

## 深層学習を用いた音楽感情認識手法の開発

山名田 恭吾（福井大学工学部） 池田 順哉（福井大学大学院工学研究科）

黒岩 丈介 小高 知宏（福井大学大学院工学研究科）

諏訪 いずみ（仁愛女子短期大学） 白井 治彦（福井大学工学部）

### 1 はじめに

楽曲には怒り、喜び、悲しみ、恐れなどのさまざまな感情が潜在している。人がそれらの感情を主観的に分類することができる。近年では、音楽情報検索の研究が行われている。しかし、一般に公開された共通のデータセットがないことや、感情に関連した音楽特徴量を選定、作成することが困難となっている。本研究では、深層学習を用いて音楽感情認識ができるようにすることを目的とする。先行研究では、感情分類精度が十分高いとは言えなかった。そのため、分類精度を高め、未知の楽曲に対しても高い精度を出せるようにする必要がある。さらに、楽曲に潜在する感情を十分に分類できれば、ユーザが所望している感情の楽曲を提供することが可能となる。以上より、本研究の目的は、高精度な分類を実現可能な楽曲の特徴量の構成法及び深層学習手法を明らかにすることである。

### 2 音楽感情分類方法

#### 2.1 使用する楽曲データ

Panda. R. らが作成したデータセット 4Q audio emotion dataset (Russell's model) (2018) を用いた。Panda. R. らは、Allmusic API から収集した 30s 間、900 曲の楽曲クリップを、ラッセルの感情の円環モデルに従って 4 種類の感情、Happy (Q1), Tense (Q2), Melancholy (Q3), Relaxed (Q4) に分類している。各感情はそれぞれ 225 曲である。その内、800 曲を学習データ、残り 100 曲をテスト用未知データとした。

#### 2.2 VGG16 モデル

今回用いたモデルは VGG16 モデルである。畳み込み層 13 層、プーリング層 5 層、全結合層 3 層、および、出力層から構成される ImageNet2014 のデータセットによって 1000 値分類用に訓練済みモデルが公開されており、これは他のタスクにも広く利用できることが知られている。この VGG16 モデルを使って学習を行う。以下に、VGG16 モデルの図を示す。

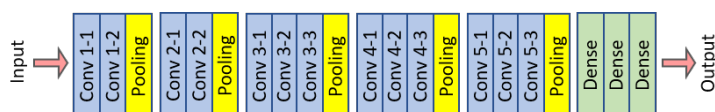


図 1: VGG16 モデル

### 3 音楽感情分類実験

#### 3.1 実験手法

音楽クリップの最小の長さが 28.5s 等だったため、全ての楽曲を 28s とした。その中の、最初の 20s は学習データ、次の 4s を検証データ、残り 4s をテストデータとした。更に、分類精度向上のため、全ての区間データを 2s ずつ二分した。その結果、1 感情あたり、学習データは 2800 個となった。学習データをそのまま入力しても分類精度が低いため、本研究では 2s 区間データをスペクトログラムに変換した。具体的には、区間データを窓サイズ 512, オーバーラップ 0.5s で SFTP (短時間フーリエ変換) を行うことで求めた。全ての区間データにこの処理を施し、これを VGG16 モデルの入力とし、学習を行った。その際のバッチサイズは 32, エポック数は 1500 とした。感情分類実験では、2s の各区間データに対して感情分類を行い、その楽曲全体での感情を多数決で決定した。

#### 3.2 実験結果

現在、VGG16 の学習を実行していて、分類精度はまだ出せていない。分類精度等については発表当日に詳細を話す予定である。

### 4 考察

音楽推薦サービスを提供するにはまだ十分な精度を得られていない。このようなサービスを提供するためにさらに精度を上げていく方法を検討することと、未知の楽曲に対しての分類精度がなぜ低いのかを考察して、当日発表する予定である。

### 5 参考文献

[1] 池田順哉 黒岩丈介 小高知宏 白井治彦 諏訪いずみ, "深層学習を用いた音楽感情分類における特徴量作成と分類精度評価", 2020 年度電気・情報関係学会北陸支部連合大会, F2-4-11