

# EM アルゴリズムを用いた旋律分類

吉原 幸輝<sup>†</sup> 三浦 孝夫<sup>†</sup>

<sup>†</sup> 法政大学 工学部 電気電子工学科 〒 184-8584 東京都小金井市梶野町 3-7-2

E-mail: †{c01d3167,miurat}@k.hosei.ac.jp

あらまし 少量のラベル付き楽曲から, EM アルゴリズムを用いて, ラベルなし楽曲を分類する手法を提案する. 主なアイデアは, 各楽曲から特徴量を抽出し, 単純ベイズ法による学習過程に基づくことにある.

キーワード 旋律分類, 旋律特徴量, 単純ベイズ分類, EM アルゴリズム

## Classifying Melodies by Using EM Algorithm

Yukiteru YOSHIHARA<sup>†</sup> and Takao MIURA<sup>†</sup>

<sup>†</sup> Dept.of Elect.& Elect. Engr., HOSEI University 3-7-2, KajinoCho, Koganei, Tokyo, 184-8584 Japan

E-mail: †{c01d3167,miurat}@k.hosei.ac.jp

**Abstract** In this investigation we propose a novel technique to classify melodies by using EM algorithm. Here we generate classifiers based on naive Bayesian classifiers during EM steps. And we discuss experimental results with several features of melodies and show we can get efficient classifiers by our approach.

**Key words** Melody Classification, Melody Features, Naive Bayes, EM Algorithm

### 1. 前書き

本稿では, 複数のテーマ (単旋律) から特徴値を抽出し, 未知の旋律を分類していずれかのテーマに帰着させる手法 (旋律の自動分類) を論じる. これらは, 本来手書き文字認識や手書き絵による類似検索と同様に, 内容検索を目的とし, 曲名や作品番号といったメタ情報を仮定しない.

旋律は主要な印象を与える役割を有しており, この機能を用いた応用は広範囲に想定できる. 例えば, 音楽作業 (作曲・編曲) に対して直接的な支援を行うことや, 同じクラスの曲を検索し著作権に抵触しそうな候補の絞込みを行うことができる. 実際, 音楽における著作権管理では, 歌詞と異なって, 旋律で判断することが多い<sup>(注1)</sup>. また, インターネットを介して得た音楽コンテンツの自動管理においても旋律による手がかりを与えることができる.

音楽情報処理分野の研究は 60 年代から開始され, マルチメディア処理技術 (信号処理からメタ情報による検索まで) や情報検索理論との関連 (N グラム, ベクトル空間モデル, 検索エンジン等) が論じられてきた.

他の情報検索分野と比較して, 音楽分野での分類問題は特に難しい点がある. まず, 音楽情報をモデル化するのにデータモデリング機能が著しく欠けていることが挙げられる. 第 2 に

旋律の特定が難しい. 実際, 旋律の特徴量を定量的に定義することができず, 改ざんによる旋律の同定を扱えことさえできない. 第 3 に, 旋律の抽出方法が不明であることも重要である. 通常, テーマ (主旋律) が中心になるが, 編曲の過程で和音, 楽器・音色などが付与される. テーマである旨の明示的な記述表現もない. 第 4 に, 旋律の類似性の判定 (何が答えか?) が明確に定義できず, クラス所属も確率やファジーメンバーシップによるのか, ランキング表現がよいのかが定まらない [18]. 最後に, 効率よいデータ管理・データ操作に関して, どのような技術が必要かが不明であることも重要である [15].

このような音楽情報処理の問題は, 他の人工知能関連のテーマと同様に, 分類問題に帰着させて定式化し, 情報検索手法や機械学習手法により, 自動計算するのが現実的である.

機械学習手法を用いた帰納推論では, 分類クラス (ラベル) が明記された幾つかの訓練データから一般的なパターン (分類器) を抽出し, これを利用して未知データを分類する. 正確に分類するには大量の訓練データを必要とするが, ラベルをつける過程は通常人間が行うため, 時間がかかり主観的で信頼性にかけたものになりがちである [11]. インターネットを介して得るデジタル音楽のラベルの多くは誤っているか無記名であり, ラベルを正確につける手間は困難な状況になっている.

近年これに対応するために, 不完全データによる分類器生成手法 (ブートストラップ手法) が提案され, とくに EM アルゴリズムが代表的である [24]. ここでは少量のラベルつき訓練データとラベル無し訓練データから分類器が構成される.

(注 1): 歌詞と切り離れた, 音楽そのもの「旋律, 和音, 節奏, 形式の 4 要素が一体となった楽曲」(東京地方裁判所昭和 43 年 5 月 13 日) も独立に著作物となり, 保護の対象となる.

この方式の利点は、少量の訓練データから自己増殖的に学習するために未知データを活用できることにある。すなわち、インターネット上に点在する音楽データを効果的に学習できることにあり、以下で実験によりその有効性を検証する。

本研究の目的は、このような少量のラベル付き訓練データから自動分類を行う旋律分類手法の提案にある。ここでは分類器の生成自体が目的ではなく、ラベル無しデータの推定精度を確かめることにある。本研究では、旋律の自動分類の精度を向上するため、旋律（楽譜）から特徴量を抽出し、機械学習手法を用いて分類に利用する。具体的には、旋律特徴量を基に単純ベイズ手法で分類し、さらに EM アルゴリズムによって計算途中の分類結果を利用する。

本稿では、音楽に関する基本知識 [23]、および情報検索技術の基本知識を仮定する [5], [26]。

第 2 節で、準備と必要な定義を述べ、第 3 節で音楽情報の特徴量を要約し、本研究で扱う特徴量を定義する。第 4 章では、主として分類に利用する単純ベイズ法および EM アルゴリズムの適用方法を示し、この特徴や利点を述べる。第 5 節では実験結果を述べ、本手法の有効性を示す。

## 2. 旋律の表現

楽曲は複数の音から構成される [23]。ピッチ (pitch) とは音の調子、即ち音の高さを表す。これは周波数で規定され、代表的に記号で表現される。例えば、A (ラ) の音は 440Hz であり、その倍の高さの音を 1 オクターブ (octave) という。例えば 880Hz は A の 1 オクターブ高い a を表す<sup>(注 2)</sup>。ピッチを音 (tone) とも呼ぶ。

