

区間線形動的システムによる時系列イベントの学習と識別

堤 公孝[†] 川 嶋 宏 彰[†] 松 山 隆 司[†]

本研究の目的は音声や映像といった時系列データを解析することで人の発話や表情といった動的イベントの認識を可能とすることである。従来、動的イベントの認識には大きく分けて二つの方法が用いられてきた。線形動的システムのように連続状態モデルに基づく方法と HMM のように離散状態モデルに基づく方法である。連続状態モデルに基づく方法は信号の時間的な変化を詳細に記述出来るが、複雑な構造を持つイベントの識別には適さない。一方、離散状態モデルに基づく方法は複雑な構造を持つイベントを識別できるが信号の時間的な変化を詳細に記述することは出来ない。そこで、本研究では連続状態モデルを用いて信号の変化を詳細に記述しながら離散状態モデルを用いて複雑な構造を持つイベントの識別を可能とするモデルを提案する。これら二つのモデルは、特定の動的システムに従う時間区間をもとに統合する。これにより、入力信号の時間的な変化を詳細に記述しながら複雑な構造を持つ動的イベントの認識が可能となる。本研究では提案モデルを発話イベントを用いて検証した。

Time-varying Event Recognition by An Interval Linear Dynamical System

KIMITAKA TSUTSUMI,[†] HIROAKI KAWASHIMA[†]
and TAKASHI MATSUYAMA[†]

Time-varying event recognition is one of the most important problems to realize intelligent machines, which can understand man's behavior, utterance, intention, and expression. Two models have been separately proposed to deal with this problem. One is based on a continuous state model such as linear dynamical systems, and the other is based on a discrete state model such as hidden Markov models. While a continuous state model can describe continuous changes of time-varying patterns, it is difficult for this model to represent complicated and structured event. A discrete state model can describe structured event, but it is difficult for this model to represent continuously varying patterns. To solve these disadvantages, we propose a hybrid model named an Interval Linear Dynamical System. This system integrates continuous and discrete state models based on temporal intervals represented by the same linear dynamical system.

1. はじめに

本研究の目的は音声や映像といったデータを解析することで人の表情や発話といった動的イベントを認識することである。これを応用すればコンピュータが人の顔の表情やしぐさ、発話などから、その意図を理解したり、逆に人工的に表情を作り出したりという機能を実現可能である。このようなマルチメディアに基づいて動的イベントの記述、認識を行なう機能を備えた認識システムを実現することで人とより自然につき合うことの出来る情報システムの実現を目指す。

従来、時系列パターン認識の手法は大きく分けて以下の二つの手法が用いられてきた。

- (1) 信号処理的な時系列解析を用いる方法
- (2) 離散状態モデルの状態遷移を用いる方法

信号処理的な時系列解析には連続 DP⁹⁾ やリカレントニューラルネットワーク^{5),6)} などがある。これらは時系列信号の詳細な変化をカメラのサンプリングレートなどの物理的な時間に対して詳細に記述できるが、一方で複雑な構造を持った時系列パターンは表現できないという欠点がある。また、離散状態モデルを用いる方法には、パターンをクラスタリングなどにより、シンボル化し、オートマトンや HMM などの離散状態モデルへの入力とする手法がある^{1),2),4)}。これらは、文法など複雑な構造を持つ時系列パターンをモデル化できるが、一方で連続的な変化を記述することが出来ない。両者の問題を解決する方法として、基本的には連続状態を持った動的システムを用いて信号レベルでパターンの変化を記述しておき、さらにその変化の構造を離散状態遷移モデルにより、記述するようなモデルも提案されている^{3),8)}。これは、複数の動的システムを用意して、それぞれを記号的な状態に割り当て、それらをマルコフ過程に基づいて切替えてゆくことに

[†] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

よって、一つの動的システムでは表現できないような時系列パターンを扱うというものである。しかし、本来、物理時間とは切り離すべきである離散状態モデルが物理時間を軸にして状態遷移が記述されているため、同じ状態に留まる時間が短いほど確率的には高くなるという不自然なモデルとなっている。これは通常のHMMについても同じことが言える⁷⁾。

そこで、本研究では時間的な区間に基づいて連続状態モデルと離散状態モデルを統合した区間線形動的システムを提案する。これは離散状態モデルが持つ時間を物理的な時間ではなく、特定の動的システムに従う区間を単位にして状態遷移を記述するモデルである。このようなモデルの有効性を音声などの時系列パターンの識別によって評価した。

2. 区間線形動的システムによる時系列イベントの識別

区間線形動的システム (Interval Linear Dynamical System, ILDS) は、観測される時系列イベントを、線形な動的システムで表現できるような複数の区間からなると仮定して表現するモデルである。システムの動作時には、一つの線形な動的システムに従う区間に区切りながら、区間同士の確率的関係によって観測される時系列信号の尤度を計算できる。この尤度をもちいて、観測される時系列イベントのクラスを識別することが可能である。まず2.1節でILDSの構造および、どのような状態を持つかについて述べる。続く2.2節、2.3節では状態遷移の具体的なモデルについて定式化を行ない、2.4節ではILDSによる時系列イベントの識別方法について述べる。

