

音源分離を用いた作曲家同定

Musical composer classification with music source separation

西大和学園高等学校 彌永琢斗 永坂慶一郎 長井晶哉

背景

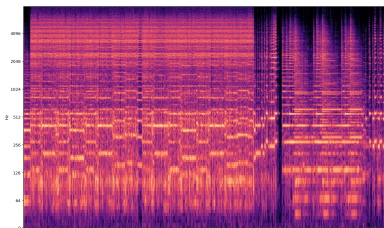
近年、SNS上に投稿される楽曲は増加し、SNS上に投稿される楽曲も増加している。そんな中、多くの検索エンジンではユーザーに多く再生されている音楽の統計情報からおすすめの音楽を提案するアルゴリズムが用いられている。こうしたアルゴリズムでは、仮に好みの作風の作曲家がいても一定数以上再生されていないと「おすすめ」されないという現状がある。そこで我々は似た作曲家を分析する、つまり作曲家らしさを分析することで、真にユーザーの好みにあった音楽を提案できると考え、一定の候補の中から初見の音楽がどの作曲家のものかを同定（作曲家同定）することにした。作曲家の楽曲にある程度共通して現れる特徴に音色、リズム各楽器の音声を分析に用いられれば分析に役立つと考え音源分離技術を用いることとした。

目的

音源分離技術の作曲家分析における有用性の検証

実験

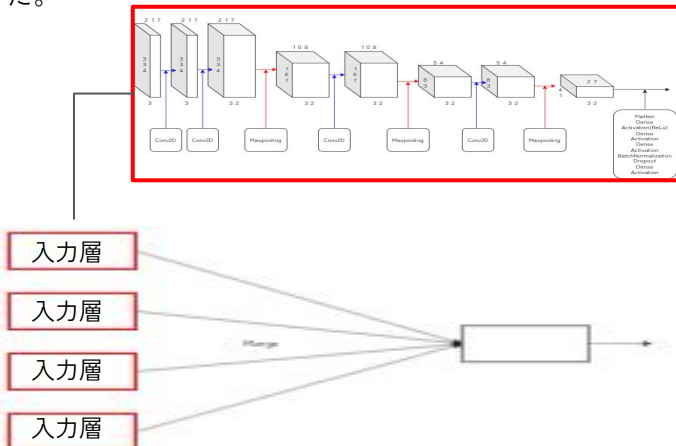
3人の作曲家の楽曲をそれぞれ12曲の計36曲を準備し、訓練用データと評価用データの比が3：1となるように分けた。それらをトリミングやノイズ付与によってデータ拡張した後、メルスペクトログラムに変換し、画像分類問題として扱った。(先行研究)



横軸：時間
縦軸：周波数
色の濃さ：音圧(dB)

FSER: Deep Convolutional Neural Networks for Speech Emotion Recognition Bonaventure F. P. Dossou Jacobs University, Germany Yeno K. S. Gbenou Drexel University, USA

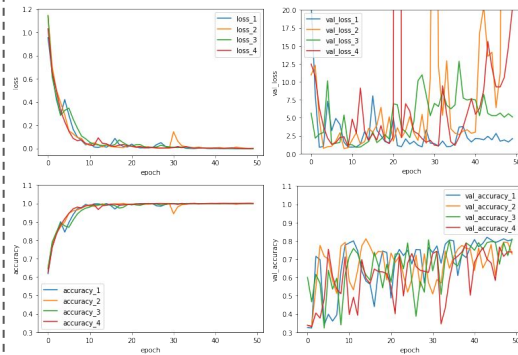
得られたメルスペクトログラムを分類するCNN(畳み込みニューラルネットワーク)を構築した後、そのモデルを並列して入力層を複数にすることで、音源分離によって得られる特徴量を用いられるようにした。こうして分離音源を用いなかった場合と同定精度を比較した。



音源の分離にはDemucsを用いた。これによって分離した音源の中から、どの楽器・音源を使用するかの組み合わせを下図に示す。図中のIndexは結果のグラフの番号と対応する。

index	1	2	3	4
ドラム	無	有	有	有
ベース	無	有	有	無
その他	無	有	無	無

結果・考察



損失のグラフ
左：訓練データ
右：評価データ

精度のグラフ
左：訓練データ
右：評価データ

いずれの場合でも評価用データの判定精度の最大値は同程度であり、損失の最小値についても同程度であった。

また、実験で得られたうち最も損失が小さくなったモデルに関して、それぞれの作曲家の氷菓用データを入力した精度と損失を、平均と母分散を用いて分析した。

		index				
		1	2	3	4	
作曲家	母分散	損失	0.562	0.041	0.285	0.275
		精度	0.031	0.002	0.014	0.015
	平均	損失	0.898	0.720	0.906	1.032
		精度	0.715	0.770	0.757	0.711

これより、今回使用したすべての分離音源を用いた場合に最も損失の母分散が小さくなったことから、音源分離を用いれば、ある特定の作曲家だけが判別しづらいというような偏りを防ぐことができると考えられる。

結論

実験では、分離音源を用いることでは判定精度に有意な差が現れなかったが、分離音源を使用しなかった場合と比べて、判定精度の偏りが小さくなった。このことから、特定の作曲家だけの汎化が進まなくなることを防げるのではないかと考えられる。分離音源は作者同定における有効な特徴量であると考えられる。

謝辞

本研究を進めるにあたり、奈良先端科学技術大学院大学データ駆動形生物学(作村研究室)作村諭一教授には、終始熱心なご指導を頂きました。心から感謝いたします。