論　文

自己増殖型ニューラルネットワークを用いた時系列データの学習・認識

岡田　将吾[†]　　長谷川　修[††]

Learning and Recognition of Time-Series Data Based on Self-Organizing Incremental Neural Network

Shogo OKADA[†] and Osamu HASEGAWA[††]

あらまし　本研究では，オンライン教師なし学習手法である Self-Organizing Incremental Neural Network (SOINN) を用いて各状態の出力分布を自己組織的に近似可能な時系列データの学習モデルを提案する．提案手法は従来手法であるストキャスティック DP 法 [12] を拡張した新規の手法である．ストキャスティック DP 法では各状態を一つの多次元正規分布で近似しているのに対し，提案手法では各状態の出力分布が SOINN によって自己組織的に近似される上，各状態の出力分布が詳細に近似されるため，時系列データの頑健なモデル化が可能となる．提案手法の有効性を検証するために，動画像から得られる動作及び音素を用いた認識実験を行った．HMM (Hidden Markov Model) 及びストキャスティック DP 法と認識精度を比較することで提案手法の有効性を示す．

キーワード　時系列データ，データ認識，自己増殖型ニューラルネットワーク，DP マッチング

1. まえがき

時系列データの認識・モデル化は動画像処理，音声情報処理，DNA 解析などの分野において重要な基盤技術である．一般に時系列データは特徴空間内での変動と時間方向の伸縮を含んでいるため，これを頑健に認識するためには，これらの特徴を吸収可能なモデル及び学習器を構築する必要がある．このため，あらかじめグラフ構造を保持したモデルをもつことで，時系列データの学習・認識を行うモデルに基づく手法が頻繁に用いられている．

モデルに基づく手法の中で HMM (Hidden Markov Model) [1] は，音声認識の分野における標準的な手法として大きな成功を収めている．HMM は音声認識以外にも，話者適応技術 [2] や音声合成技術 [3] などに用いられており，音声情報処理全般における標準的手法となっている．この音声情報処理における成功事例や，統計的理論の裏付けがあることから，HMM は動画像及び動作の認識にも多く用いられてきた．音声認識や動画像認識における手法としては離散 HMM (Discrete HMM) [4], [5] を用いるものや，連続分布 HMM (Continuous HMM) [6], [7] を用いるものがある．また各状態の持続長分布を明示的にモデル化したりフレーム間の相関をモデル化した Segment model [8] も提案されている．

通常 HMM では，パラメータの推定の容易性の理由で，音声データでいえば 1 音韻に対して 3〜5 状態のマルコフモデルを用いる場合が多い．しかしこのような少数の状態では，過渡的な時系列データの時系列に沿った特徴量の変化を詳細にモデル化できない可能性がある．

これに対し動的計画法の一種である DP マッチングは，短時間の特徴パラメータ（各フレーム）同士の局所距離に基づいて，過渡的な時系列データ間の距離を算出することが可能である．DP マッチングは音声認識 [9]，動作認識 [10] のほか，時系列データの検索などに用いられている [11]．一方で DP マッチングでは

[†] 東京工業大学大学院総合理工学研究科知能システム科学専攻，横浜市
Department of Computational Intelligence and System Science, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, 226–8503 Japan
[††] 東京工業大学像情報工学研究施設，横浜市
Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, 226–8503 Japan

標準データそのものをモデル（テンプレート）とするため，HMM に比べて特徴空間の分布を詳細にモデル化することが困難である．

これらの背景から，DP マッチングの利点と HMM の頑健性の両方を生かしたストキャスティック DP 法 [12] が中川によって提案されている．ストキャスティック DP 法では DP マッチングの局所距離の尺度に確率の尺度を用い，パスコストの代わりにパス遷移確率を用いている．またストキャスティック DP 法はテンプレートモデルの 1 フレームを 1 状態に対応させており，状態数を多くした HMM の連続出力分布をもつ left to right モデルに相当する．各状態の出力分布には単一の多次元正規分布が用いられている．一般に各状態の出力分布は，特徴量の次元数及び特性に応じて異なるため，単一の多次元正規分布で近似できない可能性がある．

この問題に対し，本研究では各状態の出力分布を特徴量に応じて自己組織的，かつ詳細に近似可能な手法を提案する．提案手法ではテンプレートモデルにおける各状態の出力分布を Self-Organizing Incremental Neural Network (SOINN) により詳細に近似する．Self-Organizing Incremental Neural Network (SOINN) [13] は Shen と Hasegawa によって提案されたオンライン教師なし学習手法である．SOINN は非定常的な入力を学習可能であり，分布に複雑な形状をもつクラスに対してもノイズを除去し，適切なクラス数及びデータの分布を近似可能である．本研究では，SOINN のノイズ除去及び分布の近似機能に着目し，この機能を各状態の出力分布の近似に用いる．SOINN の機能を用いることで，各状態の出力分布は複雑さに応じて自己組織的に近似される．提案手法において，状態数はテンプレートモデルのフレーム数に対応し，各状態の出力分布は SOINN によって自己組織的に近似される．したがって，提案手法では HMM のように最適な状態数及び連続分布の混合数を事前に決定する必要がない．

総じて本研究では，ストキャスティック DP 法を出力分布のモデル化の観点から SOINN を用いて拡張した，時系列データの学習・認識手法を提案する．この提案手法を SOINN-DP 法と定義する．

以下で，本研究で扱う時系列データについて述べる．

1.1 本研究で扱う時系列データ

本研究では，動作から得られる時系列データと音声から得られる時系列データを認識対象として，HMM 及びストキャスティック DP 法との比較実験を行う．また本論文では，始点，終点の与えられた動作・音声データを扱うものとする．

動作データには，単眼カメラにより撮影した人間の全身運動を用いた．全身を使った動作を行う際，バランスのとり方に個人差が出るため，動作から得られる時系列データは各部分で時間伸縮を含み，特徴空間上の分散も含む．また動作の中には,「全身で円を描く」動作や「全身で四角を描く」動作といった類似した軌跡をもつ動作が含まれている．これらの動作を識別・認識する際には，時系列に沿った特徴量の変化を詳細にモデル化する必要がある．

ここで SOINN-DP 法では多数の状態を保持することで，上記の性質をもつ動作データの頑健なモデル化が可能であると考えられる．上記の動作データから得られる時系列データの認識実験を行い，多数の状態をもつ SOINN-DP 法と少ない状態をもつ HMM の認識性能を比較することで，SOINN-DP 法の有効性を示す．

音声データには，発話された英語文章及び英単語から音素境界を用いてセグメントした音素データを用い，これらの認識実験を行う．ここで HMM は音声認識の分野における標準的手法であり，ストキャスティック DP 法も音声認識を目的として提案されている．したがって，これらの手法と提案手法の性能を比較する上で，音声から得られる時系列データを用いて比較実験を行うことは重要である．このため本研究では動作認識に関するタスクだけでなく，音声認識に関するタスク（音素認識実験）を行う．

以上をまとめて本研究では，性質の異なる 2 種類の時系列データを用いて認識実験を行い，この結果を HMM 及びストキャスティック DP 法と比較することで，提案手法の認識精度及び時系列データの学習性能を検証する．

1.2 関 連 研 究

提案手法と同様に，状態の出力確率分布をニューラルネットワークで表現する手法は [14]〜[16] で提案されている．まず [14] では，HMM の出力確率に MLP (Multi Layer Perceptron) の出力値を用いる手法が提案されている．[16] では [14] のアーキテクチャと [15] で提案された MLP の結合荷重 w の学習法の利点を統合した hybrid HMM/ANN system が提案されている．この研究では，MLP の結合荷重 w の学習法として，Soft-Weight-sharing ML と呼ばれる最ゆう法,

ベイズ基準の学習法，事後確率最大化基準の学習法，の三つの学習法が提案されている．連続発話された数字を用いた自動音声認識実験の結果，三つの学習法のうちいずれを用いた場合にも，[14] の手法及び連続型 HMM の認識精度を上回った．

hybrid HMM/ANN system では各状態の出力確率分布の表現に MLP を用いたのに対し，本研究では各状態の出力確率分布の表現及び近似に SOINN を用いる．SOINN を用いた場合，データ分布は自己組織的にクラスタリングされ，適切な数のクラスで近似される．クラスタリングされた後の各クラスの分布は，Parzen の窓関数で近似される．この結果，状態の出力確率分布は SOINN の学習結果から得られる，クラス数個の Parzen の窓関数で近似されることとなる．

ここで SOINN におけるクラス数は出力確率分布の近似性能に影響を与える値であり，連続型 HMM における各状態の混合正規分布の混合数，HMM/ANN hybrid における MLP の中間層の層数及びユニット数と同じ働きをもつパラメータと考えられる．提案手法では SOINN を用いることで，このクラス数（Parzen の窓関数の数）を自動的に決定可能であるが，連続型 HMM や [16] の手法では上記のパラメータを，認識対象によってあらかじめ設定しておく必要がある．

2. 提案手法

SOINN-DP 法では，DP マッチングと SOINN を用いて各クラスのモデル（以下ではテンプレートモデルと呼ぶ）を構成する．まず **2.1** で DP マッチングの理論を，また **2.2** で SOINN の理論を説明した後に，**2.3** で SOINN-DP 法のアルゴリズムの詳細を述べる．本章では入力される一つのベクトルをサンプルと呼称し，入力されるサンプルの集合を入力データと呼称する．また時系列データに関しては各フレームのベクトルをサンプルと呼称し，時系列データそのものを指し示す場合はデータと呼称する（例：訓練データ，テストデータ，標準データ）．

