

別刷

# 計測自動制御学会 論文集

年 第 卷 第 号

---

(P. ~P. )

---

---



社団法人 計測自動制御学会

## 確率ニューラルネットを木構造に導入した階層型クラスタリング†

岡本 勝\*・ト 楠\*・辻 敏夫\*

### Hierarchical Clustering with Tree Structure Based on Probabilistic Neural Networks†

Masaru OKAMOTO\*, Nan BU\* and Toshio TSUJI\*

This paper proposes a novel hierarchical clustering method. The radical distinction from traditional methods is that the proposed method requires no specific knowledge of the number of classes for classification. In this method, at each node of a hierarchical classification tree, a log-linearized Gaussian mixture network<sup>3)</sup> is utilized for clustering, and a newly invented learning law is applied to train the LLGMN unsupervisedly to classify data into two subclasses based on statistical characteristics. This method performs a binary classification hierarchically and, finally conducts a classification with a suitable number of classes. Also, unnecessary structure of the classification tree can be avoided using cross-validation. Validity of the proposed method is demonstrated with classification experiments on artificial data and electromyogram (EMG) signals.

**Key Words:** unsupervised learning, neural networks, hierarchical clustering, EMG signals

#### 1. ま え が き

筋電位信号や脳波などの生体信号は、人間の生理的な内部状態を反映しており、これらの信号のパターンから人間の意図する動作が推定できれば、自然な感覚で操作できるマン・マシンインタフェースを構築できる可能性がある。近年、このような動作の推定を行なうために EMG 信号のパターン識別に関する研究が進められている<sup>1)~3)</sup>。パターン識別は、学習データと対応する教師信号との写像関係を学習することによって行なわれるが、高齢者や障害者の生体信号はクラス間の特徴の違いがあいまいであるため、教師信号の信頼性が低くなり、識別精度が大幅に低下してしまう場合がある。

一般に教師信号が利用できない、あるいは信頼性が低い場合には、学習データの分布特性などからデータの類似関係を推定し、教師なしでクラスタリングを行なう方法が用いられる。クラスタリング手法は、K-平均法<sup>4)</sup>と階層型クラスタリング<sup>5)</sup>の2種類に大別することができ、これまでにそれぞれ多くの手法が提案されている。

K-平均法は、入力空間を与えられたクラス数と等しい数の部分空間に切り分けていく手法で、クラスタリング結果は局所的な最適性が保証される。一方、階層型クラスタリングは、サブクラスへの分類を階層的に繰り返していき、最終的に各クラスに分類する。この手法の木構造の構築法として

は、学習データひとつひとつを別クラスとみなした状態からクラスの融合を繰り返すボトムアップ型や、ID3 や C4.5, C5.0 など<sup>7)</sup>のように、ひとつのクラスから各階層で最適な分割を繰り返して、最終的に望ましいクラスタリングを実現するトップダウン型がある。さらに、階層型クラスタリングでは木を構成した後、枝刈りを行なうことによって指定されたクラス数でのクラスタリングが可能である。

しかしながら、これらの手法を用いる際に、分類すべきクラス数や各データが属するクラスの情報が必要となり、十分な情報が得られない問題に対しては、クラスタリングを行なうことは難しい。このような問題を防ぎ、データが属するクラス数を自動的に決定する手法として、自己組織化特徴マップ<sup>6)</sup>(Self Organization Maps:以下、SOM と略記)を用いたものや<sup>8)</sup>統計力学のアナロジーに基づいた自己組織化手法<sup>9)</sup>が提案されている。これらの手法では、SOM の各ユニットに分類されるデータ数に基づいたクラス数の推定や自由エネルギーと呼ばれる基準をもとに入力空間をいくつかの部分空間に分割することによってクラス数を決定する。さらに、Balanced Interactive Reducing and Clustering using Hierarchies (以下、BIRCH と略記)と呼ばれる手法も提案されている<sup>10)</sup>。BIRCH は、線形識別子を階層的に組み合わせることで、与えられたデータのクラスタリングと構築された木構造の終端ノード数に対応するクラス数の推定を同時に行なう。

