

楽曲全体における特徴量の傾向に基づいた類似検索手法

A Music Retrieval Method based on Distribution of Feature Segments

大野 和久 ♡
川越 恭二 ▲

鈴木 優 ◆

Kazuhisa OONO
Kyoji KAWAGOE

Yu SUZUKI

本論文では、楽曲中における曲調の変化を考慮した楽曲類似検索手法について提案を行う。一般的に、文書類似検索では、文書に現れる単語の出現頻度に基づき文書間類似度を算出する。この手法に対し、我々は、楽曲の特徴を表す特徴部分を楽曲から抽出し、その部分を単語として扱うことにより、ベクトル検索モデルといった文書類似検索手法を用いて楽曲の類似検索を行うことが可能となることを考える。文書類似検索に用いられるベクトル検索モデルを利用することにより、楽曲に対する特徴部分の重要度を算出することが可能となり、曲調の変化を考慮した類似検索を行うことができる。評価実験の結果、提案手法による類似検索は既存手法による類似検索より高い精度を持つことを確認した。

In this paper, we propose a music retrieval method based on the segment distributions of melodies in the music. Generally, when a text retrieval system retrieves textual documents, the system calculates retrieval status values based on the frequencies of terms in the documents. We assume that when the retrieval target music can divide into meaningful segments and the retrieval system treat the meaningful segments as terms, the system can retrieve music using textual information retrieval techniques, such as vector space retrieval model. Using vector retrieval model in text retrieval, we can consider importance of feature segments in music, and we can retrieve music based on the changes of melodies. In our experimental evaluation, we can confirm that our proposed system has better accuracy than the baseline method.

♡ 非会員 立命館大学大学院理工学研究科
oono@coms.ics.ritsumei.ac.jp

◆ 正会員 立命館大学情報理工学部
suzuki@ics.ritsumei.ac.jp

▲ 正会員 立命館大学情報理工学部
kawagoe@is.ritsumei.ac.jp

1. はじめに

日常の中で、店内やTV番組、映画のBGMなどによって、楽曲を受動的に聞く機会がある。このとき視聴者は、その楽曲に興味を持ち、興味を持った楽曲と類似している楽曲も聴きたいといった要望を抱く場合が考えられる。ところがこの場合、視聴者が楽曲に対して持つ印象は曖昧であり、視聴者の判断によって膨大な楽曲の情報から類似している楽曲を探すことができない。そのため、楽曲の類似検索に対する需要が高まっているといえる。一般的に類似している楽曲の定義として、楽曲に含まれるフレーズの一部が他の楽曲に含まれるフレーズの一部と類似しているという定義がある。この定義に対し本研究では、楽曲全体を通して印象の変化が類似しているという定義を考える。その理由として、視聴者が楽曲に興味を持つ際、楽曲に含まれる一部の印象だけに興味を持つのではなく、楽曲全体を通しての印象の変化や展開に興味を持つと考えるためである。

既存の楽曲類似検索手法 [1],[2] では、楽曲の印象を特徴付ける特徴部分を楽曲から抽出し、その特徴部分を類似比較することにより楽曲の類似検索を行っている。このとき、楽曲に含まれる特徴部分の一部だけをを用いた類似比較を行っており、一つの楽曲全体に対して類似比較を行っていない。ところが、楽曲には様々な印象が含まれており、様々な種類の特徴部分が含まれていることが考えられる。このことより、特徴部分の一部を用いた類似比較をするだけでは楽曲全体を通しての印象の変化を考慮することができない。そのため、楽曲全体を通しての印象の変化を考慮した類似検索を行うことができないといった問題が考えられる。

そこで本研究では、楽曲に含まれる特徴部分の傾向を考慮することによって、楽曲全体の印象を考慮した類似検索を行う手法の提案を行う。特徴部分の傾向を抽出する方法として、文書における類似検索に用いられるベクトル検索モデルの考え方を用いる。この考え方は、楽曲の構造と文書の構造との対応関係に着目したに基づいている。文書は様々な種類の単語によって構成されており、文書における類似検索では、文書の特徴付けている単語の出現傾向に基づいて類似検索を行っている。一方で、楽曲は様々な種類の特徴部分によって構成されている。そこで、楽曲の類似検索において、楽曲を特徴付けている部分を文書における単語として扱うことを考える。その結果、特徴部分の出現傾向を考慮することにより、楽曲全体の印象を考慮した類似検索を行うことが可能となると考える。