音程 (interval) はピッチ間の距離であり、1 オクターブ離れた音程を 12 等分した単位を半音 (semitone) という。例えば A から a までは A, A#, B, C, C#, D, D#, E, F, F#, G, G# と表す。G#, A#, C#, D#, F# はそれぞれ Ab, Bb, Db, Eb, Gb と表す。全音階的 (diatonic) な音の並びとは、A, B, C, D, E, F, G, a または C, D, E, F, G, A, B, c のように、半音がそれぞれ 2, 1, 2, 2, 1, 2, 2 または 2, 2, 1, 2, 2, 2, 1 ずつ離れた配置を言う。前者を短調 (minor) 後者を長調 (major) という。また A, C をそれぞれの主音 (key), +7 音を属音 (dominant) という。全音階的でない並びを無調的 (atonal) という。移調 (transposition) とは、主音を短調・長調の並びを保ったまま相対的に移動することをいう。

各ピッチには長さ (duration) がある。旋律 (melody) とは (長さを持つ) ピッチの順次配置である。和音 (chord) は複数のピッチの並列配置であり、これらは同時に演奏されることを想定している。音楽の意図する主要な旋律ではない部分を伴奏 (accompaniment)、伴奏の付いた音楽を多声音楽 (polyphony)、伴奏が無い音楽を単旋律 (monophony) という。

五線譜 (five horizontal lines あるいは score) は、音の高さと長さを記入した記号 (音符) を順次的あるいは並列的に記述した表現である。旋律は音符の並びで、和音は音符の重なりで記述される。音符の長さの合計を拍 (beat) という。小節 (bar)

(注 2): 以下では 1 オクターブ高いピッチを小文字で表す。

とは同じ拍数を持つ (五線譜の) 部分記述であり、拍子 (time signature) によって小節当たりの拍数と拍質を表す。特に、リズム (rhythm) とは拍の規則的な並びをいう。

音楽を表現するためには、多くの表情つけがなされている。これらは大きく 4 つに分けられ、速度・強弱に関するもの、発想に基づくもの、奏法によるもの、装飾音に関するものから成り、五線譜上に直接記載される。

これらを計算機で表現するには、これまで信号レベル (CD 等でのアナログ的变化を表現)、操作レベル (MIDI など、輝度・時間のゆれを表現) 及び楽譜レベル (表現・解釈の多様性を表現) でなされてきた。

## 3. 旋律特徴量

旋律を分類するためには、当該音楽の有する特徴値の表現を考慮する必要がある。旋律を分類するために指定される条件は、不完全である場合が多く、その大半は部分的である。

本研究では旋律から特徴量を抽出し、ベクタ空間モデル [5], [26] を用いて特徴項目ベクトル空間へ写像する。各次元は項目に対応するため、どの旋律も空間内の点で表せる。旋律質問の解はベクトル空間内の類似した点に対するランク付けで得られる [20]。空間内では伝統的な余弦質問で定義する。即ち  $\cos \theta = \frac{X \cdot Y}{|X| |Y|}$  を計算し、1.0 に近いほど類似していると判断する。ここで  $X, Y$  は空間ベクトルを、 $X \cdot Y$  は内積を、 $|X|$  はベクトル長を表す。

音楽情報を分類するためには、クラスがあらかじめ判明している旋律 (ラベル付き訓練データ) と未知旋律の関連を評価する尺度が必要である。

本稿ではピアノ曲に限定して考察する。ピアノ曲では通常主旋律を奏するのは右手である。右手に対応した第 1 声に限定し、和声があった場合は (高音とは小指や薬指になるので限定しにくいことから) 和声記述の先頭音を旋律音と見なす発見的な方法を取る。

### 3.1 ピッチ輪郭

これまで提案された旋律情報のための特徴量記述では、旋律輪郭 (Melody Contour) が代表的である [1], [7], [18]。また、近年は楽譜化技法として、XML に基づくものや五線譜利用による抽象化記述がある。旋律の輪郭情報のうち、ピッチ輪郭 (pitch contour) 表現は多くの研究で検討されている。ピッチ輪郭表現とは、単旋律にだけ有効であり、直前の音と比べてピッチが相対的に高い、低い、同じという状態をそれぞれ文字 U, D, S で表し、輪郭を文字列として記述する<sup>(注 3)</sup>。この表現は、ピッチに対して相対的なので移調に強いが、雑音に反応して文字列が変わりやすく、旋律検索は不完全文字列一致処理の実行を意味する。

[例 1] 以下では、“カエルの歌” を統一例として考察する。五線譜は図 1 で示されるが、2 段譜の右手パートに主旋律が生じている。このピッチ輪郭は-UUU DDD UUUU DDD DSSS SSUSUSUS DDD である。あるいは、長さを無視することはピッチが同じか

(注 3): 開始文字は相対位置を定めることができないので - で表す。

ぎり意味をなさないと考えることに相当するので、記号 s を取り去り、-UUU DDD UUUU DDD D UUU DDD と表すこともある。



図1 カエルの歌五線譜

しかし、ピッチ輪郭方式は数多くの問題を含む [17]。大規模楽曲の記述から主旋律を抽出することは難しいため、多声音楽から各声を旋律と見なした輪郭を記述し、これらの結合から主旋律を想定する、という組み合わせ的探索を行わざるを得ない [2], [3]。旋律記述の記述に拍子・音符長を含まないため、判別のため多くの情報を必要とする。例えば -SSD USSD は運命交響曲 (Beethoven) も雪山賛歌 あるいは Oh! My Darling, Clementine! (米国民謡) も表すためより多くの情報を必要とする。また、不完全一致検索手続きは多項式時間でよいことが知られるものの複雑で、(全走査を除けば) 2 次記憶域上での処理効率を向上させる方法は知られていない。輪郭抽出には雑音除去のための”しきい値”の設定が容易ではなく、生成される輪郭を規定する。