2.1 DP マッチング

ここではフレーム数 I の時系列データ $\boldsymbol{X} = \{x_1, x_2, \cdots, x_i, \cdots, x_I\}$ とフレーム数 J の時系列データ $\boldsymbol{Y} = \{y_1, y_2, \cdots, y_j, \cdots, y_J\}$ の DP マッチングを考え，この二つの時系列データの累積距離を算出する．ここで i，j はそれぞれ時系列データ $\boldsymbol{X}$，$\boldsymbol{Y}$ のフレーム番号を示す．また本論文では時系列データ $\boldsymbol{X}$ の各フレームの特徴ベクトル x_i を，i フレーム目のサンプルと呼称する．

本研究では，時系列データ $\boldsymbol{X}$ と時系列データ $\boldsymbol{Y}$ の累積距離 $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ の算出に，以下のような対称型漸化式を用いた．

初期条件 ($i = 1$，$j = 1$)：$\boldsymbol{g}(1,1) = \boldsymbol{d}(1,1)$

漸化式 ($i > 1$，$j > 1$)：

$$\boldsymbol{g}(i,j) = \min \left\{ \begin{array}{l} \boldsymbol{g}(i, j-1) + \boldsymbol{d}(i,j) \\ \boldsymbol{g}(i-1, j-1) + 2\boldsymbol{d}(i,j) \\ \boldsymbol{g}(i-1, j) + \boldsymbol{d}(i,j) \end{array} \right\} \quad (1)$$

上記の漸化式を用いて，累積距離 $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ は次式となる．

$$\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\boldsymbol{g}(I, J)}{I + J} \quad (2)$$

上述のように，DP マッチングでは累積距離に現時点の局所距離を累積する演算を漸化的に繰り返すことで累積距離 $\boldsymbol{D}(\boldsymbol{X}, \boldsymbol{Y})$ が求まる．また DP マッチングでは $\boldsymbol{X}$ の第 i フレーム目のサンプル x_i と $\boldsymbol{Y}$ の第 j フレーム目のサンプル y_j との最適な対応付け $j = w_i$ $(i = 1, 2, \cdots, I)$ も得られる．

2.2 SOINN

本節では，提案手法の基礎となる SOINN [13] の概要について述べる．SOINN は Growing Neural Gas (GNG) [17] を拡張した自己増殖型ニューラルネットワークと呼ばれる教師なし追加学習手法である．

2.2.1 学習アルゴリズム

SOINN の主な働きは，オンラインで入力されるサンプル集合に対し，ノードを徐々に増殖させ，各ノード間の隣接関係をエッジを用いて構成し，そのサンプル集合の分布を近似することである．SOINN ではノードの位置の更新及びエッジの挿入・削除を必要に応じて行うことで，入力データの分布を適応的に近似する．入力データの分布を近似するために，入力に対してノードの挿入とノードの位置ベクトルの更新処理が行われる．ノードの挿入は，近似されていない領域への入力に対して実行される．挿入の必要性の判断は，既存のネットワークの各ノードがもつ類似しきい値 T に基づいて行う．ノード挿入の例を図 1 に示す．入力サンプルと勝者ノード及び第 2 勝者ノードとの距離がそれらのノードの類似しきい値 T を超える場合，入力サンプルは新たなクラスタに属すると判断する．その際，入力サンプルを新ノードとしてネットワークに挿入する．ここで勝者ノードとは入力サンプルの最近傍

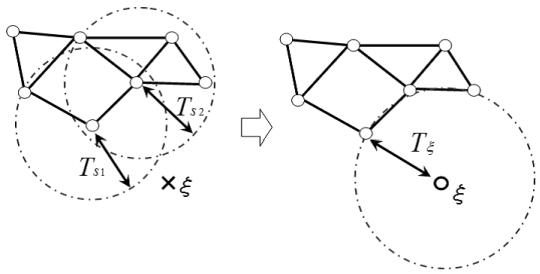

図 1　ノードの挿入処理（入力サンプル ξ と勝者ノード s_1 及び第 2 勝者ノード s_2 との距離が類似しきい値 T_{s_1}，T_{s_2} より大きい場合（左），入力 ξ を新たなノードとして挿入する（右）．新ノードの類似しきい値 T_ξ は勝者ノードとの距離で表される．）
Fig. 1　Between-class insertion process.

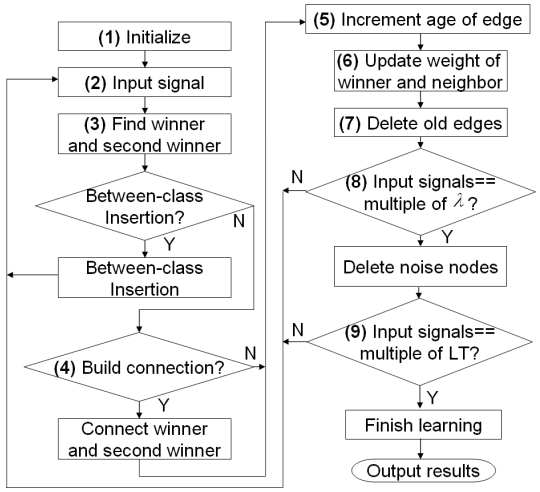

図 2　SOINN のアルゴリズムのフローチャート
Fig. 2　Flowchart of SOINN's algorithm.

ノードを意味し，第 2 勝者ノードとは，入力サンプルの第 2 近傍ノードを意味する．

入力の分布を近似するため，入力に対してノードの位置ベクトルの更新処理（後述）が行われる場合がある．そのため，ノード i の位置は随時変化するので，類似しきい値 T_i の値もそれに伴い変化させる．類似しきい値 T_i の算出方法をアルゴリズム 2.1 に示す．以下で，W_i はノード i の位置ベクトルを表す．

アルゴリズム 2.1: 類似しきい値 T の計算方法

（1） ノード i の類似しきい値を，ノードの生成（挿入）時に $+\infty$ に初期化する．

（2） ノード i が勝者ノードまたは第 2 勝者ノードである場合，T_i を更新する．

- ノード i に隣接ノード（ノード i とエッジでつながれたノード）が存在する場合，T_i をノード i から最も遠い隣接ノードとの距離値に更新する（$T_i = \max_{c \in N_i} \|W_i - W_c\|$，ただし N_i は，ノード i の隣接ノード集合を表す）．
- ノード i に隣接ノードが存在しない場合，T_i をノード i から最も近い他のノードとの距離値に更新する（$T_i = \min_{c \in A \setminus \{i\}} \|W_i - W_c\|$，ただし A は全ノード集合を表す）．

上記のノードの挿入処理のほかに，SOINN ではエッジの削除過程において，加齢処理 (*edge aging* scheme [18]) が用いられる．各エッジは「年齢」という 0 以上の整数値を保持している．具体的には各入力に対して，勝者ノードに連結するすべてのエッジの年齢を加齢し，その一方で勝者ノードと第 2 勝者ノードの間のエッジの年齢を 0 に更新する．そして，事前定義するしきい値 a_d を超える年齢になったエッジを削除する．ノードの移動によって不適切となったエッジは，隣接関係が成り立たないため，エッジの年齢が 0 に更新されずに削除される．

上記の処理を踏まえ，SOINN の処理手順をアルゴリズム 2.2 に示す．ここで，アルゴリズムの各ステップはフローチャート（図 2）の各サンプルの番号に対応している．

アルゴリズム 2.2: SOINN の処理手順

（1） ノード集合 A を，学習サンプル群からランダムに選択した二つのノード ($A = \{c_1, c_2\}$) に初期設定する．また初期設定時に，エッジ集合 C ($C \subset A \times A$) は空集合とする．

（2） $\xi \in R^n$ を入力サンプルとする．R^n は SOINN に入力される全サンプル集合とする

（3） 入力サンプルに対する勝者ノード (winner) s_1 と第 2 勝者ノード (second winner) s_2 を以下の式に従い決定する．

$$s_1 = \arg\min_{c \in A} \|\xi - W_c\| \tag{3}$$

$$s_2 = \arg\min_{c \in A \setminus \{s_1\}} \|\xi - W_c\| \tag{4}$$

入力サンプル ξ とノード（s_1 または s_2）との距離が類似しきい値（T_{s_1} または T_{s_2}）より大きい場合，入力サンプルを新ノードとして A に追加する．その後，新しい入力サンプルの学習のためにステップ（2）に戻る．類似しきい値 T はアルゴリズム 2.1 により算出される．

（4） s_1 と s_2 との間のエッジが存在しなければ，

新たに作成して C に追加する．存在する場合は該当するエッジの年齢を 0 にリセットする．

（5） s_1 につながるすべてのエッジの年齢を加算する．

（6） 勝者ノードと勝者ノードに隣接するノードの位置ベクトルを，以下の式を用いて更新する．ただし，係数 ϵ_1 及び ϵ_2 を，$\epsilon_1(t) = 1/t$，$\epsilon_2(t) = 1/100t$，また，t を該当ノードが勝者ノードに選択された回数，と定義する．

$$\Delta W_{s_1} = \epsilon_1(t)(\xi - W_{s_1}) \quad (5)$$

$$\Delta W_i = \epsilon_2(t)(\xi - W_i) \quad (\forall i \in N_{s_1}) \quad (6)$$

（7） しきい値 a_d を超える年齢のエッジを削除する．その結果，隣接関係をもたないノードが現れた場合は，該当するノードを削除する．

（8） 入力サンプル数が λ の倍数となった場合，隣接ノードが存在しない孤立したノードを削除する．この操作を行うことで，入力サンプルの外れ値によって挿入されたノードを削除する．[13] では，ノードの削除と同時に低密度領域へのノード挿入を行っている．本研究では低密度領域へのノード挿入は SOINN の学習性能にそれほど関与しないことを確認したため，この操作は行わず，ノード削除のみを行った．

（9） 学習が十分に行われるまで，ステップ（2）に戻り学習を繰り返す．図 2 の (9) において LT は学習の終了する回数を示す．すなわち LT 回学習 ((1)〜(9)) を繰り返した後に学習を終了する．学習終了時点で特徴空間上に存在するノード集合 A の中で，エッジによりつながっているノード集合が一つのクラスに対応する．