2. 関連研究

本研究に関連する研究として、楽曲の類似検索に N -gram 手法を適用した研究がある。Shyamala ら [3] は、楽曲に含まれる音符の列を文字列へ変換し、その文字列に対して N -gram 手法を適用している。音符の列を文字列としてみなしていることから、楽曲の構造を文書の構造に置き換えるという点において、本研究と類似している。ただし、索引付けの手法として N -gram 手法を用いていることから、単語とみなす区間を固定長としてとらえており、特徴部分によるフレーズの長さを考慮していない。

そのため、特徴部分として意味のまとまりが無い。この点において、特徴部分を可変長としてとらえることによって特徴部分としての意味を考慮する本研究と異なっている。

一方、音声学に基づいて、楽曲構造と文書構造が対応関係にあることに着目した研究がある [4]。この研究では、楽曲を可変の区間ごとに区切り、その区間を一つの単語とみなしている。可変の区間を楽曲の特徴としてとらえるという点において、可変の特徴部分を抽出する本研究と類似している。しかし、楽曲に含まれるすべての区間を単語としてとらえており、その単語を連続させることによって楽曲の構成を表している。この点において、複数の楽曲間における類似部分から楽曲に含まれる特徴部分を抽出する本研究と異なっている。

3. 特徴部分の傾向を考慮した類似検索

既存の類似検索手法では、楽曲の一部から得られた特徴部分を用いることによって類似検索が行われている。ところが、楽曲には様々な種類の特徴部分が含まれるため、一部分の特徴部分だけでは楽曲全体を考慮した類似検索を行うことができない。

例として、問合せに楽曲 A を用い、検索対象楽曲群に楽曲 B が含まれるときの類似検索を考える。ここで、楽曲 A は静かな印象を持つ楽曲とする。また、楽曲 B は静かな印象と明るい印象を持つ楽曲とする。楽曲 A と楽曲 B は楽曲全体を通して類似していないとする。このとき、楽曲に含まれる一部分の特徴部分を用いた類似比較による類似検索システムでは、楽曲 A に含まれる静かな印象の特徴部分と楽曲 B に含まれる静かな印象の特徴部分が類似していると判断する。そして、楽曲 A と楽曲 B の一部分が類似していることにより、楽曲 A と楽曲 B は類似していると判断する。しかし、楽曲 A における曲調の変化と楽曲 B における曲調の変化が類似しているかどうかについては考慮していないため、楽曲 A と楽曲 B が楽曲全体を通して類似しているかどうかを正しく判断することができない。

そこで本研究では、楽曲全体の特徴を考慮するために、楽曲全体における特徴部分の傾向を考慮する。このことにより、楽曲がどのような特徴部分を持っているかという判断を行うだけでなく、楽曲にはどのように特徴部分が含まれるかという判断を行う。例として、前段落において述べた例を用いる。楽曲 A は静かな印象を持つ楽曲であるため、静かな印象を表す特徴部分だけが現れる。一方、楽曲 B は静かな印象を表す特徴部分に加え、明るい印象を表す特徴部分も現れる。ここで、楽曲に含まれる特徴部分の出現傾向を考慮すると、楽曲 A における特徴部分の出現傾向と楽曲 B における特徴部分の出現傾向は異なっているといえる。このことにより、楽曲 A における曲調の変化と楽曲 B における曲調の変化が類似していないと判断し、楽曲 A と楽曲 B が楽曲全体を通して類似していないことを判断することができる。

3.1 処理の概要

提案手法による類似検索システムの概要を図 1 に示す。本システムでは、問合せに含まれる特徴部分の傾向と各検索対象楽曲に含まれる特徴部分の傾向をベクトル検索モデルによって類似比較