2.1 区間線形動的システムの状態と観測

ILDSはセグメンテーション層、識別層の二層からなる。(図1) セグメンテーション層は N 個の動的システム $D_n (n = 1, \dots, N)$ からなり、観測データに対して内部状態（以下での述べる連続状態）を変化させる。各動的システムは、観測データの変化を常に予測しようとする。このとき、観測データを特になく表現できる動的システムが活性化する。識別層はどの動的システムが活性化しているかをマクロな状態としてもつ。このとき、観測データの変化および、あらかじめ学習された遷移確率に基づいて動的システムの間を遷移していくことができる。以下、観測および各層の状態について定義する。

観測: マイクやカメラでキャプチャーしたデータを特徴抽出することで時系列特徴ベクトルが得られる。この特徴ベクトルを以下では観測ベクトルと呼ぶ。観測ベクトルの次元を k 次元とすると、観測ベクトルが表現されるベクトル空間 $\mathcal{R}^k$ を観測空間と呼ぶ。観測ベクトルの得られるタイミングは、カメラのサンプリングレートやスペクトル解析の処理レートなどに従う

ものとする。以下では時系列を全てこの離散時間で考え、整数 t ($t \geq 0$) を用いて数列表現するものとする。また、この離散時刻 t を単に時刻と呼ぶことにする。時刻 t で得られる観測ベクトルを y_t のようにあらわせば、時刻1から時刻 L までの観測ベクトル系列は $y_1, \dots, y_L$ のように表される。

セグメンテーション層の連続状態: セグメンテーション層は複数の動的システムからなり、これら動的システムはひとつの共通な n 次元実ベクトル空間 $\mathcal{R}^n$ を状態空間としてもつものとする。セグメンテーション層の状態は、この状態空間中のベクトル $X \in \mathcal{R}^n$ として表現される。セグメンテーション層の状態を「連続状態」、状態空間を特に「連続状態空間」と呼ぶことにする。

識別層の離散状態: 識別層は、セグメンテーション層がもつ動的システムと一対一に対応した状態をもつ。動的システム $D_i (i = 1, \dots, N)$ に対応する状態は記号 $q_i (i = 1, \dots, N)$ で表現する。セグメンテーション層において動的システムのうちのひとつが活性化したとき、識別層の状態は集合 $Q = \{q_1, \dots, q_N\}$ のある要素 $q_i \in Q$ を取るようになる。これを「離散状態」と呼び、集合 Q を「離散状態空間」と呼ぶことにする。

区間線形動的システム全体の状態: 区間線形動的システムはこれら連続・離散という2つの状態を持つが、システム全体として考えた場合の状態空間はこれらの直積空間であり、状態は連続・離散状態を結合したシステム状態 $(X, q_i) (X \in \mathcal{R}^n, q_i \in Q)$ によって定義する。

区間線形動的システム全体の状態分布: 区間線形動的システムによる認識時には、システムの状態は確率分布で表現し、状態遷移は確率過程として表現する。

時刻 t におけるセグメンテーション層の連続状態の確率変数を x_t とすると、これは連続状態空間中の1点となる。

$$x_t = X (X \in \mathcal{R}^n) \quad (1)$$

このとき、連続状態分布 $P(x_t)$ を考えることができる。時刻 t における識別層の離散状態の確率変数を s_t とすると、

$$s_t = q_i (q_i \in Q) \quad (2)$$

である。このとき、離散状態分布 $P(s_t)$ を考えることができる。これは、セグメンテーション層において活性化している動的システムの活性度分布のようなものである。システム全体では、時刻 t において状態分布 $P(x_t, s_t)$ をとることになる。

なお、連続状態の分布は以下の様に表すことができる。

$$P(x_t) = \sum_i P(x_t | s_t = q_i) P(s_t = q_i) \quad (3)$$

この式の意味するところは、連続状態分布とは、各動的システム D_i が寄与する状態分布 $P(x_t | s_t = q_i)$ によつ

て構成されるような混合分布であり、その混合比を決めるのが離散状態分布である。以下では、 $P(x_t|s_t = q_i)$ の分布としてガウス分布を仮定する。したがって、連続状態の分布は混合ガウス分布となる。

