

# kmeans を用いた音楽フレーズのクラスタリング

桑島洋<sup>†</sup> 池田剛<sup>†</sup> 但馬康宏<sup>†</sup> 小谷善行<sup>†</sup>

<sup>†</sup>東京農工大学工学部情報コミュニケーション工学科

## 1 序論

本研究では、kmeans を用いて音楽フレーズをクラスタリングする手法を提案する。本手法において、音楽フレーズとは主旋律を小節線で機械的に分割したものであり、すべて長さは一小節である。また、本研究では旋律のみに着目し、リズム、音の強さなどの情報は利用しない。

本研究の意義であるが、まず、音楽フレーズのクラスタリングとして、コードやスケールではない新たな分類が、自動的に得られることは重要である。例えば、ある楽曲のコードを得ることは熟練した専門家でないとい難しいが、本手法を用いれば、機械的にフレーズの種類ができる。また、文字列のクラスタリングとして本手法の有効性が確認されれば、文字列の汎用性から、他の分野への応用範囲が広く、非常に有意義である。

これまで、Thom[1] などによって音楽フレーズのクラスタリング手法が研究されている。Thom の方法では、小節単位に分割されたフレーズに対して各音程の頻度に基づいてクラスタリングした。志田 [2] による HMM によるコード列の学習も、小節のクラスタリングと考えることができる。

本研究では、和音、音階が未知の状態各小節をクラスタリングすることで、新たな分類を自動的に取得することを目的としている。音楽フレーズは文字列として扱うので、文字列のクラスタリングと考えることもできる。

最も一般的なクラスタリングアルゴリズムの一つである kmeans は、絶対座標を持たない文字列に対してクラスタの重心が計算できない問題がある。本研究はこの問題を解決し、絶対座標を持たない音楽フレーズのクラスタリングを kmeans を用いて行う手法を提案する。なお、3.1 以外の実験には RWC 研究用音楽データベース (ポピュラー音楽)[3] を利用した。

## 2 kmeans を用いたクラスタリング手法

### 2.1 アルゴリズム

本手法は、DP マッチングによって計算された編集距離を用いてクラスタリングを行う。DP マッチングとは、文字列の相違度である編集距離を計算するアルゴリズムである。

絶対座標が与えられている一般的な場合、クラスタと要素の距離はクラスタの重心と要素の距離、と考えるのが自然であるが、本研究の場合、クラスタの重心が分からない。そこで、要素  $e$  とクラスタ  $c$  の距離  $d(e, c)$  は、 $e$  と  $c$  に属する全ての要素の編集距離の平均であ

る、とする関数を新たに提案した。式 (1) に示す。ただし、 $dp(e_1, e_2)$  は DP マッチングによって計算される  $e_1$  と  $e_2$  の編集距離で、 $|c|$  は  $c$  に含まれる要素数である。

$$d(e, c) = \sum_{element \in c} dp(element, e) / |c| \quad (1)$$

kmeans において、要素とクラスタの距離を式 (1) によって計算すると、文字列のクラスタリングができる。この方法は、結局、 $n$  個の要素を任意の要素間の編集距離を満たすように  $n$  次元空間に配置し、その空間内で kmeans を適用した場合と同義である。本手法の疑似コードを図 1 に示す。

```

begin
  initialize assign;
  repeat
    for each  $e$  in element
       $assign_e := \operatorname{argmin}_c d(e, c)$ 
    where
       $d(e, c) = \sum_{element \in c} dp(element, e) / |c|$ ;
  until assign not change;
  return assign;
end

```

図 1: クラスタリング手法疑似コード

### 2.2 初期化

一般的な kmeans アルゴリズムでは、適当な要素や値でクラスタの中心を簡単に初期化することができるが、本アルゴリズムでは初期値としてクラスタリングを与える必要がある。一番簡単な方法として、ランダムに所属クラスタを決め、初期クラスタリングを与えることが考えられるが、この方法では、全てのクラスタの中心がほぼ一致してしまうと考えられる。

また、次のような初期化も考えられる。

- 1 要素からなる  $k$  個のクラスタをランダムに作成
2.  $initAssign(e) := \operatorname{argmin}_c d(e, c)$  w.r.t.  $e$

この方法では、十分に中心が分散したクラスタリングが得られると考えられるが、kmeans において適当な  $k$  個の要素の座標をクラスタ重心の初期値としたのと同義である。

他に、 $0 < m < n/k$  個の要素 ( $n$  は要素数) からなる  $k$  個のクラスタを作るとい、上記二つの方法の中間的な初期化方法も考えられるが、本研究では後者の初期化方法を採用した。

a Clustering Method for Music Phrases using kmeans  
 H. KUWAJIMA<sup>†</sup>, T. IKEDA<sup>†</sup>, Y. TAJIMA<sup>†</sup> and Y. KOTANI<sup>†</sup>  
<sup>†</sup>Dep. of Comp., Info. & Comm. Sci., Fac. of Eng., Tokyo Univ. of Agri. & Tech. (TUAT)  
 {kuwajima, ikeda}@fairy.ei.tuat.ac.jp,  
 {ytajima, kotani}@cc.tuat.ac.jp