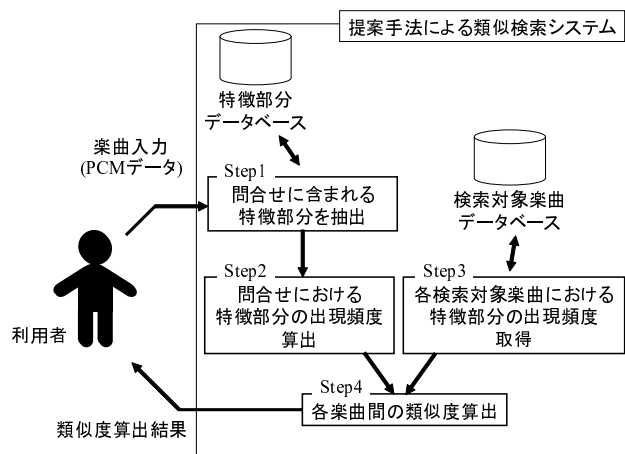


図 1 提案手法による類似検索システムの概要

Fig. 1 Overview of our music retrieval system

する。このことによって、楽曲全体における特徴部分の傾向を考慮した類似検索を行う。

類似検索のために、各検索対象楽曲に含まれる特徴部分の傾向を前処理によって取得しておく。そして、特徴部分データベースおよび検索対象楽曲データベースの構築を行う。特徴部分データベースには、各検索対象楽曲から抽出した複数の特徴部分を格納する。また、検索対象楽曲データベースには、各検索対象楽曲における各特徴部分の出現頻度を格納する。

利用者は、類似比較したい楽曲の PCM データを問合せとして入力する。システムは、Step1 として問合せと特徴部分データベースを照合し、問合せに出現する特徴部分を抽出する。そして、Step2 では、問合せにおける各特徴部分の出現頻度を算出する。一方、Step3 として各検索対象楽曲における各特徴部分の出現頻度を、検索対象楽曲データベースから取得する。問合せおよび各検索対象楽曲における各特徴部分の出現頻度を取得すると、Step4 として各出現頻度を用いて各楽曲間の類似度を算出する。最後に、利用者に対し類似度算出結果を出力する。

3.2 特徴量の類似部分抽出

楽曲に出現する各特徴部分は一定の長さを持たず、楽曲の構成によって異なる長さを持つことが考えられる。例えば、一つの楽曲中に短いフレーズと長いフレーズが混在する場合である。この場合、各特徴部分の長さを一定とすると、フレーズが無視されることとなる。そこで、各特徴部分の長さを可変長としてとらえるために、各特徴部分の長さを考慮した特徴部分の取得を行う。

ただし、本研究で扱う PCM データは時系列データである。文書において単語の可変長をとらえる場合、文字という最小単位があり、単語辞書を用いた形態素解析を行うことが可能である。一方、時系列データは連続したデータであるため、文字のような単位が無く、特徴部分の辞書は存在しない。そのため、時系列データを区分化し、可変長の特徴部分を抽出することは困難である。

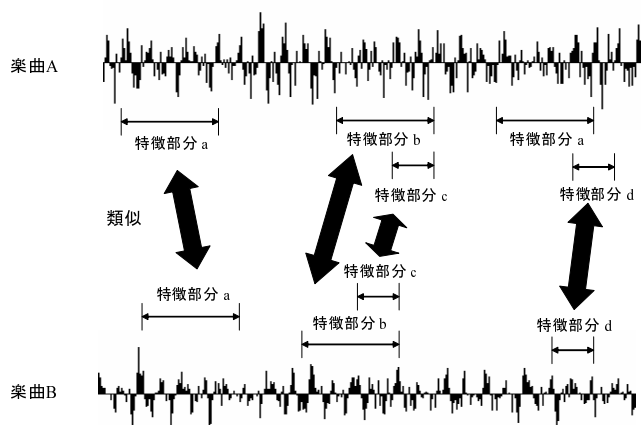


図2 類似部分による特徴部分の定義

Fig. 2 Definition of feature segments using similarity segments

ここで、楽曲を未知の言語によって書かれた文書と考える。未知の言語によって書かれた文書を解釈する場合、単純な解釈方法として、多くの文字列中において類似する文字列のパターンを抽出し、そのパターンを単語として定義する方法がある。そこで、本研究ではこの方法に基づいて、楽曲における連続した時系列データを文書における文字列とみなし、複数の楽曲間において時系列データから得られる特徴量が類似する部分を抽出する。そして、その類似部分を一つの特徴部分と定義することにより、ある特徴部分を可変長としてとらえる。

本研究における類似部分の概念図を図2に示す。図2では楽曲Aと楽曲B間の類似部分を抽出し、それぞれの類似部分を特徴部分としている。ここで、特徴部分cのように他の特徴部分に含まれている場合や、特徴部分dのように他の特徴部分と重なる場合についても考慮する。類似部分を抽出することによって、波形における最小単位を考慮する必要がなくなり、波形の連続性を保ったまま、可変長の特徴部分を取得することが可能となる。