アルゴリズム 2.2 では，二つのパラメータ (a_d，λ) の設定が必要である．まず λ はノイズとおぼしきノードを削除する周期である．λ を小さな値に設定すると頻繁にノードの削除が行われるが，極端に小さくすると実際はノイズではないノードを誤って削除してしまう．逆に λ を極端に大きな値に設定するとノイズの影響で挿入されたノードを適切に取り除くことができない．

次に a_d はノイズなどの影響で誤って挿入されたエッジを削除するために用いられる．a_d を小さな値に設定するとエッジが削除されやすくなりノイズによる影響を防ぐことができるが，極端に小さくすると頻繁にエッジが削除され学習結果が不安定になる．逆に a_d を極端に大きな値に設定すると，ノイズの影響で挿入されたエッジを適切に取り除くことができない．

以上の特性を考慮して，パラメータ (a_d，λ) の設定を行う必要がある．本論文の実験で用いるパラメータの決定方法は **3. 2** で述べる．

2. 2. 2 SOINN の学習機能の検証

ここで SOINN の機能を検証するために行った，人工データセットを用いた実験を示す．この実験では，図 3 に示す二次元の人工データから 1 点ずつサンプルをオンラインで入力した場合の SOINN の挙動を検証した．データセットは二つのガウス分布，二つの同心円，及び Sin 曲線の合計五つのクラスによって構成されている．また，実世界の環境を想定して，五つのクラスから生起するデータに 10%の一様ノイズが加えられている．このデータセットをオンラインで追加的に入力し，SOINN に教師なし分類を行わせた．

この入力データが SOINN によって分類された後の出力結果を図 4 に示す．図 4 より入力データに含まれるノイズは削除され，入力データのクラスタ数とその分布が正しく近似されていることが分かる．SOINN のアルゴリズムの詳細については [13] に記載されている．

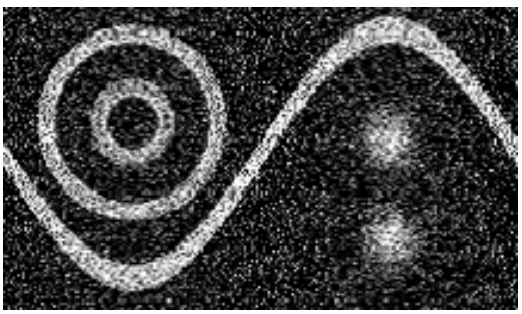

図 3 ノイズを含む二次元の入力データ
Fig. 3 2D artificial data set with noize pollution.

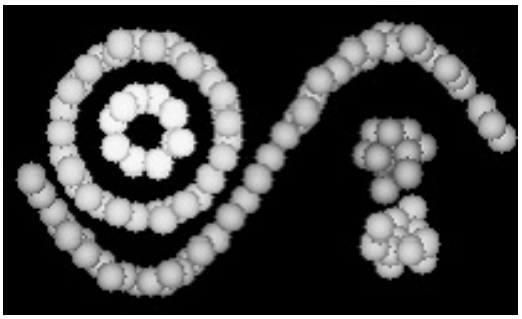

図 4 SOINN によるクラスタリング結果
Fig. 4 Result of clustering.

2.2.3 SOINN-DP 法における SOINN の役割

SOINN-DP 法では各状態の出力分布を推定するために SOINN を用いる．また一つの状態が一つの SOINN に対応している．各状態に分配されたデータ（サンプル集合）を入力として SOINN により学習が行われた後，SOINN から入力データの分布は複数のクラス（ノードとエッジの集合）として出力される．この複数のノードとエッジの集合から出力分布を推定する．ここで各状態へのデータの分配方法及び出力分布の推定方法については **2.3** で述べる．

2.3 SOINN-DP 法

SOINN-DP 法は，訓練データ間において DP マッチングを行うことでテンプレートモデルを作成する．また，各クラスの訓練データから構成されたテンプレートモデルと入力データを DP マッチングすることで，入力データの認識を行う．

2.3.1 テンプレートモデルの作成

SOINN-DP 法では以下の三つの STEP に従って時系列データのテンプレートモデルが作成される．以下では，クラス $\mathcal{C}$ に属する N 個の訓練データが与えられたとし，この N 個の訓練データからテンプレートモデルを作成する手順を説明する．

[STEP 1：標準データの決定]

訓練データ群から，テンプレートモデルの中心となる標準データを決定する．クラス $\mathcal{C}$ 内のある訓練データ $\boldsymbol{P_m}$ と，クラス $\mathcal{C}$ 内の $\boldsymbol{P_m}$ 以外の訓練データ $\boldsymbol{P_n}$ との間で DP マッチングを行う．この操作を，クラス $\mathcal{C}$ 内の訓練データの全組合せで（総当りで）行う．DP マッチングの結果から得られるデータ間同士の累積距離の和を求め，最も累積距離の和が小さいデータを以下の式で選択する．

$$m^* = \arg\min_m \left\{ \sum_{n=1}^{N} \boldsymbol{D}(\boldsymbol{P_m}, \boldsymbol{P_n}) \right\} (\{\boldsymbol{P_n}, \boldsymbol{P_m}\} \in \mathcal{C}) \tag{7}$$

式 (7) において arg は，各訓練データ間の累積距離の和が最も小さい訓練データの番号 m^* を返す．クラス $\mathcal{C}$ の m^* 番目の訓練データを，テンプレートモデルの中心となる標準データ $\boldsymbol{P^*}$ と決定する．ここで $\boldsymbol{P^*}$ のフレーム数 T^* をテンプレートモデルの時系列長とする．

[STEP 2：データを各状態に分配]

標準データ $\boldsymbol{P^*}$ と，その他 $N-1$ 個の訓練データとの間で DP マッチングを行った結果，その他の全訓練データの時系列長は，標準データ $\boldsymbol{P^*}$ の時系列長に正規化される．また標準データ $\boldsymbol{P^*}$ の各フレームのサンプルと，その他 $N-1$ 個の訓練データの各フレームのサンプルとの対応付けが得られる（**2.1** を参照）．ここで対応関係にあるサンプル群を各 SOINN 空間（各状態）に入力する．

標準データ $\boldsymbol{P^*}$ の第 j フレーム目のサンプルを p_j^*，訓練データ $\boldsymbol{P_n}$ $(n \in \mathcal{C})$ の第 i フレームのサンプルを p_i^n とし，この p_j^* と p_i^n との最適な対応付け w^n を以下のように定義する．

$$i = w_j^n \ (j = 1, 2, \cdots, T^*) \tag{8}$$

式 (8) に従い，訓練データの i フレームのサンプルを j フレームの状態（SOINN 空間）に分配する．

上記の操作を，標準データとその他 $N-1$ 個の訓練データとの間で行った後に，$N-1$ 個の最適経路 w^n $(n = 1, \cdots, N-1)$ が得られる．この $N-1$ 個の最適経路に従い，各状態に訓練データの各サンプルを分配する．ここで j 番目の状態に分配されたサンプル集合を $\mathcal{Z}_j$ と定義する．

SOINN-DP 法ではストキャスティック DP 法と同様に，1 フレームを 1 状態に対応させているため，一つの状態に分配されるサンプルは少量となる．ここで分配されるデータが少量の場合，SOINN の学習性能（分布を近似する機能）が低下する．そこで十分な学習性能を得るためには，特徴量の次元数に相応のデータ量が必要である．

この問題に対し，ストキャスティック DP 法では共分散行列をある状態間で共有する手法がとられている．この手法は，隣接する状態間のサンプル群は類似する，つまり時刻の近い状態 j のサンプル集合と状態 $j+L$ のサンプル集合同士は空間的に近接しているという仮定のもとに成り立っている．本研究でも SOINN-DP 法にこの仮定を用いることで，上記の問題を解決する．SOINN-DP 法では，ある時間の範囲（状態間）に分配されたサンプル集合を，一つの SOINN に入力する．具体的には，$\mathcal{Z}_j$ から $\mathcal{Z}_{j+L-1}$ までのサンプル集合を，j 番目の状態 (SOINN) に入力する．この j 番目の SOINN 空間に入力するサンプル集合を ${\mathcal{Z}^*}_j$ と定義し，以下で表す．

$${\mathcal{Z}^*}_j = \{\mathcal{Z}_j, \mathcal{Z}_{j+1}, \cdots, \mathcal{Z}_{j+L-1}\} \tag{9}$$

ここで L は SOINN-DP 法のパラメータであり，このパラメータを Segment 数と定義する．このパラメー

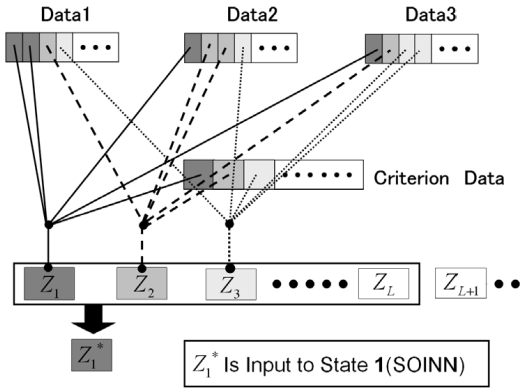

図 5 STEP 2. の処理過程（図 5 において Criterion Data は標準データを示し，Data1〜3 は訓練データを示す．また Data 及び Criterion Data の中の各ブロックは各フレームのサンプルを示す．また各ブロックにおいて同色の部分は，DP マッチング後の最適経路における対応箇所を示す．ここで Criterion Data（標準データ）の 1 フレーム目のサンプル（黒色）に対応したサンプルは，Data1 の 1，2 フレーム目のサンプル，Data2，3 の 1 フレーム目のサンプルであり，これらのサンプル群が $\mathcal{Z}_1$ となる．黒線は対応するサンプル群 ($\mathcal{Z}_1$) 同士を結ぶ線である．黒破線は $\mathcal{Z}_2$ のサンプル群，黒点線は $\mathcal{Z}_3$ のサンプル群同士を結ぶ線である．式 (9) より $\mathcal{Z}_1$ から $\mathcal{Z}_L$ までのデータ集合 ${\mathcal{Z}^*}_1$ が，状態 1 の SOINN 空間に入力される．）

Fig. 5 Process of STEP 2. (In DTW, optimal path between criterion data and training data is determined. Corresponding data in optimal path are input to each SOINN.)