しかしながら、扱う問題の複雑さに応じて、ネットワークのユニット数などの最適なパラメータを事前に決定する必要があり、分布構造が未知なデータに対して適切なクラスタリングを実現することは難しい。

† 第6回情報論的学習理論ワークショップで発表(2003・11)

\* 広島大学 大学院 工学研究科 東広島市鏡山 1-4-1

\* Hiroshima Univ., Faculty of Engineering, Kagamiyama 1-4-1, Higashi-Hiroshima

(Received June 24, 2004)

(Revised November 8, 2004)

そこで本論文では、クラスタリング時に木構造を構築することによって、木構造の終端ノード数としてクラス数を決定する手法を提案する。提案手法では混合正規分布モデル (Gaussian mixture model:以下、GMM と略記) を内包したニューラルネットワークのひとつである Log-Linearized Gaussian Mixture Network<sup>3)</sup> (以下、LLGMN と略記) を用いることにより、複雑な分布構造への対応を可能にする。本論文ではまず、LLGMN でデータのクラスタリングを行なうため教師なし学習則を導出し、LLGMN を階層型クラスタリングの各非終端ノードにおける分類識別子として用いたクラスタリングアルゴリズムを示す。LLGMN を用いてデータの分布構造の統計的性質を学習することにより、出力である事後確率を利用して分類の曖昧さの検証ができ、データの統計的特性の観点から不要な分類の生成を避けることが可能となる。本論文では、各分岐点で 2 クラス分類を行なうクラスタリング手法について報告する。

以下、2. で各非終端ノードにおける分類に用いる LLGMN について説明し、3. で LLGMN を識別子とした提案手法の説明を行なう。そして、4. で本手法と従来法との比較を行ない、本手法の有効性を示す。

## 2. LLGMN

LLGMN は GMM を対数線形化してネットワークに展開したもので、学習データの持つ事後確率を学習によって推定する<sup>3)</sup>。本手法では、LLGMN を階層型クラスタリングの各非終端ノードにおける分類識別子として用いる。以下に、LLGMN の構造、教師あり学習則、および新しく提案する教師なし学習則について説明する。

### 2.1 構造<sup>3)</sup>

LLGMN のネットワーク構造を Fig. 1 に示す。まず、GMM の各コンポーネントに対応する正規分布をネットワークの重み係数として表現するために、入力ベクトル  $\mathbf{x} (\in \mathbb{R}^D)$  を次式のように非線形変換する。

$$\mathbf{X} = [1, \mathbf{x}^T, x_1^2, x_1x_2, \dots, x_1x_D, x_2^2, x_2x_3, \dots, x_2x_D, \dots, x_D^2]^T \quad (1)$$

ネットワークの第 1 層は、変換された入力ベクトル  $\mathbf{X}$  の次元数  $H = 1 + D(D+3)/2$  にあわせて  $H$  個のユニットから構成され、各ユニットは入力  $X_h$  をそのまま出力する。第 1 層の出力は重み  $w_h^{(c,m)}$  を掛け合わされて第 2 層に伝えられる。ただし、 $h = 1, 2, \dots, H$ ;  $c = 1, 2, \dots, C$ ;  $m = 1, 2, \dots, M_c$  である。 $C$  は出力ユニット数、 $M_c$  は出力  $c$  に対応するコンポーネント数を表す。第 1 層の出力を  $^{(1)}O_h$ 、第 2 層のユニット  $\{c, m\}$  の入力を  $^{(2)}I_{c,m}$ 、出力を  $^{(2)}O_{c,m}$  とすると、

$$^{(2)}I_{c,m} = \sum_{h=1}^H ^{(1)}O_h w_h^{(c,m)} \quad (2)$$

$$^{(2)}O_{c,m} = \frac{\exp[^{(2)}I_{c,m}]}{\sum_{c'=1}^C \sum_{m'=1}^{M_{c'}} \exp[^{(2)}I_{c',m'}]} \quad (3)$$