3.2.1 楽曲の特徴量

楽曲の特徴量として、時系列データから得られる二つの値を用いる。一つは、類似部分に対するFFT(Fast Fourier Transform)値の算出によって得た周波数分布である。周波数分布とは周波数領域を n 等分し、各区間において振幅値を総和した値である。周波数分布を特徴量とすることによって、音色や音質の違いを考慮することができる。考える。 n 個の周波数分布 F_1, F_2, \dots, F_n を得るとき、 h 番目($h = 1, 2, \dots, n$)の区間における F_h の算出式を(1)式に示す。ここで、周波数領域中の周波数点 i における振幅値を f_i とし、一つの区間における周波数点の数を N とする。

$$F_h = \sum_{i=(h-1)N+1}^{hN} f_i \quad (1)$$

もう一つの特徴量は、類似部分の平均音量である。平均音量は、

類似部分において、各標本化点の振幅値を平均化した値である。平均音量を特徴量とすることによって、音の強弱による印象の違いを考慮することができる。平均音量 A の算出式を(2)式に示す。ここで、類似部分において j 番目($j = 1, 2, \dots, L$)の標本化点における振幅値を a_j とし、類似部分の長さを L とする。

$$A = \frac{\sum_{j=1}^L a_j}{L} \quad (2)$$

3.2.2 楽曲の特徴部分を取得する範囲

3.2.1節で述べた特徴量の取得には、 L が必要である。しかし、類似部分が未抽出の段階では、類似部分の長さが未知であるため、類似部分全体に対してFFT値を算出することができない。そこで、時系列データを細かなフレームへ分割し、特徴量が類似するフレームを連結することにより、類似部分を取得する。

類似部分を取得する処理として、まず、特徴量を取得する各楽曲において、一つのフレームに対して特徴量の取得を行う。そして、各楽曲のフレームにおいて、各特徴量を要素としたベクトルを生成し、ベクトル間の類似度を算出する。このとき、算出した類似度が閾値 θ を超えていれば、後続のフレームへ処理を移す。後続のフレームへ処理を移す中で、 θ を下回るフレームが現れると、処理を終了する。そして、類似部分として、処理を開始したフレームから処理を終了したフレームの一つ前にあたるフレームまでを類似部分とする。処理の開始時に θ を下回った場合は、後続のフレームから類似部分の抽出処理を改めて開始する。

なお、フレームの長さは $25[msec]$ とし、フレームをずらす周期は $10[msec]$ とする。これらの値は音声認識で広く用いられている値であることから、本研究においてもこれらの値を用いる。また、ベクトル間類似度の定義としてコサイン尺度を用いた。

3.2.3 楽曲から取得した特徴部分の分類

楽曲から取得した特徴部分は、平均音量と周波数分布の組合せによって識別している。しかし、平均音量と周波数分布は様々な値を持っていることから、多数の組合せが存在することが考えられる。そのため、同じ組合せの特徴部分が出現しにくくなり、特徴部分の出現頻度を考慮することが困難となる。そこで、組合せが類似している特徴部分を同一の特徴部分とみなし、特徴部分の種類を限定するための分類を行う。各特徴部分間の比較は、類似部分の長さ L が等しい特徴部分間において行われる。比較処理として、各特徴部分に含まれる各フレーム間の類似度を算出する。このとき、算出したすべてのフレーム間類似度が閾値を超えていれば同一の特徴部分であるとみなし、すべての類似度が閾値を超えていなければ、別の特徴部分とする。なお、このときの閾値は3.2.2節と同様に、閾値 θ を用いた。

3.3 ベクトル検索モデルの適用

本研究では、文書類検索モデルの一つであるベクトル検索モデル[5]において利用されているTF-IDF法のアルゴリズムを、楽曲の類似検索に適用する。文書が様々な単語の出現頻度によって特徴付けられていることと同様に、楽曲も様々な特徴部分の出現頻度によって特徴付けられており、その出現頻度を考慮するこ

とによって、特徴部分の傾向を考慮した類似検索が可能となると考えるためである。本研究における TF の定義として、一つの楽曲中に一つの特徴部分が出現する頻度を用い、IDF の定義として、一つの特徴部分が出現する楽曲数の逆数を用いる。