### 3 実験と考察

#### 3.1 合成フレーズのクラスタリング

ccccccc のようなフレーズを 3 種類用意し、それぞれのフレーズから編集距離が非常に近い caccccc のようなフレーズを 8 個ずつ、合計 24 個人工的に生成した。合成されたフレーズで  $k = 2, 3, 4$  についてクラスタリングの実験を行った。

$k = 2$  のとき、片方のクラスタに二種類のフレーズが全て入った。 $k = 3$  のとき、適切にクラスタリングされた。 $k = 4$  のとき、ある一種類のフレーズが二つのクラスタに分割された。

この結果より、 $k$  が適切な場合、クラスタリングが妥当に行われていることがわかった。

#### 3.2 クラスタに含まれるフレーズ例

$k = 7$  で 98 曲 (7639 小節) の学習実験を行った。ある二曲 (曲 A, 曲 B) に着目し、各クラスタからそれぞれの曲に含まれる異なるフレーズを二つずつ取り出した結果を表 1 に示す。

表 1: フレーズ例

	曲 A のフレーズ		曲 B のフレーズ	
$C_1$	eddegg-e	fgb-ce-db-c	afdcd	ffdfgag
$C_2$	ggabb-bb-b	bagdgg	aaaagfg	gagagfd
$C_3$	ce-db-cgb-	ce-db-cgg	cdec	cdedfed
$C_4$	ddddeb	bddddd	dededcag	ddecgce
$C_5$	b-b-cb-e-gg	b-b-b-cb-e-gg	ab-ab-ab-ab-	-
$C_6$	gb-ce-db-c	dddd	gggacdc	fffedc
$C_7$	gb-	dc	fg	da

表 1 から、同じ曲の中では非常に似通ったフレーズが同一クラスタに入っているが、曲が変われば同じクラスタでもフレーズは大幅に変わることがわかる。このことから、3.1 で観察された、 $k$  が小さい場合に別のフレーズが同じクラスタに入ってしまうのと同じ状態が起きているものと考えられる。

#### 3.3 各音程の出現頻度の比較

$k = 8$  で 98 曲 (7639 小節) の学習実験を行い、各クラスタに含まれるフレーズの音程の出現頻度 (ド, レ, ミなどの音がそれぞれのクラスタで何回現れたか) を比較した。図 2 にヒストグラムを示す。それぞれの Figure は一つのクラスタに対応し、その順番は意味を持たない。図 2 によると、それぞれのクラスタの音程頻度は特

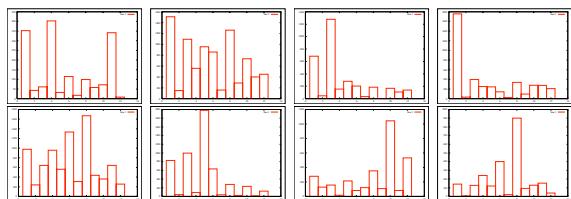


図 2: 各クラスタの音程の頻度の比較 (縦軸: 頻度, 横軸: 音程)

徴を持っていることが分かる。クラスタリングを全くのランダムで行った場合、ヒストグラムは全て同じような形になることが考えられる。この実験により、各ク

ラスタに含まれるフレーズがランダムではなく、それぞれ傾向を持っていることが確かめられた。クラスタリングは音程の頻度に関する情報は使用せずに行ったが、音程の頻度についても特徴を抽出できたとと言える。

#### 3.4 クラスタリングと楽譜の対応

$k = 7$  で 98 曲 (7639 小節) の学習を行った。その中からある曲の一部分を取りだし、クラスタリング結果と楽譜の対応を図 3 に示す。



図 3: クラスタリングと楽譜と対応

図 3 から、明らかに同じフレーズ、似ているフレーズが同じクラスタに属していることがわかる。例えば、8 小節目は 4 小節目を少し変形したものと考えられるが、同じクラスタに入っており、クラスタリングが成功している。また、テーマとサビを繰り返す歌謡曲や、あるモチーフを少しずつ変形していくクラシックの楽曲のように、音楽は繰り返しの構造を持つことは重要であるが、この例では、5-6-0 というシーケンスが 2 回観察され、繰り返し構造が学習できたことがわかる。

これより、音楽特有の特徴もうまく抽出できていると考えられる。

### 4 結論

音程の頻度と、繰り返し構造取得の観点から、音楽の特徴を学習することに成功した。また、本研究では  $k$  の値を適切に与える方法について考察はしなかったが、音楽フレーズという性質上、完璧にクラスタリングする必要はなく、 $k$  の値によって細かくクラスタリングしたり、大まかにクラスタリングを行っている、と考えることもできる。

以上から、本手法が音楽のフレーズをクラスタリングするのに有効であることが確認された。

### 参考文献

- [1] Thom, B. 2000. Unsupervised Learning and Interactive Jazz/Blues Improvisation. In *Proceedings of the AAAI2000*. AAAIPress.
- [2] 志田裕樹, 乾伸雄, 小谷善行. 2002. 隠れマルコフモデルを用いたコード進行のベースラインからの推定 in 第 64 回情報処理学会全国大会講演論文集.
- [3] Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R. 2002. RWC Music Database: Popular, Classical, and Jazz Music Databases in *Proc. IS-MIR, 2002*, pp.287-288.