

確率過程に基づく複旋律分類

吉原 幸輝[†] 三浦 孝夫[†]

[†] 法政大学 工学研究科 電気工学専攻 〒184-8584 東京都小金井市梶野町 3-7-2

E-mail: †{i05r3247,miurat}@k.hosei.ac.jp

あらまし 本稿では、複旋律を分類することを提案する。主なアイデアは、各小節から和音の特徴量を抽出し、和音間の類似性を考慮しつつ、マルコフモデルに基づいて分類基準を設定することにある。提案手法は、小節の長さごとに実験で検証され、高い正解率を得ることを示す。

キーワード 旋律分類, 旋律特徴量, マルコフ過程, マルコフモデル分類

Classifying Polyphonic Melodies based on Probability Stochastic Processes

Yukiteru YOSHIWARA[†] and Takao MIURA[†]

[†] Dept. of Elect. & Elect. Engr., HOSEI University 3-7-2, KajinoCho, Koganei, Tokyo, 184-8584 Japan

E-mail: †{i05r3247,miurat}@k.hosei.ac.jp

Abstract In this investigation we propose a new and novel approach for classifying polyphonic melodies. Our main idea comes from *Probability Stochastic Processes* using Markov models where the characteristic features of polyphonic melodies are extracted from each bar. The similarity among harmonies can be considered by means of the features. We show the effectiveness and the usefulness of the approach by some experimental results.

Key words Melody Classification, Melody Features, Markov process, Markov Modeling

1. 前書き

本稿では、複旋律で構成された楽曲から特徴値を抽出し、未知の複旋律を分類していずれからラベルを付与する手法（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^(注1)。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。

しかし、音楽情報検索において多くの手法は単旋律を元にした手法であるのに対し、広く利用可能であり、また面白い音楽の大部分は複旋律で構成される。複旋律の分類が困難な点は、移調や転調、拍子の変化がパートごとに異なり、複旋律の構成は、その異なった音同士が重なり合い構成されるため、何が主旋律かという要素が不明瞭になる点にある。

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うことさえできない。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、主たる単旋律（テーマ）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、例えばクラス所属を確率、ランキング表現など、いずれにすべきか基準が定まらない [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]。

(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[23],[25]。

しかし、今までの手法は単旋律分類には効果的であるが、音の構成が不明瞭である複旋律楽曲に直接適用することができない。それに対し、広く利用可能である音楽データは複旋律楽曲である。従って、複旋律楽曲を扱えるような手法が必要となる。本稿では、あらかじめ楽譜データを処理することにより、複旋律楽曲の分類が確率的で計算できるよう、新しい特徴値を提案する。特徴値の抽出には、ベクトル空間に射影した音のスペクトル表示を用いる。

本研究の目的は、複旋律楽曲において確率過程を適用するのに適した特徴量を抽出し、確率過程に基づき旋律を分類することにある。具体的には、複旋律楽曲において小節単位で特徴量を抽出し、旋律特徴量からマルコフモデルを作成し、生成確率を計算することで旋律の分類を行う。

本稿では、音楽に関する基本知識[24]、および情報検索技術の基本知識を仮定する[5],[27]。

第2章で、準備と必要な定義を述べ、第3章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マルコフ過程およびマルコフモデルの適用方法を示し、この特徴や利点を述べる。第5章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

本稿における旋律(melody)とは、全ての音に対して開始時間や高さ(pitch)、長さ(duration)を表すシンボリックなものである。単旋律(monophony)とは、同時に発生する音はせいぜい1音しかないものを指す。ホモフォニー(homophony)とは、同時に発生する音はせいぜい1組の和音しかないものを指し、和音間の音の重なりはないものとする。複旋律(polyphony)とは、これらの制限がないものとする。

楽曲は複数の音から構成される[24]。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^(注2)。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とも表す。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。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。

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)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。しかしこれらを組み合わせても楽曲の記述は容易ではない。例えば、先頭の数小節を与え何らかの類似曲が想定されたとしても、リズムや表情付け、速度等で大きい差があるかもしれない。

楽曲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)でなされている。本稿では、特徴量の記述を目的とするため、旋律を五線譜を用いた楽譜レベルの表現を利用する。

3. 旋律特徴量

旋律を分類するためには、当該音楽の有する意味表現を考慮する必要がある。しかし旋律を分類するために記述される特徴量としてどのようなものが有効なのであろうか? 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、ベクトル空間モデル[5],[27]を用いて特徴項目ベクトル空間へ写像する。各次元は特徴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。

(注2):以下では1オクターブ高いピッチを小文字で表す。

3.1 単旋律における特徴量

これまで提案された旋律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]。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(注3)。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。

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。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。

旋律が五線譜による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たヒストグラムを言う。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]。ピッチスペクトルは (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。

3.2 複旋律における特徴量

複旋律は、単旋律やホモフォニーのような制限事項がないため、表現される旋律の構成は、無限に近い組み合わせが考えられる。音の構成要素が複雑に関係しているため、単旋律における手法を直接適用することができない。そこで、楽譜をあらかじめ処理することによって対応する方法が考えられる [14]。前処理を行うことで、確率的に計算することが可能となる。