そして、これらの定義により TF-IDF 値を算出し、各楽曲における各特徴部分の重みを決める。検索対象楽曲における特徴部分の重みを算出する式を、(3) 式に示す。ここで、 r 個の検索対象楽曲 $S = \{S_1, S_2, \dots, S_r\}$ から m 個の特徴部分 w_1, w_2, \dots, w_m が抽出されるとし、 l 番目 ($l = 1, 2, \dots, r$) の楽曲 S_l における k 番目 ($k = 1, 2, \dots, m$) の特徴部分 w_k の出現頻度を t_{kl} 、 w_k を含む楽曲数を p_k 、 S_l における w_k の重みを s_{kl} とする。

$$s_{kl} = t_{kl} \log \frac{r}{p_k} \quad (3)$$

また、問合せ中の特徴部分 w_k の重み q_k には、問合せにおける w_k の出現頻度を用いた。

最後に、各重みを要素とした特徴ベクトルを生成し、各特徴ベクトル間類似度を算出する。本研究では、コサイン尺度を特徴ベクトル間類似度の定義とする。検索対象楽曲の特徴ベクトル $s_l = [s_{1l} \ s_{2l} \ \dots \ s_{ml}]$ と問合せの特徴ベクトル $q = [q_1 \ q_2 \ \dots \ q_m]$ 間のコサイン尺度を算出する式を、(4) 式に示す。

$$\cos(s_l, q) = \frac{\sum_{k=1}^m s_{kl} q_k}{\sqrt{\sum_{k=1}^m s_{kl}^2} \sqrt{\sum_{k=1}^m q_k^2}} \quad (4)$$

コサイン尺度は、一つの楽曲全体における各特徴部分の出現頻度を考慮することにより、一つの楽曲全体における各特徴部分の重要性を比較した結果である。このことより、楽曲全体における特徴部分の傾向を考慮した楽曲間類似度の算出が可能となる。

4. 予備実験

評価実験のために、閾値の決定、不要な連結フレーム数の除去、検索対象楽曲数と特徴部分数との相関関係について予備実験を行った。

予備実験では、検索対象楽曲数を 100 曲とし、検索対象楽曲群には、クラシックやヘビーメタルといった様々な印象の曲を数曲ずつ含めた。楽曲データはモノラルの WAVE 形式とし、標準化周波数を 44100Hz、量子化ビット数を 16bit とした。

4.1 閾値の決定

本研究では、楽曲から特徴部分を抽出するために、楽曲から抽出したフレーム間の類似比較を行う。このとき、フレーム間の類似性を判断するために、閾値 θ を用いる。閾値により、楽曲から抽出される特徴部分の数は変化する。そのため、閾値は、フレーム間の類似度に対して適切に設定される必要があると考える。そこで、閾値の変化と、類似すると判断されるフレーム数の変化との相関関係について確かめ、閾値の決定を行う。

閾値の変化による抽出フレーム数の変化について、図 3 に示す。抽出フレーム数は、類似していると判断されたフレームの数を示す。本予備実験では、フレーム間の組合せ数を 73818 個と