タの設定方法は **3. 2. 3** で述べる．またテンプレートモデルの状態数は Segment 数 L と標準データの時系列長 T^* を用いて，$T^* - L - 1$ と決定される．

STEP 2. の処理過程を図 5 に示す．

［STEP 3：SOINN の学習］

各状態 j において，サンプル群 ${\mathcal{Z}^*}_j$ を SOINN 空間に入力する．ここで ${\mathcal{Z}^*}_j$ を入力する際，${\mathcal{Z}^*}_j$ の各サンプルを一つずつランダムに入力する．これは，SOINN がオンライン学習用の手法であるため，このような入力方法で行う．また ${\mathcal{Z}^*}_j$ は **2. 2** のアルゴリズム 2.2 の R^n に相当する．

サンプル集合が SOINN 空間に入力されると，SOINN 空間ではノード及びエッジの挿入，削除が繰り返され，最終的にノード集合 A が出力される（SOINN による学習過程は，**2. 2** を参照）．ノード集合 A の位置ベクトル W_i から出力分布を推定する．

出力分布の推定方法は **2. 3. 2** で述べる．また後の評価実験で用いた SOINN のパラメータについては **3. 2** で述べる．

2. 3. 2 確率密度関数の推定

訓練データ群からテンプレートモデルが構成された後，テンプレートモデルの各状態には SOINN により出力されたノード集合が存在する．このノード集合から確率密度関数（状態の出力確率の分布）を推定する．

ここでノード集合の中で，同じクラスに属するノード同士はエッジで連結されている．ノード集合の中で一つのクラス（エッジで連結されたノード集合）を一つの内部クラスと定義する．SOINN-DP 法では 2 種類の確率密度関数を推定し，これらの確率密度関数から 2 種類のゆう度を算出する．2 種類のゆう度をそれぞれ大域的ゆう度，局所的ゆう度と定義する．

［大域的ゆう度の算出］

大域的ゆう度の算出には j 番目の状態 S_j の SOINN 内に存在する全ノードを用いる．まず SOINN 内に存在する全ノードの位置ベクトル (W) から多次元正規分布の確率密度関数を推定する．確率密度関数 $P_{whole}(\boldsymbol{x_i}|S_j)$ を以下の式で表す．

$$P_{whole}(\boldsymbol{x_i}|S_j) = \frac{1}{(2\pi)^{M/2}|\boldsymbol{\Sigma_j}|^{1/2}} \times \exp\left\{-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu_j})^t\boldsymbol{\Sigma_j}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu_j})\right\} \quad (10)$$

式 (10) において M はサンプル x_i の次元数，$\boldsymbol{\mu_j}$ は状態 S_j の SOINN 内に存在する全ノードの位置ベクトルの平均，また $\boldsymbol{\Sigma_j}$ は共分散行列である．この二つのパラメータは最ゆう推定により算出される．$P_{whole}(\boldsymbol{x_i}|S_j)$ から得られる対数ゆう度 $\log(P_{whole}(\boldsymbol{x_i}|S_j))$ を，大域的ゆう度と定義する．ここでストキャスティック DP 法では $\mathcal{Z}$（**2. 3** の STEP 2. を参照）の平均 $\boldsymbol{\mu_j}$ を算出する（方法 A）．一方 SOINN-DP 法では $\mathcal{Z}^*$ を SOINN へ入力した後，学習結果として出力された全ノードの位置ベクトルから $\boldsymbol{\mu_j}$ を算出する（方法 B）．SOINN-DP 法では，後の予備実験 (**3. 2. 1**) において方法 A より方法 B で平均 $\boldsymbol{\mu_j}$ を算出した場合の方が認識精度が良好であったため，$\boldsymbol{\mu_j}$ の算出に方法 B を用いた．

［局所的ゆう度の算出］

局所的ゆう度は，SOINN によってクラスタリングされた，複数の内部クラスの情報を用いて算出される．図 6 において，class1〜3 が内部クラスを示す．これらの各内部クラスの分布をノンパラメトリックの手法である Parzen 窓 [19] を用いて推定する．Parzen 窓の窓関数にはガウス核関数を用いた．ここで Parzen

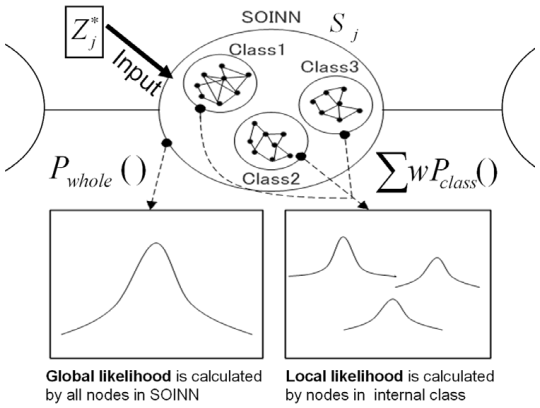

図 6　状態 S_j において SOINN から形成される 2 種類の確率分布の様子（図左下より大域的ゆう度は SOINN の全ノードから算出される．図右下より局所的ゆう度は SOINN により構成された内部クラスから算出される．ここで class1，2，3 は内部クラス U_{jk} を表している．）

Fig. 6　Two kinds of probability distribution formed with result of SOINN. (nodes and edges)

窓を用いた理由は，各内部クラスが保有するノード数は少数の場合（最低で 2 個）が多く，このような少数データから多次元正規分布（特に共分散行列）を推定することが困難なためである．SOINN の k 番目の内部クラスを U_{jk} と定義し，U_{jk} から推定される $P_{class}(\boldsymbol{x_i}|U_{jk})$ は以下のように表される．

$$P_{class}(\boldsymbol{x_i}|U_{jk}) = \frac{1}{(2\pi h_{jk}^2)^{M/2}} \exp\left\{-\frac{|\boldsymbol{x_i}-\boldsymbol{x_{jk}}|^2}{2h_{jk}^2}\right\} \quad (11)$$

式 (11) で M はサンプル $\boldsymbol{x_i}$ の次元数であり，$\boldsymbol{x_{jk}}$ は SOINN 内の内部クラス U_{jk} に存在する全ノードの平均位置ベクトルである．また h_{jk} は核関数の領域の大きさを示すパラメータであり，以下のように算出した．

$$h_{jk} = \frac{1}{N_{jk}} \sum_{l=1}^{N_{jk}} |a_l - \boldsymbol{x_{jk}}| \quad (12)$$

ここで a_l は内部クラス U_{jk} のノード l の位置ベクトルを示し，N_{jk} は内部クラス U_{jk} に含まれるノードの総数を示す．この $P_{class}(\boldsymbol{x_i}|U_{jk})$ から得られる対数ゆう度 $\log(P_{class}(\boldsymbol{x_i}|U_{jk}))$ を，局所的ゆう度と定義する．

最終的に，$\log(P_{whole}(\boldsymbol{x_i}|S_j))$，$\log(P_{class}(\boldsymbol{x_i}|U_{jk}))$ を用いて，状態 S_j に対する入力サンプル x_i のゆう度 $C(\boldsymbol{x_i}, S_j)$ は次式で示される．

$$C(\boldsymbol{x_i}, S_j) = \frac{1}{2}\left(\log\left(\sum_k^K \omega_{jk} P_{class}(\boldsymbol{x_i}|U_{jk})\right) + \log(P_{whole}(\boldsymbol{x_i}|S_j))\right) \quad (13)$$

式 (13) において，$\omega_{jk} = \frac{N_{jk}}{N_j^{all}}$ とした．ここで N_j^{all} は状態 S_j の SOINN 内に存在する全ノードの総数を示し，K は状態 S_j の SOINN 内の内部クラスの数を示す．

2.3.3　入力データの認識

SOINN-DP 法は，各クラスの訓練データから作成されたテンプレートモデルと入力データとを DP マッチングすることで，入力データがどのクラスに属するかを認識する．

入力データの認識は次式に従って行う．

$$c^* = \arg\max_c \boldsymbol{E}(\boldsymbol{IP}, \boldsymbol{TM_c}) \quad (14)$$

式 (14) の右辺は，入力データ IP と最も累積一致度 $\boldsymbol{E}(\boldsymbol{IP}, \boldsymbol{TM_c})$ の大きいテンプレートモデルのクラス番号を出力する関数であり，この場合入力データ IP の帰属クラスは c^* であると認識される．ここで DP マッチングで用いる漸化式及び累積一致度 $\boldsymbol{E}(\boldsymbol{IP}, \boldsymbol{TM_c})$ の算出方法は次項で述べる．

2.3.4　SOINN-DP 法の漸化式

クラス c のテンプレートモデル TM_c と入力データ IP との DP マッチングには，式 (1) と同様の対称型漸化式を用いる．SOINN-DP 法に対称型漸化式を用いた理由は，実データの認識実験において非対称型漸化式を用いた場合より対称型漸化式を用いた場合の方が認識精度が向上したためである．

SOINN-DP 法で用いる対称型漸化式を以下に定義する．

$$Q(i,j) = \max\begin{cases} Q(i,j-1) + C(\boldsymbol{x_i}, S_j) \\ Q(i-1,j-1) + 2C(\boldsymbol{x_i}, S_j) \\ Q(i-1,j) + C(\boldsymbol{x_i}, S_j) \end{cases} \quad (15)$$

