

ピアノ譜から吹奏楽譜への楽器編成を指定可能な自動編曲

鍋岡 琢渡¹

中村 栄太^{2,3}

寺尾 萌夢²

吉井 和佳^{2,4}

¹ 京都大学 工学部情報学科

² 京都大学 大学院情報学研究所

³ 京都大学 白眉センター

⁴ JST さきがけ

1. はじめに

吹奏楽は幅広い世代に人気の演奏形態である。しかし、吹奏楽譜は一般的に高価であり、入手できたとしても所属するバンドの楽器編成の制約上、そのまま演奏することは難しい場合がある。本研究では、比較的入手が容易なピアノ譜を、任意の楽器編成の吹奏楽譜に自動編曲する問題に取り組む。これまで、大編成の吹奏楽譜の小編成化 [1] や、ピアノ譜のオーケストラ譜への自動編曲 [2] が試みられてきたが、本研究はこれらの課題を一挙に行うものとして位置づけられる。

吹奏楽譜の編曲では、数十にもなる楽器パートの音域や楽器特性に加えて、楽器パート間の同時的・経時的な関係性を考慮する必要があるため、高度な専門性が要求される [3]。この作業を自動化するため、既存の編曲データを用いて、ピアノ譜を吹奏楽譜に変換する深層ニューラルネットワーク (DNN) を教師あり学習することを考える。しかし、吹奏楽譜とピアノ譜の完全な対応が取れたペアデータは入手が難しい。また、既存の吹奏楽譜では楽器毎の使用頻度に偏りがあるため、使用率が少ない楽器の編曲規則が十分に学習されないという懸念がある。

本研究では、自動ピアノ編曲手法 [4] を用いて既存の吹奏楽譜からピアノ譜を生成し、それらを学習用ペアデータとして用いる方法を試みる。そのうえで、U-Net [5] を用いて、楽器パート毎にピアノ譜の各音符を割り振るかどうか決定するマスクを推定する。また、使用頻度が低い楽器パートの品質向上のため、学習・推論時に楽器編成情報を有効利用する手法を提案する。

2. 提案法

2.1 問題設定

提案手法の入力は両手 2 パートからなるピアノ譜であり、出力はユーザーが指定する楽器編成の吹奏楽譜である。ただし、ユーザーは最大楽器編成から任意数のパートを選択することで、楽器編成を指定するものとする。最大楽器編成は、文献 [3] を参考に、音高をもたない打楽器は排除し、よく用いられる 28 楽器 43 パートとした。つまり、指定の楽器編成情報 L は 43 次元のバイナリベクトルで表される (使用楽器を 1 で表す)。

ピアノ譜の両手各パートと吹奏楽譜の各パートは、音高毎の発音時刻と継続時間を表す 2 層のバイナリ行列で表す (行数は MIDI と同じ音高数 128、列数は 16 分音符の $1/3$ を単位とする曲の長さ)。後述の U-Net の入出力には、これらの行列を $4/4$ 拍子で 4 小節に当たる長さ (192) で分割したものを用いる (端数はゼロ埋めする)。

2.2 ペアデータの作成

まず、電子楽譜の公開サイト musescore.com から吹奏楽譜を収集した。入手した楽譜において今回の最大編成 43 パートのみを抽出した楽譜を吹奏楽譜として用いる。次に、これらの吹奏楽譜に対して自動ピアノ編曲法 [4] を適用してピアノ譜を得た。

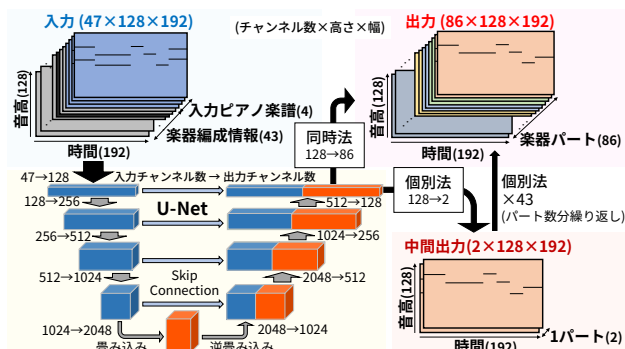


図 1: 提案法: U-Net に基づく自動編曲

この手法は、基本的に合奏曲に含まれる音符の一部を選択することでピアノ譜を生成するため、ピアノ譜の音符は吹奏楽譜の音符の一部となっている。これはピアノ譜の音符を各楽器に割り振ることで吹奏楽譜を生成する、本手法と整合性が良い。以下では、こうして得られたペアデータを学習及び評価データとして用いる。