本稿で提案する特徴量は、各小節内で出現する音すべてをベクトルに射影し、音長のスペクトル列で表現することを提案する。この手法による特徴量は、現れる全ての音から構成され、また小節内の音の構成も見ることができるといった利点がある。

[例 1] カエルの歌を複旋律情報で表現すると次のようになる。

$$\{C:1, D:1, E:1, F:1\}, \{C:2, D:2, E:2, F:1\}, \\ \{C:1, D:1, E:2, F:1, G:1, A:1\}, \{E:2, F:2, \\ G:2, A:1\}, \{C:2, E:1, F:1, G:1\}, \{C:4\}, \\ \{C:3, D:1, E:1, F:1\}, \{C:2, D:2, E:2, F:1\}$$



図 1 カエルの歌五線譜

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、各小節内で出現する音を和音とみなす手法を導入し、マルコフモデルによって分類精度を向上できるかを論じる。

各小節内で出現する音すべてをベクトルに射影し、音長のスペクトル列で表現する方法を提案したが、さらに、音長上位 n 音にしぼり、その n 音を和音としたものを特徴量 (特徴量音) と定義する。小節単位の和音を取ることで、主旋律の進行や特徴を得られると考えられる。ここで構成された特徴量音とは、音楽理論とは無関係の新しい和音の要素である。

[例 2] カエルの歌の先頭 4 小節を、小節内の出現音の長さ上位 3 音を和音とすると次のようになる。

$$1 \text{ 小節目: } \{C:1, D:1, E:1, F:1\} = \{CDE, CDF, CEF\} \\ 2 \text{ 小節目: } \{C:2, D:2, E:2, F:1\} = \{CDE\} \\ 3 \text{ 小節目: } \{E:1, F:1, G:1, A:1\} = \{EFG, EFA, FGA\} \\ 4 \text{ 小節目: } \{E:1, F:1, G:1\} = \{EFG\}$$

4. マルコフ過程とマルコフモデル

以下では、あらかじめクラスが与えられている旋律 d を用いて、未知旋律をいずれかのクラスに分類する。本研究では、マルコフモデルに基づいた分類手法を用いる。 d は、特徴量 w_j を用いた多重集合 $\{w_1, \dots, w_m\}$ と仮定する。

4.1 マルコフ過程

旋律 $d = \{w_1, \dots, w_m\}$ 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。旋律 d の同時確率 $P(d)$ は、音符列の出現する確率であり、条件付確率の積で表現することができる。しかし、全ての音符の組み合わせに対して条件付確率を $P(w_j | w_1 \dots w_{j-1})$ を推定することは現実的には難しいため、ある事象の確率がその直前の N 個の事象だけに依存すると仮定すると、近似的な式で表すことができる。これを N 重マルコフ過程という。 N グラムモデルにおける生成確率は

$$P(d_n | d_{n-1}) = P(w_j | w_{j-N}) \quad (1)$$

(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。

となる．つまり，条件付確率 $P(d_n|d_{n-1})$ が求めれば，旋律全体の生成確率の計算が可能となる．しかし，全ての発生する音符列の組み合わせを考えたとき， $P(d_n|d_{n-1})$ の計算は現実的に困難である．例えば，主な三和音のコード列（長・短・増・減の 48 種類）の生成確率を考えた場合，16 小節で考えたとしても，1 旋律あたり 48^{16} 個の値を推定しなければならない．そこで本稿においては，もっとも単純な $N = 1$ のときの単純マルコフ過程を考える．単純マルコフ過程であれば，計算量は特徴量数のみに限定され，計算量は少量で済む．

4.2 マルコフモデルによる文書分類

マルコフモデルとは，現在の状態を限られた過去の状態の積計算で求められる生成確率を導く手順である．一般に，与えられた単語列を $w = \{w_1, \dots, w_m\}, m > 0$ としたとき，求まる生成確率は

$$P(w_1^m) = \prod_{j=1}^m P(w_j|w_1^{j-1}) \quad (2)$$

である．本稿では，過去の状態は 1 つ前までの単語列，つまり直前で発生した特徴量しか影響されない単純マルコフモデルを用いる．単純マルコフモデルは以下のように表現できる：

$$P(d) = P(w_j|w_{j-1}) \quad (3)$$

単純マルコフモデルによる分類は，まず d がクラス c に属する確率 $P(c|d)$ の条件付確率を求めることである．この条件付確率は，生成確率を求めることと同等である．

