

MIDI データと自己組織化マップを組み合わせた演奏曲調分類の研究

T120425 垣田 詩鶴香

指導教員 三好 力 教授

1. はじめに

一般的に人が何らかの用途で楽曲を選ぶ際は、歌手・作曲家・ジャンル・曲調等の特徴を利用していると考えられる。曲調による分類に注目すると例えば、街中を歩いていると周辺の店舗から音楽が聞こえてくるが、お店の雰囲気に合った曲調の楽曲が流れていることが多いように感じる。また、曲調分類は楽器を練習するための練習曲を選ぶ際にも有用である。曲調分類が可能になると、知らない曲からも選ぶことが出来、より好みの曲や作曲家を見つけることの出来る機会が増える。このように曲調は選曲の上で重要度が高いと考えられるが、歌手や作曲家で選別することに比べて、ジャンルや曲調は感性による分類であるため、それらで選別するのは容易ではない。そこで本研究では自己組織化マップ(Self-Organizing Map:以下 SOM)と MIDI 形式のデータ(以下 MIDI データ)を用いた曲調分類を検討した。音楽におけるジャンルや曲調は定義付けが曖昧であるため、本研究では”ポップス”や”ロック”などの音楽様式のことをジャンル、”勇ましい”、”華やか”などの楽曲の調子や雰囲気のことを曲調と定義する。

SOM は教師なし学習の手法の一つであり、汎化能力が高く結果もわかりやすいため、容易に曲調で分類するために適している。

2. MIDI (Musical Instrument Digital Interface)

MIDI データはヘッダ部とトラックデータ部からなり、トラックデータ部はシステムコマンド、音高、音の長さ、音の強さのデータ列からなる。情報単位は 16 進数 2 桁からなり、音高、長さ、強さは 0x00~0x7F の値を、システムコマンドは 0x80~0xFF の値をとる。MIDI データは音響波形情報と比べてデータ量が少なく、処理が容易であることから本研究では音響波形情報ではなく MIDI データを用いる。

3. 実験 1

MIDI データから得られる、全ての楽器の音高、長さ、強さを総合した音の流れと、前述のジャンルの間に関係性が見られるかどうかを確認する実験を行った。MIDI データのヘッダはトラックデータと比べてデータ量が遥かに小さいため、無視できる程度の誤差である。また、トラックデータのシステムコマンドは 0x80~0xFF の範囲で表されるため、0x00~0x7F のデータを用いることによってシステムコマンドを除去する。メロディーとジャンルの間に関係性がある場合、それぞれのジャンルによって音高と長さ、強さの特徴があると考えられるが、音高、長さ、強さは独立した情報であるために特徴は異なり、3 つの分布を足し合わせて 1 つの分布として見てもジャンルごとの特徴が失われないため、情報を分離せずに用いて分類を行ってもある程度の性能で分類が可能であると考えた。実験には様々なジャンルと曲調の楽曲 102 曲を用意し、それらを入力ベクトル、出力は 10×10 ユニットとして SOM を作

成した。なお、今回用いた楽曲は RWC 研究用音楽データベースから得た 101 曲 30 ジャンルを用いた。

3.1. 実験 1 の結果

30 分類中 23 分類、101 楽曲中 61 楽曲が同ジャンルの他楽曲と近い位置に配置されたことが確認された。同ジャンルの楽曲と同じユニットに出力されたのは 101 楽曲中 22 楽曲である。ヒストグラム作成には約 1.128 秒の処理時間を要した。

4. 実験 2

さらなる計算量削減を考える。音高、長さ、強さはそれぞれ 16 進数 2 桁で表現されるため実験 1 では 128 次元のベクトルを用いた。しかし、多くの楽曲はその一部分の値のみを用いている事に着目し、桁を無視して MIDI データを 16 進数 1 桁の集合とみなして読み込み、16 次元のヒストグラムをデータベクトルとしても楽曲の特徴を保持しているのではないかと考えた。これにより、MIDI データからのヒストグラムの作成時の処理を低減することができる。さらに、128 次元の 8 分の 1 の 16 次元データを用いて SOM の学習を行うため、学習時間が短縮される。

4.1. 実験 2 の結果

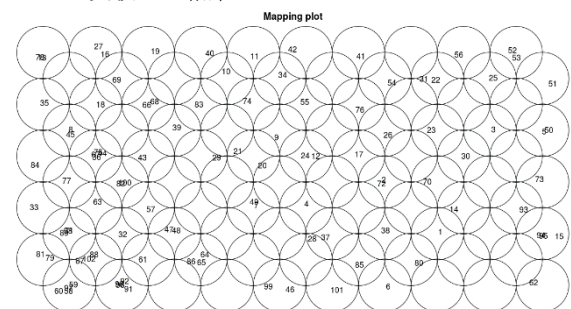


図 1: 実験 2 の出力結果

図より 30 分類中 22 分類、101 楽曲中 63 楽曲が同ジャンルの他楽曲と近い位置に配置されたことが確認された。同ジャンルの楽曲と同じユニットに出力されたのは 101 楽曲中 20 楽曲である。ヒストグラム作成には約 1.058 秒の処理時間を要した。

5. 考察

128 次元は約 60.4%、16 次元は約 62.4%の楽曲が分類できることが確認された。ヒストグラムの作成時間は約 6.28%減少した。精度の観点では、同じユニットに出力された数の多い 128 次元の方が優れているが、分類数の観点では 16 次元の方が優れていると言える。これは 128 次元から 16 次元へ変化した際に音楽と関係のない 0x80~0xFF のデータが増加したことにより、汎用性が高くなる代わりに精度が減少したと推測される。しかし、精度・分類数ともにほとんど差のないこと、本研究の目的においては精度よりも汎用性を重視することを考慮すると、16 次元の方が適している。過半数の楽曲が分類可能なことから、MIDI データから得られるメロディーとジャンルとの間に何らかの関係性があることが示唆される。