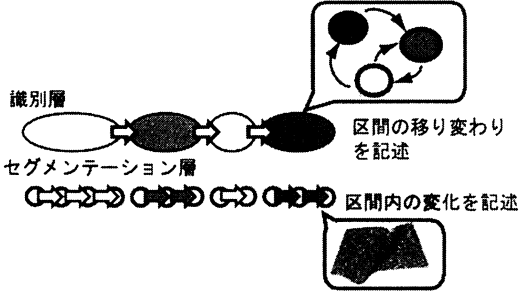


図1 ILDSの構造

2.2 セグメンテーション層の状態遷移

セグメンテーション層を構成する各動的システムは、観測空間から入力する時系列特徴ベクトルと予め学習されたダイナミクスに基づいて連続状態を遷移させる。動的システムの遷移、観測行列はいずれも線形であるとすれば、区間 i を表す動的システムの状態方程式および観測方程式は次式で表される。

$$x_t = F^{(i)}x_{t-1} + \omega_t^{(i)} \quad (4)$$

$$y_t = Gx_t + v_t^{(i)} \quad (5)$$

ここで $\omega_t^{(i)}$ 、 $v_t^{(i)}$ はプロセスノイズおよび観測ノイズである。これらはそれぞれ平均値 0、共分散行列 $Q^{(i)}$ および $R^{(i)}$ の正規分布に従う。すなわち、次式のような確率密度関数を定義することとなる。

$$P(x_t|x_{t-1}, s_t = q_i) = \mathcal{N}(F^{(i)}x_{t-1}, Q^{(i)}) \quad (6)$$

$$P(y_t|x_t, s_t = q_i) = \mathcal{N}(Gx_t, R^{(i)}) \quad (7)$$

すると、通常のカルマンフィルタと同様にガウス・マルコフ過程を仮定することになるため、 $t-1$ まで観測が得られた条件の下で、一期先の状態および観測を推定することができる。

$$P(x_t|y_1^{t-1}, s_t = q_i) = \mathcal{N}(x_{t|t-1}^{(i)}, V_{t|t-1}^{(i)}) \quad (8)$$

$$P(y_t|y_1^{t-1}, s_t = q_i) = \mathcal{N}(Gx_{t|t-1}^{(i)}, GV_{t|t-1}^{(i)}G^T + R^{(i)})$$

ここで、 $x_{t|t-1}^{(i)}$ と $V_{t|t-1}^{(i)}$ はカルマンフィルタの更新式に従う。

区間 $[b, e]$ で観測された系列 $y_b^e = y_b, \dots, y_e$ が同じ動的システム D_i に従うと仮定すると、区間 $[b, e]$ において識別層では離散状態 q_i をとることになる。すなわち、 $s_b^e = q_i$ である。このとき、 y_b^e の尤度は以下

のように計算できる。

$$d_{[b,e]}^{(i)} = P(y_b^e|s_b^e = q_i) = \prod_{t=b}^e P(y_t|y_{t-1}, s_t = q_i) \quad (10)$$

ただし、各動的システムの初期分布として正規分布 $\mathcal{N}(\bar{x}_1^{(i)}, V_1^{(i)})$ を仮定している。

2.3 識別層の状態遷移

あるひとつの動的システムに対応する区間を I_m とし、各区間の始まる時刻を b_m 、終りの時刻を e_m とする。このとき、区間の長さは $\tau_m = e_m - b_m + 1$ となる。いま、全ての区間 I_m ($m = 1, \dots, M$) がいずれかの動的システム D_i ($i = 1, \dots, N$) に従い、観測ベクトル y_t ($t = b_m, \dots, e_m$) が生成されると仮定する。

ひとつの区間がひとつの動的システムに対応し、ある区間から別の区間に遷移する確率は、離散状態遷移モデルの遷移確率によって決まるものとする。すなわち、 m 番目の区間を I_m とし、 $m-1$ 番目の区間を I_{m-1} とすると、その遷移確率は次式で与えられる。ここでは、これを「区間遷移確率」と呼ぶ。区間遷移確率を表す記号として、以下の式で定義される A_{ij} を用いる。

$$P(I_m = q_j | I_{m-1} = q_i) = A_{ij} \quad (11)$$

同じ動的システムに従う区間でも、長さが異なれば、その出現確率も異なると考えられる。そこで、ある動的システムに従う区間 I の継続時間に注目する。この区間が時間 l だけ継続する確率が分布 $P(l|I)$ に従うと考え、これを「区間長分布」と呼ぶ。区間長分布を表す記号として、以下の式で定義される $B_{i\tau}$ を用いる。

$$P(l_m = \tau | I_m = q_i) = B_{i\tau}(\tau) \quad (12)$$

2.4 時系列イベントの識別アルゴリズム

時系列イベントの識別とは、得られた観測信号系列 $y_1^T = y_1, \dots, y_L$ が、クラス $c = 1, \dots, C$ のいずれに属するかを決定する処理である。ここでは、ベイズの決定則にしたがって識別を行うものとする。すなわち、観測された系列に対して各クラスの事後確率 $P(c|y_1^L)$ を求め、最も高い確率を与えるクラスを結果として出力することにする。ここで、 c はクラスのラベルである。このとき、ベイズの定理より事後確率は次式のように求まる。

$$P(c|y_1^L) = P(y_1^L|c) \frac{P(c)}{P(y_1^L)} \quad (13)$$

今は、 $P(c)$ が各クラスについて等確率であると仮定する。この場合には $P(y_1^L|c)$ のみが必要となる。この $P(y_1^L|c)$ を計算するために、クラス毎に予め学習された区間線形動的システムを用意し、系列 y_1^L が得られたときに、それぞれが $P(y_1^L|c)$ を計算するとすると、観測系列 y_1^L を最もうまく表現出来る ILDS が最も高い値となり、それがどの ILDS であったかを見ることで識別結果が得られる。