マルコフモデル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：クラスあり旋律集合，未知旋律集合
- (2) あらかじめクラスの与えられている旋律集合から，マルコフモデル（遷移確率）を生成
- (3) (2) で生成されたマルコフモデル（遷移確率）を用いて，未知旋律のクラス所属確率 $P(c_i|d)$ を推定する
- (4) 出力：各未知旋律のラベル c_k は次のように定める：

$$c_k = \text{ArgMax}_{c \in C} P(c|d)$$

マルコフモデルでは，状態遷移が存在しない確率（0 確率）が選ばれると解が求まらない．

4.3 旋律分類への適用

複旋律楽曲において，マルコフモデルを旋律分類に適用する方法を論じる．

以下では $C = \{c_1, \dots, c_n\}$ はクラス集合を表す．あらかじめクラスを与えられた旋律 e_c はいずれかのクラスに含まれる： $c \in C$ ． D はすべての未知旋律を表し， 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 に属する確率を表す．

$$P(c|d) = \prod_{i=1}^{|m|} P(w_i|w_{i-1}) \quad (4)$$

旋律 d_j が属するクラスは c_k に属する確率 $P(c_k|d_j)$ が最大となるものを選ぶ，即ち最尤法に基づくとする．

本稿では，主題と変奏曲を比較する際，変奏曲内の和音で主題に現れない和音が存在する可能性が十分にある．これまでの手法を適用した場合，得られる確率が 0 になる可能性が非常に高い．この問題を解決するために，通常修正のいずれかの種類が紹介される，しかし，確率値が積計算によってより小さい値になりがちであるため，我々が予想しない分類結果になるかもしれない．従って，容易に値の修正を決定することができない．かわりに本稿では，それぞれの 1 対の特徴量と和音の間に類似性の考えを導入し，確率値を調整する．

2 つの特徴量と和音 w, w' という条件のもとで，我々は類似性 $\text{sim}(w, w')$ を余弦値と定義する：

$$\text{sim}(w, w') = \frac{w \cdot w'}{\|w\| \|w'\|} \quad (5)$$

そして，遷移確率 $P(w_i|w_{i-1})$ に w_{i-1} から w_i までに起こるすべての可能な和音の上に類似性をかけあわした値の和集合を $P'(w_i|w_{i-1})$ と定義する．

$$P'(w_i|w_{i-1}) = \sum_g P(w_i|g) \times \text{sim}(g, w_{i-1})$$

この値は，和音 g だけに関する， w_i を通過し，そして 0 以外の類似性を持っているものであり，マルコフモデルより容易に選ぶことが可能である．また，重みを加えられた生成確率は次で与えられる．余弦値の計算は，特徴を取り出す全ての旋律に対し行う必要がある．

$$P'(c|d) = \prod_{i=1}^{|m|} P'_c(w_i|w_{i-1}) \quad (6)$$

ここで重み付けされた確率 $P'(c|d)$ は通常確率値とは異なるが，最尤推定を応用した分類であることから，あらかじめ分類をするすべての音楽について処理し，最尤推定によって所属クラスを決定する．

[例 3] 特徴量表記によるマルコフモデル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) をクラス分類する ($L_2 = \{d_3\}$) ．ここで利用する特徴量と和音は 3 和音とする．



図 2 モーツァルト:キラキラ星変奏曲



図 3 ベートーベン:交響曲第 9 番

ABC 表記による先頭 8 小節，もしくは 9 小節の特徴値はそれぞれ次のようになる．

d_1 : {CDEF,-}, {EDC,CDEF}, {EFGA,EDC},
 {GFE,EFGA}, {CC,GFE}, {CC,CC},
 {C/2C/2D/2D/2E/2E/2F/2F/2,CC},
 {EDC,C/2C/2D/2D/2E/2E/2F/2F/2}, {CDE}
 d_2 : {CCGG,CCEC}, {AAGG,FCEC}, {FFEE,DBCA},
 {DD3/4E/4C2,FGC}, {CCGG,CCEC}, {AAGG,FCEC},
 {FFEE,DBCA}, {DD3/4E/4C2,FGC}
 d_3 : {FFGA}, {AGFE}, {DDEF}, {F3/2E/2E2},
 {FFGA}, {AGFE}, {DDEF}, {E3/2D/2D2}

小節単位で出現音長をまとめ、その小節における特徴量と和音を決定する。特徴量と和音となり得る候補が複数ある場合は、音の高いものから構成される和音を選択する。また音の進行から、 d_1 と d_2 についてマルコフモデルを作成する。

小節	d_1	d_2	d_3
1	{C:1, D:1, E:1, F:1}	{C:5, E:1, G:2}	{F:2, G:1, A:1}
2	{C:2, D:2, E:2, F:1}	{C:2, E:1, F:1, G:2, A:2}	{E:1, F:1, G:1, A:1}
3	{C:1, D:1, E:2, F:1, G:1, A:1}	{C:1, D:1, E:2, F:2, A:1, B:1}	{D:2, E:1, F:1}
4	{E:2, F:2, G:2, A:1}	{C:13/4, D:7/4, E:1/4, F:1, G:1}	{E:5/2, F:3/2}
5	{C:2, E:1, F:1, G:1}	{C:5, E:1, G:2}	{F:2, G:1, A:1}
6	{C:4}	{C:2, E:1, F:1, G:2, A:2}	{E:1, F:1, G:1, A:1}
7	{C:3, D:1, E:1, F:1}	{C:1, D:1, E:2, F:2, A:1, B:1}	{D:2, E:1, F:1}
8	{C:2, D:2, E:2, F:1}	{C:13/4, D:7/4, E:1/4, F:1, G:1}	{D:5/2, E:3/2}
9	{C:1, D:1, E:1}		