式 (15) において $C(\boldsymbol{x_i}, S_j)$ はテンプレートモデル TM_c の j 番目の状態 S_j に対する，入力データ IP の i フレーム目のベクトル $\boldsymbol{x_i}$ のゆう度を示す．

SOINN-DP 法では，このゆう度の和が最大になるように DP マッチングが行われる．DP マッチングの結果，テンプレートモデル TM_c と入力データ IP の累積一致度 $\boldsymbol{E}(\boldsymbol{IP}, \boldsymbol{TM_c})$ は次式で表される．

$$\boldsymbol{E}(\boldsymbol{IP},\boldsymbol{TM_c}) = \frac{\boldsymbol{Q}(I_{IP},J_c)}{I_{IP}+J_c} \tag{16}$$

式 (16) において I_{IP} は入力データ IP の時系列長，J_c はテンプレートモデル TM_c の時系列長を表す．

3. 実　　験

本章では，SOINN-DP 法の学習機能，認識精度を検証するために実データを用いた認識実験を行った．

SOINN-DP 法の時系列データの汎用的学習機能を評価するために，本研究では動画像から得られる動作データと音素データの 2 種類のデータセットを用いた．

3.1 比 較 手 法

SOINN-DP 法との比較手法には，HMM，ストキャスティック DP 法を用いた．

3.1.1 HMM (Hidden Markov Model)

HMM はシンボル出力確率の計算方法によって，離散型 HMM と連続分布型 HMM に分類される．ここで本研究では，音声認識・動作認識では連続分布型 HMM が多く用いられるため，比較手法には連続分布型 HMM を用いた．

また HMM はトポロジー（状態の接続関係）によって，ある状態からすべての状態に遷移できる全遷移型（Ergodic）モデルや，状態遷移が一定方向に進む left to right モデルなどに分類される．一般に音声認識や動作認識の分野では，left to right モデルが多く用いられるため，比較手法には left to right モデルの HMM を用いた．

HMM のパラメータ推定法には Baum-Welch アルゴリズムを用いた．また Baum-Welch アルゴリズムのパラメータ推定精度を向上させるため，パラメータの初期値設定に Segmental K-means 法を用いた．

3.1.2 ストキャスティック DP 法

ストキャスティック DP 法 [12] で用いられた漸化式を次式に示す．

$$Q(i,j) = \max \begin{cases} Q(i-2,j-1) + \log P(\boldsymbol{a}_{i-1}|j) \\ \quad + \log P(\boldsymbol{a}_i|j) + \log P_{DP1}(j) \\ Q(i-1,j-1) + \log P(\boldsymbol{a}_i|j) \\ \quad + \log P_{DP2}(j) \\ Q(i-1,j-2) + \log P(\boldsymbol{a}_i|j) \\ \quad + \log P_{DP3}(j) \end{cases} \tag{17}$$

漸化式（式 (17)）は非対称型の漸化式を基盤に構成されている．漸化式（式 (17)）における条件確率 $P(\boldsymbol{a}_i|j)$ と状態遷移確率 $P_{DP1,2,3}(j)$ は [12] に記載された手法で算出した．ここで条件確率 $P(\boldsymbol{a}_i|j)$ は多次元正規分布である．[12] では $P(\boldsymbol{a}_i|j)$ の共分散行列に関してはある範囲で同じものを使っている．例えば 10 個の状態で同じ共分散行列を用いる場合，状態 1〜10 に分配されたデータすべてから一つの共分散行列 σ を最ゆう推定により算出し，状態 1〜10 の各状態で同じ σ を用いる（状態 11〜20，21〜30 でも同じ操作を行う）．

3.2 パラメータ設定

SOINN のパラメータ及び SOINN-DP 法のパラメータを予備実験により設定した．

3.2.1 予備実験によるパラメータ設定

SOINN-DP 法のパラメータを設定するために，孤立単語を用いた予備実験を行った．実験には，男性話者 3 人により 50 回ずつ 5 単語を発話したデータ（1 単語につき 150 個，計 750 個）を用いた．単語は「こんばんは」，「こんにちは」，「またあした」，「おはよう」，「さようなら」の 5 単語である．音声特徴量には，後の本実験と同様の特徴量（**3.3.2** を参照）を用いた．一クラスにつき訓練データを 50，テストデータを 100 として，テストデータと訓練データを交換しながら計 20 回のクロスバリデーション実験を行った．20 回のクロスバリデーション実験から各実験でテストデータに対する認識率を求め，その平均値を求めた．この平均認識率が最大となったパラメータを，後の音素認識実験及び動作認識実験に用いた．

3.2.2 SOINN のパラメータ

SOINN を用いて学習する際，二つのパラメータ (a_d，λ) の設定が必要となる．ここで学習回数 LT は Baum-Welch アルゴリズムの再推定の繰返し回数と同様のパラメータであり，十分な学習回数 LT を設定する必要がある．予備実験の結果，SOINN の学習回数を $LT = 30000$ に設定し $\lambda = 10000$ と設定した．すなわち学習中に 3 回のノイズとおぼしきノードの削除を行った．

次に SOINN-DP 法では，一つの状態に分配されるサンプルが少数であるため，a_d を小さくすると学習結果が不安定であった．したがって，本研究では $a_d = 10000$ と設定した．すなわち学習中に 3 回のエッジの削除を行った．

3.2.3 SOINN-DP 法のパラメータ

式 (9) のセグメント数 L の設定方法を説明する．L を大きくした場合，各状態の SOINN への入力データ

が多くなるため，SOINN の学習精度が向上すると考えられる．しかし L を極端に大きい値に設定すると，時系列的に離れたデータを一つの状態に入力することになり，時系列を無視することになる．この結果，過渡的な時系列データの特徴をモデル化できず，テストデータに対する認識率の低下を招く．逆に L を極端に小さい値に設定すると，SOINN のネットワーク（ノードとエッジの集合）が構築されない．ノード数が少数の場合，式 (10) における共分散行列 Σ を求めることが困難となる．したがって式 (10) における共分散行列 Σ を求めることが可能なサンプル数を一つの状態 (SOINN) に入力すべきである．[12] では，特徴量の次元数 p に対して，少なくとも $p \times 4$〜5 倍以上のサンプル数が必要であり，p^2 個以上が望ましいとされている．

ここで訓練データ N 個をモデルの学習に用いた場合，各状態に分配されるサンプル数を平均 N 個と仮定する．この場合，一つの SOINN に入力されるサンプル集合 $\mathcal{Z}^*{}_i$ のサンプル数 $N \times L$ は以下で定義する．

$$(\text{訓練データ数 } N) \times (\text{セグメント数 } L) \geq 4p \sim p^2$$

したがってセグメント数 L は以下の範囲となる．

$$L \geq \frac{4p}{N} \sim \frac{p^2}{N} \tag{18}$$

予備実験を通して，上記の範囲内における最適なセグメント数を $L \geq \frac{6p}{N}$ を満たす最小の値と決定した．ここでセグメント数 L はストキャスティック DP 法の共分散行列を共有する範囲に対応すると考えられる．ストキャスティック DP 法を用いて同様の予備実験を行ったところ，ストキャスティック DP 法で最大の認識率が得られる共分散行列を共有する範囲は，セグメント数 L と等しいことを確認した．これより後の本実験において，ストキャスティック DP 法の共分散行列を共有する範囲は L とした．

このセグメント数 L の SOINN-DP 法への寄与を **4.2** で考察する．

3.3 音素認識実験

英語音素を対象とした認識実験を行った．

3.3.1 音素データ

本実験では特定話者認識タスクを行う．実験に用いたデータベースは以下の 2 種類であり，これらの詳細は表 1 に示す．

（1） KED TIMIT [20]

- 1 回の実験に用いる 1 クラス当りのデータ数は (訓練データ, テストデータ) $=(40,60),(80,20)$ とした．
- 訓練データとテストデータを入れ換え，10 回のクロスバリデーション実験を行った．

（2） Resource Management1 [21]

- 男性話者 2 名 (BEF03，DTB03)，女性話者 2 名 (CMR02，DAS12) の計 4 人によって発話された英単語データを，音素境界でセグメントし，音素データを収集した．
- 1 回の実験に用いる 1 クラス当りのデータ数は (訓練データ, テストデータ) $=(80,40)$ とした．表 1 より，1 クラス当りの 1 人の話者のデータ数は 30 であるため，このうち 20 データを訓練，10 データをテストに用いた．したがって 1 回の実験で訓練データ数は 20 データ $\times$ 4 人で 80，テストデータ数は 10 データ $\times$ 4 人で 40 とする．この操作を毎回の実験で行った．
- 訓練データとテストデータを入れ換え，10 回のクロスバリデーション実験を行った．

表 1 音素認識実験に用いたタスク
Table 1 Task of phone classification experiment.