以下では、観測系列 y_1^L が得られたときに $P(y_1^L|c)$ を計算する方法について述べる。なお、以下ではひと

つのクラスに注目したときの尤度を考えるため、クラス c のラベル c は省略する。

まず、時刻 t において同じ離散状態が続いている持続時間を確率変数 l_t で表す。すると、時刻 t において離散状態 q_j が時間 τ だけ続いている確率は $P(s_t = q_j, l_t = \tau)$ のように表すことができる。

今、時刻 1 から t までに観測系列 $y_1^t = y_1, \dots, y_t$ が得られたとする。このとき、時刻 t において離散状態 $s_t = q_i$ が時間 $l_t = \tau$ だけ続いているとすると、この結合確率は $P(y_1^t, s_t = q_i, l_t = \tau)$ で表される。クラス c の ILDS に対するこの観測系列の尤度は、時刻 t でとりうる全ての s_t, l_t の値に関して、この結合確率を足すことで計算できる。

$$P(y_1^t) = \sum_{q_i} \sum_{\tau} P(y_1^t, s_t = q_i, l_t = \tau) \quad (14)$$

したがって、尤度計算には上式右辺の結合確率を求める必要があるが、これは動的プログラミングを用いて以下の様な漸化式で計算することができる。

$$P(y_1^t, s_t = q_j, l_t = \tau) = \sum_{\substack{q_j \in Q \\ (j \neq i)}} \sum_{\tau} A_{ij} B_j(\tau) d_{[t-\tau+1, t]}^{(j)} P(y_1^{t-\tau}, s_{t-\tau+1} = q_i, l_{t-\tau+1} = \tau) \quad (15)$$

ここで、 $d_{[t-\tau+1, t]}$ は区間 $[t-\tau+1, t]$ における観測データの尤度である (2.2 節を参照のこと)。 $A_{ij}, B_j(\tau)$ は、それぞれ、離散状態 i が続く区間から離散状態 j が続く区間への区間遷移確率、および離散状態 j が時間 τ 続く確率である (2.3 節を参照のこと)。

3. セグメンテーション層の学習

3.1 線形動的システムのパラメータ決定法

学習時には、与えられたデータから、用意すべき動的システムの数および、各動的システムのパラメータを求める必要がある。しかし、動的システムのパラメータを求めるためには同一の動的システムに従う区間を予め区切っておく必要があるが、区間を区切るためには動的システムのパラメータが求まっている必要がある。この問題を解決するために次のような方法を用いた。

最初は時系列を出来るだけたくさんの区間に区切ってそれぞれの区間を表現する動的システムを用意する。この結果、求まった動的システムをダイナミクスの近さに基づいて併合していく。最終的には動的システムの数が 1 個になるが、併合の途中で 3.4.1 節で述べる基準が満たされる時の動的システムの数とパラメータを記録する。そのようにして得られた動的システムの組のうち、数が最小となるものを取り出し、動的システムの数とパラメータを決定する。ここで、これらの動的システムをダイナミクスの近さに基づいて併合していくには、ダイナミクス間の類似度を計る尺度が必要となる。動的システム間の類似度の基準として、3.3 節で動的システム間距離尺度を定義する。ま

た、これを用いたクラスタ併合アルゴリズムについて 3.5 節で述べる。なお、以下では簡単化のため、観測行列 $G^{(i)} = I, R^{(i)} = O$ であると仮定する。この時システムのパラメータは $\theta^{(i)} = (F^{(i)}, Q^{(i)})$ となる。

3.2 同一の動的システムに従う場合のパラメータの算出法

特徴ベクトル系列 $x_t, \dots, x_{t+L-1}$ が一つの動的システムに従うと仮定する。この時、各時刻の列ベクトルを横一列に並べて出来る行列をそれぞれ $X_{t+1} = (x_{t+1}, x_{t+2}, \dots, x_{t+L-1})$, $X_t = (x_t, x_{t+1}, \dots, x_{t+L-2})$ とする。このとき、隣り合う時刻において

$$x_{t+1} = Fx_t + \omega_t \quad (16)$$

が成り立つが、この ω_t が最も小さくなるような F は、以下の式によって計算できる。

$$F = X_{t+1} X_t^\dagger \quad (17)$$

ここで、 $X_t^\dagger$ は、 X_t の疑似逆行列である。

3.3 動的システム間の距離尺度

動的システムの併合のためには、二つの動的システム間の距離を定義する必要がある。しかし、パラメータ同士の単純な比較をするだけでは不十分である。なぜなら同じ変化を表現するパラメータが複数存在する可能性があるからである。そこで、ここでは動的システム間の距離を、一方の動的システムを求める際に用いたデータを他方の動的システムでどれだけよく表現出来るかによって定義する。

