The 39th Annual Conference of the Japanese Society for Artificial Intelligence, 2025

深層生成モデルと作品選択モデルに基づく歌詞付きメロディの評価値予測

Evaluation Score Prediction for Melodies with Lyrics Based on Deep Generative Model and Artwork Selection Model

西村 草介 *1　　中村 栄太 *1
Sosuke Nishimura　　Eita Nakamura

*1九州大学
Kyushu University

自動音楽生成が盛んに研究される中，楽曲の自動評価の研究が重要になっている．本研究では，歌唱曲を対象として，歌詞に対するメロディの評価値を予測する方法を研究する．具体的には，芸術制作過程を表す作品選択モデルを考え，メロディの深層生成モデルを用いて推定するメロディの適合度から鑑賞者にとっての平均評価値を予測する方法を提案する．性能評価実験の結果，提案手法により聴取実験の結果を高い精度（相関値 0.643）で予測できた．楽曲データのみから作品の自動評価を可能とする本手法は多様な芸術データに適用可能であり，人間の価値基準を反映した音楽生成や創作支援を通して芸術の発展に寄与することが期待される．

1. はじめに

近年，音楽生成の研究が活発化している [1]．機械学習を用いることで既存の音楽を模したデータが生成できるが，生成確率のみで音楽の芸術的価値を評価することは難しく，高品質の音楽生成には楽曲の自動評価の問題を考える必要がある．大量の主観評価値データを基に楽曲の評価値を推定する研究が既にあるが [2]，評価値データを集めるコストは大きく，多様な形式・ジャンルの音楽に適用する上で課題も多い．ここでは，楽曲データのみから構築可能な楽曲の自動評価手法について考える．本研究では，多数の鑑賞者による楽曲の平均的評価値を楽曲の特徴から予測する問題を考え，楽曲データから推定する創作者にとっての楽曲の適合度を用いた予測手法を構築する．この適合度の推定方法として，芸術制作過程を表す作品選択モデルに基づく方法を提案し，その有効性を実験を通して示す．

具体的な対象として，日本ポピュラー音楽の歌唱曲に注目し，歌詞に対するメロディの評価値を扱う．歌詞の特徴量とメロディ要素の対応関係は，曲の聴きやすさや歌いやすさを通して，曲の評価に影響すると考えられる．山田耕筰は，高低アクセントを持つ日本語の性質に着目し，歌詞抑揚と旋律輪郭の一致原理（歌詞のアクセント遷移とメロディの音高遷移の方向を一致させる作曲原理）を提唱した [3]．近年では，この一致原理に関する童謡と唱歌のコーパス分析 [4] や歌詞のアクセントを制約とするメロディ生成手法 [5]，メロディに対する歌詞生成 [6] の研究があるが，楽曲の評価値予測までは行っていない．メロディ要素には音高とリズムがあり，それらに影響を与え得る歌詞特徴量はアクセントの他にも単語境界や意味情報などが考えられる．本研究では，歌詞とメロディの複雑な関係性を機械学習に基づく手法で捉えることを目指し，影響が大きい歌詞特徴量の調査も行う．

2. 研究方法

2.1 歌詞付きメロディデータの構築と基礎分析

本研究のために収集した歌詞付きメロディデータについて述べる．歌詞付きメロディデータは，メロディの音楽情報と歌詞の言語情報からなる（図 1）．メロディ $(p_l, \tau_l)_{l=1}^{L}$ の各音符

図 1: 歌詞付きメロディデータの例（ヨルシカ「春泥棒」）

は，MIDI ノート番号で表す音高 $p_l \in \{0, \ldots, 127\}$ と，1 小節を 48 分割したティック単位で表す発音時刻 $\tau_l \in \mathbb{N}$ のペアで表す．音高は，楽曲を自然調（ハ長調またはイ短調）に移調してから計算する．歌詞は，通常の日本語表記の「平文」と，各音符に割り当てられた振り仮名で表される「音符ルビ」の二つの形態で表す．平文は後述の特徴量の抽出に用いる．音符ルビは仮名 3 文字までのものが全体の 99.92% であったため，仮名の文字数は最大 3 とし，1 音符に 3 文字を超える仮名を持つ歌詞はデータに含めていない．音符をまたぐ仮名の継続（楽譜では横棒で表される）は「〜」で表す．データは，平文の改行箇所で区切られたフレーズを単位として整備する．なお，英語や数字，特殊記号などが含まれるフレーズはデータには含めなかった．ポピュラー音楽のヒット曲を主に含む 1988 曲から 33,751 フレーズ，音符総数 287,396 のデータを収集・整備した．