### 3.2 ピッチスペクトル

音楽データベース情報は、特定されるべき楽曲であるから五線譜情報を有するほうが有利である。実際、これらは高度の抽象性を有し、厳格な指示 (五線譜上の記号による指示) に忠実に従っている。不完全な旋律指定 (不正確・欠落・旋律リズムの近似、移調・長単調の変更、作為的な表情付け、こぶし、ずれ、ポルタメントといった揺らぎ) に対して、楽譜記述を用いれば、リズム・音色・発想・速度・奏法・強弱・拍子などの観点で問題を局所化する。反面、移調・装飾音・揺らぎという問題については、楽譜化では容易には解決できない。

旋律が五線譜・楽譜で記述されるとき、小節単位にピッチスペクトル (pitch spectrum) を導入し特徴量とすることができる [10]。ピッチスペクトルとは、小節内に生じる各音符を長さで集積しそのヒストグラムをとるもので、以下では 4 オクターブ  $12 \times 4 = 48$  音を対象に、48 次元ベクトルを生成する<sup>(注4)</sup>。小節ごとにピッチスペクトルを算出し、これを対象楽曲の規定小節数だけ用意する。このアイデアは信号処理では提案されているが、楽譜レベルで特徴量化するというアプローチはない [22]。[例 2] カエルの歌 (2 声) から抽出した旋律 (第 1 声) に対する 16 小節分のピッチスペクトル (図 2) を示す。

ピッチスペクトルは各ピッチを長さで合計したヒストグラムであり、(ギターコードのような) 小節全体に和音を指定したものと異なる。ピッチ輪郭とは異なって、小節単位の特徴を集約しているため、旋律の輪郭を表してはいない。むしろ、小節

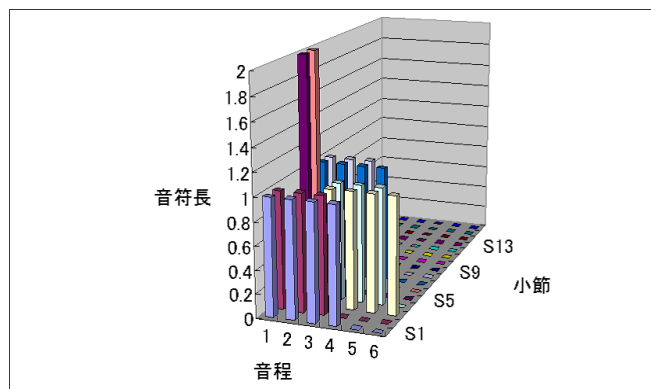


図2 カエルの歌ピッチスペクトル

間の変化を記述しているため、旋律の時系列表現とも異なる。ピッチスペクトルは不完全な旋律に生じる問題を解消することができる。実際、装飾音・旋律・長短調・揺らぎがピッチスペクトルで局所化する。

ピッチスペクトルの問題は 移調に変動しやすいこと、および長調・短調の変化に対応しづらいことであることが知られている [10]。前者については、旋律から主音 (キー) を決定しテーマへの移調を行えばよい。しかし後者の問題については対応が難しく、より高度な対応が必要である。

### 3.3 本稿で提案する旋律特徴量

本稿では、旋律分類を行うための特徴量を、ピッチ輪郭およびピッチスペクトルに加え、平滑化ピッチ輪郭 (Mean Contour)、ピッチ細分輪郭 (Pitch Fine Contour) および輪郭スペクトル (Contour Spectrum) を導入し、EM アルゴリズムによってどの程度まで分類精度を向上できるかを論じる。

ピッチ輪郭手法には幾つかの重要な欠点がある [7]。その第 1 は、旋律内のピッチの微妙な変化 (雑音) に影響を受け易いことにある。例えばトリルのような細かなピッチの変動は、印象をほとんど変えないにもかかわらず、ピッチ輪郭では大きな差となる。そこで、平均フィルタ (Mean Filter) と同様に、各音を前後と合わせて平均化し、この結果に対してピッチ輪郭で表現する手法を、平滑化ピッチ輪郭 (Mean Contour) と定義する。[例 3] カエルの歌を平滑化すると、第 1 小節最終音 (“ファ”) がその前後と平均化され “ミ” に代わる。同様に第 3 小節最終音 (“ラ” が “ソ” に) に代わり、その結果、この平滑化ピッチ輪郭は -UU DD UUU DD D UUU DDD である。

ピッチ輪郭手法の第 2 の問題は、旋律内のピッチの変化を上昇 (U) または下降 (D) でしか表現できないことにある [7]。滑らかな音階進行もオクターブ移動のような激しい進行もピッチ輪郭では差になって生じない。そこで、ピッチの移動幅 (半音を 1 として) を付加しピッチ輪郭を拡張表現する手法を、ピッチ細分輪郭 (Pitch Fine Contour) と定義する。

[例 4] カエルの歌をピッチ細分輪郭で表現すると次のようになる。

-U2U2U2 D1D2D2 U4U1U2U2 D2D2D1 D4 U2U2U1

(注 4): 倍精度実数で表せば  $8 \times 48 = 384$  バイト長となる。

しかし、この方法では、明らかに雑音の影響を（ピッチ輪郭以上に）受けやすい。このため、ピッチスペクトルと同様に小節ごとにスペクトルを取りベクトル化する。これを輪郭スペクトル (Contour Spectrum) という。ピッチ輪郭は長さ情報を考慮していないため、ベクトル長さは 1 に正規化する。輪郭スペクトルは上昇下降それぞれ 24 までと仮定して 48 次元ベクトルで表す。例えば D2 が 3 回生じるとき、D2 次元に 3 を設定し、最後に長さが 1 になるように調整する。以下では D2 が 3 回生じるときに D2:3 と表現する。

[例 5] カエルの歌を輪郭スペクトルで表現すると 8 個のベクトル列になる。

$$\{ U2:3 \}, \{ D1:1, D2:2 \}, \{ U1:1, U2:2, U4:1 \}, \{ D1:1, D2:2 \}, \{ D4:1 \}, \{ \}, \{ U2:2, U1:1 \}, \{ D1:1, D2:2 \}$$