タスク：	特定話者認識
データ (1)：	KED TIMIT データベース
認識対象：	英語文章からセグメントした音素：39 クラス (aa,ae,ah,ao,ax,ay,bcl,ch,dcl,dh,dx,eh,er,ey, f,gcl,h,ih,iy,jh,k,kcl,l,m,n,ng,ow,p,pcl,r,s,sh, t,tcl,uw,v,w,y,z)
話者：	男性 1 名
サンプル数：	計 3900 サンプル（1 クラスにつき 100 サンプル）
データ (2)：	Resource Management1 データベース
認識対象：	英単語からセグメントした音素：27 クラス (aa,ae,ax,ay,b,ch,d,eh,el, ey,f,iy,jh,k,l,m,n,ow,p,r,s,sil,t,uw,v,w,y)
話者：	男性 2 名 (BEF03，DTB03), 女性 2 名 (CMR02，DAS12)
サンプル数：	計 3240 サンプル（1 クラスにつき 120 サンプル (4 人 $\times$ 30 サンプル)）

3.3.2 音声からの特徴抽出

実験で用いた音声データの特徴抽出時のパラメータ，及び特徴量は以下のとおりである．

- サンプリング周波数：16 kHz
- フレーム長：15 ms
- フレーム周期：5 ms
- 特徴量：12 次元 MFCC (Mel-Frequency Cepstrum Coefficient) 特徴量，対数パワー，12 次元 ΔMFCC 特徴量，Δ 対数パワー，12 次元 ΔΔMFCC 特徴量，ΔΔ 対数パワーからなる計 39 次元の特徴量

実験で用いた SOINN-DP 法のパラメータについてセ

グメント数 L は式 (18) より，訓練データ 40 個の場合 $L = 6$，訓練データ 80 個の場合 $L = 3$ と決定した．

HMM の各状態の出力確率は全共分散行列をもつ混合正規分布とした．ここで，最大の認識率を得る HMM の最適なパラメータ（状態数及び混合正規分布の混合数）を探索する必要がある．このため，それらのパラメータを変化させながら実験を行い，最適なパラメータを探索し，そのパラメータを用いた場合の認識率を求め，これを HMM による認識率とした．

次にストキャスティック DP 法については，非対称型漸化式（式 (17)）を用いる場合以外に，対称型漸化式を用いた場合の実験も行った．これは対称型漸化式を用いている SOINN-DP 法との比較を行うためである．対称型漸化式には式 (15) における $C()$ を $P(\boldsymbol{a}_i|j)$ に交換した式を用いた．共分散行列を共有する範囲は，SOINN-DP 法のセグメント数 L と同様に訓練データ 40 個の場合 6 状態の間，訓練データ 80 個の場合 3 状態の間とした．

3. 3. 3 音素認識実験の結果

10 回のクロスバリデーション実験の結果から得られたテストデータに対する平均認識率を表 2 に示す．表 2 において 1 段目は KED TIMIT データベースで訓練データ数 40 の場合の平均認識率，2 段目は KED TIMIT データベースで訓練データ数 80 の場合の平均認識率，3 段目は Resource Management1 データベースで訓練データ数 80 の場合の平均認識率をそれぞれ示す．「SO-DP」は SOINN-DP 法，「ST-DP(1)」は非対称型漸化式を用いたストキャスティック DP 法，「ST-DP(2)」は対称型漸化式を用いたストキャスティック DP 法をそれぞれ示す．HMM の認識率の下の () 内は，最大の認識率を得たときのパラメータ（S：状態数，M：混合数）を示す．表 2 より，いずれのタスクにおいても SOINN-DP 法の平均認識率は，ストキャスティック DP 法，HMM のそれに比べ，良好であった．

HMM を用いた実験については，比較のため状態数を 1 個から 13 個まで変動させ実験を行った結果，状態数 3〜7 の付近において認識率が最大となったため，状態数 3〜7 が本実験で使用した音素データに対する最適状態数であると推定した．

そこで状態数 3〜7 において，各状態に割り当てられている出力確率を混合連続確率分布に変更し，混合数を変化させ実験を行った．実験の結果，KED TIMIT データベースで訓練データ 40 個の場合，5 状態 2 混合，KED TIMIT データベースで訓練データ 80 個の場合，5 状態 4 混合，また Resource Management1 データベースで訓練データ 80 個の場合，3 状態 3 混合において認識率が最大となった．またストキャスティック DP 法については，[12] で提案された非対称型漸化式を用いるより，対称型漸化式を用いた認識結果の方が良好であった．これは，対称型漸化式を用いる方が時系列の伸縮を吸収しやすいためと考えられる．

以上をまとめて音素認識実験の結果，SOINN-DP 法では，ストキャスティック DP 法，HMM で得られる最大の認識率より，良好な認識率を得られることを示した．

3. 4 動作認識実験

本節では，動画像から得られる動作を対象とする認識実験を行った．実験には単眼カメラから直接とらえた人間による 7 種類の全身運動（動作）を用いた．実験に用いた 7 種類の動作の内容を図 7 に示す．動画のフレーム率は 29 フレーム毎秒とし，各動作の時間長は最小で 110 フレーム，最大で 440 フレームであ

表 2 音素実験におけるテストデータに対する認識率 [%]（表は 10 回のクロスバリデーション実験の結果から得られたテストデータに対する平均認識率を示している．k-TIMIT は KED TIMIT データベースを示し，RM1 は Resource Management1 データベースを示す．また TD40 は訓練データ数が 40 の場合，TD80 は訓練データ数が 80 の場合をそれぞれ示している．）

Table 2 Classification rate in phone classification task [%].

	SO-DP	ST-DP(1)	ST-DP(2)	HMM
k-TIMIT TD40 [%]	56.36	30.81	51.71	47.69 (5S, 2M)
k-TIMIT TD80 [%]	62.55	33.83	55.46	51.90 (5S, 4M)
RM1 TD80 [%]	71.85	47.52	68.55	63.49 (3S, 3M)

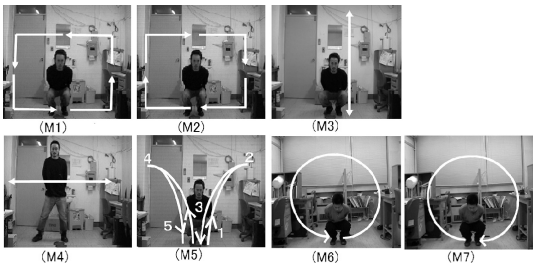

図 7 実験に用いた動画像例（実験で用いた動作 M1〜M7 の様子を示す）

Fig. 7 Examples of moving images used for experiments.

る．入力データには個人差を含み，動作の各部分において伸縮性も含まれている．また「反時計回りに四角形を描く動作 (M1)」と「反時計回りに円を描く動作 (M6)」及び M2 と M7 は類似しており，これらの動作の識別・認識には，得られる時系列データの変化量を詳細にモデル化する必要がある．上記の動作を用いて認識実験を行うことで，SOINN-DP 法のモデル化の性能を評価する．

3.4.1 動画像からの特徴抽出

本研究では，位置不変特徴である局所自己相関特徴 [22] を学習に用いて，動的特徴を抽出した．

[Step1.] まずフレーム間差分画像を算出する．次に差分画像の RGB 値を輝度値に変換し，輝度値にしきい値を設定して，2 値化する．

[Step2.] 差分画像間において，時間方向の自己相関特徴 [22] を抽出する．自己相関特徴の算出には 3×3 サイズ（計 9 次元）のマスクを用いた．ここで自己相関特徴は各フレーム間の時系列方向のみ抽出した．中央のマスクの値を除いて，各フレームで計 8 次元の入力ベクトルを得る．

各フレーム間において，自己相関特徴を抽出した．入力ベクトルは 3×3 サイズ（計 9 次元）のマスクの値を用いた．ただしマスクの中央位置のマスクの値には，「動き」の方向性特徴が現れないため，この値を除いた．結果的に各フレームで計 8 次元の入力ベクトルが得られた．

3.4.2 実 験 条 件

実験条件の詳細を表 3 に示す．表 1 のデータセットについて，（訓練データ 15・テストデータ 10）の場合，3 人が行った 5 回の動作データを訓練データに，別の 2 人が行った 5 回の動作データをテストデータに用いる．また（訓練データ 20・テストデータ 5）の場合，4 人が行った 5 回の動作データを訓練データに，別の 1 人が行った 5 回の動作データをテストデータに用いる．

表 1 の評価方法について，（訓練データ 15・テストデータ 10）の場合，5 人から 3 人（または 2 人）を選ぶ組合せで 10 回，（訓練データ 20・テストデータ 5）の場合，5 人から 4 人（または 1 人）を選ぶ組合せで 5 回のクロスバリデーション実験を行った．

SOINN-DP 法のパラメータには音素認識実験と同じものを用いた．ただしセグメント数 L は式 (18) より訓練データ 10 個，入力次元 8 なので $L = 5$ とした．

HMM の各状態の出力確率は全共分散行列をもつ多次元正規分布とした．ここで音素認識実験と同様に，

表 3 動作認識実験の条件
Table 3 Condition of motion classification experiment.

撮影条件：	複数の室内環境において，単眼カメラを用いて撮影
認識対象：	7 種類（クラス）の動作（全身運動）（図 7） （「反時計回りに四角形を描く動作 (M1)」, 「時計回りに四角形を描く動作 (M2)」, 「上下 2 往復の屈伸運動 (M3)」, 「左右 2 往復の移動 (M4)」, 「座った状態から体を斜めに開く動作 (M5)」, 「反時計回りに円を描く動作 (M6)」, 「時計回りに円を描く動作 (M7)」）
被験者の数：	5 人
データ数：	1 人が各動作を 5 回ずつ行い，各動作につき 25 データ（5 人 × 5 回）を収集した．7 クラスの合計データ数は 175（7 クラス × 25 データ）
データセット：	1 回の実験に用いる 1 クラス当りのデータ数 (訓練データ数, テストデータ数) = $(15, 10), (20, 5)$
評価方法：	訓練データとテストデータを入れ換えながら，クロスバリデーション実験
特徴量：	局所自己相関特徴計 8 次元

表 4 動作実験におけるテストデータに対する認識率 [%]（表は 100 回のクロスバリデーション実験の結果から得られたテストデータに対する平均認識率を示している．TD15 は訓練データ数が 15 の場合，TD20 は訓練データ数が 20 の場合をそれぞれ示している．）
Table 4 Correct classification rate in motion classification task [%] (SOINN-DP was compared to stochastic DP and HMM).