いま、学習データの時系列特徴ベクトルが適当な区間に区切られているとする。すると、前節で述べた方法で各区間において、動的システムのパラメータを求めることが出来る。ただし、一般には離れた区間が同じ動的システムで表現出来る場合がある。そこで、区間ではなく区間集合から動的システムを計算することにする。動的システム D_i を計算するのに用いた区間集合を $\mathcal{I}_i (i = 1, \dots, N)$ とすると動的システム D_i によって区間集合 $\mathcal{I}_j$ を予測した時の予測誤差は以下の式で与えられる。

$$L(D_i, \mathcal{I}_j) = \sum_{x_t \in \mathcal{I}_j} \frac{1}{N_j} \|x_{t+1} - F^{(i)} x_t\|^2 \quad (18)$$

ただし、 $F^{(p)}$ は D_i のパラメータである。区間集合のうち、似たダイナミクスをもつ二つの区間集合を選ぶことを考える。 $L(D_i, \mathcal{I}_j)$ が小さい時ほど、 D_i のパラメータは $\mathcal{I}_j$ の系列をうまく予測出来ることから $L(D_i, \mathcal{I}_j)$ の最小値を与えるペア (i, j) を見つければ良さそうである。しかし、 $L(D_i, \mathcal{I}_j)$ は (i, j) に対して非対称である。そこで動的システム D_i と D_j の距離として次式を用いることとする。

$$Dist(i, j) = L(D_i, \mathcal{I}_j) + L(D_j, \mathcal{I}_i) \quad (19)$$

この値が小さい動的システムは似たダイナミクスを持つと考えられるため、併合する動的システムの組 (i^*, j^*) は次式により決定する。

$$(i^*, j^*) = \arg \min_{(i^*, j^*)} Dist(i, j) \quad (20)$$

このとき、区間集合 I_i と I_j を最も似たダイナミクスを持つものとして併合することになる。(図 2)

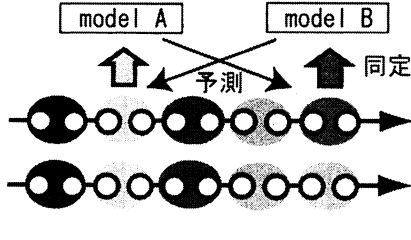


図 2 動的システム間の距離は互いによく系列を予測出来るかによって定義する。

3.4 動的システム数の決定方法

前節では動的システム間の併合方法について述べた。動的システム間の併合は動的システム間距離尺度が最小になる動的システムの組を併合することによって行なわれるが、あらゆる動的システムの組合せを用いて求められた動的システム間距離尺度には常に最小値が存在する。従って動的システムの数が 1 になるまで動的システムの併合が停止しない。本節では動的システムの数を決定するための基準を定義し、この基準のもとで動的システムの数を決定するアルゴリズムについて述べる。

3.4.1 動的システム数の決定基準

各動的システム内での予測誤差の平均 $G_W(N)$ の値はなるべく小さくかつ、異なる動的システム間での距離の平均値 $G_B(N)$ はなるべく大きくなるような基準で動的システムの数を決めることが望ましい。そこで判別分析の考え方と同様に、これらの $G_B(N)$ 、 $G_W(N)$ の比をとった $J(N)$ を、動的システム数の決定基準とする。 $J(N)$ が大きいほど各動的システムが入力区間系列をよく表現していると言える。

$$J(N) = \frac{G_B(N)}{G_W(N)} \quad (21)$$

ここで、 $G_B(N)$ および $G_W(N)$ は次の式で定義される値である。

$$G_B(N) = \sum_{i=1}^N P(I_i) L(D_i, I_i) \quad (22)$$

$$G_W(N) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N P(I_i) P(I_j) L(D_i, I_j) \quad (23)$$

ここで、 $P(I_i)$ は入力区間系列のうち、区間集合 I_i が現れる確率を表す。

3.5 動的システムの併合アルゴリズム

動的システム間の併合アルゴリズムは次の通りである。

- (1) 観測信号系列を一定の長さの短い区間で区切り、各区間で動的システムのパラメータを求める。

- (2) 全ての区間集合 (ただし、一回目の反復では区間) と動的システムについて $L(D_i, I_j)$ を求める。
- (3) 全ての動的システムの組合せについて、動的システム間距離尺度を求め、最小の値を与える動的システムの組を見つけ、両システムの表現する区間集合を併合する。
- (4) 動的システム数決定基準 $J(N)$ を計算する。 $J(N)$ が最大の時の動的システム数とそのパラメータを記録する。
- (5) 併合された区間集合が一つの動的システムに従うとして、新たにパラメータを計算する。
- (6) 併合された動的システムを D_i とすると $L(D_i, I_j)$ および $L(D_j, I_i)$ ($j = 1, \dots, N$) を計算する。
- (7) 動的システムの数が 1 でなければ (3) に戻り、併合を繰り返す。
- (8) 求めた動的システムパラメータのうち小さな区間に対応する動的システムを除く。

