

# 文書検索手法を適用した音楽データの類似検索手法

大野 和久<sup>†</sup> 鈴木 優<sup>‡</sup> 川越 恭二<sup>‡</sup>

<sup>†</sup>立命館大学 理工学部

<sup>‡</sup>立命館大学 情報理工学部

## 1 はじめに

近年、携帯型音楽プレーヤーの普及に伴い、音楽データの扱いが身近なものとなっている。そのため、個人が扱う音楽データ数が膨大化し、データ検索の需要が高まっている。その需要を満たすものとして、曲の一部から、その曲の情報を知るための検索システム [1] や、曲の一部及び全体を入力し、その曲と類似している曲を出力するシステム [2] が公開されている。

従来の類似検索手法では、問合せ波形と時間窓ごとに区切った検索対象波形を比較している。しかし、区切られた波形の出現頻度は考慮しておらず、問合せ波形と検索対象波形相互の特徴量の比較だけを行っているため、類似検索結果が正確であるとは考えにくい。一方で、音符を文字に見立て、音符列に対して数字列のパターンを割り当てることによって、文字列検索手法を用いる研究がされている [3]。しかし、この手法では、類似検索のための適切な値が取れないため、類似検索の実現は困難であると考えられている。

これらの従来手法に対し、本研究では、楽曲の構造が文書の構造と似ていることに着目した。文書は単語によって構成され、各単語が文書の特徴付けている。文書における TF-IDF 法は、この各単語の出現頻度を用いることによって、文書内容を考慮した類似検索を行い、精度を高めている。同様に、楽曲は波形によって構成され、各波形が楽曲を特徴付けている。そこで、文書における TF-IDF 法と同様に、各波形の出現頻度を用いることによって、楽曲の内容を考慮した類似検索を行うことができると考えた。

本稿では、文書検索における索引語への重み付けとして用いられる TF-IDF 法を用いることによって、分割した波形が波形全体をどのように特徴付けているかを考慮して音楽データの類似検索を行い、その有効性

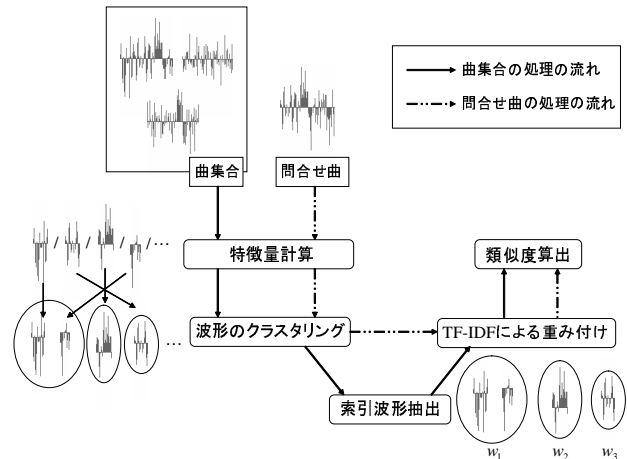


図 1: 類似検索処理の流れ

について述べる。

## 2 楽曲の内容を考慮した音楽データの類似検索

類似検索処理の流れを図 1 に示す。TF-IDF による重み付けのため、問合せ曲と検索対象曲集合から、文書における索引語にあたる要素を抽出する。本研究では、この要素を索引波形と呼ぶ。まず、検索対象曲集合の各曲に対して、特徴量の計算と波形のクラスタリングを行うことによって索引波形を抽出する。そして、各検索対象曲の索引波形と、問合せ曲に現れる索引波形に対して、TF-IDF による重み付けを行う。最後に、各検索対象曲の索引波形と問合せ曲の索引波形に与えた重みを用いて、類似度算出を行う。

### 2.1 索引波形抽出

索引波形抽出のために、曲を複数の波形に分割し、分割した各波形に対して索引付けを行う。本稿では、 $t$  秒ごとに波形の分割を行う。ところが、文字の種類が有限であり、1 文字という明確な区切りが存在する文字列に対し、波形は 1 文字という単位が無く、分割した波形の種類数が無限にあるため、等間隔に分割するだけでは、分割要素数だけ索引が作成される。そこで、分

A Similarity Search for Music Data Based on Text Retrieval Method

Kazuhiro OONO<sup>†</sup>, Yu SUZUKI<sup>‡</sup> and Kyoji KAWAGOE<sup>‡</sup>

<sup>†</sup>College of Science and Engineering, Ritsumeikan University

<sup>‡</sup>College of Information Science and Engineering, Ritsumeikan University