輪郭スペクトルは、ピッチ輪郭の利点をすべて有している。更に雑音にも強い。実際、ピッチの移動幅が多少異なってもベクトルの類似性 (余弦値) は消えない (0 にはならない)。小節単位で同期を取るために、対応が取れない部分があってもそれに降に影響を与えないという意味で、ピッチ輪郭の問題に対応している。

#### 4. 単純ベイズ法と EM アルゴリズム

以下では、出典の知られている旋律  $d$  をラベル付き訓練データとして、未知旋律をいずれかのクラスに分類する。本研究では EM アルゴリズムと単純ベイズ分類を組合わせた手法を用いる [12]。  $d$  は、特徴量  $w_j$  を用いて多重集合  $\{w_1, \dots, w_m\}$  として表されると仮定する。

##### 4.1 単純ベイズ法による文書分類

旋律  $d$  に、クラス集合  $C = \{c_1, \dots, c_n\}$  のいずれかの値を割り当てる (分類する) ことを考える。ベイズ規則による分類は、 $d$  がクラス  $c$  に属する確率  $P(c|d)$  の確率分布を求めることである。排他的な分類の場合、最大事後確率をとるクラス  $c \in C$  を分類することで分類エラーが最小になると考える。ここでクラスラベルは次のように定める:  $c_k = \operatorname{argmax}_{c \in C} P(c|d)$ 。ベイズ規則は以下のように与える:  $P(c|d) = P(c) \times \frac{P(d|c)}{P(d)}$ 。すなわち、ベイズ規則での分類規則生成 (訓練) は訓練データ集合から、確率分布  $P(c)$ ,  $P(d)$ ,  $P(d|c)$  を推定することである。

しかし、旋律  $d = \{w_1, \dots, w_m\}$  に関して  $P(d|c)$  や  $P(d)$  の推定が問題であるため、一般に、特徴量  $w_j$  の出現は、統計的に他の  $w_i$  の出現とは独立であるという仮定をおき、各旋律を特徴量の集合とみなす単純ベイズ手法を使うことが多い。単純ベイズでは、 $P(d|c)$  を以下の形式に分解して考える。 $P(d|c) = \prod_{i=1}^{|d|} P(w_j|c)$ 。これにより、排他的な分類の場合ではベイズ規則は以下のように書くことができる。

$$P(c|d) = P(c) \times \prod_{i=1}^{|d|} P(w_j|c) \quad (1)$$

また、ここでは旋律内での各特徴量は、その定義に従って与えられる。ベイズ推定手法は特に文書分類で有効であることが知られている [25], [27]。

##### 4.2 EM アルゴリズム

EM アルゴリズムとは、データの欠損部分 (直接観測できない確率変数) を最尤推定により求める、欠損部が分かればその形は単純で解析的に表現できるという仮定を置く。

Expectation (期待値), Maximization (最大化) は、それぞれ欠損値の推定、期待値を得る過程を与えるパラメタの推定に対応している。不完全な観測データ  $x_1, \dots, x_N$  が、モデル  $P_\theta(x)$  から得られたとする。EM アルゴリズムとは、この不完全な観測データ  $x_1, \dots, x_N$  から、未知のパラメタ  $\theta$  の値を推定することである。現時点での  $\theta$  を使用して、条件付確率モデルから、完全データにおけるサンプル数、期待値を求める (E ステップ)、続いて、ここで求めたサンプルから期待値を最大化するパラメタ  $\hat{\theta}$  を求める (M ステップ)。この E ステップと M ステップを繰り返すことにより、モデルの対数尤度を最大化するパラメタを求める。

EM アルゴリズムを分類操作に適用するアルゴリズムは以下のようになる。

- (1) 入力: ラベル付旋律集合, ラベルなし旋律集合
- (2) ラベル付旋律集合から、単純ベイズ分類規則  $\hat{\theta}$  を生成
- (3) 以下のステップを一定回数、またはパラメタが収束するまで繰り返す

(E-step) 現在の分類規則  $\hat{\theta}$  を使用してラベルなし旋律のクラス所属確率 ( $P(c_j|d_i; \hat{\theta})$ ) を推定する

(M-step) 推定された事後確率 (分類結果) を利用して、分類規則  $\hat{\theta} = P(D|\theta)P(\theta)$  を再度生成する。

(4) 出力: 分類規則  $\hat{\theta}$  を用いて、各ラベル無し旋律のラベル  $c_k$  は次のように定める:  $c_k = \operatorname{argmax}_{c \in C} P(c|d)$

EM アルゴリズムでは各繰り返しステップで尤度が単調増加するため、必ず解を有する。反面、収束が遅く局所解に至る可能性があるため、初期値の選定が難しい [24]。

##### 4.3 旋律分類への適用

様々な特徴量を使用することで単純ベイズ法に基づく EM アルゴリズムを旋律分類に適用する方法を説明する。

$C = \{c_1, \dots, c_m\}$  はクラス集合を表す。ラベル付き旋律  $e_c$  はいずれかのクラスに含まれる  $c \in C$ 。  $D$  はすべての訓練データを表し、 $w_i$  はデータ内の各特徴量を表す。また  $N(w_i, d_j)$  は旋律  $d_j$  における特徴量  $w_i$  の特徴値である。さらに、 $P(c_k|d_j)$  は前述の旋律  $d_j$  がクラス  $c_k$  に属する確率を表し、ラベル付き訓練データに関しては、そのラベル付けられたクラス  $c_m$  においては、 $P(c_m|d_j) = 1$  であり、それ以外のクラスに対しては 0 と取る。これに対して、ラベルなし訓練データは、当初は全クラスに対して 0 であるが、最初は通常のベイズ分類により、その後は EM アルゴリズムにより、徐々に適切な値へと更新される。

$$P(w_i|c_k) = \frac{1 + \sum_{j=1}^{|D|} N(w_i, d_j)P(c_k|d_j)}{|V| + \sum_{i=1}^{|V|} \sum_{j=1}^{|D|} N(w_i, d_j)P(c_k|d_j)} \quad (2)$$

$$P(c|d) = P(c) \times \frac{\prod_{i=1}^{|d|} P(w_j|c)}{\prod_{i=1}^{|d|} P(w_j)} \quad (3)$$

