2P-03

ポピュラー音楽に対する難易度に応じた深層ピアノ編曲

寺尾 萌夢[1]　石塚 崚斗[2]　錦見 亮[2]　吉井 和佳[2]

[1]京都大学 工学部情報学科　[2]京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

ピアノ編曲とは，複数楽器で演奏されている楽曲をピアノ演奏用に変換するタスクを指す．ポピュラー音楽のピアノ楽譜は必ずしも入手可能であるとは限らず，練習目的でピアノ演奏する際には個人の技量に即した譜面が必要になるため，難易度に応じたピアノ編曲は有用である．従来，隠れマルコフモデル (HMM) を用いて難易度別にピアノ編曲を行う手法 [1] が提案されていたが，ピアノ譜のもつ複雑な同時的・経時的構造を十分に表現できるわけではなかった．最近，深層ニューラルネットワーク (DNN) を用いて音楽（MIDI データ）の生成や編曲，スタイル変換を行う研究が盛んになりつつあるが [2]，統計的に妥当なピアノ譜を難易度別に推定しようとする試みは未だ行われていない．

本研究では，まず，ポピュラー音楽のバンド譜と対応する難易度ラベル付きのピアノ譜を収集し，バンド譜のメロディ・伴奏パートとピアノ譜の右手・左手パートの関係（図 1）や，難易度毎のピアノ譜の特徴を統計的に調査する（図 2）．この結果に基づき，バンド譜に含まれる音符およびそれらをオクターブシフトした音符の集合から，音符を削除することでピアノ譜が得られるという妥当な仮定をおく．バンド譜のメロディ・伴奏パートからマスクを推定する DNN を難易度条件付きで学習する際には，推定されるピアノ譜が統計的に適切な性質を持つように誘導を行う．具体的には，正解のピアノ譜との誤差に加えて，実際のピアノ譜が持つ統計量との誤差を同時に最小化する．

2. 提案法

本章では，ピアノ譜の難易度ごとの統計的性質に基づくピアノ編曲の手法（図 3）について述べる．

2.1 問題設定

いま，難易度ラベル $\mathbf{L} = \{L_0, L_1\}$ とバンド譜のメロディ・伴奏パート $\mathbf{B} \in \{0,1\}^{2\times P\times N}$ が与えられたときに，ピアノ譜の右手・左手パートを推定したい．ここで，$L_0 = \{0\}^{P\times N}$ は初級ラベル，$L_1 = \{1\}^{P\times N}$ は上級ラベル，P はバンド譜の MIDI データに含まれる音高数，N は 16 分音符数を表す．

2.2 モデルの学習

バンド譜の局所的・全体的な特徴に基づいてピアノ編曲を行うため，U-net [3] を用いる．バンド譜からピアノ譜への変換に加えてオンセットを同時に学習するため，U-net は $\mathbf{L}$ と $\mathbf{B}$ に加えてバンド譜のメロディ・伴奏パートのオンセットを表す行列をスタックしたオンセット付きバンド譜 $\mathbf{X} \in \{0,1\}^{5\times P\times N}$ を入力とし，オンセット付きアクティベーション $\boldsymbol{\phi} \in [0,1]^{4\times P\times N}$ を出力する．バンド譜に含まれる音符及びそれらをオクターブシフトした音符の集合を $\mathbf{Z} \in \{0,1\}^{4\times P\times N}$ とする．正解のオンセット付きピアノ譜を $\hat{\mathbf{Y}} \in \{0,1\}^{4\times P\times N}$ とすると，$\boldsymbol{\phi}$ と $\mathbf{Z}$ の要素積が音符の削除に対応していることに注意して，以下の損失関数を最小化するように U-net を学習する．

$$\mathcal{L}^{\mathrm{aran}}(\boldsymbol{\phi}) = -\sum_{c=1}^{4}\sum_{p=1}^{P}\sum_{n=1}^{N}\Big(a\hat{Y}_{c,p,n}\log\phi_{c,p,n}Z_{c,p,n} + (1-\hat{Y}_{c,p,n})\log\left(1-\log\phi_{c,p,n}Z_{c,p,n}\right)\Big) \quad (1)$$

ここで，$a \in \mathbb{R}_+$ は正例に対する重みである．

統計的に妥当なピアノ譜を推定するため，$\boldsymbol{\phi}$ を Gumbel-sigmoid trick により微分可能な形で二値化して，オンセット付きピアノ譜を $\mathbf{Y} =\in \{0,1\}^{4\times P\times N}$ 得る．ただし，$\mathbf{Y}$ にスタックされているピアノ譜とオンセットを表す行列をそれぞれ $\mathbf{S} \in \{0,1\}^{2\times P\times N}$，$\mathbf{O} \in \{0,1\}^{2\times P\times N}$ とする．オンセット付きピアノ譜 $\mathbf{Y}$ の同時発音数，同時発音における度数幅，小節あたりの音符密度のヒストグラム $\mathbf{Q}^{\mathrm{poly}}(\mathbf{S})$，$\mathbf{Q}^{\mathrm{deg}}(\mathbf{S})$，$\mathbf{Q}^{\mathrm{dens}}(\mathbf{O})$ を以下のように微分可能な形で計算する．

$$f_i(\mathbf{S}) = \sum_{n=1}^{N}\mathrm{ReLU}\left(-\sum_{p=1}^{P}S_{p,n}+i-1\right) \quad (2)$$

$$g_i(\mathbf{S}) = \sum_{n=1}^{N}\mathrm{ReLU}\Big\{\min\left(\mathbf{S}_n\odot\mathbf{p}\right) - \max\left(\mathbf{S}_n\odot\mathbf{p}\right)+i-1\Big\} \quad (3)$$