4. 識別層の学習

本節では、セグメンテーション層のひとつ上の識別層において、これら動的システムの間の関係をモデル化する際に必要なパラメータの学習方法について述べる。学習が必要なパラメータは、2 節で述べたように、直前の動的システムに対して次の区間でどのような動的システムが出現しやすいかを定める区間遷移確率、同じ動的システムに従う区間はどのくらいの時間継続するかを定める区間長分布、および始めの区間ではどの動的システムが起りやすいかを定める初期区間分布である。

4.1 区間長分布の学習方法

学習データにおいて、ある一つの動的システムに従う区間を、その長さとともに全て数えることによってヒストグラムを作成する。これによってある動的システムに従う区間が継続する確率を決定する。

学習データ中に同一の動的システム A に従う区間 I_{A_m} ($m = 1, 2, \dots, M$) が現れたとする。動的システム A の区間長分布 $P(\tau)$ は次式で与えられる。

$$P(\tau) = \frac{1}{M} \sum_{t=T}^M \text{hist}(\tau) \quad (24)$$

ここで、 $\text{hist}(t)$ は学習データに現れる区間 I_{A_m} のうち時間長 τ の区間の個数である。ただし、ヒストグラムに基づく上述の手法は、非常に多くのサンプル数を要するという問題点がある。この解決方法として、評価実験においては密度関数としてガウス関数を当てはめることで対処した (図 3 参照。)

4.2 区間遷移確率の学習方法

ある動的システムに従う区間から別の動的システムに従う区間への遷移確率も学習データにおいて現れる

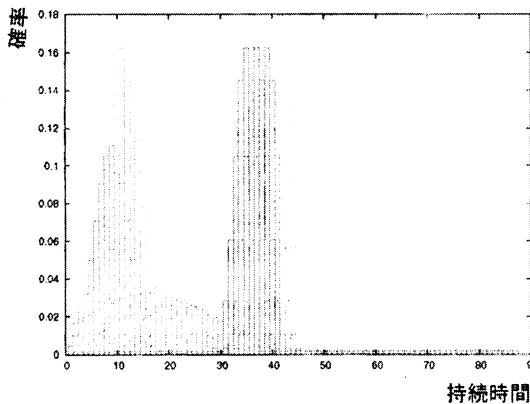


図3 観測系列から求めたヒストグラムから求めた区間長分布

区間同士の遷移を全て数えあげることで決定する。ここで、学習データは N 個の区間からなる区間系列であるとする。 $C(B|A)$ は学習データにおいて区間 A が現れた上で、次の遷移によって区間 B が現れた回数である。

$$P(B|A) = \frac{C(B|A)}{N-1} \quad (25)$$

ここで、 $C(B|A)$ は学習データに現れる区間 A から区間 B への区間遷移の回数である。

5. 実験による検証

まず最初に、3.3 節で定義した動的システム間の距離尺度の有効性を検証した。全ての動的システムを一つずつ用いて、全ての区間に対して求めた予測誤差を図4に示す。図において青く表示されている部分は予測誤差が小さいことを表す。図は横軸方向が時間方向であり、縦軸方向が動的システムの違いである。時間方向に見て、領域が複数現れる場合、その動的システムでうまく表現出来る時間領域が複数現れることを意味する。対角方向には8個の領域ができた。上から1番目と7番目、2番目と5番目、6番目と8番目はそれぞれ同じダイナミクスを表現出来る動的システムであり、その性質は似ていると考えられる。

次に最初は、短時間の区間に時系列を区切っておき、順に距離尺度をもとに動的システムを併合していった結果を図5に示す。併合が進むと、5個の動的システムで表現される、まとまった大きさの8個の区間が得られることがわかる。一つ動的システムに対応する区間は同一の色が付いている。学習に用いたサンプルは8個の母音からなる発話イベントであり、1番目と7番目、2番目と5番目、6番目と8番目が同じ母音である。図5より、同じ母音が同じ動的システムで表現されていることから、動的システム間距離尺度をもとに併合を行えば同一の動的システムに従う区間を正しく求めることが可能である。

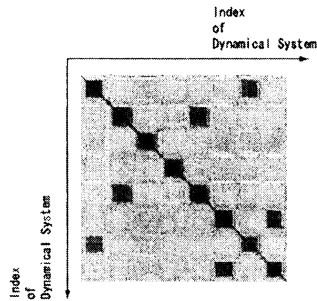


図4 発話/あえいうえお/に対して全ての動的システムを用いて全ての区間に対して求めた予測誤差

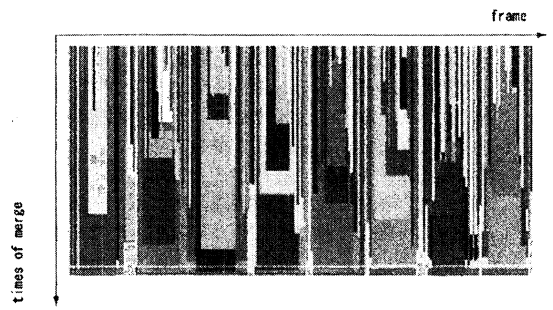


図5 同じ色の区間は同じ動的システムによって表現される区間である

次に、第3.4.1節で定義した動的システム数決定基準の妥当性について検証を行なった。動的システムの数を最初は入力時系列特徴ベクトルの個数だけ用意し、併合により、次第に減らしながら評価値 $J(N)$ を併合毎に求めた。結果は図6ようになった。図5において評価値が最大値をとる動的システム数の部分には縦線を示した。図5において区間系列の変化をみると、途中までは比較的大きな区間が8個とその他の小さい区間が存在することがわかる。したがって、評価値が最大の時に動的システムを取り出せば、それらは線形なダイナミクスを表現する動的システムであるといえる。ただし、区間の持続長が比較的小さいものは除くものとする。