Method	SOINN-DP	ST-DP(1)	ST-DP(2)	HMM
TD15 [%]	97.29	92.14	94.14	89.86(S9)
TD20 [%]	98.29	97.14	97.71	90.29(S10)

状態数を変化させながら実験を行い最適なパラメータを探索し，そのパラメータを用いた上での認識率を求めた．

またストキャスティック DP 法については非対称型漸化式（式 (17)）を用いる場合以外に，対称型漸化式を用いた場合の実験も行った．ストキャスティック DP 法の，共分散行列を共有する範囲は 5 状態の間とした．

3.4.3 動作認識実験の結果

実験結果として得られた認識率を表 4 に示す．HMM の認識率の右の () 内は，最大の認識率を得たときのパラメータ（S：状態数）を示す．表 4 から，訓練データ 15 の場合 (TD15) も訓練データ 20 の場合 (TD20) も SOINN-DP 法の認識率は，比較手法の認識率に比べ良好であることを示した．

HMM を用いた実験において，比較のため状態数を 1 から 15 まで変動させ実験を行った結果，状態数 11 において認識率が最大となった．

ストキャスティック DP 法について，音素認識実験と同様に, [12] で提案された非対称型漸化式を用いた場合より，対称型漸化式を用いた場合の認識結果の方が良好であった．

音素認識実験と同様に動作認識実験でも，SOINN-DP 法では，ストキャスティック DP 法，HMM で得られる最大の認識率より良好な認識率が得られた．

3.5 実験から得られた知見

ストキャスティック DP 法及び HMM との比較実験の結果から得られた知見をまとめる．

［ストキャスティック DP 法との比較］

ストキャスティック DP 法と SOINN-DP 法の各タスクでの認識率を比較した結果，SOINN-DP 法はストキャスティック DP 法よりも認識率の点で優れていた．この結果から，状態を SOINN によって詳細に近似する SOINN-DP 法はストキャスティック DP 法よりも時系列データの頑健なモデル化を行うことが可能であることを示した．

［HMM との比較］

SOINN-DP 法では各状態の出力分布を SOINN によって自動的に決定することが可能である．また状態数は標準データの時系列数と決定されるため，状態数もあらかじめ設定する必要がない．一方，HMM で時系列データを学習する際，状態数と状態の出力分布（混合正規分布の場合，混合数）を事前に決める作業が必要である．

このため実験では HMM の状態数及び混合数を変化させながら，認識率が最も高くなる場合を探索した．この探索結果から得られた HMM の最大認識率より，SOINN-DP 法の認識率は良好であった．以上の結果より，SOINN-DP 法では事前に状態数及び出力分布のパラメータを設定せずに，高い認識率が得られることが示された．

4. 考　　察

本章では，提案した SOINN-DP 法の性能に関する考察及び SOINN-DP 法に関する今後の課題について議論する．

4.1 自由パラメータ数の比較

本節では SOINN-DP 法のパラメータ数と比較手法（連続型 HMM，ストキャスティック DP 法）のパラメータ数を比較する．SOINN-DP 法では，SOINN のパラメータである (a_d, λ) とセグメント数 L の三つのパラメータを設定する必要がある．また連続型 HMM では，状態数及び状態の出力分布に用いる混合正規分布の混合数の二つのパラメータを設定する必要がある．またストキャスティック DP 法では，共分散行列を共有する範囲（SOINN-DP 法におけるセグメント数 L の機能と等しい）を設定する必要がある．したがって SOINN-DP 法のパラメータ数は一番多く，二つの比較手法より設定すべきパラメータの数が多い．

ただし，SOINN-DP 法では予備実験（本実験とは異なるタスク）により SOINN のパラメータを決定し，このパラメータを本実験で用いている．このため音素データや動作データといった認識対象に応じて毎回パラメータを設定する必要はない．これに対し，連続型 HMM では認識対象に応じてパラメータを設定する必要がある．また [12] で提案されたストキャスティック DP 法（実験における ST-DP(1)）については，パラメータ数が少ない点はメリットであるが，上記の実験において十分な認識率は得られなかった．

4.2 セグメント数 L と認識率の関係

SOINN-DP 法のパラメータであるセグメント数 L は，SOINN-DP 法の認識性能に影響を与える．ここでは，セグメント数が増えることによる認識精度への影響について検証実験を行った．

この実験ではセグメント数 L の変化に対する，認識率の変移を検証する．検証は，KED TIMIT データベースを用いて，**3.3** と同様の条件下で音素認識実験を行った．1 回の実験に用いる 1 クラス当りの訓練データ数は 40 とした．またセグメント数を 1〜10 まで変化させて，それぞれのセグメント数を用いた場合について計 10 回のクロスバリデーション実験を行った．実験を行った結果を図 8 に示す．

図 8 から，セグメント数を $L = 1$ から増加させる

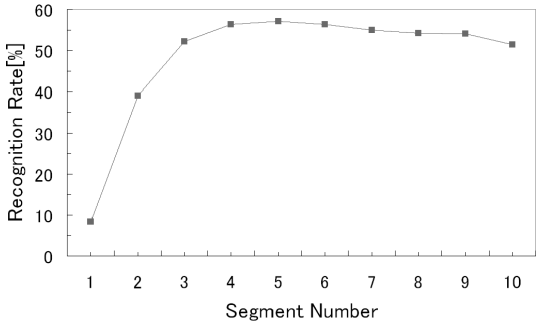

図 8　セグメント数と認識性能の関係
Fig. 8　Relation between number of segments and classification rate.

ごとに徐々に認識率が上昇し，$L = 5$ で最大認識率 (57.04%) を得た．更にセグメント数を増加させると認識率は下降することが確認できる．

一方，予備実験の結果から求めたセグメント数は，訓練データ 40 個の場合には $L = 6$ であった．図 8 より，$L = 6$ の場合は全体で 3 番目に認識率が高いことが分かる．この結果から，本論文で用いたセグメント数の推定方法（**3. 2. 3**）が妥当であったことが示された．

4. 3 SOINN の認識性能への寄与

SOINN-DP 法は，各状態の出力分布を SOINN によって詳細に近似する点で，ストキャスティック DP 法を拡張した手法となっている．このため本節では，SOINN-DP 法の認識性能に SOINN がどのように寄与しているかを議論する．

ここでは SOINN を用いて分布を近似しない手法を二つ定義し，SOINN-DP 法との比較を行った．SOINN の学習結果を用いない手法と比較を行うことで，SOINN の認識精度への寄与を検証した．SOINN の学習結果を用いない比較手法には以下の二つの手法を用いた．

［手法：1］ サンプル群 $\mathcal{Z}^*{}_j$ を SOINN に入力せず，$\mathcal{Z}^*{}_j$ から直接，最ゆう推定により多次元正規分布 $P(\boldsymbol{x_i}|S_j)$ を求めた．ゆう度 $C(x_i, S_j) = \log(P(\boldsymbol{x_i}|S_j))$ とし，このゆう度 $C(x_i, S_j)$ を用いた漸化式によって入力データの認識を行った．

［手法：2］ サンプル群 $\mathcal{Z}^*{}_i$ を SOINN に入力し，SOINN の分類結果から $P_{whole}(\boldsymbol{x_i}|S_j)$ を求めた．ただし SOINN のクラスタリング結果から得られる $P_{class}(\boldsymbol{x_i}|U_{jk})$ を入力データの認識に用いなかった．これはゆう度 $C(\boldsymbol{x_i}, S_j)$ を式 (19) で表し，$\alpha = 0$ とおくことに等しい．

$$C(\boldsymbol{x_i}, S_j) = \alpha \log \left(\sum_k^K \omega_{jk} P_{class}(\boldsymbol{x_i}|U_{jk}) \right) + (1-\alpha) \log(P_{whole}(\boldsymbol{x_i}|S_j)) \quad (19)$$

ここで検証実験は KED TIMIT データベースを用いて，**3. 3** と同様の条件下で行い，1 回の実験に用いる 1 クラス当りの訓練データは 80 個，セグメント数 $L = 3$ とした．α を 0〜1.0 まで 0.05 ずつ変化させながら，それぞれの α を用いた場合について計 10 回のクロスバリデーション実験を行った．

検証実験の結果から得られた，［手法 1］及び［手法 2］で得られた認識率と SOINN-DP 法で得られた認識率を表 5 に示す．SOINN-DP 法の認識率は［手法 1］，［手法 2］の認識率を約 4%上回っている．この結果より，SOINN の学習結果として得られる内部クラスの情報を用いる SOINN-DP 法がこの情報を用いない［手法 1］，［手法 2］に対して認識率の点で優位性があることが示された．ただし表より，［手法 1］と［手法 2］で得られるテストデータに対する認識率はほとんど等しい．この結果は SOINN の学習結果として得られる内部クラスの情報を用いない限り，SOINN の効果が見られないことを意味する．総じて，SOINN の学習結果として得られる内部クラスの情報を用いることで，認識率が向上することを確認した．

表 5 SOINN を用いない場合の認識率と SOINN-DP 法で得られる認識率の比較
Table 5 Comparison between classification rate obtained by SOINN-DP and classification rate obtained by method that doesn't use SOINN.

	SOINN を用いない手法		SOINN を用いた手法	
Method	［手法 1］	［手法 2］	SOINN-DP	$\alpha = 0.45$
[%]	58.41	58.86	62.55	63.22

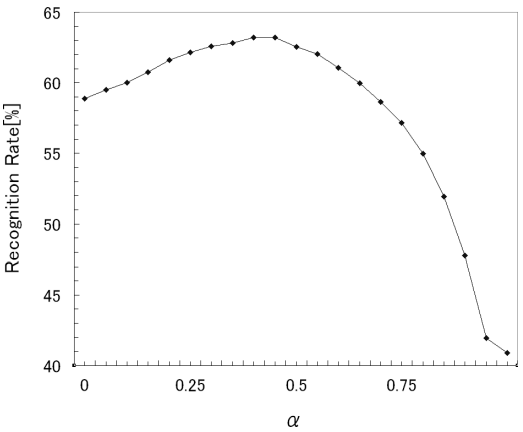

図 9 α の変化に対する認識率の変化（$\alpha = 0.0$ での認識率が［手法 2］で得られた認識率である．$\alpha = 0.45$ において認識率が最大 (63.22%) となる．）
Fig. 9 Changes of the classification rate when α is changed.