式3と式2によりEMアルゴリズム内で分類規則を生成する。同様に  $P(c_j)$  は以下のように与えられる。

$$P(c_j) = \frac{1 + \sum_{j=1}^{|D|} P(c_k|d_j)}{|C| + |D|} \quad (4)$$

この式2,4はそれぞれスムージングを行っている。

旋律の特徴量に関して、ピッチ輪郭、ピッチ細分輪郭、平滑化ピッチ輪郭、輪郭スペクトル、ピッチスペクトルの各場合においての適用方法を定義をする。

特徴量としてピッチ輪郭、平滑化ピッチ輪郭およびピッチ細分輪郭を用いる場合、旋律  $d$  に生じる上昇 (U, U2, ...) や下降 (D, D3, ...) の発生個数がそのまま特徴量となる。一定数の小節 (例えば先頭4小節) での出現数をその特徴値と見なす。

ピッチスペクトルあるいは輪郭スペクトルの場合、旋律  $d$  はベクトル列  $(w_1, \dots, w_k)$  であり、各ベクトル  $w_i$  は第  $i$  小節に対応している。テーマ  $e = (e_1, \dots, e_k)$  に対しても同様の計算を行っているため、内積値  $(w_i, e_i)$  を第  $i$  小節の特徴値と見なす。ラベル付きテーマが複数存在する場合、その数だけの特徴値があるとみなせばよい。

ピッチ細分輪郭、平滑化ピッチ輪郭も (U1, U2, ..., D1, D2, ... などのように) ベクトルの次数が増えるが考え方は同じである。また、ピッチスペクトル、輪郭スペクトルも小節数を限定して類度を計算するため同様に考えてよい。例えば、ピッチスペクトルでは各小節を48次元で考えると、先頭4小節で判断すれば  $48 \times 4 = 192$  次元のベクトル空間で考えればよい。

[例6] ピッチ輪郭 (U, D 表記) によるEMアルゴリズムを用いたクラス分類例を示す。"カエルの歌 (A)"、"キラキラ星変奏曲 (B)" をラベルとし、その旋律が  $d_1, d_2$  として与えられている ( $L_1 = \{d_1, d_2\}$ ) として、"歓喜の歌" ( $d_3$ ) をクラス分類しよう ( $L_2 = \{d_3\}$ )。



図3 Morzart KV265



図4 Beethoven Symphony No.9, Op.125

ピッチ輪郭による先頭4小節の特徴値はそれぞれ次のようになる。

- $d_1$ : -UUUUDDUUUUUUDD
- $d_2$ : -UUD
- $d_3$ : -UUUUUUUU

melody	(U,D)	U+D	class	P(U *)	P(D *)
$d_1$	7, 6	13	A	7/13	6/13
$d_2$	2, 1	3	B	2/3	1/3
$d_3$	4, 5	9			

(1) 単純ベイズ法によりクラス割り当てを行う。定義より  $P(A|d_1) = P(B|d_2) = 1$ ,  $P(B|d_1) = P(A|d_2) = 0$  and  $P(A|d_3) = P(B|d_3) = 0$ .

である。ここでは Bayes 定理を用いて

$$P(C) \times P(d_3|C).$$

を最大化する  $C$  を求めたい。この式は  $P(C|d_3) \times P(d_3)$  (注5) を意味している。

$D = L_1 \cup L_2$  には3件のデータが含まれ、クラス A, B にはそれぞれ1件しか含まれていないから、 $P(A) = P(B) = 1/3$  である。

単純 Bayes の仮定から  $P(d_3|C) = P(U|C) \times P(D|C)$  である。ここで  $P(A|d_3), P(B|d_3)$  をそれぞれ推定する。

$$P(A) \times P(d_3|A) = P(A) \times P(U|A) \times P(D|A)$$

$$= \frac{1}{3} \times \frac{7}{13} \times \frac{6}{13} = 0.083$$

$$P(B) \times P(d_3|B) = P(B) \times P(U|B) \times P(D|B)$$

$$= \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} = 0.073$$

この結果、 $P(C|d_3)$  を最大化する  $C$  は A であり、"歓喜の歌" は "カエルの歌" に割り当てられた。

旋律  $d_3$  が A に割り当てられた結果、 $P(A) = 2/3$ ,  $P(B) = 1/3$  に変わる。

また、 $d_3$  のクラスが割り当てられた結果条件確率  $P(U|A), P(D|A)$  が変わる。

class	P(U *)	P(D *)
A	11/22	11/22
B	2/3	1/3

更に  $d_3$  のクラス所属確率  $P(C|d_3)$  は  $P(C|d_3) = P(C) \times P(d_3|C)/P(d_3)$  で求められ、それぞれ次の値となる。

$$P(A|d_3) = P(A) \times P(d_3|A)/P(d_3) = \frac{2}{3} \times \frac{11}{22} \times \frac{11}{\frac{13}{25} \times \frac{12}{25}}$$

$$P(B|d_3) = P(B) \times P(d_3|B)/P(d_3) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{\frac{13}{25} \times \frac{12}{25}}$$

(2) 新たに定まった確率値を用いて、再度  $d_3$  の所属確率を求める。 $P(C|d_3)$  を最大化する  $C$  を求めるため、次式を計算する。

(注5):  $P(d_3) = P(U) \times P(D) = (13/25) \times (12/25)$

$$P(A) \times P(d_3|A) = P(A) \times P(U|A) \times P(D|A)$$

$$= \frac{2}{3} \times \frac{11}{22} \times \frac{11}{22} = 0.167$$

$$P(B) \times P(d_3|B) = P(B) \times P(U|B) \times P(D|B)$$

$$= \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} = 0.073$$

この結果、 $P(C|d_3)$  を最大化する  $C$  は  $A$  のままであり、” 歓喜の歌 ” は ”カエルの歌” に割り当てられた状態を保つ。 $d_3$  の所属確率を求めるための状況は先と変更が無いため、ここで繰り返しを停止する。