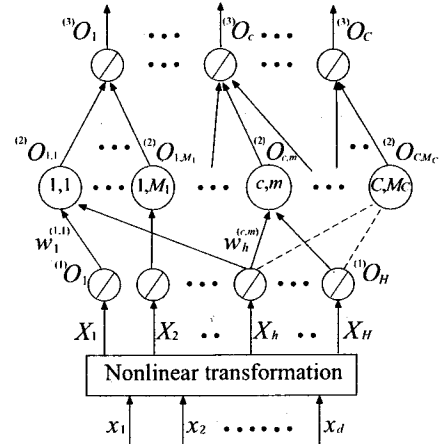


Fig. 1 The structure of LLGMN<sup>3)</sup>

となる。ただし、 $w_h^{(C,M_c)} = 0$  である。さらに第 3 層のユニット  $c$  は第 2 層の  $M_c$  個のユニットと結合しており、第 3 層のユニット  $c$  の入力  $^{(3)}I_c$  と、出力  $^{(3)}O_c$  は、

$$^{(3)}I_c = \sum_{m=1}^{M_c} ^{(2)}O_{c,m} \quad (4)$$

$$^{(3)}O_c = ^{(3)}I_c \quad (5)$$

で与えられる。第 3 層の出力  $^{(3)}O_c$  は、入力ベクトルが対応するクラス  $c$  に分類される事後確率  $P(c|\mathbf{x})$  であり、分類結果の曖昧性を評価する指標として用いることが可能である。

このネットワークはデータの持つ統計的特性に応じて高い適応学習能力を持ち、混合正規分布モデルを内包しているため、従来のパラメータが制限された正規分布を利用した手法<sup>11)</sup>と比較して複雑な分布構造を持つデータの識別が可能である。

### 2.2 教師あり学習則<sup>3)</sup>

入力ベクトルの集合  $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$  に対して、教師信号  $\mathbf{T}^{(n)} = (T_1^{(n)}, \dots, T_C^{(n)}, \dots, T_C^{(n)})$  ( $n = 1, \dots, N$ ) が与えられた場合を考える。 $\mathbf{x}^{(n)}$  がクラス  $c$  のときは  $T_c^{(n)} = 1$  とし、それ以外は  $T_c^{(n)} = 0$  ( $\hat{c} \neq c$ ) とする。ここで学習データ数を  $N$  とすると、ネットワークは次式の評価関数を最小化、つまり各学習データが教師信号の示すクラスに識別される尤度を最大化するように学習を行なう。

$$J_{SV} = - \sum_{n=1}^N \sum_{c=1}^C T_c^{(n)} \log ^{(3)}O_c^{(n)} \quad (6)$$

このとき、重みの修正量  $\Delta w_h^{(c,m)}$  は

$$\Delta w_h^{(c,m)} = -\eta \sum_{n=1}^N \frac{\partial J_{SV}^{(n)}}{\partial w_h^{(c,m)}} \quad (7)$$

$$\begin{aligned} \frac{\partial J_{SV}^{(n)}}{\partial w_h^{(c,m)}} &= \frac{\partial}{\partial w_h^{(c,m)}} \left( - \sum_{c=1}^C T_c^{(n)} \log ^{(3)}O_c^{(n)} \right) \\ &= ^{(2)}O_c^{(n)} - \frac{^{(2)}O_c^{(n)}}{^{(3)}O_c^{(n)}} T_c^{(n)} X_h^{(n)} \end{aligned} \quad (8)$$

で与えられる。 $\eta > 0$  は学習率である。

### 2.3 教師なし学習則

上述したように教師信号を用いた学習により LLGMN は高い精度での識別が可能となるが、生体信号のような教師信号の信頼性が低いデータの識別では、識別精度が低下してしまうといった問題点がある。このような問題に対しては教師信号を必要としない教師なし学習を行なう必要があるが、LLGMN では、入力信号の特性からクラスタリングを行なう学習則は提案されていない。そこで、本論文では入力信号が各クラスに属する事後確率のあいまいさに基づいた教師なし学習則を新たに導出する。

クラス数  $C$  が既知である場合には、ネットワークは  $N$  個のデータが各クラスに分類されるあいまいさの尺度として次式のようにエントロピー  $J_{SO}$  を求める。

$$J_{SO} = - \sum_{n=1}^N \sum_{c=1}^C {}^{(3)}O_c^{(n)} \log {}^{(3)}O_c^{(n)} \quad (9)$$