2.3 ネットワークの設計

ピアノ譜から吹奏楽譜への変換には U-Net [5] を用いる。U-Net は Encoder-Decoder モデルの一種で、畳み込み演算と逆畳み込み演算を繰り返して特徴抽出を多階層で行い、各階層において Encoder 側の特徴を Decoder 側に連結する (図 1)。これにより、音高・時間方向の複数の解像度における性質を捉えた処理が可能になると考えられる。U-Net の出力は、吹奏楽譜 (発音時刻及び継続時間) を表すバイナリ行列に対応する存在確率を表す確率値の行列であり、学習にはクロスエントロピー損失関数を用いる。推論時には、この行列に閾値処理他を行い、吹奏楽譜を得る (後述)。本研究では、U-Net の入出力形式が異なる以下の 3 つの手法を試みる。

単純法では、ピアノ譜 (4 チャンネル) のみを入力として、最大編成の吹奏楽譜 (86 チャンネル) を出力とする。学習時には、各曲で使用されない楽器も含めて損失関数を計算する。推論時には、楽器編成情報 L で指定された楽器パートのみを抽出する。この手法では、楽器編成情報を入力に用いないため、代替楽器の適切な使用等、編成に応じた処理は原理的に難しい。

同時法では、単純法に比べて、入力に楽器編成情報を表す 43 層の行列を加える (図 1)。この各層は最大編成に含まれる各楽器パートに対応し、楽器パートが用いられている場合には音高方向は全て 1 を代入し、用いられていない場合は 0 を代入する。楽器編成情報は、(1) 曲単位での楽器パートの有無を記述する方法と (2) 小節単位で音符の有無を記述する方法の 2 通りを考える。前者では休符区間を含めた音符配置の学習、後者では休符区間は指定された状況で音符配置の効率的な学習が期待される。同時法の出力データ形式及び損失関数は単純法と同じである。この方法では、楽器パート間のバランス等、楽器編成情報に依存する性質を捉えた処理の学習に加えて、学習データ内での楽器パートの使用頻度の偏りに対

表 1: 各提案手法の評価結果 (%)

手法	楽器編成情報	F 値	適合率	再現率
単純法	—	28.8	29.2	30.1
同時法	曲単位	3.8	22.0	2.3
個別法	曲単位	36.8	33.9	41.3
同時法	小節単位	2.6	17.9	1.5
個別法	小節単位	39.2	34.8	45.9

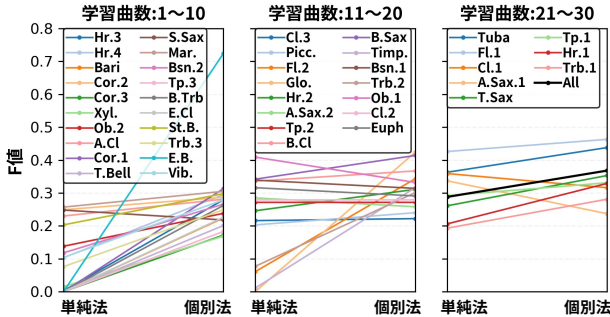


図 2: 単純法と個別法（曲単位）に対する楽器毎の F 値. 各楽器パートが使用されている学習データの曲数に応じて、3つのグループに分けた結果が示されている。

してより頑健な学習が行えると考えられる。

個別法では、出力を1つの楽器パートの楽譜（2チャンネル）として、楽器パート毎に共通のU-Netにより処理を行う。同時法と同様に、入力には楽器編成情報を表す43層の行列を加えるが、処理する楽器パート以外は全てゼロ埋めして用いる。処理する楽器パートについては、ここでも曲単位及び小節単位で音符の有無を記述する2通りの方法を考える。学習時には、各曲に含まれる楽器パート毎に損失関数を計算する。推論時には、指定の楽器編成に含まれる楽器パート毎の出力を結合して吹奏楽譜を生成する。この手法では、楽器編成情報に応じた楽器パート間のバランス調整は難しいが、使用頻度が低い楽器の効率的な学習が期待される。