後述するように、”カエルの歌” と ” 歓喜の歌 ” は、ピッチ輪郭を用いた類似度は 90.3% であり、かなりの確度で分類したことになる。

## 5. 実 験

以下の実験では幾つかの変奏曲を扱う。変奏曲は 1 つのテーマと複数の変奏曲から構成されるため、ラベル付き訓練データはテーマの数だけあると考える。これ以外はラベル無しデータとみなし、旋律分類の対象とする。

### 5.1 準 備

本稿では、ピアノ変奏曲 (単旋律) を用いる。使用する楽曲は、Mozart キラキラ星変奏曲 K.265, Schubert 即興曲 Op.142-3, および Beethoven トルコ行進曲による 6 つの変奏曲 Op.76 の 3 つの変奏曲であり、それぞれ主題 (テーマ) と 12 曲, 5 曲, 6 曲の変奏 (パリエーション) からなる。これらは MIDI データとして表現されているものを ABC フォーマット [21] に変換して実験に使用する。

まず最初に、すべての単旋律の ABC ファイルから特徴量を取り出す。特徴量は、曲の先頭 4 小節の音を対象とする。取り出した特徴量を、ラベル付旋律集合とラベルなし旋律集合に分ける。本実験では、ラベル付旋律集合は変奏曲のテーマ (主旋律) とし、残りの変奏曲はラベルなし旋律集合とする。

また、ここで抽出する特徴量は、

- ピッチ輪郭: 上昇 (U), 下降 (D), 同等 (S) の 3 アイテム
- ピッチ細分輪郭: オクターブを 6 音とした, D6 から U6 の 13 アイテム
- 平滑化ピッチ輪郭: オクターブを 6 音とした, D6 から U6 の 13 アイテム
- 輪郭スペクトル: オクターブを 6 音とした, D6 から U6 の 13 アイテム
- ピッチスペクトル: 半音を除いた 7 音 + 休符 の 8 アイテム

とする。

### 5.2 評価方法

特徴値ごとに、2 つの旋律の類似度について次のように定義する。

輪郭情報 (pitch contour, mean contour, pitch fine contour) の場合は、旋律の先頭 4 小節について、最終的にアイテムごとに集計された頻度ベクトルとなる。例えば pitch contour では U,D,S の個数からなる 3 次元ベクトルを得る。このとき旋律  $E1, E2$  の類似度を 特徴値ベクトル  $V1, V2$  の余弦と定義する。

スペクトル情報 (pitch spectrum, contour spectrum) の場合は、各小節ごとにスペクトル (ベクトル) が生成される。テーマおよび変奏曲 T, E のスペクトル列  $\langle A1, A2, A3, A4 \rangle$ ,  $\langle B1, B2, B3, B4 \rangle$  に対して小節ごとに余弦値を求め特徴値ベクトル  $V = \langle V1, V2, V3, V4 \rangle$  とする。ここで  $V1$  は  $A1, B1$  の 2 つのベクトルの余弦値を示す。このとき、2 つの旋律  $E1, E2$  の類似度を、特徴値ベクトル  $V, W$  の余弦と定義する。特にテーマ T と変奏曲 E の類似度は  $\langle 1, 1, 1, 1 \rangle$  と特徴値ベクトルの余弦として求められる。

本実験では、単純ベイズ法だけを用いた分類と EM アルゴリズムを用いた分類による正解率の変化や、特徴量の抽出方法による正解率の変化を比較することで、どの抽出方法がよいか、または EM アルゴリズムの繰り返し回数が関係するのかを検証する。正解率は、次式で定義される。

$$\frac{\text{類似度以上で正しく分類された総変奏曲数}}{\text{総変奏曲数}} \times 100(\%) \quad (5)$$

次に、ラベルなし旋律集合を分類する。最初に持っている分類規則は、ラベル付旋律の特徴量とする。4.2 で述べた方法で、この分類規則を用いて EM アルゴリズムを適用する。その分類規則から、ラベルなし旋律のクラス所属確率を、ラベル付旋律の特徴量とラベルなし旋律の特徴量から推定する (E-Step)。また、類似度 (余弦値) を求め、その値が 30%, 50%, 60%, 66.7% 以上の時、ラベル付旋律集合に分類されたときとみなし、それ未満の場合は、分類されたと考えない (正解率の対象ではない)。同じ変奏曲に分類された場合を正解とする。そして、所属確率に基づいて、分類規則を学習することで、新しいラベル付旋律の特徴量を得る (M-Step)。

これらの、E-Step と M-Step を一定回数繰り返し、その時の正解率を見る。本実験では、繰り返し回数を、0 回 (単純ベイズ法のみ)、5 回、10 回、15 回、20 回、25 回の 6 パターンについて結果をまとめる。

### 5.3 実験結果

特徴量の抽出に、ピッチ輪郭を使用した際の正解率を表 1 に、ピッチ細分輪郭を使用した際の正解率を表 2 に、平滑化ピッチ輪郭を使用した際の正解率を表 3 に、輪郭スペクトルを用いた使用した際の正解率を表 4 に、ピッチスペクトルを用いた使用した際の正解率を表 5 に示す。

		EM 回数					
		0	5	10	15	20	25
類似度 (%)	30	34.8	34.8	52.2	52.2	52.2	52.2
	50	34.8	34.8	52.2	52.2	52.2	52.2
	60	34.8	34.8	47.8	47.8	47.8	47.8
	66.7	34.8	36.4	47.8	47.8	47.8	47.8

表 1 ピッチ輪郭 (pitch contour)

また EM アルゴリズムによる改善の割合を示す。類似度 30% 以上の場合を図 5 に、類似度 50% 以上の場合を図 6 に、類似度 60% 以上の場合を図 7 に、類似度 66.7% 以上の場合を図 8 に示す。

		EM 回数					
		0	5	10	15	20	25
類似度 (%)	30	39.1	31.8	31.8	31.8	31.8	31.8
	50	39.1	26.1	26.1	26.1	26.1	26.1
	60	31.8	26.1	26.1	26.1	26.1	26.1
	66.7	31.8	26.1	26.1	26.1	26.1	26.1

表 2 ピッチ細分輪郭 (pitch fine contour)