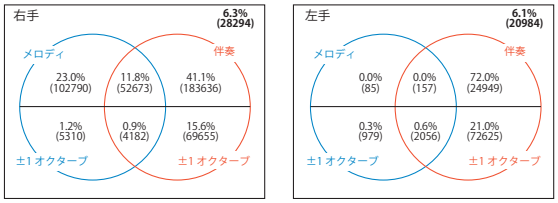

図 1: ピアノ譜に含まれる音符の由来

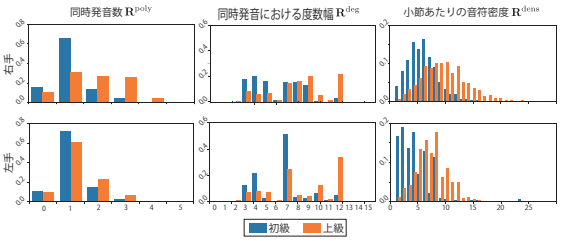

図 2: 収集したピアノ譜から得られた $\mathbf{R}^{\mathrm{poly}}$, $\mathbf{R}^{\mathrm{deg}}$, $\mathbf{R}^{\mathrm{dens}}$

Difficulty-Aware Deep Piano Arrangement of Popular Music: *Moyu Terao, Ryoto Ishizuka, Ryo Nishikimi, and Kazuyoshi Yoshii (Kyoto Univ.)*

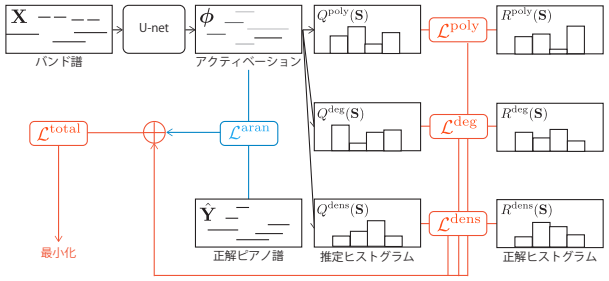

図 3: 難易度別ピアノ譜の推定モデル

$$h_i(\mathbf{O})=\sum_{m=1}^{N/16}\mathrm{ReLU}\left(-\sum_{p=1}^{P}\sum_{n=1}^{16}O_{p,16m+n+i-1}\right) \quad (4)$$

$$Q_i^{\mathrm{poly}}(\mathbf{S}) = f_{i+2}(\mathbf{S}) - 2f_{i+1}(\mathbf{S}) + f_i(\mathbf{S}) \quad (5)$$

$$Q_i^{\mathrm{deg}}(\mathbf{S}) = g_{i+2}(\mathbf{S}) - 2g_{i+1}(\mathbf{S}) + g_i(\mathbf{S}) \quad (6)$$

$$Q_i^{\mathrm{dens}}(\mathbf{O}) = h_{i+2}(\mathbf{O}) - 2h_{i+1}(\mathbf{O}) + h_i(\mathbf{O}) \quad (7)$$

ただし，$i \geq 0$ はヒストグラムの階級，ReLU はランプ関数，$\mathbf{p} \in \mathbb{R}^P$ は要素にピッチ番号 $1 \leq p \leq P$ をもつベクトル，$\odot$ は要素積を表し，$Q_0^{\mathrm{deg}}(\mathbf{S}) = 0$ とする．収集したピアノ譜から得られたヒストグラムをそれぞれ $\mathbf{R}^{\mathrm{poly}}$，$\mathbf{R}^{\mathrm{deg}}$，$\mathbf{R}^{\mathrm{dens}}$ とすると，以下の JS ダイバージェンス D_{JS} を最小化するように U-net を学習する．

$$\mathcal{L}^{\mathrm{poly}} = D_{JS}\left(\mathbf{Q}^{\mathrm{poly}} \parallel \mathbf{R}^{\mathrm{poly}}\right) \quad (8)$$

$$\mathcal{L}^{\mathrm{deg}} = D_{JS}\left(\mathbf{Q}^{\mathrm{deg}} \parallel \mathbf{R}^{\mathrm{deg}}\right) \quad (9)$$

$$\mathcal{L}^{\mathrm{dens}} = D_{JS}\left(\mathbf{Q}^{\mathrm{dens}} \parallel \mathbf{R}^{\mathrm{dens}}\right) \quad (10)$$