表 1 出現和音

小節	1	2	3	4	5	6	7	8	9
d_1	DEF	CDE	EGA	EFG	CFG	C	CEF	CDE	CDE
d_2	CEG	CGA	EFB	CDG	CEG	CGA	EFB	CDG	
d_3	FGA	FGA	DEF	EF	FGA	FGA	DEF	EF	

表 2 特徴量と和音

	CDE	CEF	CFG	DEF	EFG	EGA	C
CDE	0.5	0	0	0	0	0.5	0
CEF	1	0	0	0	0	0	0
CFG	0	0	0	0	0	0	1
DEF	1	0	0	0	0	0	0
EFG	0	0	1	0	0	0	0
EGA	0	0	0	0	1	0	0
C	0	1	0	0	0	0	0

表 3 マルコフモデルの遷移確率：カエルの歌

マルコフモデルにより、 d_3 にクラス割り当てを行う。 d_3 のクラス所属確率は上記のそれぞれのマルコフモデルに類似度を考慮した値から求められ、それぞれ次の値となる。

$$P(A|d_3) = 0.0181 \times 0.0181 \times 0.0375 = 1.23 \times 10^{-5}$$

	CDG	CEG	CGA	EFB
CDG	0	1	0	0
CEG	0	0	1	0
CGA	0	0	0	1
EFB	1	0	0	0

表 4 マルコフモデルの遷移確率：キラキラ星

$$P(B|d_3) = 0.0417 \times 0.0347 \times 0.0170 = 2.46 \times 10^{-5}$$

この結果、 d_3 が最も近いクラスは B であり、“歓喜の歌” は “キラキラ星” に割り当てられた。

なお、本稿で用いられる旋律内での各特徴量は、その定義に従って与えられる。

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、複旋律で構成されたピアノ変奏曲を用いる。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（バリエーション）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。

最初に、各楽曲をラベル付旋律集合とラベルなし旋律集合に分ける。ラベル付旋律集合は変奏曲テーマ（主旋律）3 曲からなるとし、残り 23 の変奏曲はラベルなし旋律集合を構成する。各楽曲を表す ABC ファイルからラベル付旋律集合は全小節分の特徴量、ラベルなし旋律集合は先頭 4 小節分の特徴量、または先頭 8 小節分の特徴量を取り出す。

本実験で抽出する特徴量は、小節内の出現音の長さ上位 3, 4, 5 音を、その小節の特徴量と和音とする。

5.2 評価方法

本実験では 3 つのクラス “Mozart”, “Schubert”, “Beethoven” を設定し、マルコフモデルを用いた分類器は全 23 変奏曲をいずれかのクラスに分類する。変奏曲が “正しく分類” できたとはそのテーマがラベルの作曲者によって作曲されたときとする。形式的に正解率を次のように定義する：

$$\frac{\text{正しく分類された変奏曲数}}{\text{総変奏曲数}}$$

本稿における総変奏曲数は 23 となる。

5.3 実験結果

変奏曲テーマ（主旋律）から作成したマルコフモデルを図 4 から図 9 に示す。また、特徴量を先頭 4 小節を用いた際と先頭 8 小節を用いた際のマルコフモデルで計算された生成確率を表 5、表 6、図 13 から図 16 に、各テーマに対する正解数、総正解数、正解率を表 7 に示す。

	CDG	CEG	CGA	DEG	EFG	EFB
CDG	0	1	0	0	0	0
CEG	0	0.25	0.5	0	0.25	0
CGA	0	0	0	0	0	1
DEG	0	0	0	0	1	0
EFG	0	0.5	0	0.5	0	0
EFB	1	0	0	0	0	0

図 4 マルコフモデル : Mozart 3 和音

	CFG	CFA	CFB	DFB	DGB	DAB	EGB	FGA
CFG	0	0	0	1	0	0	0	0
CFA	0	0	0	0	1	0	0	0
CFB	0	0	0	1	0	0	0	0
DFB	0	0.444	0.111	0.111	0.111	0	0.222	0
DGB	0	0	0.333	0	0	0.333	0.333	0
DAB	0	0	0	0.5	0.5	0	0	0
EGB	0	0	0.5	0	0	0	0	0.5
FGA	1	0	0	0	0	0	0	0

図 5 マルコフモデル : Schubert 3 和音

	CDA	CEA	CAB	DFA	DFB	DGA	EGA	DF	DB
CDA	0	0	0	0	0	0	0	1	0
CEA	0	0	0	0	0	0	0	1	0
CAB	0	0	0	1	0	0	0	0	0
DFA	0.125	0	0	0.625	0.125	0.125	0	0	0
DFB	0	0.5	0.25	0	0	0	0	0	0.25
DGA	0	0	0	1	0	0	0	0	0
EGA	0	0	0	1	0	0	0	0	0
DF	0	0	0	0	1	0	0	0	0
DB	0	0	0	0	1	0	0	0	0