メロディへの影響が考えられる歌詞特徴量として，アクセント，単語境界，品詞，意味情報，仮名ラベルを扱う．アクセントと単語境界，品詞情報は MeCab を用いて抽出した．先行研究 [4–6] に従い，アクセントは音高の高低で表す．音高が低いならば 0 を，高いならば 1 を割り当てる．単語境界も二値のラベルで表し，単語の先頭の仮名に (1)，その他に (0) のラベルを付与する．品詞情報は，仮名が属する単語の品詞（16 種類）に応じて，1 から 16 までのラベルで表す．意味情報の抽出には，日本語用 Sentence BERT を用いた *1．この手法により，フレーズは 768 次元の埋め込みベクトルとして表される．冗長な次元による過学習を避けるため，主成分分析（PCA）による次元削減で得られた n 次元のベクトルを意味情報の表現

連絡先: 西村草介, 九州大学工学部電気情報工学科, nishimura.sosuke.638@s.kyushu-u.ac.jp

*1 https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2

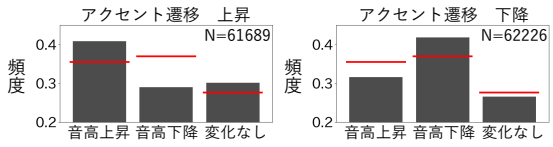

図 2: 歌詞のアクセント遷移によるメロディの音高遷移の頻度分布の変化．赤線は全体平均を示す．

として最終的に用いる．

ポピュラー音楽において，歌詞抑揚と旋律輪郭の一致原理 [3] がどの程度成り立っているかを確認するため，隣り合う音符に関する歌詞のアクセント遷移タイプごとの音高遷移の頻度分布を図 2 に示す．アクセントが上昇/下降する場合，全体平均に比べ，音高が上昇/下降する頻度が増加している．これはアクセントと音高の遷移方向の一致度がメロディの評価基準として使えることを示している．一方で，他の歌詞特徴量とメロディ要素の間には，この様に単純な対応関係を見出すことは難しく，複雑な要素間の関係性を捉えたメロディの評価値予測には機械学習を用いる方法が有効であると考えられる．

2.2　作品選択モデルに基づく適合度関数の構成

機械学習を用いたメロディ評価値予測を行うため，適合度の推定に基づく方法を考える．この方法では，まず歌詞に対するメロディの適合度を推定して，推定された適合度からメロディの評価値を予測する．下記の作品選択モデルを用いることで，創作者にとってのメロディの適合度をメロディの生成確率と歌詞条件付き確率の比によって表すことができ，この関数は歌詞付きメロディデータを用いて学習できる．作品選択モデルの仮定では，創作者は様々な作品候補を生成して，その中から作品として残すものを，作品の適合度を重みとして確率的に選択する．具体的には，候補のメロディ X の生成確率を $\widetilde{P}(X)$，歌詞 Y に対するメロディ X の適合度を $W(X;Y) \in \mathbb{R}_{>0}$ とすると，Y に対する X の生成確率 $P(X|Y)$ は次の様に表される．

$$P(X|Y) = \widetilde{P}(X)W(X;Y) \tag{1}$$

ここで，不定性がある W のスケールを上の等式が成立つように定めた．次に，$\widetilde{P}(X)$ がメロディ X の周辺化確率 $P(X) = \sum_Y P(X|Y)P(Y)$ と一致するという「学習と生成過程の整合性の仮定」を用いると，適合度は歌詞条件付き確率と条件なし確率の比として表せる．

