

CNN を用いた楽曲の感情分類

Emotional classification of songs using CNN

学籍番号 175720C 氏名 松田理美 指導教員：山田孝治

1 はじめに

音楽を聴いた時、人は楽しい気分になったり、その曲に荘厳さを感じたりする。これは [1] に述べられているように、音楽と感情が深い関係にあるからである。音楽を聴いて何か特定の感情を得たいという際に、その感情を得ることができる楽曲を選ぶことになるが、その為には既にその楽曲を聴いたことがある前提となる。そのため、新しい楽曲を開拓していくことは難しい。

現在の Apple Music などにある楽曲検索システムでは、オススメと表示される楽曲は、最近聴いていた楽曲のアーティストの別の楽曲などが表示されるが、同じアーティストでも曲の雰囲気が違うことはあり得るので、求める楽曲が並ばないことがある。オススメされた楽曲をお気に入り楽曲として登録していくことで、自分の求める楽曲を集めていくことはできるが、その楽曲を一度聴いてから判断する必要があるので、時間がかかってしまう。

[2],[3] では、楽曲の歌詞に注目して、歌詞情報に基づく楽曲聴取者の気分に応じた楽曲推薦システムの実現を目指していた。歌詞から取得した感情単語を元に、楽曲の感情分類を行っていたが、評価実験を行うにあたって、楽曲を実際被験者に聴取させ、5 種類の感情の中から 1 つ以上を選択する形を採っていた。しかし、実際に楽曲を聴くことになるので、楽曲の持つメロディで判断している可能性があった。

そこで本研究では、楽曲のメロディのみに注目し、メロディのみでも感情分類が行えるかを調査するために、メロディ情報に基づく楽曲聴取者の気分に応じた楽曲推薦システムの実現を目指す。

先行研究で行われた、音声ファイルの波形から特徴量を抽出し、CNN を利用して分類する研究を元に、楽曲聴取者が感じた感情と比較し、メロディを情報を元にした教師あり学習による楽曲の分類を行う。

2 先行研究

2.1 楽曲の特徴量

[4] では、楽曲に使われている楽器を判別する為に、メル周波数スペクトログラムを抽出し、離散コサイン変換 (DCT) をかけて、各楽器のメル周波数ケプストラム係数 (MFCC) を抽出し、CNN を利用して分類を行っていた。しかし、DCT

を行うと、特定の周波数帯の特徴量のみを残し、他の周波数帯の特徴量を消すため、楽曲内に使われている特定の楽器を判別することなどは可能だが、楽曲の特徴の大部分が削除されてしまうため、楽曲全体を利用した学習が不可能になってしまう。

下の図 1、図 2 は、Runner Runner の楽曲、「So Obvious」という楽曲の冒頭から 1 分間を、メル周波数スペクトログラムの抽出までを行なった画像と、MFCC の抽出までを行なった画像である。

横軸は時間、縦軸は周波数、色の濃さはその時間のその周波数帯での音圧を表しており、色が薄いほど音圧が高いことを表している。

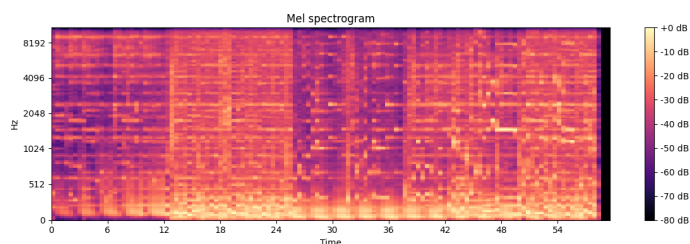


図 1: メル周波数スペクトログラム

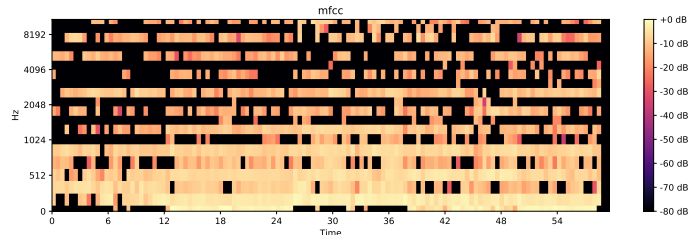


図 2: 図 1 に DCT を行なって抽出した MFCC

この楽曲は、最初の約 12 秒間は落ち着いたメロディーだが、そこから約 12 秒間少し激しい曲調へと変わり、また落ち着いたメロディーへと戻っていく。

図 1 では、最初の約 12 秒間は全体的に低い音圧から始まり、そこからまた約 12 秒間高い音圧になっていることを表しており、その後また低い音圧へと戻っている。

対して、図 2 では、ほぼ常に高い音圧が出続けているような状態を表しており、主に高い周波数帯で音圧がほぼ 0 に近い状態になっている。

本研究では,メル周波数スペクトログラムにDCTを行う手順を省き,メル周波数スペクトログラムをそのままCNNに利用する.

2.2 感情カテゴリー

[2],[3]の研究では,楽曲を被験者に実際に聴取してもらい,5つの感情カテゴリーに分けて,その際に感じた感情を正解ラベルとして学習し,感情の分類を行っていた.

感情カテゴリーは,喜び,安らぎ,悲しみ,昂り,好意の5つである.

3 提案手法

[4]での特徴量の抽出の手法と,[2],[3]での感情分類に利用された感情カテゴリーを利用し,以下の手順で波形分類を行う.

3.1 アンケート

被験者が普段聴取している楽曲からランダムに50曲を選択.それらの楽曲のメロディーのみを実際に聴いてもらい,その楽曲の印象を感情カテゴリーの5つの中から選択してもらう.選択された感情を,その楽曲の正解ラベルとする.

3.2 特徴量の抽出

楽曲の波形から,メル周波数スペクトログラムを抽出する.データ量を増やす為に,1曲を1分ごとに分割し,楽曲のメル周波数スペクトログラムを抽出する.

3.3 モデルの作成

アンケートによって集計したデータセットから,CNNの学習モデルを作成する.1分ごとに分割した50曲分の画像データのうち,40曲分の画像を訓練データとして学習し,10曲分の画像をテストデータとして使用する.

4 今後の課題

本研究を行う為に,被験者へのアンケート調査による楽曲の収集及び正解ラベルの作成,また,学習を行う為のCNNの実装などを早急に行う必要がある.

参考文献

- [1] 音楽と感情, 大串健吾, 2006
- [2] 歌詞情報を用いた楽曲の推薦システム, 具志堅大輝, 山田孝治, 2019
- [3] 歌詞情報を用いた歌詞の楽曲分類手法, 具志堅大輝, 山田孝治, 2020
- [4] DEEP CONVOLUTIONAL NETWORKS ON THE PITCH SPIRAL FOR MUSICAL INSTRUMENT RECONGNITION, Vincent Lostanlen and Carmine-Emanuele Cella,École normale supérieure, PSL Research University, CNRS, Paris, France, 2017
- [5] Deep Learning for Audio Signal Processing, Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuo-yiin Chang, Tara Sainath, 2019
- [6] Experimental Studies of the Elements of Expression in Music, Kate Hevner, 2020
- [7] 古屋瑞生・黄宏軒・川越恭二(2014)「歌詞情報に基づく聴取目的に応じた楽曲推薦システムの提案」, 情報処理学会第76回全国大会
- [8] 聴取目的に応じた音楽推薦のための歌詞からの音楽印象分類方法, 古屋瑞生・黄宏軒・川越恭二, 2015