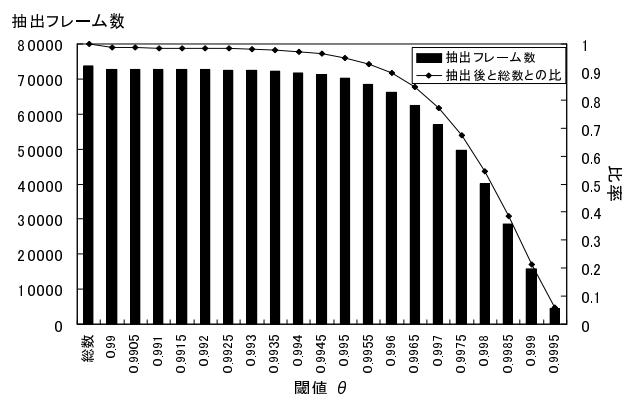


図 3 閾値 θ の変化による抽出フレーム数の変化

Fig. 3 Change of extracted frame number based on change of threshold θ

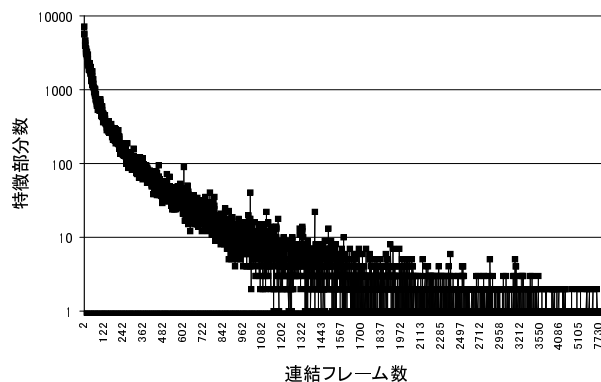


図 4 連結フレーム数と特徴部分数との相関関係

Fig. 4 Correlation connected frame number with feature segment number

し、各フレーム間に対して類似度の算出を行った。 $\theta = 0.99$ のとき、抽出フレーム数は 72842 個となり、総数に対して 98% であった。そこで、閾値の変化は $\theta = 0.99, 0.9905, \dots, 0.999$ とした。閾値が 0.9965 までは総数に対する比率が 80% 以上であり、緩やかな曲線を示している。ところが、閾値が 0.997 を超えると、急な傾きの曲線となる。この算出結果に基づき、本研究では三つの閾値を決定する。一つ目は、閾値の変化において最も値の小さい $\theta = 0.99$ とする。二つ目は、総数に対する比率が 67% となる $\theta = 0.9975$ とする。三つ目は、総数に対する比率が 38% となる $\theta = 0.9985$ とする。ここで決定した三種類の閾値を評価実験において用いることにより、抽出するフレームの数はどの程度が適切であるかということについて検証を行う。

4.2 連結フレーム数と特徴部分数との相関関係

楽曲に含まれる特徴部分は、固定長のフレームを連結することによって抽出されている。このとき、連結するフレームの数には

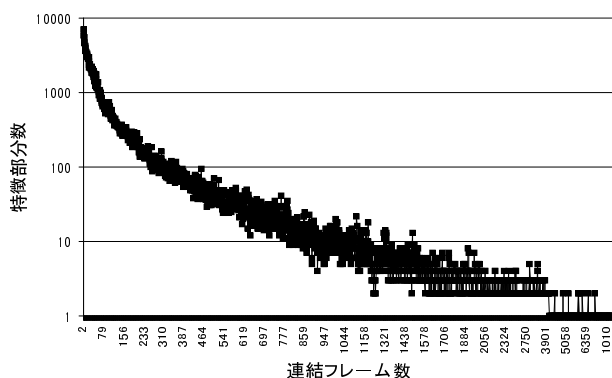


図5 適切ではない連結フレーム数を除去した相関関係

Fig. 5 Correlation except useless connected frame number

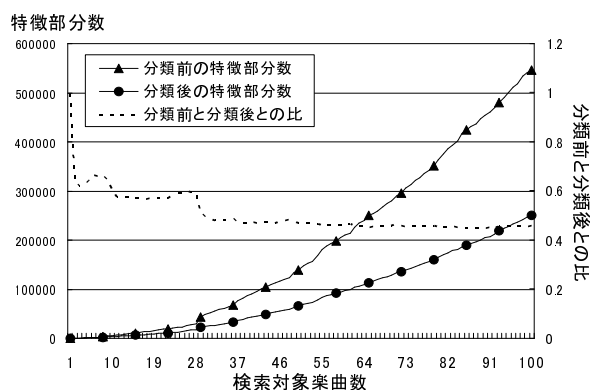


図6 分類前と分類後における特徴部分数の変化

Fig. 6 Change of feature segment number based on clustering

様々な種類があると考えられる。ところが、文書検索における不要語のように、適切ではない連結フレーム数が存在することが考えられる。そのため、連結フレーム数と、各連結フレーム数における特徴部分数との相関関係を算出し、適切ではない連結フレーム数を除去することを考える。

各連結フレーム数と、各連結フレーム数における特徴部分数との相関関係について算出した結果を図4に示す。この結果は $\theta = 0.99$ として算出した。この図から、相関関係は Zipf の法則 [6] に基づくと考えられる。しかし、連結フレーム数によっては Zipf の法則から外れており、全体としてばらつきがあると考えられる。そこで、Zipf の法則に従うように、適切ではない連結フレーム数を除去した。除去後の相関関係を図5に示す。この除去により、連結フレーム数の種類は 3173 種類から 2025 種類となり、特徴部分の数は 249894 個から 247548 個となった。