$$W(X;Y) = P(X|Y)/P(X) \tag{2}$$

また，以下の式変形により，適合度 $W(X;Y)$ と等価な音符単位のクロスエントロピー差（CE 差）$\Delta\mathrm{CE}(X;Y)$ を定義する．

$$\log_2 W(X;Y) = -\log_2 P(X) - [-\log_2 P(X|Y)] \tag{3}$$

$$= L(X)\Delta\mathrm{CE}(X;Y) \tag{4}$$

$L(X)$ はメロディ X の音符数である．

2.3　データからの適合度の推定

2.3.1　生成モデルに基づく手法

メロディの生成モデルを構築することにより，式 (2) の右辺の $P(X|Y)$ と $P(X)$ を計算することを考える．メロディデータには様々な音域や楽譜上の位置のメロディが混ざっているが，ここではメロディの適合度はこれらの属性に依存しないものと考える．そこで，音高のオクターブ移調と発音時刻の小節単位の平行移動について対称性を持つ生成モデルを考えるため，メロディ音符を拡張音高クラスと拍節位置で表す．拡張音高クラス $q_l \in \{0,\dots,35\}$ は，音高クラスの定義 $p_l\%12$（% は剰余演算）を拡張して，上 18 半音，下 17 半音までの音程を忠実に表現できる音高の表現法である．これは，$q_1 = p_1\%12$ と以下の漸化式で定義される．

$$q_l = [p_{l-1}\%12 + \mathrm{Clip}_{-17}^{18}(p_l - p_{l-1})]\%36 \tag{5}$$

ただし，$\mathrm{Clip}_a^b(x)$ は音程 x を必要最小限のオクターブ移動を行って範囲 $\{a,\dots,b\}$ 内にマップする関数である．拍節位置 $m_l \equiv \tau_l\%48 \in \{0,\dots,47\}$ は，発音時刻の小節内での相対位置を表すもので，最大 1 小節の音価を持つ発音時刻の系列を忠実に表現できる．以上により，メロディを拡張音高クラスと拍節位置の系列 $X = (q_l, m_l)_{l=1}^L$ として表す．l 番目の音符に対する歌詞特徴量を y_l と記す時，歌詞条件付きと条件なしの生成確率は以下のように表される．

$$P(X) = P(q_{1:L}, m_{1:L}) \tag{6}$$

$$P(X|Y) = P(q_{1:L}, m_{1:L}|y_{1:L}) \tag{7}$$

ただし，$q_{k:l}$ は，変数の集合 $(q_k, q_{k+1}, \dots, q_l)$ を表す．

具体的な生成モデルの単純な例として，まず Markov モデルを考える．歌詞条件なし Markov モデルでは，拡張音高クラスと拍節位置の生成確率は独立とし，l 番目の音符の拡張音高クラス q_l は直前の拡張音高クラス q_{l-1} のみに依存する（拍節位置も同様）と仮定することで，生成確率は次の式で表される．

$$P(X) = P(q_1)\left[\prod_{l=2}^{L} P(q_l|q_{l-1})\right] P(m_1)\left[\prod_{l=2}^{L} P(m_l|m_{l-1})\right] \tag{8}$$

歌詞条件付き Markov モデルでは，生成確率を $P(X|Y) = P(q_{1:L}|y_{1:L})P(m_{1:L}|y_{1:L})$ と展開する．l 番目の音符の歌詞特徴量としてアクセントラベル $\bar{a}_l \in \{0,1\}$ と単語境界ラベル $\bar{b}_l \in \{0,1\}$ のみ考える．これらは l 番目の音符内の各仮名 k のアクセントラベル $a_{lk} \in \{0,1\}$ と単語境界ラベル $b_{lk} \in \{0,1\}$ から次の式で定義する．

$$\bar{a}_l = \max\{a_{l1}, a_{l2}, a_{l3}\}, \quad \bar{b}_l = \max\{b_{l1}, b_{l2}, b_{l3}\} \tag{9}$$