図 6 マルコフモデル : Beethoven 3 和音

	CDEG	CDFG	CFG	DEFG	DEGB	EFAB	CEG
CDEG	0	0	0	0	0	0	1
CDFG	0	0	0	0.667	0	0	0.333
CFG	0	0.2	0	0	0	0.8	0
DEFG	0.5	0	0	0	0.5	0	0
DEGB	0	0	0	1	0	0	0
EFAB	0	1	0	0	0	0	0
CEG	0	0	1	0	0	0	0

図 7 マルコフモデル : Mozart 4 和音

	CDFA	CDFB	CEFA	CEFB	CEGB	CFG	CFGB	DEGB	DFA	DFGB	DFAB	DGAB	DFB
CDFA	0	0	0	0	0	0	0	0	0	0	0	0	0
CDFB	0	0	0	0	0	0	0	0	0	0	0	0	0
CEFA	0	0	0	0	0	0	0	1	0	0	0	0	0
CEFB	0	0	0	0	0	0	0	0	0	0	0	0	1
CEGB	0	0	0	0	0	0	0.5	0	0.5	0	0	0	0
CFG	0	0	0	0	0	0	0	0	0	1	0	0	0
CFGB	0	0	0	0	0	0	0	0	0	0	1	0	0
DEGB	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0
DFA	0	0	0	0	0	1	0	0	0	0	0	0	0
DFGB	0	0	0	0	1	0	0	0	0	0	0	0	0
DFAB	0	0.2	0	0	0	0	0	0	0	0	0.2	0.4	0.2
DGAB	0	0	0	0	1	0	0	0	0	0	0	0	0
DFB	0.5	0	0.5	0	0	0	0	0	0	0	0	0	0

図 8 マルコフモデル : Schubert 4 和音

全ての楽曲において、4 小節より 8 小節の特徴量を用いた場合に、同等あるいは高い正解率を示した。特に 3 和音のときに "Schubert" の正解率が 60 % も向上した。また、4 小節においても "Mozart" は高い正解率を得た。

5.4 考察

結果から明らかのように、特徴量の小節数を増やすことで結果を向上させることができる。実際、4 小節と少ない小節数で 73.9 % の正解率だったのに対し、8 小節においては 91.3 % と 17.4 % の向上が見られた。この理由として、特徴値の取り方が考えられる。変奏曲が他の主題と類似していることから間違った分類が行われた。具体的には、"Schubert" テーマの第 3 変奏曲は、{CFB}{CEF}{CEF,CFB}{CFB}{CFB}{CEF}{CEB}{CFB} という和音進行をとる。この和音進行で出現する音は C, E, F, B のみである。先頭 4 小節では、"Mozart" テーマの方が "Schubert" テーマと同じ和音が多く現れたため、正解との判別

	CDFB	CDGA	CEGA	CEAB	DEGA	DFA	DFA	DFB	DF	DB
CDFB	0	0	0	0.5	0	0	0	0	0	0.5
CDGA	0	0	0	0	0	0	0	0	0	1
CEGA	0	0	0.5	0	0	0	0.5	0	0	0
CEAB	0	0	0	0	0	0	1	0	0	0
DEGA	0	0	0	0	0	1	0	0	0	0
DFA	0	0	0	0	0	0	1	0	0	0
DFA	0.143	0.143	0	0	0.143	0	0.571	0	0	0
DFB	0	0	1	0	0	0	0	0	0	0
DF	0	0	0	0	0	0	0	1	0	0
DB	0	0	0	0	0	0	0	0	1	0

図 9 マルコフモデル : Beethoven 4 和音

	CDEFG	CDEGB	CEFGA	DEFAB	DEFG	CEG
CDEFG	0	0	0	0	0.667	0.333
CDEGB	0	0	0	0	0.5	0.5
CEFGA	0	0	0	1	0	0
DEFAB	1	0	0	0	0	0
DEFG	0	1	0	0	0	0
CEG	0	0	1	0	0	0

図 10 マルコフモデル : Mozart 5 和音

	CDEFA	CDEGB	CDFA	CDFA	CEFGA	CEFGB	CEGAB	DEFGB	DFGAB	DFGB	DFB
CDEFA	0	0	0	0	0	0	0	1	0	0	0
CDEGB	0	0	1	0	0	0	0	0	0	0	0
CDFA	0	0	0	0	1	0	0	0	0	0	0
CDFA	0	0	0	0.4	0	0	0	0	0.4	0	0.2
CEFGA	0	0	0	0	0	0	0	0	0	1	0
CEFGB	0	0	0	0.5	0	0	0	0	0	0	0.5
CEGAB	0	0	0	0	0	1	0	0	0	0	0
DEFGB	0	0	0	0.5	0	0.5	0	0	0	0	0
DFGAB	0	1	0	0	0	0	0	0	0	0	0
DFGB	0	0	0	0	0	0	1	0	0	0	0
DFB	1	0	0	0	0	0	0	0	0	0	0

