

メロディからの自動インデクシングと 音楽検索への応用

4 V-5

柳瀬 隆史[†]高須 淳宏[‡]安達 淳[‡][†] 東京大学大学院工学系研究科[‡] 学術情報センター研究開発部

1 はじめに

近年のマルチメディアデータベースの大規模化に伴い、ユーザからの多様な検索要求に応えるための研究が多くなされている。これらのうち、マルチメディアコンテンツとして音楽情報を対象とした音楽検索には、書誌的情報から検索するものと演奏内容から検索するものがある。このうち内容検索についてはカラオケ検索 [2] や感性語検索 [3] など、多様な検索が行えるシステムが研究されているが、その際検索を効率よく行うためのインデクシングは手作業で行っているものが多いのが現状である。

そこで我々は現在、内容検索におけるインデクシングを自動で行う手法を検討している。具体的には、楽曲の主旋律（メロディ）を分割して得られる、フレーズをインデクシングの単位とし、メロディからフレーズへの分割を自動で行うことにより、インデクシングを行う。フレーズへの分割では、実際に歌詞曲のMIDIデータを用いて分割を行い、評価を行ったので報告する。

2 フレーズによるインデクシング

フレーズとは、メロディを自然に区切ることでできる一つの流れを指し、1つのメロディは複数のフレーズで構成されているのが普通である。本研究では、このフレーズを基としてインデクシングを行い、音楽検索に応用することを考える。そのメリットとしては、次のようなことが挙げられる。

- 検索者の記憶の単位に合致している

カラオケ検索など、ユーザが記憶しているメロディを問い合わせとして入力する場合、メロディに関する記憶は断片的であっても、フレーズ単位であることが多い。従って問い合わせ入力もフレーズ単位となっていることが予測される。

- 楽曲の特徴がより正確に抽出できる

感性語検索など、特徴抽出による検索を考えた場合、曲の途中で曲の特徴（イメージなど）が大き

く変わってしまう場合なども多く、曲単位で特徴を抽出してもそれが正確ではない恐れがある。一方フレーズの途中では曲の特徴が変わることはほとんどないと考えられるので、インデクスを用いてフレーズ単位で特徴抽出を行うことにより、より正確な特徴抽出が期待できる。

- 検索コストの低減が可能

1つのメロディ中に同じ、あるいは類似したフレーズが繰り返し現れることが多いので、繰り返し現れるフレーズを1つにまとめることにより、検索コストの低減が可能である。

3 自動フレーズ認識

前章のようなインデクシングを自動で行うには、メロディからフレーズへの分割（フレーズ認識）を自動で行う必要がある。ここでは、その手法について説明する。

3.1 自動フレーズ認識手法

楽曲中のメロディにおけるフレーズは、以下のような特徴を持つと考えられる。

- フレーズの区切れ目には、楽譜上の休符に相当する、音の鳴らない時間が存在することが多い。
- 1つの楽曲中に同じ、あるいは類似した演奏内容（音符列）を持つフレーズが繰り返し出現することが多い。
- 異なる演奏内容を持つフレーズどうしても、先頭部分の演奏内容が重複している場合が多い。

本研究で提案する自動フレーズ認識手法は、上記の特徴を利用している。その手順は次のようになる。

1. サブメロディの生成

メロディを、休符をデリミタとして「サブメロディ」に分割する。本研究では、フレーズは一つあるいは複数のサブメロディから構成されるものとする。

2. サブメロディ間の類似度の計算

メロディから分割された n 個のサブメロディのうち任意の2つ N_j, N_k に対して、それぞれのサブメロディ間の類似度 $S(j, k)$ を求める。類似度の計算には DP マッチング法を用いる。サブメロディ間の類似度 $S_p(j, k)$ を次式で定義する。

$$S_p(j, k) = \frac{u(N_j, N_k)}{\max(m_s(N_j), m_s(N_k))} \quad (1)$$

ここで、 $u(N_j, N_k)$ は DP マッチングにより双方のサブメロディに一致した部分音高列長の最大値、 $m_s(N_j)$ 、 $m_s(N_k)$ はそれぞれの音高列の長さ（サブメロディを構成する音符の数）である。