各音符 l のメロディ要素 q_l, m_l に対してアクセントと単語境界の遷移の依存性を取り入れると，生成確率は次式で表される．

$$P(q_{1:L}|y_{1:L}) = P(q_1|\bar{a}_1, \bar{b}_1)\prod_{l=2}^{L} P(q_l|q_{l-1}, \bar{a}_l, \bar{a}_{l-1}, \bar{b}_l, \bar{b}_{l-1})$$

$$P(m_{1:L}|y_{1:L}) = P(m_1|\bar{a}_1, \bar{b}_1)\prod_{l=2}^{L} P(m_l|m_{l-1}, \bar{a}_l, \bar{a}_{l-1}, \bar{b}_l, \bar{b}_{l-1})$$

次に，Markov モデルでは表せないメロディ要素同士および歌詞特徴量とメロディ要素間の複雑な関係性を取り入れた生成モデルを構築するために，自己回帰型の深層生成モデルである LSTM ネットワークを考える．歌詞条件なし LSTM は，各ステップ l で (q_{l-1}, m_{l-1}) を入力として，予測確率 $P(q_l|q_{1:(l-1)}, m_{1:(l-1)})$ と $P(m_l|q_{1:(l-1)}, m_{1:(l-1)})$ を出力する．歌詞条件付き LSTM では，出力は同じであるが，入力に歌詞特徴量を含めた $(q_{l-1}, m_{l-1}, y_{(l-h):(l+h)})$ を用いる．半幅 h 内にある周辺音符の歌詞特徴量を入力に含めることで，それらによる直接的な依存性を取り入れることが可能である．歌詞特徴量 y_l として，音符 l 内の各仮名 k のアクセントラベル

$a_{lk} \in \{0,1\}$ と単語境界ラベル $b_{lk} \in \{0,1\}$，仮名ラベル c_{lk}（84 種類），品詞ラベル d_{lk}（16 種類），意味情報 e_l（n 次元）を合わせたものを用いる．$q_{l-1}, m_{l-1}, a_{lk}, b_{lk}, c_{lk}, d_{lk}$ は，各々 36, 48, 2, 2, 84, 16 次元の one-hot ベクトルで表す．対応する情報が欠損している場合は 0 埋めする．意味情報はフレーズごとに与えられたベクトルを全ての音符に対して入力する（この部分は h に依らず n 次元とする）．入力次元は $612h+n+390$ である．3. 章の実験では，入力に一部の歌詞特徴量のみ使用した場合も比較するが，この際には意味情報以外の部分の入力次元は固定して，扱わない特徴量の部分は 0 埋めして入力する．

上記の Markov モデルおよび LSTM ネットワークを用いて適合度を推定するには，歌詞付きメロディの学習データを用いて，各々のモデルパラメータを学習し，評価時には学習済みパラメータを用いて生成確率 $P(X)$ と $P(X|Y)$ を計算する．以下では，この方法で求めた適合度を $W_{\mathrm{Markov}}(X;Y)$ および $W_{\mathrm{LSTM}}(X;Y)$ と表す．

2.3.2 条件付き確率に基づく手法

生成モデルを用いて歌詞に対するメロディの適合度を計算する方法として，$P(X|Y)$ を直接用いることも考えられる．例えば，歌詞条件付き LSTM による確率 $P_{\mathrm{LSTM}}(X|Y)$ を用いて推定した適合度を次のように表す．

$$W_{\mathrm{cond}}(X;Y) = P_{\mathrm{LSTM}}(X|Y) \tag{10}$$

この適合度と作品選択モデルに基づく $W_{\mathrm{LSTM}}(X;Y)$ は，$W_{\mathrm{LSTM}}(X;Y) = W_{\mathrm{cond}}(X;Y)/P_{\mathrm{LSTM}}(X)$ の関係で表される（$P_{\mathrm{LSTM}}(X)$ は歌詞条件なし LSTM での生成確率）．

2.3.3 ルールベースの適合度

歌詞抑揚と旋律輪郭の一致原理に基づくルールベース手法で適合度を推定することもできる．この手法では，フレーズの中で歌詞のアクセントが上昇または下降する箇所の数を N_{AC}，これらの箇所でメロディの音高遷移の方向が一致している回数を N_{match} とすると，適合度は次式で表される．

$$W_{\mathrm{rule}}(X;Y) = \frac{N_{\mathrm{match}}}{N_{\mathrm{AC}}+\epsilon} + \epsilon \tag{11}$$