最終的に，次式で与えられる総合的な損失関数 $\mathcal{L}^{\mathrm{total}}$ を最小化するように U-net を学習する．

$$\mathcal{L}^{\mathrm{total}} = \mathcal{L}^{\mathrm{aran}} + \mathcal{L}^{\mathrm{poly}} + \mathcal{L}^{\mathrm{deg}} + \mathcal{L}^{\mathrm{dens}} \quad (11)$$

3. 評価実験

3.1 実験条件

実験には，YAMAHA ミュージックデータベースから収集したバンド譜とピアノ譜の MIDI データ 184 ペア（初級 85 ペア，上級 99 ペア）を使用し，4 分割交差検証を用いて評価を行なった．U-Net は 4 層の畳み込み層と 4 層の逆畳み込み層で構成される．全ての層で $\mathbf{L}$ をスタックし，バッチ正則化を行った．カーネルサイズは 4，ストライドは 2，パディングは 1 に設定した．過学習を防ぐため，全ての逆畳み込み層でドロップアウト（p = 0.5）を適用した．ネットワークの最適化には Adam（lr = 10^{-4}）を用い，$a = 4$ とした．検証データに対し，$\mathcal{F}$ 値が最大となる閾値を右手左手それぞれで採用した．評価尺度には $\mathcal{F}$ 値を利用し，推定結果の統計的妥当性を評価するため $\mathcal{L}^{\mathrm{poly}}$，$\mathcal{L}^{\mathrm{deg}}$，$\mathcal{L}^{\mathrm{dens}}$ を計算した．

3.2 実験結果

図 1 は，ピアノ譜に含まれる音符の由来を表す．バンド譜由来の音符は右手左手どちらも約 75 %であるが，上下 1 オクターブの音符を含めると約 94 %を占める．これより，バンド譜に含まれる音符及びそれらを上下 1 オクターブシフトした音符の集合から音符を削除すること

表 1: 評価結果

	$\mathcal{F}$ (右)	$\mathcal{F}$ (左)	$\mathcal{L}^{\mathrm{poly}}$	$\mathcal{L}^{\mathrm{deg}}$	$\mathcal{L}^{\mathrm{dens}}$
$\mathcal{L}^{\mathrm{aran}}$	0.50	0.35	0.008	0.0051	0.0014
$\mathcal{L}^{\mathrm{total}}$	0.40	0.22	0.015	0.0119	0.0036

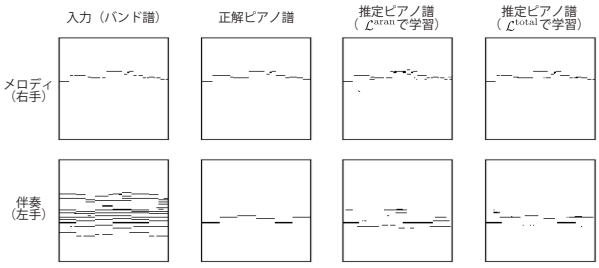

図 4: $\mathcal{L}^{\mathrm{total}}$ で学習した U-net の推定例

でピアノ譜が得られるという仮定の妥当性を確認できる．図 2 は，収集したピアノ譜から得られた難易度別の同時発音数，同時発音における度数幅，小節あたりの音符密度のヒストグラムを表す．各項目において，初級と上級は異なる分布をもつことが確かめられる．

表 1 に，U-net を $\mathcal{L}^{\mathrm{aran}}$ で学習した場合と，$\mathcal{L}^{\mathrm{total}}$ で学習した場合の評価値を示す．ヒストグラムに関する制約を入れることで，右手と左手の $\mathcal{F}$ 値が低下した．この理由として，多様な損失関数を加えたことで学習の収束が遅くなっていることが考えられる．さらに，推定結果の統計的妥当性に関する評価値もすべて増加した．この理由として，正解ピアノ譜に基づく損失関数と JS ダイバージェンスに基づく損失関数のスケールが異なり，それぞれ重みを設定していなかったために，ヒストグラムに関する制約が十分に学習に寄与しなかったことが考えられる．

図 4 を見ると，損失関数に $\mathcal{L}^{\mathrm{total}}$ を用いた場合，右手に関しては余分な音符が減少しており，左手に関しては損失関数に $\mathcal{L}^{\mathrm{aran}}$ を用いた場合に推定し損ねた音符を推定できている．

4. おわりに

本稿では，深層学習を用いて，ピアノ譜の難易度ごとの統計的性質に基づくピアノ編曲の手法を提案した．今後の課題としては，歌声が存在しない部分で主旋律を正しく推定することに加え，移調やテンポに代表されるような難易度を定める要素を統計的に調査し，初級・上級という離散的な難易度の枠組みを超えて連続的に難易度を制御することが挙げられる．

謝辞本研究の一部は，JST ACCEL No. JPM-JAC1602，JSPS 科研費 No. 16H01744, No. 19H04137 の支援を受けた．

参考文献

[1] E. Nakamura *et al.*: “Statistical Piano Reduction Controlling Performance Difficulty,” *APSIPA*, 2018.
[2] Z. Wang *et al.*: “POP909: A Pop-Song Dataset for Music Arrangement Generation,” *ISMIR*, 2020.
[3] O. Ronneberger *et al.*: “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *MICCAI*, 2015.