学習サンプルから求めた動的システムを用いて音声信号の識別を行なった。発話イベント/あえいうえお/を用いて、このイベントの識別に必要な動的システムの数とパラメータを求めた。そして、それを用いて、発話イベント/あえいうえお/と発話イベント/あいうえお/の識別を行なった。結果は図8のようになった。横軸方向が時間を表し、縦軸方向の離散的な値は異なる動的システムに対応する。これより、同一の音声区間は同じ動的システムに従っていることがわかる。同様に、発話/あいうえお/に対して識別を行なったものを図9に示す。いずれのサンプルも図8の場合と同様に、それぞれの発話に対応する区間は対応

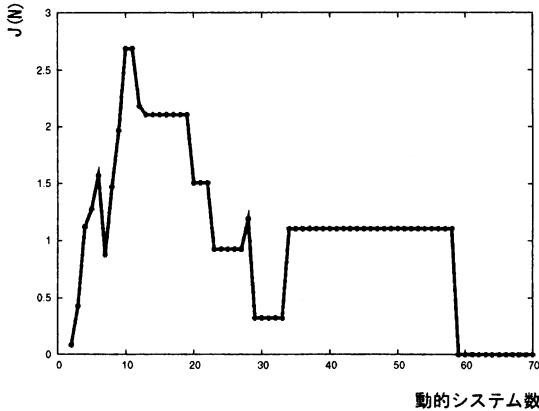


図 6 動的システム数決定のための評価値の様子

する動的システムによって表現出来ている。

学習によって求まった動的システムのパラメータを用いて、入力する動的イベントを識別したときに得られた事後確率を図 7 に示す。発話イベント/あえいうえお/を用いて学習したパラメータは、発話イベント/あいうえお/と/あえいうえお/の二つのイベントを識別出来ていることを示す。

以上の結果より、次のことが確認できた。

距離尺度による併合の妥当性 距離尺度を用いて併合を行なった際、併合の最後の方で得られた比較的大きな区間が、同じ性質のダイナミクスに従うことが予想される区間と一致した。これは距離尺度を基準に併合を行なったとき、時系列を同じダイナミクスをもつ区間に区切っていくことが可能なことを意味する。

動的システム数の決定基準の妥当性 動的システム数決定基準を用いると、その評価値が最大となるとき、得られた比較的大きな区間集合は一つの母音に対応する区間となった。決定基準によって得られた動的システム数よりも動的システムの数が少ないと、本来異なる動的システムによって表現されるべき複数の区間がひとつの動的システムで表現されてしまうことがあった。従って、動的システム決定基準を基に動的システムの数を決定すると、最も適当な区間分割が可能となると考えられる。

一方、次のような欠点も明らかとなった。

距離尺度による区間分割のあいまいさ 学習サンプルから併合により動的システムを求める際、併合により得られる動的システムのパラメータは用いるサンプルにより、ばらつきが生じる。これにより、同じ時系列イベントでもうまく表現できないことがわかった。併合により、求まる動的システムのパラメータのばらつきの原因は同じダイナミクスに従うわけでもないが誤って併合されてしまう区間が存在することが原因の一つであると考

えられる。似たダイナミクスをもつ動的システムは時間的に近い場所に存在すると考えられるが、併合が進むにつれて、同じ動的システムで表現される区間がどのように現れるかが予測出来なくなる。よって、最初は時間的に近い場所にある区間同士を併合し、区間がある程度の大きさになってきたところに時間的に離れたところに存在する区間との併合も行なうなど、動的システム同士の併合には、動的システム間の距離尺度のみならず区間の時間的な近さを考慮する必要がある。

どの動的システムにも対応しない区間の扱い 識別時の観測信号から区間系列への変換の曖昧さの原因は、時系列信号に含まれるどの動的システムでも表現できない区間が存在するためである。このような区間に属する観測信号には異なる動的システムのラベルがばらばらにつけられ、これにより時系列信号がこのモデルにより表現できる確率の値が大きく変化してしまう。

2 節では、次のような仮定をおいていた。

- (1) 入力する時系列信号は全て線形動的システムで表現できる。
- (2) 一つの区間は一つの動的システムで表現できる。しかし、子音や無音区間のように線形動的システムのみを用いて時系列信号全てを表現するのは困難であることがわかる。この問題を解決するためには次のような方法が考えられる。
- (1) 非線形なモデルを導入する。
- (2) 線形動的システムを複数組み合わせ、非線形な区間を表現する。