次に，α の変化に対して認識率がどのように変化していくかを議論する．α の変化に対する認識率の変化を図 9 に示す．

図 9 では，x 軸方向が α の値を示しており，y 軸が各々の α に対応する認識率を示している．図 9 より，認識率は $\alpha = 0.45$ で最大となり，それ以上 α を大きくすると認識率は低下する．最終的に $\alpha = 1.0$ で認識率は最低の値となる．$\alpha = 1.0$ の状態とは式 (19)

より，右辺の第 2 項のみでゆう度の算出を行うことに相当する．この場合，SOINN の学習結果として得られる内部クラスの情報のみを用いてゆう度を計算している．各内部クラスは核関数により近似されているため，次元間の相関を多次元正規分布のようにモデル化できない．このため各内部クラスによる情報のみを用いてゆう度を計算した場合，テストデータに対する認識率が低下したと考えられる．ただし表 5 の結果からも分かるように，各内部クラスによる情報と大域的情報（SOINN の全ノード）の両方を用いることで認識率を向上させることができた．また図 9 より，$\alpha = 0.45$ で最大の認識率が得られている．これは α をデータにフィッティングさせることによって，更に SOINN-DP 法の認識精度を向上させることが可能であることを示唆している．今後，データセット及び特徴量に応じてこの重みパラメータ α を推定する手法を検討する．

4.4 SOINN-DP 法の計算量

入力データの認識に要する計算量は入力データの時系列長とモデルの状態数に依存する．ここで同じ時系列長の入力データの認識を行う場合，SOINN-DP 法では多数の状態を保持するため少数の状態数を保持する HMM より計算時間がかかる．この議論では，1 状態の出力確率の計算に要する時間が HMM と SOINN-DP 法で等しいと仮定した．

この問題への対処として，DP マッチングの際に要する計算量を削減する方法が考えられる．この方法は様々な研究で提案されている．[23] では，DP の漸化式に整合窓を設けることで計算量の削減が見込まれるとされている．[24] では，解の最適性を保ったまま DP マッチングの高速化を図る手法が提案されている．つまりこれらの手法を SOINN-DP 法に取り入れることで計算量の削減が見込める．

4.5 SOINN-DP 法の拡張について

今後 SOINN-DP 法を，実際の音声認識及び動作認識のタスクに応用するための課題について議論する．

4.5.1 音声認識への応用

今後，特定話者認識だけでなく不特定話者認識にも応用可能な手法として，SOINN-DP 法を拡張する予定である．これに際し，以下の 2 点を今後の課題とする．

まず不特定話者認識に応用するためには未知の話者に対しての認識精度を保障するため，多数の話者による音声をモデル化する必要がある．このため個人差の影響を受けた，クラス内分散が大きい時系列データ群をモデル化する機能が不可欠である．この問題の解決策として，従来手法であるマルチテンプレート DP 法 [25] の利用が考えられる．具体的には，SOINN-DP 法のテンプレートモデルを一つのクラスで複数用意することで，クラス内分散が大きい時系列データ群のモデル化に対処する．SOINN-DP 法においてマルチテンプレート化を行う方法論の提案は今後の課題とする．

本実験では，比較手法の HMM には連続型 HMM を用いた．実際の音声認識システムでは，異なるモデル・状態間で混合正規分布のための正規分布を共有することで，効率的なモデルを構成するタイドミクスチャ形の HMM が用いられる．したがって今後，連続型 HMM だけでなくタイドミクスチャ形の HMM [1] との比較も行い，SOINN-DP 法の音素のモデル化性能を検証する予定である．

4.5.2 動作認識への応用

今後，ハンドジェスチャやヘッドジェスチャなどの他の動作について SOINN-DP 法を適用する予定である．また SOINN-DP 法は DP マッチングの拡張手法と考えられるため，連続動作・連続音声認識の手法である 2 段階 DP マッチング [26] を適用することが可能である．したがって，2 段階 DP マッチングを用いて連続動画像の認識（及び連続音声認識）への応用も検討する．

5. むすび

本研究では，一つのフレームを一つの状態に対応させ，各状態の出力分布を Self-Organizing Incremental Neural Network (SOINN) によって近似することで，時系列データを頑健にモデル化可能な SOINN-DP 法を提案した．SOINN-DP 法はストキャスティック DP 法の拡張手法であり，各状態の出力分布を SOINN が自動的に推定する点に特徴がある．

SOINN-DP 法の有効性を検証するため，動画像から得られる動作と音素を用いて認識実験を行った．実験の結果，連続型 HMM 及びストキャスティック DP 法より認識精度の点で提案手法の有効性を示した．今後，**4.4** で述べた計算量の問題を解決し，**4.5** で述べた応用問題に SOINN-DP 法を適用する．

謝辞 本研究の実施にあたり NEDO 産業技術研究助成事業から支援を頂きました．記して感謝致します．また音声認識についての指導をして頂いた山岸順一氏（The Centre for Speech Technology Re-

search，University of Edinburgh 研究員）に深く感謝致します．

文　献

[1] L.R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proc. IEEE, pp.257–286, 1989.

[2] 篠田浩一，“確率モデルによる音声認識のための話者適応化技術，” 信学論（D-II），vol.J87-D-II, no.2, pp.371–386, Feb. 2004.

[3] 益子貴史，徳田恵一，小林隆夫，今井　聖，“動的特徴を用いた hmm からの音声パラメータ生成アルゴリズム，” 音響誌，vol.53, no.3, pp.192–200, 1997.

[4] J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov models,” Proc. IEEE International Conference on Computer Vision, pp.379–387, 1992.

[5] A. Elgammal, V. Shet, Y. Yacoob, and L.S. Davis, “Learning dynamics for exemplar-based gesture recognition,” Proc. IEEE International Conference on Computer Vision and Pattern Recognition, vol.1, pp.16–22, 2003.

[6] A. Wilson and A. Bobick, “Learning visual behavior for gesture analysis,” Proc. IEEE International Symposium on Computer Vision, vol.5A, Motion2, 1995.

[7] R. Hamdan, F. Heits, and L. Thoraval, “Gesture localization and recognition using probabilistic visual learning,” Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.98–103, 1999.

[8] M. Ostendorf, V. Digalakis, and O. Kimball, “From hmms to segment models: A unified view of stochastic modeling for speech recognition,” IEEE Trans. Speech Audio Process., vol.4, no.5, pp.360–378, 1996.

[9] 岡　隆一，“連続 dp を用いた連続単語認識，” 音響学音声研資，SP78-20, 1978.

[10] 西村拓一，向井理朗，野崎俊輔，岡　隆一，“低解像度特徴を用いた複数人物によるジェスチャの単一動画像からのスポッティング認識，” 信学論（D-II），vol.J80-D-II, no.6, pp.1563–1570, June 1997.

[11] 川嶋宏彰，西村拓一，“コンピュータビジョンにおける時系列パターン認識，” 信学技報，CVIM2006-154, 2006.

[12] 中川聖一，“ストキャスティック dp 法および統計的手法による不特定話者の英語子音の認識，” 信学論（D），vol.J70-D, no.1, pp.155–163, Jan. 1987.

[13] F. Shen and O. Hasegawa, “An incremental network for on-line unsupervised classification and topology learning,” Neural Netw., vol.19, no.1, pp.90–106, 2006.

[14] H. Bourlard and N. Morgan, “Continuous speech recognition by connectionist statistical methods,” IEEE Trans. Neural Netw., vol.4, no.6, pp.893–909, 1993.

[15] Y. Bengio, R.D. Mori, G. Flammia, and R. Kompe, “Global optimization of a neural network-hidden Markov model hybrid,” IEEE Trans. Neural Netw., vol.3, no.2, pp.252–259, 1992.

[16] E. Trentin and M. Gori, “Robust combination of neural networks and hidden Markov models for speech recognition,” IEEE Trans. Neural Netw., vol.14, no.6, pp.1519–1531, 2003.

[17] B. Fritzke, “A growing neural gas network learns topologies,” Advances in Neural Information Processing Systems (NIPS), pp.625–632, 1995.

[18] T. Martinetz and K. Schulten, “Topology representing networks,” Neural Netw., vol.7, no.3, pp.507–522, 1994.

[19] R. Duda, P. Hart, and D. Stork, Pattern Classification, second ed., John Wiley & Sons, Canada, 2001.

[20] University of Edinburgh, “CSTR US KED TIMIT,” http://www.teu.ac.jp/media/˜earth/FK/

[21] P. Price, W. Fisher, J. Bernstein, and D. Pallett, “Resource management complete set 2.0,” Linguistic Data Consortium, Philadelphia, 1993.

[22] 大津展之，“パターン認識における特徴抽出に関する数理的研究，” 電子技術総合研究所研究報告，vol.818, 1981.

[23] 内田誠一，“Dp マッチング概説—基本と様々な拡張，” 信学技報，CVIM2006-166, 2006.

[24] C. Raphael, “Coarse-to-fine dynamic programming,” IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.12, pp.1379–1390, 2001.

[25] L. Rabiner and B. Juang, Fundamentals of Speech Recognition, PTR Prentice-Hall, 1993.

[26] H. Sakoe, “Two-level dp-matching—A dynamic programming-based pattern matching algoritm for connected word recognition,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-27, no.6, pp.588–595, 1979.

（平成 19 年 4 月 19 日受付，10 月 9 日再受付）

岡田　将吾　（学生員）

2003 横浜国大・工・知能物理卒，2005 東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程了．現在，同大学院博士課程在学中．

長谷川　修　（正員）

1993 東京大学大学院電子工学専攻博士課程了．博士（工学）．同年電子技術総合研究所入所，1999 から 1 年間米国カーネギーメロン大学客員研究員，2001 産業技術総合研究所主任研究員，2002 東京工業大学像情報工学研究施設助教授．