以上の3手法に共通して、推論時には、各楽器パートの発音時刻の確率推定値を閾値処理をした後、入力ピアノ譜に含まれる音符のみを選択して、音域と単旋律の制約を課すことで推定楽譜を得る。音域は文献[3]等を参考に設定した。また、各時刻で確率推定値が最大となる音高を選択することで単旋律になるようにする。求めた各音符の音長は、入力ピアノ譜を参照して決定する。

3. 評価実験

収集した110曲のペアデータから無作為に選んだ80曲を学習データ、残りの30曲を評価データとして用いた。評価尺度として、全楽器パート及び各楽器パート毎に対する、音符の音高・発音時刻の一致に対するF値、適合率、再現率の平均を用いた。

表1に評価結果を示す。個別法では、単純法に比べていずれの尺度でも精度が向上しており、入力に楽器編成情報を用いる方法の有効性が確認できた。単純法と個別法に対する楽器毎のF値の比較から、個別法では特に使用頻度が低い楽器に対してF値が大きく改善したことがわかる（図2）。また、楽器編成情報の与え方（曲単位または小節単位）の比較により、曲中で楽器パートの使用区間を指定しなくても、精度は大きくは低下しないことがわかった。一方で、最も表現力が高いと期待された



図 3: 出力楽譜例 (Joplin の The Entertainer より)。楽譜は全楽器とも実音で表記している。

同時法では精度が著しく低く、複雑なネットワーク構造により学習効率が低下した可能性がある。同時法に関しては、最適化手法の改良やデータ量の増加による学習方法の改善が課題であることがわかった。

個別法（楽器編成情報は曲単位）による、既存のピアノ譜に対する4楽器7パートでの出力結果を図3に示す。概ね演奏可能な吹奏楽譜になっており、木管の3パートはピアノ譜の右手パート、金管の4パートは主に左手パートを担当しており、各々の楽器群内でパートの自然な順序に沿った和音の声部配置が行われている。これは本手法により各楽器パートの音域だけではなく、楽器特性やパート間の同時的關係を捉えた編成が可能であることを示唆している。一方で、個別法では楽器編成に応じて、楽器パートの役割を適応的に変化させたり、ピアノ譜の音符を漏らさないように各楽器パートに割り当てる等の処理はできないという限界がある。また、Flute 2の2小節目ではメロディーの一部の音のみを割り当てており、不適切な結果になっている。このように、滑らかに演奏可能な編成の生成に必要な、楽器パート内での系列依存性の扱いの改善が必要であることがわかった。

4. おわりに

本研究では、ピアノ編曲手法を用いたペアデータ生成により、既存の吹奏楽譜のみから、ピアノ譜から吹奏楽譜への自動編成を実現できる可能性を示した。また、U-Netを用いて、楽器パート間の関係性を捉えた和音の声部配置を学習できる可能性が示された。

実際の演奏に適した吹奏楽譜の生成に向けて多くの課題が残されている。不自然な跳躍進行やリズム等を含む演奏が困難な音符列の出力を抑制するには、時間的な依存性を考慮できる長短期記憶(LSTM)ネットワークの利用が考えられる。また、幅広い音域を生かした吹奏楽譜を生成するには、ピアノ譜の音符を上下にオクターブ移動させたものを含める方法が考えられる。編曲の専門家による評価や、実際の演奏を通じた出力結果の品質評価にも今後取り組む必要がある。

謝辞 JST PRESTO No. JPMJPR20CB 及び科研費 No. 19H04137, 21H03572, 21K02846, 21K12187, 22H03661 の支援を受けた。

参考文献

- [1] H. Maekawa et al.: Proc. ICMPC, pp. 278–283, 2006.
- [2] L. Crestel et al.: Proc. SMC, pp. 434–442, 2017.
- [3] 柳田孝義: 吹奏楽の編曲入門～楽器の選択からオーケストレーションまで～, 音楽之友社, 2021.
- [4] E. Nakamura et al.: APSIPA Transactions on Signal and Information Processing, vol. 7, no. e13, pp. 1–12, 2018.
- [5] O. Ronneberger et al.: Proc. MICCAI, pp. 234–241, 2015.