この式を評価関数として、次式の重み修正量  $\Delta w_h^{(c,m)}$  を用いて  $J_{SO}$  を最小化することにより、クラスタリングのあいまいさを小さくするような学習が可能である。

$$\Delta w_h^{(c,m)} = -\eta \frac{\partial J_{SO}}{\partial w_h^{(c,m)}} \quad (10)$$

$$\begin{aligned} \frac{\partial J_{SO}}{\partial w_h^{(c,m)}} &= \frac{\partial}{\partial w_h^{(c,m)}} \left( - \sum_{n=1}^N \sum_{c=1}^C {}^{(3)}O_c^{(n)} \log {}^{(3)}O_c^{(n)} \right) \\ &= - \sum_{n=1}^N \sum_{c'=1}^C \frac{\partial}{\partial {}^{(3)}O_{c'}^{(n)}} ({}^{(3)}O_{c'}^{(n)} \log {}^{(3)}O_{c'}^{(n)}) \\ &\quad \times \sum_{m'=1}^{M_{c'}} \frac{\partial {}^{(3)}O_{c'}^{(n)}}{\partial {}^{(2)}O_{c',m'}^{(n)}} \frac{\partial {}^{(2)}O_{c',m'}^{(n)}}{\partial {}^{(2)}I_{c,m}^{(n)}} \frac{\partial {}^{(2)}I_{c,m}^{(n)}}{\partial w_h^{(c,m)}} \\ &= -(J_{SO} - \log {}^{(3)}O_c^{(n)}) {}^{(2)}O_{c,m}^{(n)} X_h^{(n)} \quad (11) \end{aligned}$$

このように重みを修正していくことによって、 $C$  クラスに分類するクラスタリングが可能となるが、設定した重みの初期値によっては教師なし学習の最初から全データを1つのクラスとして分類してしまい、ローカルミニマに収束するといった問題点がある。K-平均法などでは、データの平均値などを利用して初期パラメータの設定を行なうことによりこの問題を回避できるが<sup>12)</sup>、LLGMN のネットワーク構造では、各コンポーネントの平均値や分散値などの統計的パラメータがネットワークの重みに陽に現れないため、学習に適した重みの初期値を事前に設定することが困難である。また、この学習則では分類するクラス数  $C$  に関する情報が必要となってしまうといった問題点が残る。

そこで、このようなローカルミニマを回避し、クラス数未知の問題のクラスタリングを行なうため、本手法では LLGMN を分類識別子とした階層型クラスタリングを提案する。これにより、各 LLGMN での分類が単純な構造となるため、ローカルミニマの回避が可能となる。また、階層構造を利用することにより、任意のクラス数でのクラスタリングが可能となる。

### 3. 階層型クラスタリング

本手法では分類木の構築時に学習データの持つ統計情報に基づいて枝の生成を判定する。本論文では、分岐木として2進決定木を用い、各ノードで2クラスへの分類に限定した教師なしクラスタリングアルゴリズムを提案する。

#### 3.1 分類木

分類木の初期構成として、ルートノードのみの状態を用意し、LLGMN を用いて学習データを2クラスに分類する。分割結果をサブクラスとし、対応するノードを木構造に追加する。さらに LLGMN を用いてサブクラスを詳細に分類していく。以上のサブクラスの分割とノードの追加を繰り返すことにより、データの統計的特徴に基づいた分類木の構築が可能となる。

提案手法による木構造の構築アルゴリズムを示す。

- (1) すべてのデータを含むクラスとしてルートノードを用意する。
- (2) 終端ノードの中からひとつのノードを選択し、そのノードに分類されたデータを用いて LLGMN の教師なし学習を行ない、データを2つのサブクラスに分類する。
- (3) 前ステップで学習した LLGMN を用いたデータ分割の判定を行なう。基準を満たしていれば、木構造にサブクラスに対応するノードを追加する。満たしていなければ分岐生成を停止し、ひとつのクラスとする。
- (4) すべての終端ノードにおいて分岐生成が停止するまで(2)、(3)を繰り返す。
- (5) 構築された木から外れ値に対応するノードを削除する(枝刈り)。