以上のような方法により、非線形な区間を含むような時系列信号から、学習と識別が出来るシステムが可能となると考えられる。

6. おわりに

本研究では、時系列イベントを識別するモデルとして区間線形動的システムを提案した。このシステムは連続状態を持つセグメンテーション層と離散状態を持つ識別層の二つの層から構成され、以下のような特徴を持つ。

- (1) 複雑な時系列イベントを単純な変化を表す区間の組合せとして表現する。
- (2) 時系列イベント中の非線形な伸縮を区間の継続時間の分布としてモデル化できる。
- (3) 観測信号系列を与えるだけで、必要な動的システムの数とパラメータを推定することが可能である

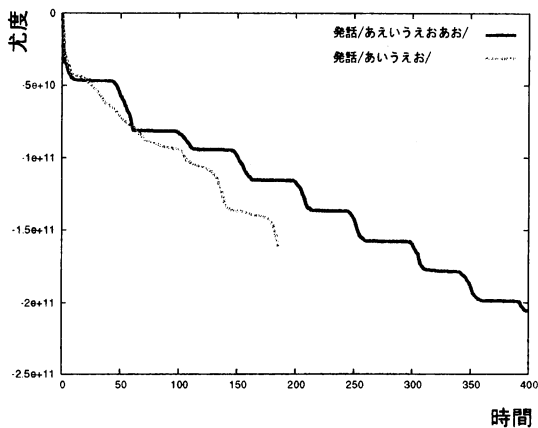


図 7 サンプルを変えたときの事後確率

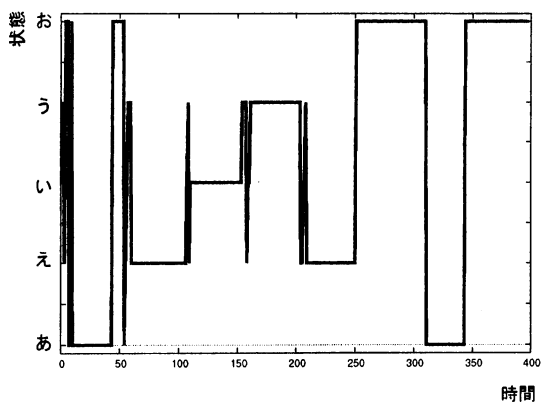


図 8 動的システムにより、時系列データを識別する様子

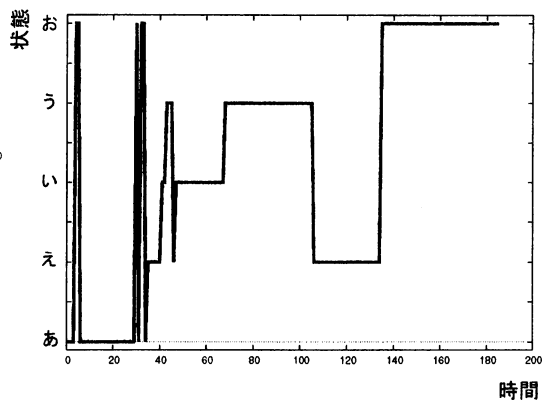


図 9 異なるクラスを識別する様子

参考文献

- 1) 大和淳司, 大谷淳, 石井健一郎: 隠れマルコフモデルを用いた動画画像からの人物の行動認識, 電子情報通信学会論文誌, Vol. J76-D-II, No. 12, pp. 2556-2563 (1993).
- 2) Bobick, A.F. and Wilson, A.D.: A State-Based Approach to the Representation and Recognition of Gesture, *IEEE transaction on pattern analysis and machine intelligence*, Vol. 19, No. 12 (1997).
- 3) Bregler, C.: Learning and recognizing human dynamics in video sequences, *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 568-574 (1997).
- 4) Huang, H. D., Ariki, Y. and Jack, M. A.: *Hidden Markov Models for Speech Recognition*, Edinburgh Univ. (1990).
- 5) J. L. Elman: Finding structure in time, *Cognitive Science*, Vol. 14, pp. 179-211 (1990).
- 6) 森田昌彦, 村上聡: 非単調神経回路網による時系列パターンの認識, 電子情報通信学会論文誌, Vol. J81-D-II, No. 7, pp. 1679-1688 (1998).
- 7) Ostendorf, M., Digalakis, V. and Kimball, O. A.: From hmms to segment models, *IEEE Trans. Speech and Audio Process*, Vol. J81-D-II, pp. 360-378 (1996).
- 8) Pavlovic, V., Rehg, J. and MacCormick, J.: Learning switching linear models of human motion, *Proc. of Neural Information Processing Systems* (2000).
- 9) 高橋勝彦, 関進, 小島浩, 岡隆一: ジェスチャー動画画像のスポッティング認識, 電子情報通信学会論文誌, Vol. J77-D-II, No. 8, pp. 1552-1561 (1994).