図 11 マルコフモデル : Schubert 5 和音

	CDEGA	CDEAB	CDFA	DEFGA	DEGAB	CEGA	DFA	DFB	DF	DB
CDEGA	0	0	0	0	0	0	0	0	1	0
CDEAB	0	0	0	0	0	0	1	0	0	0
CDFA	0	0.5	0	0	0	0	0	0	0	0.5
DEFGA	0	0	0	0	0	0	1	0	0	0
DEGAB	0	0	0	1	0	0	0	0	0	0
CEGA	0	0	0	0	0	0.5	0.5	0	0	0
DFA	0.143	0	0.143	0	0.143	0	0.571	0	0	0
DFB	0	0	0	0	0	1	0	0	0	0
DF	0	0	0	0	0	0	0	1	0	0
DB	0	0	0	0	0	0	0	0	1	0

図 12 マルコフモデル : Beethoven 5 和音

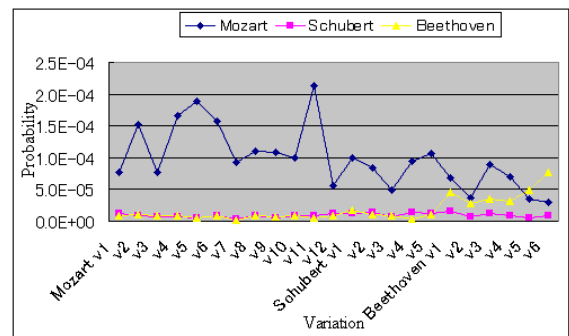


図 13 マルコフモデルによる生成確率: 4 和音, 4 小節

が難しかったと考えられる。8 小節に拡張した場合、"Mozart" テーマには "Schubert" のテーマとは類似度が低い和音が現れ、生成確率が大きく減少したことで正解へと向上したと考えられる。なお、表 8 における下線部は一致する和音を指す。

正しく分類されなかった変奏曲の特徴として、いくつかの原因が挙げられる。1 つは、変調によるものである。4 小節、8 小節において間違いとなっている "Mozart" テーマの第 12 変奏曲は、主題が $\frac{4}{4}$ 拍子に対し、変奏曲は $\frac{12}{8}$ 拍子となっている。調子が変わることで、小節単位による特長量の取り方では、小節をまたがった場合に音のずれをカバーできていないことが分かる。

	Mozart	Schubert	Beethoven
Mozart v1	7.497×10^{-5}	1.292×10^{-6}	6.384×10^{-7}
v2	4.895×10^{-5}	3.066×10^{-6}	2.140×10^{-6}
v3	5.604×10^{-5}	2.910×10^{-6}	3.034×10^{-6}
v4	2.257×10^{-4}	1.923×10^{-6}	9.237×10^{-7}
v5	8.335×10^{-5}	1.898×10^{-6}	1.530×10^{-6}
v6	1.827×10^{-4}	2.331×10^{-6}	7.621×10^{-7}
v7	6.125×10^{-5}	2.970×10^{-6}	9.249×10^{-7}
v8	6.417×10^{-5}	1.782×10^{-5}	1.057×10^{-6}
v9	4.434×10^{-5}	1.229×10^{-5}	9.714×10^{-7}
v10	3.936×10^{-5}	8.525×10^{-6}	1.003×10^{-6}
v11	6.801×10^{-5}	9.203×10^{-6}	1.814×10^{-6}
v12	2.415×10^{-5}	8.412×10^{-6}	1.069×10^{-6}
Schubert v1	1.961×10^{-5}	1.608×10^{-5}	1.155×10^{-5}
v2	6.319×10^{-6}	3.563×10^{-5}	8.079×10^{-6}
v3	1.548×10^{-5}	2.418×10^{-5}	2.429×10^{-6}
v4	6.221×10^{-5}	1.854×10^{-5}	4.496×10^{-7}
v5	3.212×10^{-5}	2.091×10^{-5}	2.068×10^{-6}
Beethoven v1	5.943×10^{-6}	9.357×10^{-6}	6.138×10^{-5}
v2	7.563×10^{-6}	1.380×10^{-5}	6.987×10^{-6}
v3	4.065×10^{-5}	1.003×10^{-5}	2.370×10^{-5}
v4	5.372×10^{-5}	2.674×10^{-6}	7.352×10^{-6}
v5	6.174×10^{-6}	6.447×10^{-6}	2.851×10^{-5}
v6	2.026×10^{-6}	1.878×10^{-5}	5.476×10^{-5}