4.3 検索対象楽曲数と特徴部分数との相関関係

検索対象楽曲数を増やした場合の特徴部分数の変化について調べるために、検索対象楽曲が 2 曲である場合から 100 曲である場合まで、それぞれの場合において特徴部分を抽出した。このこ

とにより、抽出した特徴部分が、文書における単語のような収束性の性質を持っているかどうかということについて確かめる。

検索対象楽曲数と抽出した特徴部分数の変化について算出したグラフを図6に示す。閾値は、4.2 節と同様に、 $\theta = 0.99$ とした。この図より、分類前の特徴部分数に対して、分類後の特徴部分数は緩やかに増加していることがわかる。分類前の特徴部分数は、検索対象曲から抽出した段階での数である。一方、分類後の特徴部分数は、分類によって形成されたクラスタ数によるため、検索対象曲が増加するにしたがってクラスタの種類が限定される。そのため、このような緩やかな増加となったことが考えられる。ただし、文書における単語数の収束性のように、特徴部分数自体が収束するといった関係は見られない。また、図6における破線は、分類前の特徴部分数と分類後の特徴部分数との比を示す。この比率より、分類前の特徴部分数と分類後の特徴部分数の関係については収束性があると考えられる。なお、検索対象楽曲数が 100 曲であるとき、分類前の状態では 545237 個の特徴部分を得られ、分類を行うと、247548 個の特徴部分を得ることができた。

5. 評価実験

5.1 実験方法

提案手法による類似検索結果の再現率、精度と、既存手法による類似検索結果の再現率、精度を比較することにより、提案手法の評価を行う。既存手法として、楽曲に含まれる特徴部分の一部だけの類似比較により、楽曲の類似検索を行う手法を用いる。このことにより、楽曲における曲調の変化を考慮した類似検索を行うという点において、提案手法が既存手法に比べて有用であることを実証する。また、特徴部分を取得する際に用いる閾値 θ を変更し、各閾値において 11 点平均精度を算出する。閾値は、4.1 節により決定した、 $\theta = 0.99, 0.9975, 0.9985$ とする。

本実験では、予備実験と同じ検索対象楽曲を用いた。問合せ曲には、クラシック楽器に由来する静かな印象とヘビーメタルに由来する強い印象を含む曲である Silent Jealousy を用いた。また、正解集合は 6 人の人手により作成した。問合せに対して正解であるかどうかの判断基準として、楽曲全体を通して、楽曲に含まれる曲調が類似しているかどうかを基準とした。正解集合の曲数は、検索対象楽曲 100 曲のうちの 10 曲とした。

5.2 実験結果

各閾値および既存手法において算出した再現率精度曲線について、図7に示す。ここで、各閾値および既存手法の各曲線において 11 点平均精度を算出すると、それぞれ 0.256, 0.244, 0.26, 0.181 といった値を得られた。これらの値より、提案手法では、既存手法の平均精度に比べて最大で 0.079 の向上を得ることができ、提案手法は既存手法に比べて有用性が高いといえる。

検索結果として、提案手法では、クラシック楽器に由来する静かな印象とヘビーメタルに由来する強い印象を含む曲である紅や、クラシック楽器に由来する静かな印象とロックに由来する強い印象を含む曲である Karn Evil 9 が上位に現れた。一方、既存手法では、紅や Karn Evil 9 が上位に現れたとともに、ピア

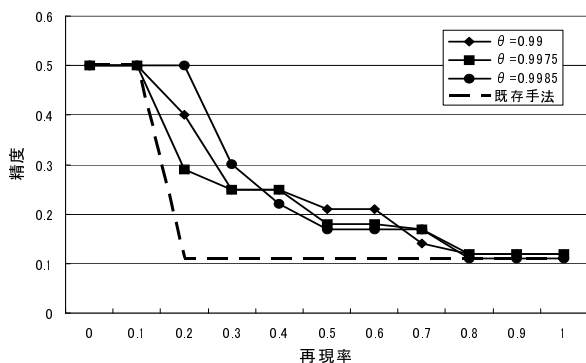


図7 提案手法および既存手法による再現率精度曲線

Fig. 7 Precision recall curve based on proposed method and current method

ノ曲であるラ・カンパネラや、バイオリンが主体となっている Riverdance も上位に現れた。この結果は、問合せに含まれるピアノやバイオリンによる曲調だけが、検索対象楽曲に含まれる曲調と類似していると判断された結果であると考えられる。このことより、既存手法では、楽曲に含まれる一部分の曲調だけによる類似検索を行っていることに対し、提案手法では、楽曲に含まれる曲調の変化を考慮した類似検索を行っていると考えられる。