<sup>†</sup>oono@coms.ics.ritsumei.ac.jp,

<sup>‡</sup>{yusuzuki, kawagoe}@is.ritsumei.ac.jp

割した各波形に対し、各波形の特徴量によるクラスタリングを行い、形成された各クラスを一つの索引波形とする。

各波形のクラスタリングのために、周波数分布と平均音量を特徴量として用いる。周波数分布は、分割した波形にFFT(Fast Fourier Transform)を行って得られた周波数領域を、 $n$ 個の等間隔な区間に分割し、各区間において振幅値を合計したものである。平均音量は、分割した波形の平均振幅値である。

検索対象曲をそれぞれ $S_1, S_2, \dots, S_r$ とし、検索対象曲集合を $S = \{S_1, S_2, \dots, S_r\}$ とする。 $S$ に対して索引抽出を行い、 $m$ 個の索引波形 $w_1, w_2, \dots, w_m$ を抽出する。

## 2.2 TF-IDF による重み付け

$S_j$ における $w_i$ の重み $s_{ij}$ を求める式を以下に示す。

$$s_{ij} = t_{ij} \frac{F_i}{p_i} \quad (1)$$

ただし、 $S_j$ における $w_i$ の出現頻度を $t_{ij}$ 、 $S$ における $w_i$ の出現頻度を $F_i$ 、 $w_i$ を含む曲数を $p_i$ とする。

問合せ曲における $w_i$ の重み $q_i$ は、問合せ曲における $w_i$ の出現頻度とする。

## 2.3 類似度算出

検索対象曲の重みベクトル $s_j = [s_{1j} \ s_{2j} \ \dots \ s_{mj}]^t$ と問合せ曲の重みベクトル $q = [q_1 \ q_2 \ \dots \ q_m]^t$ 間のコサイン尺度を求める式を、以下に示す。

$$\cos(s_j, q) = \frac{\sum_{i=1}^m s_{ij} q_i}{\sqrt{\sum_{i=1}^m s_{ij}^2} \sqrt{\sum_{i=1}^m q_i^2}} \quad (2)$$

ここで得られたコサイン尺度を、問合せ曲と検索対象曲との類似度とする。

## 3 評価実験

提案手法による類似検索の精度を検証するため、11点平均適合率算出による評価実験を行った。

表 1: 本手法の 11 点平均適合率

	$t = 0.05$	$t = 0.1$	$t = 0.2$	$t = 0.3$
$n = 4$	0.24	0.25	0.33	0.24
$n = 5$	0.28	0.28	0.29	0.28
$n = 6$	0.29	0.25	0.32	0.28

## 3.1 実験条件

検索対象曲集合には、クラシックやロック、J-POPといった複数の種類の曲を30曲用い、問合せ曲には、松谷卓氏の「TAKUMI/匠」を用いた。なお、問合せ曲は、検索対象曲集合には含まれていない。音楽データの形式は、いずれもPCMである。図1に示す処理の流れに基づいて、 $n, t$ の値の組合せを変えながら、問合せ曲とすべての検索対象曲の間で類似度を算出した。そして、人手によって作成した正解曲集合をもとに、11点平均適合率の算出を行った。なお、正解曲は、検索対象曲30曲中の5曲とした。

## 3.2 実験結果と考察

実験によって得られた11点平均適合率を、表1に示す。本実験では、 $n = 4, \dots, 6, t = 0.05, 0.1, 0.2, 0.3$ の範囲内で値の変更を行った。この場合、 $n = 4, t = 0.2$ の組合せの結果が最良であることがわかる。また、11点平均適合率は、 $n$ 及び $t$ の値に比例せず、いずれの $n$ の値についても、 $t = 0.2$ の場合が最良であることもわかる。このことより、11点平均適合率の変動に法則性は無く、索引波形の分割単位の大小にかかわらず、適切な $t$ の値が存在すると考えられる。

## 4 おわりに

本研究では、文書検索において索引語への重み付けとして用いられるTF-IDF法を、PCM形式の音楽データの類似検索に適用し、その有効性を確かめた。11点平均適合率を用いた評価実験を行った結果によると、約3割程度の精度であることがわかった。

今後の検討事項として、特徴量及びクラスタリング分類手法の改善、重み付けの検討が挙げられる。また、従来手法との精度の比較を行い、有効性について更に検討することも必要であると考えられる。これらの検討事項を解消しつつ、データベースの利用を導入することにより、システムの高速化についても考える。

## 参考文献

- [1] KDDI. 聴かせて検索. [http://www.au.kddi.com/ezweb/service/kikasete\\_kensaku/](http://www.au.kddi.com/ezweb/service/kikasete_kensaku/).
- [2] OWL Multimedia. OWL. <http://www.owlmm.com/>.
- [3] Costas S. Iliopoulos, Masahiro Kurokawa. String matching with gaps for musical melodic recognition. *Proceedings of the Prague Stringology Conference '02*, pp. 55–64, 2002.