表5 マルコフモデルによる生成確率: 3和音, 4小節

	Mozart	Schubert	Beethoven
Mozart v1	2.516×10^{-10}	3.959×10^{-14}	4.648×10^{-15}
v2	9.985×10^{-11}	2.385×10^{-13}	5.345×10^{-14}
v3	2.036×10^{-10}	1.764×10^{-13}	9.124×10^{-14}
v4	4.047×10^{-9}	3.852×10^{-14}	4.682×10^{-15}
v5	2.788×10^{-10}	6.669×10^{-14}	3.353×10^{-14}
v6	2.163×10^{-9}	1.132×10^{-13}	5.755×10^{-15}
v7	1.560×10^{-10}	4.688×10^{-14}	8.624×10^{-15}
v8	7.913×10^{-11}	9.745×10^{-12}	1.189×10^{-14}
v9	1.153×10^{-10}	3.767×10^{-12}	1.054×10^{-14}
v10	1.231×10^{-10}	7.571×10^{-13}	5.520×10^{-15}
v11	2.391×10^{-10}	2.059×10^{-12}	5.617×10^{-14}
v12	2.121×10^{-11}	2.293×10^{-12}	7.002×10^{-14}
Schubert v1	2.394×10^{-12}	1.285×10^{-11}	5.642×10^{-13}
v2	3.857×10^{-13}	4.468×10^{-11}	2.513×10^{-12}
v3	8.988×10^{-12}	1.010×10^{-11}	5.125×10^{-14}
v4	1.977×10^{-11}	2.134×10^{-11}	2.417×10^{-14}
v5	6.763×10^{-12}	1.354×10^{-11}	7.304×10^{-14}
Beethoven v1	5.455×10^{-13}	1.606×10^{-12}	3.163×10^{-11}
v2	7.135×10^{-13}	4.240×10^{-12}	1.785×10^{-12}
v3	5.780×10^{-11}	2.013×10^{-12}	2.207×10^{-12}
v4	3.143×10^{-12}	2.966×10^{-13}	5.968×10^{-12}
v5	8.236×10^{-13}	1.124×10^{-12}	2.356×10^{-11}
v6	2.051×10^{-13}	1.486×10^{-11}	1.082×10^{-10}

表6 マルコフモデルによる生成確率: 3和音, 8小節

このことから4小節という少ない小節数の場合、特徴的な差を得られるような特徴値の出現パターンが現れていないことが

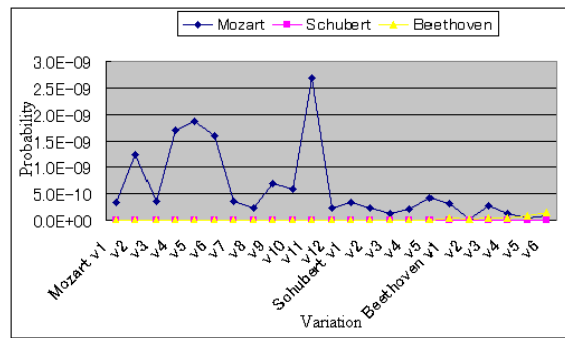


図14 マルコフモデルによる生成確率: 4和音, 8小節

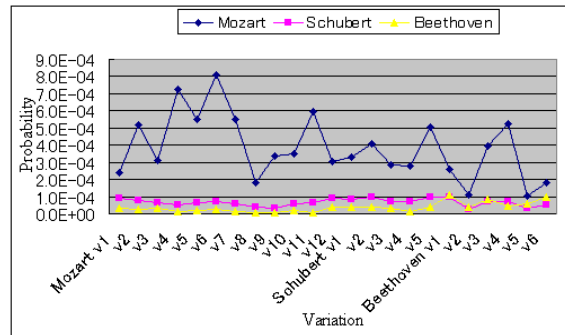


図15 マルコフモデルによる生成確率: 5和音, 4小節

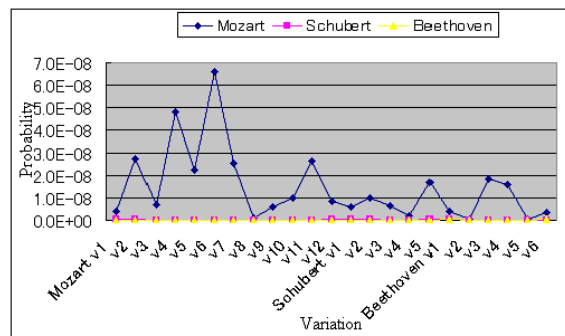


図16 マルコフモデルによる生成確率: 5和音, 8小節

()内は和音数	正解数			合計	正解率
	Mozart	Schubert	Beethoven		
(3) 4小節	12/12	2/5	3/6	17/23	73.9 %
(3) 8小節	12/12	5/5	4/6	21/23	91.3 %
(4) 4小節	12/12	0/5	2/6	14/23	60.9 %
(4) 8小節	12/12	0/5	2/6	14/23	60.9 %
(5) 4小節	12/12	0/5	0/6	12/23	52.2 %
(5) 8小節	12/12	0/5	0/6	12/23	52.2 %

表7 正解率と正解数

	4小節	8小節
Schubert v3	{CFB} {CEF}	+{CEB}
Mozart テーマ	{DFB} {CFA} {DEB} {DGB} {CEF} {CEB} {CFB}	+{DFB}{DAB}
Schubert テーマ	{CEG} {CGA} {CEF} {DEF} {EFA} {EFB} {CDF} {CDG}	追加なし

表8 出現和音一覧

ら、本稿で提案する特徴量では正解率向上の限界があると思われる。しかし、本実験において4小節でも78.3%の正解率が得られていることから、本手法は複旋律分類において非常に有効であると考えられる。