		EM 回数					
		0	5	10	15	20	25
類似度 (%)	30	26.1	17.4	17.4	17.4	17.4	17.4
	50	26.1	17.4	17.4	17.4	17.4	17.4
	60	26.1	17.4	17.4	17.4	17.4	17.4
	66.7	26.1	17.4	17.4	17.4	17.4	17.4

表 3 平滑化ピッチ輪郭 (mean contour)

		EM 回数					
		0	5	10	15	20	25
類似度 (%)	30	47.8	34.8	34.8	34.8	34.8	34.8
	50	39.1	34.8	34.8	34.8	34.8	34.8
	60	34.8	26.1	26.1	26.1	26.1	26.1
	66.7	26.1	21.8	21.8	21.8	21.8	21.8

表 4 輪郭スペクトル (contour spectrum)

		EM 回数					
		0	5	10	15	20	25
類似度 (%)	30	87	91.3	91.3	91.3	91.3	91.3
	50	78.3	91.3	91.3	87	87	87
	60	60.9	82.6	87	82.6	82.6	82.6
	66.7	52.2	78.3	78.3	69.6	65.2	65.2

表 5 ピッチスペクトル (pitch spectrum)

ピッチ細分輪郭, 平滑化ピッチ輪郭, 輪郭スペクトルを使用した結果は, 単純ベイズ法のみでの計算で終わった時が正解率が一番よく, 輪郭スペクトルを使用した結果では, 類似度 30 % 以上の場合, 47.8 % の正解率を示した. ピッチ輪郭とピッチスペクトルを使用した結果では, EM アルゴリズムを 10 回繰り返した場合が, 一番正解率が向上し, ピッチスペクトルを使用した結果では, 類似度 60 %, 66.7 % 以上の場合, 26.1 % も正解率が向上した.

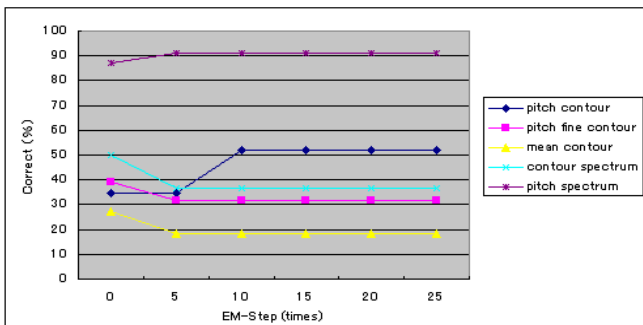


図 5 類似度 30 % 以上

#### 5.4 考 察

結果から明らかのように, ピッチ輪郭とピッチスペクトルの

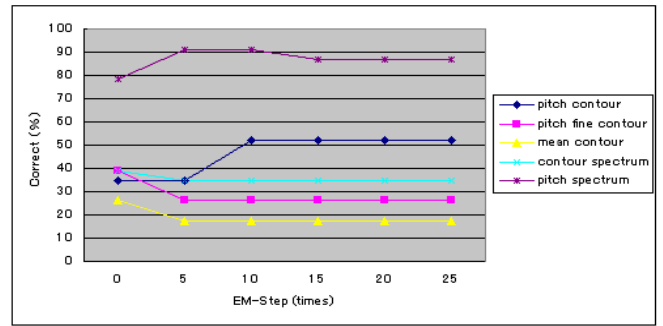


図 6 類似度 50 % 以上

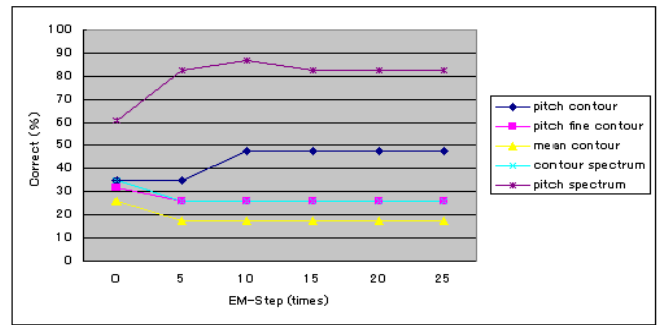


図 7 類似度 60 % 以上

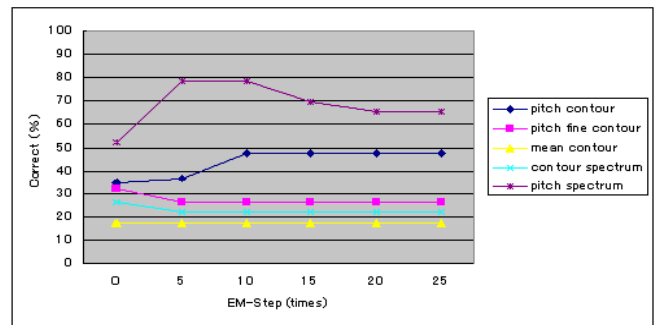


図 8 類似度 66.7 % 以上

場合において EM アルゴリズムによる学習方法は, 結果を向上させることができる. 実際, ピッチ輪郭に関しては, どの類似度の場合でも, EM アルゴリズムの繰り返しが 5 回から 10 回までの間に 15 % 前後の向上, ピッチスペクトルに関しては, 30 % 以上の類似度の場合で 0 回から 5 回までの間に 4 % 前後の向上, 50 % 以上の類似度の場合で 0 回から 5 回までの間に 12 % 前後の向上, 60 % 以上の類似度の場合で 0 回から 5 回までの間に 20 % 強の向上, 5 回から 10 回までの間に 4 % 前後の向上, 66.7 % 以上の類似度の場合で 0 回から 5 回までの間に 25 % 強の向上が見られる. この理由として, 1 つは, 1 回 EM アルゴリズムの計算をした時点で, 間違っただけラベルが付けられた変奏曲は, 正解ラベルの所属確率と, 誤ラベルの所属確率が大きくないことである. これにより, EM アルゴリズムを数回行った結果, 正解のラベルが付けられている. もう 1 つは, 正解ラベルであるにもかかわらず, 類似度が低いために正解率の対象とならなかった変奏曲が, EM アルゴリズムの繰り返し計算によって, 類似していくことになったと考えられる.

