

## DTM での音色検索を対象とした機械学習アルゴリズムの提案 Proposal of Machine Learning Algorithm for the Searching Timbre Information on DTM

齋藤 創<sup>†</sup>  
Hajime Saito

大場 みち子<sup>‡</sup>  
Michiko Oba

### 1. はじめに

近年、PC を用いて作曲を行う DTM(DeskTop Music) が発展している。これは YouTube、ニコニコ動画等の消費者生成系メディア(CGM: Consumer Generated Media)の急速な広がりが必要因として考えられる<sup>[1]</sup>。DTM での作曲におけるプロセスの一つとして、音色づくりと呼ばれる作業がある。音色づくりとは、DTM 作曲者が作成する楽曲に適した音色を作成するため、ソフトウェア音源のパラメータを調整する作業のことである。

本研究の目的は、音色づくりをするための情報(音色づくり情報)をより効率よく検索可能にすることである。一方、音色には厳密な定義が存在せず、音色を表現するための基準となる指標が存在しない<sup>[2]</sup>。DTM 作曲者が音色を表現する際、同じ音色でも作曲者ごとに表現に差異があるため、求める音色づくり情報を効率よく検索できないという問題がある。これに対して、DTM 作曲者ごとに効率よく音色を検索できるようにする機械学習アルゴリズムを用いた検索フィルタの作成を行なっている。

本論文では、検索フィルタを作成するための機械学習アルゴリズムの改善と検証について述べる。

### 2. 先行研究

先行研究<sup>[3]</sup>では、DTM 作曲者の音色づくりの効率化を図るため、TASS(Timbre Adjustment Support System: 音色づくり支援システム)の開発を行った。TASS では DTM 作曲者が音色づくりのメタデータを用いて、作成したい音色と類似の音色の検索を可能にした。検索した音色の再利用を可能にすることで、音色づくりの効率化を図った。一方、作曲者ごとの表現の差異により、作曲者ごとに適切な音色づくり情報を検索できないという課題により、TASS では DTM 作曲者それぞれに適切な音色づくり情報を提示できていないという課題がある。

#### 2.1 課題解決アプローチ

先行研究では、作曲者が求める音色づくり情報を適切に提示するため、登録されたオーディオファイルの音響特徴量と作曲者が用いた雰囲気を表すデータを用いた機械学習

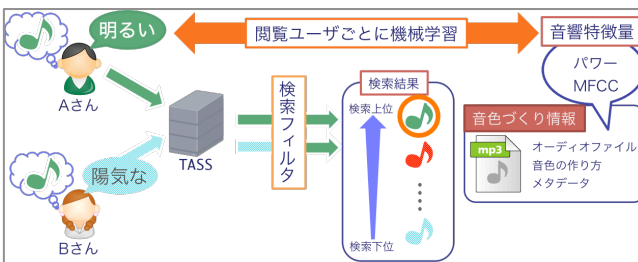


図 1 先行研究課題解決アプローチ

<sup>†</sup> 公立はこだて未来大学大学院

<sup>‡</sup> 公立はこだて未来大学

を行い、DTM 作曲者ごとに適用した検索フィルタを作成するというアプローチを取った。先行研究の課題解決アプローチを図にしたものが図 1 である。

音響特徴量として MFCC<sup>[4]</sup>を利用し、雰囲気の種類は HEVNER の形容詞クラスタ<sup>[5]</sup>を利用した。MCFF とは、人の知覚特性を反映させた音色に関する音響特徴量であり、音声認識にも用いられている。HEVNER の形容詞クラスタは、音響研究で用いられる感性を示す形容詞を C1~C8 に分類したクラスタである。検索フィルタを作成するための機械学習アルゴリズムとして回帰式(1)を用いた。

$x \in (1,2,3, \dots, 8)$ :形容詞のクラスタ  $y \in (0,1,2, \dots, 12)$ :mfcc 次元数

$f_x$ :検索フィルタ

$m_y$ :閲覧情報の mfcc

$C_x$ :各クラスタの親データ

$\vec{c} = (\vec{c}_1, \vec{c}_2, \vec{c}_3, \dots, \vec{c}_8)$

$I_x$ :入力された雰囲気の数値

$P_x$ :過去の検索フィルタ  
(初期値は $\vec{c}_x$ : $C_x$ のスカラー)

$$f_x = \begin{cases} P_x + (\bar{m} - P_x) \times \frac{\text{abs}(\bar{c}_x - \bar{m})}{\bar{c}_x} \times \frac{I_x}{100} & \text{if } \text{abs}(\bar{m}) < \bar{c}_x \\ P_x + (\bar{m} - P_x) \times \frac{I_x}{100} & \text{if } \text{abs}(\bar{m}) \geq \bar{c}_x \end{cases} \dots (1)$$