定数 ϵ は $N_{\mathrm{AC}}=0$ または $N_{\mathrm{match}}=0$ の場合に適合度を適切に定義するために導入しており，具体的には $\epsilon = 10^{-3}$ とする．

2.4 適合度に基づく評価値予測手法

本研究では，聴取実験において多数の鑑賞者によるメロディ評価の結果として得られた平均的な評価値を実測データとして扱う．具体的には，ある歌詞 Y に対して二つのメロディ X_1 と X_2 を付けた歌唱曲を対比較して，X_1 がより好ましい方として選ばれた割合 $p(X_1;X_2,Y)$ を X_1 の評価値と定義する．

前節で定式化した創作者にとっての適合度 $W(X;Y)$ からこの評価値を予測する手法として，ロジスティック回帰を用いる方法を考える．まず，鑑賞者にとっての歌詞 Y に対するメロディ X の適合度 $W'(X;Y)$ は，$W(X;Y)$ の関数として表されるとする．具体的には，ルールベース手法や条件付き確率に基づく適合度に関しては，係数 α を導入して，$W'(X;Y) = W(X;Y)^{\alpha}$ と表す．この α は適合度を補正するものであり，一般的に創作者に比べて鑑賞者が音楽的属性の違いに鈍感である効果などを表す．確率比に基づく適合度では，$W'(X;Y)$ を次式で表す．

$$W'(X;Y) = \begin{cases} W(X;Y)^{\alpha_+}, & W(X;Y) \geq 1 \\ W(X;Y)^{\alpha_-}, & W(X;Y) < 1 \end{cases} \tag{12}$$

ここで，確率比の 1 に対する大小関係に対して非対称な補正効果を含めるため，係数 α_+, α_- を導入している．

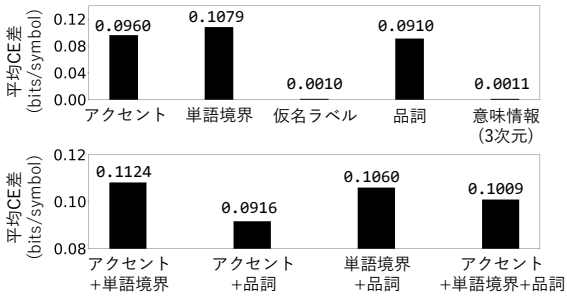

図 3: （上）単一の歌詞特徴量に対する平均 CE 差.（下）複数の歌詞特徴量に対する平均 CE 差.

次に，歌詞 Y に対する二つのメロディ X_1 と X_2 の対比較では，鑑賞者にとっての適合度 $W'(X_1;Y)$ と $W'(X_2;Y)$ の相対比に応じて各々のメロディの選択確率が決まると考えると，X_1 の選択確率は次で表される．

$$R(X_1;X_2,Y) = \frac{W'(X_1;Y)}{W'(X_1;Y)+W'(X_2;Y)} \tag{13}$$

さらに，実際の実験では，鑑賞者がノイズを含む未知の要因によるバイアスを受けてメロディの選択を行うと考え，この効果を $R(X_1;X_2,Y)$ に対する一次式の形で表すと，最終的に X_1 の選択確率は次式で表される．

$$p(X_1;X_2,Y) = \sigma(\beta_1 R(X_1;X_2,Y) + \beta_0) \tag{14}$$

ここで，$\sigma(z) = 1/(1+e^{-z})$ はシグモイド関数で，右辺が 0 から 1 の間の確率値になるように変数変換を行う．式 (14) により，この手法は $R(X_1;X_2,Y)$ を説明変数としたロジスティック回帰として理解できる．上で導入した定数 α, β_0, β_1 などは，自乗誤差を損失関数とするロジスティック回帰を用いて，実測値の予測誤差を最小化するように求める．

