段階尺度で回答してもらう．評価結果を図 4 に示す．

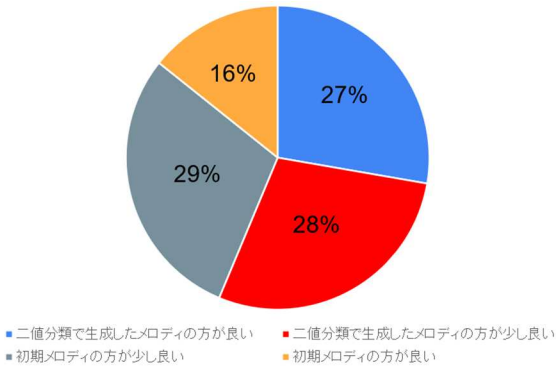

図 4：4 段階尺度調査の評価結果

結果から，二値分類モデルを用いたメロディ生成がある程度の性能を示す様子が確認でき，特に初期メロディの方が良いと判断される例が少ないことが確認できる．

８．　考察

評価実験では，初期メロディの方が良いと判断される例が最も少なかったが，17 例確認された．17 例確認された要因として，初期メロディ生成時の Pitch の出現確率から作成したメロディは，既にある程度の楽曲としての統一感を備えていることが考えられる．そのため，初期メロディの段階で良いメロディと感じるものがしばしば生成され，二値分類モデルを用いることでむしろ悪いメロディになってしまうことがあると考えられる．また，元々ある程度良いメロディをさらに良いメロディに変更できるようにするために，良いメロディ悪いメロディの二値分類モデルをさらに改良することが必要と考えられる．

９．　おわりに

本研究では，Transformer による二値分類モデルを用いてメロディを生成する手法を提案した．今後の課題として，初期メロディを生成する時の Pitch の出現確率を別の機械学習モデルにより求めることが挙げられる．また，特に Pitch の値に学習のムラがある課題の解決を目指したい．

参考文献

[1] Ashish Vaswani, Noam Shazeer et al. “Attention Is All You Need”, Neural Information Processing Systems, 2017.

[2] Cheng-Zhi, Anna Huang et al. “Music Transformer: Generating Music with long-term Structure”, International Conference on Learning Representations, 2019.

[3] Yu-Siang, Yi-Hsuan Yang. “Pop Music Transformer: Beat-based Modeling and Generation of Expressive Pop Piano Compositions”. International Conference on Multimedia, 2020.