### 2.2 検証・結果

#### 2.2.1 検証

昨年度の研究では回帰式(1)を反映させた Microsoft Excel を用いて、回帰式(1)が学習を重ねる毎に学習値が収束するかどうかを検証した。検証では被験者 3 名に 20 種類の音色を提示し、提示された音色が HEVNER の形容詞クラスタにどの程度合致するのかを 0 から 100 の値で入力させた。

#### 2.2.2 結果

検証で得られた回帰式(1)の推移の一例を図 2 に示す。横軸は学習回数、縦軸は回帰式によって得られた学習値を示し、C1~C8 は HEVNER の形容詞クラスタを示している。

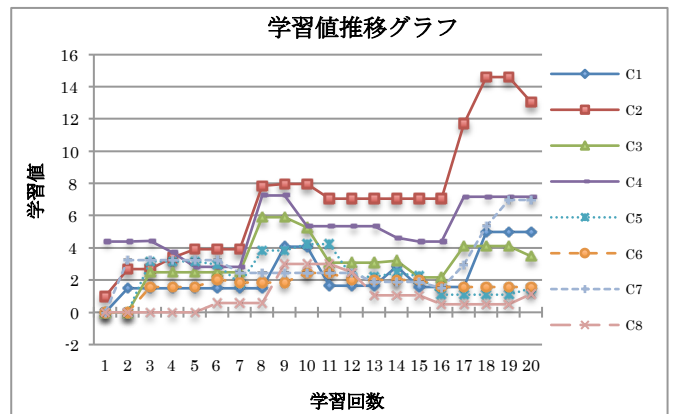


図 2 学習値の推移グラフ

図 2 から学習値が正の値に偏っており、負の値に推移していない事がわかる。これは入力値 $I_x$ が $0 \leq I_x \leq 100$ であり、回帰式(1)では $I_x/100$ を変化量に掛けあわせているため、変化量が小さくなりすぎてしまっていることが原因として

考えられる。図 2 で提示した被験者以外の学習値の推移も同様に学習値が正の値に偏っており、回帰式(1)では適切な学習ができていないことが判明した。

### 3. 研究課題・課題解決アプローチ

本研究での検索フィルタは正の値だけではなく負の値を取ることができる必要がある。一方、現状の回帰式(1)では値が正の値に偏り、負の値への学習がされていないため、検索フィルタを作成するための機械学習アルゴリズムとして適切ではないという課題がある。

回帰式(1)では、変化量に $I_x/100$ を掛けあわせたことにより、値の変化が小さくなり、負の値への学習ができないという問題が発生したと考えられる。そこで、値の変化が大きくなるように、変化量に $(1+I_x/100)$ を掛けあわせることで変化量が大きくなるように回帰式(2)を作成した。本研究では回帰式(2)と既存の機械学習アルゴリズム

である SVM(Support Vector Machine)を適用した場合を比較する。比較した結果から最も適当なアルゴリズムを TASS に適用し、TASS v2 を構築する。SVM は判別分析を行う機械学習であるので、表 1 のように判定するレベルの数も検証する必要がある。レベル数は大きいほど雰囲気の詳細に表すことができる。

$$f_x = \begin{cases} P_x & \text{if } I_x = 0 \\ P_x + (\bar{m} - P_x) \times \frac{I_x - \bar{m}}{C_x} \times \left(1 + \frac{I_x}{100}\right) & \text{if } I_x \neq 0 \cap \bar{m} < C_x \dots (2) \\ P_x + (\bar{m} - P_x) \times \left(1 + \frac{I_x}{100}\right) & \text{if } I_x \neq 0 \cap \bar{m} \geq C_x \end{cases}$$

## 4. 検証・結果・考察

### 4.1 検証

本研究では、修正した回帰式(2)と SVM を用いた場合の比較を行った。検証には先行研究で利用した 20 種類の音色を用いた。20 種類の内 15 種類の音色を学習値とし、残りの 5 種類の音色を検証値として用いた。検証時には HEVNER の形容詞クラスごとに一致しているかどうかを判定した。SVM の検証では、R 言語の kernlab<sup>[6]</sup>というパッケージの ksvm 関数をデフォルトで利用し、レベル数毎の検証を行い、レベル数が大きく一致率が高くなる場合を調査した。回帰式の検証では、学習した値を検索フィルタとして用いて、検証値を正しく検索できるか判定した。