上述の方法を拡張して，2.3 節に述べた複数の適合度を組み合わせた評価値予測もできる．例えば，ルールベースの適合度 W_{rule} と LSTM の確率比による適合度 W_{LSTM} を用いる場合は，選択確率の予測値 $p(X_1;X_2,Y)$ は次式で表される．

$$\sigma(\beta_2 R_{\mathrm{LSTM}}(X_1;X_2,Y) + \beta_1 R_{\mathrm{match}}(X_1;X_2,Y) + \beta_0) \tag{15}$$

3. 検証実験

実験では，2.1 節のデータを約 8 : 1 : 1 の割合で学習用と検証用，テスト用に分割して，LSTM と Markov モデルの学習と適合度の推定を行った．前者では，ユニット数が 512 の LSTM 3 層に，入力側に 512 次元への線型層，出力側に出力次元への線形層と softmax 関数を連結したものを用いた．予備実験では，意味情報の次元削減後の次元は $n=3$，その他の歌詞特徴量の入力の窓半幅は $h=3$ が CE 差に関する最適値として得られたため，以下ではこれらの値を用いる．

3.1 各歌詞特徴量の寄与度の分析

確率比に基づく適合度の推定における各歌詞特徴量の寄与度を調べるため，LSTM を用いて計算した歌詞特徴量ごとの平均 CE 差を図 3（上）に示す．アクセントや単語境界，品詞に対する平均 CE 差は約 0.1 bits であり，適合度に十分寄与し得る程度である．一方，仮名ラベルと意味情報に対しては平均 CE 差が小さく，適合度への寄与が無視できる程度であった．仮名ラベルに関しては高い入力次元の影響により，過学習が生じた可能性が考えられる．

適合度の推定手法	予測誤差	相関係数	p 値
ルールベース手法 W_{rule}	0.149	0.542	0.0135
条件付き確率 W_{cond}	0.175	0.143	0.5467
確率比 W_{Markov}	0.171	0.264	0.2602
確率比 W_{LSTM}	0.159	0.443	0.0496
W_{LSTM} と W_{rule}	0.134	0.653	0.0018

表 1: 適合度の推定手法ごとの評価結果

アクセントと単語境界，品詞について，複数の要素を入力に加えた場合の結果を図 3（下）に示す．アクセントと単語境界の両方を用いた場合，単独の場合より平均 CE 差は増加したものの，増加量は小さかった．多くの単語の第二モーラでアクセントが上昇するという日本語の性質により，これらの特徴量が相関を持つことが原因として考えられる．また，品詞を加えると CE 差は減少しており，種類数が多い品詞情報は過学習を招きやすいことが示された．以降では，平均 CE 差が最大となったアクセントと単語境界の組合せを歌詞特徴量として用いる．

3.2 聴取実験によるデータ取得

評価値の実データを収集するため，参加者に同じ歌詞に対する 2 つのメロディを提示してより好ましいと感じた方を選択してもらう実験を行った [7]．実験では，20 種類の歌詞に対する 20 組のメロディを用いて，NEUTRINO [8] で合成した歌声をランダムな順序で提示した．比較対象のメロディのペアは，学習済みの歌詞条件付き LSTM と歌詞条件なし LSTM を用いて生成した．対比較で差を見出し得る有益な実験とするため，前者では CE 差が正で大きく，後者では CE 差が負で小さいものを用いた．各メロディのペア i について，歌詞条件付き LSTM で生成した方（CE 差が大きい方）が選択された割合 p_i を評価値データとして用いる．参加者 186 人の計 1546 件の比較結果から得られた p_i の平均値は 0.561 であり，ランダム選択 ($p_i = 1/2$) からのずれは有意（$P = 1.35 \times 10^{-6}$）であった．また，同じ条件で生成したメロディを楽器音で提示した場合の p_i の平均値は 0.452 であり，メロディの評価値の有意差が歌詞との関係に起因することが確認できた．