3. フレーズへの分割

高い類似度を持つ2つのサブメロディはそれぞれ、同じあるいは類似したフレーズを構成するサブメロディであると考えられる。そこで、ある閾値以上の類似度をもつ2つのサブメロディにはそれぞれ同じラベルを付ける、という要領でサブメロディにラベル付けを行う。そして、曲中ではフレーズの先頭部分が重複して現れやすいという性質から、一度現れた先頭部分と同じラベルを持つサブメロディが再び現れば、そこをフレーズの区切れ目とみなす。

3.2 フレーズ認識の評価

提案手法の評価のため、あらかじめ楽曲構成の抽出およびフレーズの認識を手作業で行い、これを正解としておき、メロディからサブメロディに分割された段階で自動フレーズ認識の評価を行った。評価は正解としたフレーズごとに行い、各フレーズについて以下の3段階に分類した。

1. フレーズの先頭・末尾とも正しく認識（正解）
2. フレーズの先頭のみ正しく認識
3. フレーズの先頭が正しく認識されていない（不正解）

評価には市販の歌謡曲の MIDI データ（48 曲、653 フレーズ）を用いた。フレーズ認識は前節で述べた手法と、不必要に分割されるのを避けるために、さらにフレーズの時間的な長さを考慮し、ある下限（3 小節）以下のフレーズを作らないようにするなどの措置を講じたものと2種類の手法で行った。図1は全653フレーズのうち、正解と不正解にランクされたフレーズ数の割合である。

この図より、フレーズの時間的な長さを考慮することにより、精度が向上することが確認された。また、今回は曲ごとの最適閾値の決定は手作業で行ったが、

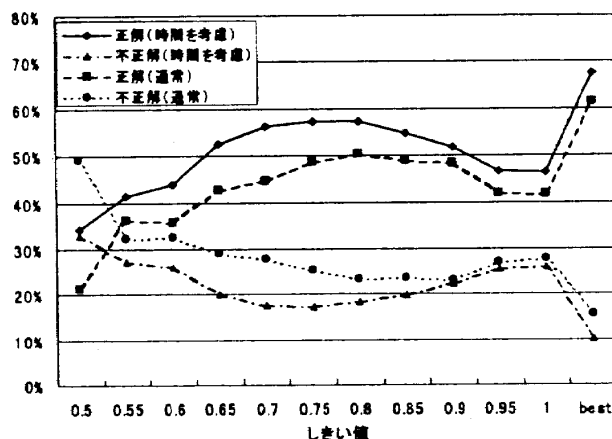


図1 評価結果（閾値の best は最適な分割が得られる閾値を曲ごとに変化させた場合）

閾値を固定した場合よりも飛躍的に向上していることから、曲ごとに最適な分割が得られる閾値にはばらつきがあると思われる。したがって、曲ごとの最適な閾値を動的に決めることができるようにする必要があると考えている。

4 おわりに

本稿では、音楽検索におけるインデクシングとしてフレーズを基とする手法を提案した。そして、楽曲中に同じフレーズが繰り返し現れることが多いことを利用して、フレーズの自動認識手法を提案し、その有効性を評価した。

今後は、フレーズ認識については先に述べた閾値を動的に決める手法を検討する他、類似度を与える式やラベル付けからフレーズの先頭を検出する手法についてさらに検討を進め、認識精度の向上を図る。また今回提案したインデクシングの実装を行って、実際の音楽検索システムでの有効性を評価していく予定である。

参考文献

- [1] 柳瀬隆史, 高須淳宏, 安達淳, "音楽検索における自動インデクシング法", 情報処理学会研究報告, 98-DBS-116(2)-42, pp.117-124, 1998.
- [2] 貝塚智憲, 後藤真孝, 村岡洋一, "歌声の旋律情報と歌詞情報をキーとした曲検索システム", 第54回情報処理学会全国大会, 7J-06, 1997.
- [3] 辻康博, 星守, 大森匡, "曲の局所パターン特徴量を用いた類似曲検索・感性語による検索", 信学技報, SP96-124, 1997.