### 4.2 結果

表 2 は各被験者および被験者全体に対して、レベル数ごとの SVM および回帰式(2)の一致率をまとめたものである。表 2 から、SVM を用いた場合、2 分割した際の全体の一致率 58.3%が一番高い値となり、3 分割した場合と 4 分割した場合の一致率は共に 51.7%となった。一方、回帰式(2)の一致率は、全体の一致率は 55.6%となった。特に被験者 A の場合 84.2%という高い一致率を確認することができたが、被験者 B,C 共に一致率は低くなった。

表 2 機械学習一致率

機械学習 アルゴリズム		一致率(%)				
		被験者			全体	
		A	B	C		
SVM	レベル数	2	52.5	70.0	52.5	58.3
		3	52.5	62.6	40.0	51.7
		4	52.5	60.0	42.5	51.7
		5	52.5	57.5	33.3	47.5
回帰式(2)		84.2	41.2	44.4	55.6	

表 1 レベル数分類例

レベル数	例
2	明るくない 明るい
3	明るくない 明るい
4	明るくない 明るい
5	明るくない 明るい

### 4.3 考察

SVM を用いた場合、レベル数が大きくなるほど一致率が低くなることが確認できた。また、SVM でのレベル数が大きく、全体の一致率が 50%以上であるレベル数が 4 の場合が一番適しているということが判明した。

回帰式(2)を用いた場合、全体での一致率は 55.6%となり、レベル数が 4 の SVM を用いた場合の一致率より大きくなるということが判明した。一方、被験者 A では 84.2%と非常に高い一致率を示したが、被験者 B,C は共に一致率が 50%を下回った。以上の結果より、現状の回帰式(2)ではユーザごとの検索結果の一致率が大きく異なってしまい、検索フィルタを作成するための機械学習アルゴリズムとして適切ではないという事が判明した。

以上の結果により、本研究では機械学習アルゴリズムの第一候補としてレベル数が 4 の SVM を利用する。一方、今回の検証では学習サンプルが少なかった等の要因により、どの結果も高い一致率を得ることはできなかった。今後は引き続き特徴量のスケール等による SVM の精度向上ならびに回帰式の修正を行う必要がある。

## 5. おわりに

本研究では、DTM 作曲者が求める音色を効率よく検索するための検索フィルタの作成を行なった。検索フィルタを作成するための機械学習アルゴリズムを選定するための検証を行った。結果、機械学習アルゴリズムの第一候補としてレベル数が 4 の SVM が適当であると判明した。

今後は 4.3 節の考察を元に、交差検証(leave-one-out cross-validation)等を利用し、学習サンプルを増やした状態での検証を行ない、引き続き機械学習アルゴリズムの精度向上を行う。最終的には、機械学習アルゴリズムを TASS に反映させた TASS v2 を構築する。構築した TASS v2 をリリースし、ユーザからのアンケートおよびアクセスログを取得し、検索フィルタおよびシステムの有効性の検証を行う。

### 参考文献

- [1] 斎藤明, “サービス・ドミナント・ロジックにおける価値共創概念と市場創造: 「初音ミク」という市場創造を中心として”, 情報文化学会誌, vol. 21, no. 1, pp. 2936, Aug. 2014.
- [2] Jason Leung, “Timbre Recognition”, MUMT, vol. 621, no. Assignment 4, pp. 0-5, Mar. 2014.
- [3] 齋藤創, 大場みち子, “メタデータを利用した機械学習による DTM(DeskTop Music)での音色づくりの効率化”, 研究報告音楽情報科学 (MUS), vol. 2016-MUS-110, no. 15, pp. 1-6, Feb. 2016.
- [4] aidiary, “メル周波数ケプストラム係数 (MFCC) - 人工知能に関する断創録”, 人工知能に関する断創録, 25-Feb-2012. [Online]. Available: <http://aidiary.hatenablog.com/entry/20120225/1330179868>. [Accessed: 13-Jun-2016].
- [5] K. HEVNER, “experimental studies of the elements of expression in music”, *American Journal of Psychology*, vol. 48, pp. 246-268, 1936.
- [6] A. Karatzoglou, “CRAN - Package kernlab.” [Online]. Available: <https://cran.r-project.org/web/packages/kernlab/>. [Accessed: 12-Jun-2016].