

## 情報処理学会第 78 回全国大会 における研究発表

垣田 詩鶴香

Shizuka KAKITA

情報メディア学専攻修士課程 1年

### 1. はじめに

私は、2016年3月10日から12日に開催された第78回情報処理学会全国大会のオーラルセッションに参加し、「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」というタイトルで発表を行いました。本報告では発表内容や発表を通して感じたことを報告します。

### 2. 研究内容

#### 2.1 研究背景

一般的に人が何らかの用途で楽曲を選ぶ際は、歌手・作曲者・ジャンル・曲調等の特徴を利用していると考えられる。楽曲の曲調による分類に注目すると、たとえば、街中を歩いているとあちこちのお店から音楽が聞こえてくるが、お店の雰囲気合った曲調の音楽を流し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、歌手や作曲者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 とする) と MIDI 形式のデータ (以下 MIDI データとする) 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とは、定義付けが曖昧である。本研究では、“クラシック”や“ロック”、“ポップス”などの音楽の様式や形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、容易に曲調で選別するために適している。

#### 2.2 MIDI (Musical Instrument Digital Interface)

MIDI データは音響波形情報 (以下波形情報とする) と比べてデータ量が少なく、また、MIDI に対応した電子楽器を演奏してその演奏情報を計算機に入力するこの形式で保存されるため、処理が容易である。波形情報から音高や音の長さを得るためにはフーリエ変換などの処理が必要になるため、全ての処理を行うには多くの時間が必要になり、計算量が大きくなる。また、演奏曲の入力データも波形情報を用いるよりも MIDI 楽器の演奏情報から得た MIDI データを用いた方がフーリエ変換等の処理が不要であるため処理が容易である。これらの理由から、本研究では波形情報ではなく MIDI データを用いる。

#### 2.3 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 データにはヘッダ部とトラックデータ部が存在するが、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、トラックデータには音高、長さ、強さ、システムメッセージの4種類のデータが存在するが、音高、長さ、強さは  $0x00 \sim 0x7F$  の範囲、システムコマンドは  $0x80 \sim 0xFF$  の範囲で表されるため、 $0x00 \sim 0x7F$  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何かしらの関係性がある場合、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、音高と長さ、強さの分布を足し合わせて一つの分布として見てもジャンルごとの特徴が失われなないと考えた。よっ

て、MIDI データの音高と長さ、強さの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として SOM を作成した。出力は  $10 \times 10$  ユニットで行った。なお、今回は 102 曲 31 ジャンルの楽曲を用いた。

この実験では、MIDI データのバイナリデータから得た  $0 \times 00 \sim 0 \times 7F$  の範囲の 16 進数 2 桁の各文字のヒストグラムを作成した。得られた 128 次元のベクトルを 1 曲分のデータとした。

出力結果より、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。

## 2.4 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。

MIDI データのバイナリデータから得た 16 進数 1 桁の各文字のヒストグラムを用いて 16 次元のベクトルを作成して 1 曲分のデータとした。なお全ての文字を用いたため、実験 1 では除いた  $0 \times 80 \sim 0 \times FF$  の範囲も含まれている。

出力結果より、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。

## 2.5 考察

実験 1 では約 60.4%、実験 2 では約 62.4% の楽曲が分類できることが確認された。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 次元へと変化した際に音楽と関係のない  $0 \times 80 \sim 0 \times FF$  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を考慮すると 16 次元で扱っても問題はないと考える。本研究での目的においては精度よりも汎用性を重視するため、今回の目的には 16 次元の方が適している。分類システムとして利用可能なレベルではないが、過半数の楽曲が分類可能ということは、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。

## 3. 発表について

今回初めて学外で発表を行い、自身の知識不足を痛感しました。この経験を今後の学生生活、並びに社会人生活に役立てていきたいと思っております。

## 4. おわりに

研究や今回の学会発表において多大なご指導を頂きました三好力教授に厚く御礼申し上げます。