3.3 評価値予測の評価実験

2.4 節の手法による評価値予測の結果を表 1 に示す．評価尺度には予測誤差（RMSE）を用いており，参考値として予測値と実測値の相関係数とその p 値も示している．単一の適合度を用いた場合，ルールベース手法の精度が最も高かった．生成モデルを用いた場合の比較では，条件付き確率よりも確率比を用いる方が有効であることと，Markov モデルよりも LSTM を用いる方が有効であることが示された．また，LSTM の確率比とルールベース手法を組み合わせた場合に予測誤差が最小となった．図 4 では，ルールベース手法による選択確率の予測値が同一になるデータ点に対して，LSTM による適合度を組合せることで予測が改善した例が観察できる．これは，深層生成モデルにより単純なルールでは捉えられない歌詞とメロディの関係性を取り入れた評価値予測が可能であることを示している．また，LSTM による適合度に関する係数の値 ($\alpha_+ = 0.001$, $\alpha_- = 0.933$) は，鑑賞者にとっての適合度が創作者にとっての適合度の低い領域により強く依存することを示している．

4. まとめと今後の展望

本研究では，ポピュラー音楽の歌詞に対するメロディの評価値予測には，歌詞のアクセントと単語境界の寄与が大きく，仮名ラベルや品詞，意味情報は寄与が小さいことが示された．

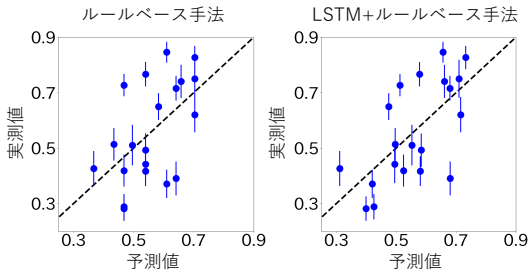

図 4: 評価値の実測値と予測値（縦棒は標準誤差）

メロディ生成モデルの確率比に基づく適合度推定は，歌詞抑揚と旋律輪郭の一致原理に基づくルールベースの適合度と組み合わせることで，聴取実験の結果を高い精度で予測できることが分かった．本手法の考え方は，ある領域のデータの生成確率とより特定された領域のデータの生成確率の比によって，後者のデータにおける価値を推定する枠組みとして一般化できる．その汎用性から，日本語以外の歌唱曲や器楽曲の自動評価への応用や，音楽以外の芸術データへの適用も考えられる．

今後の展望として，人間の価値基準を反映した音楽生成や創作支援への応用が考えられる．通常の機械学習に基づく音楽生成では既存の音楽の再現を目的関数としており，人間の価値基準を直接は学習できない．これに対して，楽曲の自動評価が可能になれば，品質を保ちつつ新しいスタイルの音楽を自動生成する可能性が生まれる．こうした研究は，音楽生成技術が社会に普及した際に永続的な芸術発展を実現するために重要だと考えられる．本手法を用いれば，楽曲編集の前後でどのように評価値が変化するかも予測できるため，音楽創作過程を効率化する効果も期待される．この様な芸術的探求を支援する人工知能技術を通して，音楽文化がさらに発展すると期待している．

謝辞

本研究は，JST FOREST No. JPMJPR226X 及び科研費 No. 23K24917 の支援を受けた．

参考文献

[1] S. Ji et al.: A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges, ACM Computing Surveys, Vol. 56, No. 1, Article 7, 2023.

[2] G. Cideron et al.: MusicRL: Aligning music generation to human preferences, Proc. ICML, pp. 8968–8984, 2024.

[3] 山田耕作: 歌謡曲作曲上より見たる詩のアクセント, 詩と音楽, Vol. 2, No. 2, 1923.

[4] 堤彩香ら: 日本語の音韻と旋律の関係について〜童謡・唱歌を中心に〜, 情報処理学会研究報告, Vol. 2014-MUS-105, No. 5, 2014.

[5] 深山覚ら: 音楽要素の分解再構成に基づく日本語歌詞からの旋律自動作曲, 情報処理学会論文誌, Vol. 54, No. 5, pp. 1709–1720, 2013.

[6] 渡邉研斗ら: メロディと歌詞の相関に基づく自動歌詞生成, 情報処理学会研究報告, Vol. 2017-NL-231, No. 16, 2017.

[7] 実験ページ: https://ice.inf.kyushu-u.ac.jp/ExpNishi2/

[8] NEUTRINO Diffusion: https://studio-neutriNO.com/