また逆に、特徴量と音の大きさを拡大していくにつれて、正解率が低下していることが分かる。そして、正解の旋律が”Mozart”クラスに集中している。これは、構成される音が多くなるにつれて、特徴的な情報までが失われてしまっていると考えられる。実際、{C:2,D:2,E:2,F:1,G:1}と{C:1,D:1,E:2,F:2,G:2}の和音では、上位3音ならば{CDE}と{EFG}となり、それぞれは違う構成であることが分かるが、上位5音ならばいずれの特徴量と音も{CDEFG}となり、異なる旋律同士がおなじレベルを与えられることになる。その結果として、本実験では類似した旋律はすべて”Mozart”とみなされてしまい、それ以外の曲の正解率があがらなかったと考えられる。これより、特徴量と音は3和音程度で構成することがよいと考えられる。

正解率の比較として、単旋律曲の分類において最も正解率の高かった特徴量を先頭4小節でピッチスペクトラムを用い、分類手法に単純ベイズ、およびEMアルゴリズムを用いた場合の正解率[23]を表9に示す。

		EM loop					
		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	87.0	87.0	87.0

表9 正解率：単旋律、ピッチスペクトル

単純ベイズ法(EM loop 0回)で最高87%、EMアルゴリズムで最高91.3%を示したが、本研究の手法においても、単旋律から複旋律へと複雑化しているにもかかわらず特徴量抽出を8小節用いた場合の正解率が91.3%と、単旋律での分類と比較しても互角である。本研究の手法は、複旋律分類において有効な抽出方法といえよう。

6. 結 論

本研究では、あらかじめクラスが与えられた少量の複旋律楽曲から多く存在する未知旋律を分類する方法を提案した。具体的には、複旋律情報の楽曲を小節単位の音情報を和音とした特徴量も用いることと、マルコフモデルに基づいて計算することにより、分類結果で高い正解率が得られた。この結果、複旋律情報においても高い分類精度を得ることを確認した。

本研究では単純マルコフモデルを用いた推定を行っているが、これは変奏曲分類の楽曲で成り立つかの検証の必要がある。

謝辞

本研究の一部は文部科学省科学研究費補助金(C)(課題番号16500070)の支援をいただいた。

文 献

[1] Dowling W.J.: Scale and Contour – two components of a theory of memory for melodies, *Psychological Reviews* 85-4, 1978, pp.341-354
 [2] Droettboom, M. et al.: An Approach Towards A Polyphonic Music Retrieval System, *Intn'l Symp. on Music Information*

Retrieval (ISMIR), 2001
 [3] Droettboom, M. et al.: Expressive and Efficient Retrieval of Symbolic Musical Data, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2002
 [4] Ghais, A. et al.: Query By Humming, *ACM Multimedia Conf.*, 1995
 [5] Grossman, D., O'Frieder: Information Retrieval – Algorithms and Heuristics, Kluwer Academic Press, 1998
 [6] Kageyama, T. et al.: Melody Retrieval With Humming, *ACM Intn'l Computer Music Conf. (ICMC)*, 1993
 [7] Kim, Y. et al.: Analysis of A Contour-based Representation for Melody, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
 [8] Lie, L. et al.: A New Approach to Query By Humming in Music Retrieval, *Intn'l Conf on Multimedia and Expo (ICME)*, 2001
 [9] Meek, C. et al.: Thematic Extractor, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2001
 [10] Miura, T. and Shioya, I.: 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] Pickens, J.: 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] Pickens, J. and Crawford, T.: Harmonic Models for Polyphonic Music Retrieval, *ACM Conf. on Information and Knowledge Management (CIKM)*, 2002
 [15] Ramakrishnan, R. and Gehrke, J.: Database Management Systems (2nd ed) 2000, McGrawHill
 [16] Sonoda, T. et al.: A WWW-based Melody Retrieval System, *Intn'l Computer Music Conf. (ICMC)*, 1998
 [17] Uitdenbogerd, A.L. et al.: Manipulation of Music For Melody Matching, *ACM MultiMedia Conf.*, 1998
 [18] Uitdenbogerd, A.L. et al.: Matching Techniques for Large Music Databases, *ACM Multimedia Conf.*, 1999
 [19] Uitdenbogerd, A.L. et al.: MUSIC IR: Past, Present and Future, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
 [20] Uitdenbogerd, A.L. et al.: Music Ranking Techniques Evaluated, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2000
 [21] Walshaw, C.: abc Version 1.6, <http://www.gre.ac.uk/~c.walshaw/abc2mtex/abc.txt>
 [22] Yang, C.: Music Database Retrieval Based on Spectral Similarity, *Intn'l Symp. on Music Information Retrieval (ISMIR)*, 2001
 [23] Yoshiwara, Y. and Miura, T.: Melody Classification Using EM Algorithm. *COMPSAC*, pp. 204-210, 2005
 [24] 石桁他：新装楽典 - 理論と実習，音楽之友社，2001
 [25] 岩崎学：不完全データの統計解析，エコノミクス社，2002
 [26] 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
 [27] 北研二他：情報検索のアルゴリズム，共立出版，2002
 [28] 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