他方,他の特徴量はEM アルゴリズムでは効果的でない.この原因として,特徴値が細かいために,先頭4小節の特徴量では類似性を見つけるのが困難で,少ない特徴量の中で雑音がテーマ(主旋律)とマッチしてしまうと,誤ラベルの所属確率が大幅に増加してしまったことが考えられる.また,正解ラベルの所属確率と誤ラベルの所属確率の差が大きく,EM アルゴリズムを20回程度の繰り返しでは,差は縮まるもののラベルの変更までは至らないこともある.

繰り返し回数がある一定数を超えると正解率が低下する.この原因として,EM アルゴリズムで求める分類規則が推定のものであるため,初期値(本実験ではテーマ)が正解のラベルなし旋律と大きく異なると,その影響を繰り返しのたびに誤りを増大させていると考えられる.実際,例としてピッチ細分輪郭を用いた分類で,キラキラ星変奏曲テーマ, Schubert 即興曲テーマ(共にラベルつき訓練データ),ラベルなし旋律のキラキラ星変奏曲2番を比較する.

	contour(回数)						
	D2	D1	S	U1	U2	U3	U4
K.265 テーマ	4	36	38	14	0	2	6
Op.76 テーマ	9	26	54	14	0	11	4
K.265 v2	0	5	8	1	0	0	1

表6 特徴量

単純ベイズ法のみでの計算では, K.265 v2 は,キラキラ星変奏曲と正しく分類された.ただ,EM アルゴリズムを1回計算すると, Schubert 即興曲に分類されてしまう. D1 : S の出現回数を見ると, K.265 テーマ はほぼ 1 : 1, Op.76 テーマと K.265 v2 はほぼ 1 : 2 であることが分かる. EM アルゴリズムによりこの形に従うように変更していく.したがって繰り返しの都度,誤分類をますます強めていく状況を生んでいる.

## 6. 結 論

本研究では,少量のラベル付き楽曲から大量のラベルなし楽曲を分類する方法を提案した.具体的には,5つの特徴量の抽出方法と,単純ベイズ法とEM アルゴリズムを組み合わせることにより,分類精度が改善できるかを評価した.この結果,ピッチ輪郭とピッチスペクトルではEM アルゴリズムの特性を上手く捉えており,分類精度を改善させることができる,ことを確認した.

今後,この手法を変奏曲以外に適用することを考えている.しかし,本研究ではEM アルゴリズムをそのまま使うためには,テーマと変奏曲の類似性がガウス分布的に推定できるという仮定を利用している.しかし,この仮定は変奏曲以外の楽曲で成り立つとは限らず,さらにガウス分布に従うという仮説も検証の必要がある.

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16500070)の支援をいただいた.

- [1] W.J.Dowling: Scale and Contour – two components of a theory of memory for melodies, *Psychological Reviews* 85-4, 1978, pp.341-354
- [2] M.Droettboom et al.: An Approach Towards A Polyphonic Music Retrieval System, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2001
- [3] M.Droettboom et al.: Expressive and Efficient Retrieval of Symbolic Musical Data, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2002
- [4] A.Ghais et al.: Query By Humming, *ACM Multimedia Conf.*, 1995
- [5] D.Grossman, O.Frieder: Information Retrieval – Algorithms and Heuristics, Kluwer Academic Press, 1998
- [6] T.Kageyama et al.: Melody Retrieval With Humming, *ACM Intn'l Computer Music Conf. (ICMC)*, 1993
- [7] Y.Kim et al.: Analysis of A Contour-based Representation for Melody, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
- [8] L.Lie et al.: A New Approach to Query By Humming in Music Retrieval, *Intn'l Conf on Multimedia and Expo (ICME)*, 2001
- [9] C.Meek et al.: Thematic Extractor, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2001
- [10] T.Miura, I.Shioya: Similarities among Melodies for Music Information Retrieval, *ACM Conf. on Information and Knowledge Management(CIKM)*, 2003
- [11] Mitchell, T.: Machine Learning, McGrawHill Companies, 1997
- [12] Nigam,K. McCallum,A. Thrun,I. Mitchell,T.: Text Classification from Labeled and Unlabeled Documents using EM, Kluwer Academic Publishers, Boston. Manufactured in The Netherlands.
- [13] J.Pickens: A Comparison of Language Modelling and Probabilistic Text Information Retrieval Approaches to Monophonic Music Retrieval, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
- [14] J.Pickens, T.Crawford: Harmonic Models for Polyphonic Music Retrieval, *ACM Conf. on Information and Knowledge Management(CIKM)*, 2002
- [15] R.Ramkrishnan, J.Gehrke: Database Management Systems (2nd ed) 2000, McGrawHill
- [16] T. Sonoda, et al.: A WWW-based Melody Retrieval System, *Intn'l Computer Music Conf. (ICMC)*, 1998
- [17] A.L.Uitdenbogerd et al.: Manipulation of Music For Melody Matching, *ACM MultiMedia Conf.*, 1998
- [18] A.L.Uitdenbogerd et al.: Matching Techniques for Large Music Databases, *ACM Multimedia Conf.*, 1999
- [19] A.L.Uitdenbogerd et al.: MUSIC IR: Past, Present and Future, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
- [20] A.L.Uitdenbogerd et al.: Music Ranking Techniques Evaluated, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
- [21] C.Walshaw: abc Version 1.6, <http://www.gre.ac.uk/~c.walshaw/abc2mtex/abc.txt>
- [22] C.Yang: Music Database Retrieval Based on Spectral Similarity, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2001
- [23] 石桁他: 新装楽典 - 理論と実習, 音楽之友社, 2001
- [24] 岩崎学: 不完全データの統計解析, エコノミクス社, 2002
- [25] 上嶋 宏, 三浦 孝夫, 塩谷 勇.: 同義語, 多義語の考慮による文書分類の精度向上, 電子情報通信学会誌 Vol.J87-D-I No.2,2004
- [26] 北研二他: 情報検索のアルゴリズム, 共立出版, 2002
- [27] 新納 浩幸, 佐々木 捻.: EM アルゴリズムの最適ループ回数の予測を用いた語義判別規則の教師なし学習, 情報処理学会論文誌 Vol.44, No.12, 2003