ただし、提案手法では、ポップ曲も上位に出現した。その原因として、問合せに含まれる特徴部分と検索対象楽曲に含まれる特徴部分を明確に区別できていないことが考えられる。クラシック楽器に由来する印象とヘビーメタルに由来する印象を区別する場合、各曲調の特徴部分における特徴量は大きく異なる。しかし、一つの曲調において複数の特徴部分を区別する場合、各特徴部分における特徴量の差は、曲調間における特徴量の差ほど大きくない。そのため、一つの曲調内において特徴部分を明確に区別することはできていないと考える。

6. おわりに

本研究では、楽曲全体の印象を考慮した類似検索を行うために、楽曲全体における特徴部分の傾向に基づいた類似検索手法の提案を行った。また、楽曲全体における特徴部分の傾向を抽出する手法として、楽曲の構造が文書の構造と対応関係にあることに着目し、文書類似検索に用いられるベクトル検索モデルを用いた。ベクトル検索モデルを用いることにより、楽曲における特徴部分の重要度を算出することが可能となり、楽曲全体における特徴部分の傾向を抽出することが可能となることを提案した。

今後の課題として、二つの事項について検討する。一つは特徴部分の構成および特徴量に対する検討である。本研究では、各楽曲における固定長のフレームを類似比較することにより、特徴部分の抽出を行っている。ただし、固定長のフレームを連結し、文書における単語とみなすようにしているだけである。そこで、ただ連結するだけでなく、楽曲における特徴部分の表現について検

討することを考える。また、本研究において用いた特徴量では、一つの曲調内における特徴量の差を明確に区別することはできていないと考えられる。そこで、楽曲の類似検索における、適切な特徴量についても検討する。

もう一つの事項は特徴部分の出現位置に対する検討である。現段階では、特徴部分の出現位置を考慮することを行っていない。このことに対し、視聴者が楽曲全体から受ける印象は、特徴部分の出現位置によって異なる場合が考えられる。この課題に対し、楽曲に含まれる特徴部分間の遷移関係について検討する。

謝辞

本研究の一部は、文部科学省科学研究費補助金（課題番号：20500104, 20700101）によるものです。ここに記して謝意を表します。

[文献]

- [1] Michael Clausen and Frank Kurth. A Unified Approach to Content-Based and Fault-Tolerant Music Recognition. *IEEE TRANSACTIONS ON MULTIMEDIA*, Vol. 6, No. 5, pp. 717-731, 2004.
- [2] Esko Ukkonen, Kjell Lemstrom, and Veli Makinen. Geometric Algorithms for Transposition Invariant Content-Based Music Retrieval. *Proc. of ISMIR 2003*, 2003.
- [3] Shyamala Doraisamy and Stefan Ruger. A Polyphonic Music Retrieval System Using N-Grams. *Proc. of ISMIR 2004*, 2004.
- [4] Jeremy Reed and Chin-Hui Lee. A Study on Music Genre Classification Based on Universal Acoustic Models. *Proc. of ISMIR 2006*, 2006.
- [5] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Pub., 1999.
- [6] G.K.Zipf. Human Behavior and the Principle of Least Effort. *Addison-Wesley*, 1949.

大野 和久 Kazuhisa OONO

2007年立命館大学工学部情報学科卒業。現在、立命館大学大学院理工学研究科博士前期課程在学中。情報処理学会学生会員。

鈴木 優 Yu SUZUKI

立命館大学情報理工学部講師。マルチメディア電子文書検索に関する研究に従事。ACM SIGMOD, IEEE Computer Society, 情報処理学会, 日本データベース学会各会員。

川越 恭二 Kyoji KAWAGOE

立命館大学情報理工学部教授。情報システム, ネットワークサービス, データベースに関する研究に従事。情報処理学会, IEEE, ACM SIGMOD, 日本データベース学会, 電子情報通信学会各会員。