このように木構造の構築を LLGMN の教師なし学習とともに進めていくことにより、統計的情報に応じたクラス数でのクラスタリングが可能となる。

以下に、LLGMN の教師なし学習、分割判定、そして枝刈りについて説明する。

#### 3.2 各ノードの教師なし学習

前述したように LLGMN の教師なし学習則だけでは、初期値によっては正確なクラスタリングが行なえない場合があるため、ここでは、分類クラス数を  $C=2$  とした教師なしクラスタリング法を提案する。本手法では学習を行なうノードに分類された一部のデータを選択して重みの初期化を行ない、クラスタリングの進行に合わせて学習データを段階的に追加していく。これにより、学習初期には単純な分布のクラスタリングを行ない、徐々に複雑な分布のクラスタリングへと移行していくことが可能となる。

以下、学習データを2クラス  $C_1, C_2$  に分類する LLGMN の教師なし学習則を示す。ここで、 $\mathbf{A}$  は学習を行なう LLGMN に対応する学習データの集合、 $\mathbf{B}$  は学習データのうちクラスタリングに用いるデータの集合を示す。まず、初期化のため学習データの集合  $\mathbf{A}$  から2点のデータ  $\mathbf{x}_1, \mathbf{x}_2$  を選択し、 $\mathbf{B} = \{\mathbf{x}_1, \mathbf{x}_2\}$  とする。これには、乱数等によりラ

ランダムに選択する方法と何らかのヒューリスティックを用いる方法が考えられるが、ここでは(12)式に従い、2点間のユークリッド距離が最大となるデータ  $\mathbf{x}_1, \mathbf{x}_2$  を選択する。

$$(\mathbf{x}_1, \mathbf{x}_2) = \underset{\mathbf{x}^{(i)}, \mathbf{x}^{(j)} \in A}{\operatorname{argmax}} (\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\|) \quad (12)$$

選択した2点  $\mathbf{x}_1, \mathbf{x}_2$  がそれぞれ  $C_1, C_2$  に属するとし、 $\mathbf{x}_1, \mathbf{x}_2$  が属するクラスに正確に分類されるように(6)式の教師あり学習則を用いて学習を行なう。このとき仮想教師信号として  $\mathbf{x}_1, \mathbf{x}_2$  に対してそれぞれ  $\mathbf{T}^{(\mathbf{x}_1)} = (1, 0)^T, \mathbf{T}^{(\mathbf{x}_2)} = (0, 1)^T$  を与える。このように初期化を行なうことによって、すべてのデータをひとつのクラスとして分類するような初期値を避けることができる。

つぎに、上記の手法により設定した初期重みをもとに、教師なし学習を行なう。まず、集合  $B$  に含まれるデータのうち、クラス  $C_1, C_2$  に属するデータの平均値  $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$  を求め、どちらか一方とのユークリッド距離が最小となるデータ  $\mathbf{x}^{(k)} \in A - B$  を選択する。 $\mathbf{x}^{(k)}$  は集合  $B$  に含まれないデータの中でどちらかのクラスに最も当てはまりやすいデータであるため、ネットワークの教師なし学習によって、推定パラメータが大きく変化することはない。そこで、 $B = B + \{\mathbf{x}^{(k)}\}$  とし、 $B$  に属するデータを用いて(10)式によりLLGMNの教師なし学習を行なう。教師なし学習を一定回数またはエントロピーの平均値が閾値  $H_T$  以下になるまで学習を行なった後、さらに  $B$  に追加するデータを選択し、データを追加した集合  $B$  を用いて教師なし学習を行なう。このデータの追加と教師なし学習の繰り返しを  $A - B = \{\emptyset\}$  となるまで行なうことにより、データの分布構造に基づいた分類が可能となる。

### 3.3 分割判定

3.2で述べる方法を用いて学習したLLGMNを用いてデータの分割を行なうことで、データの統計的特徴に基づいたクラスターリングが可能となるが、分類木の階層が深くなりすぎると、過剰な分類を生じてしまう危険性がある。そこで本手法では、木構造の分岐生成をLLGMNでの事後確率から判定することにした。判定のために、対応するノードに分類された教師信号の存在しないデータの集合  $N_a$  を2分割し、ひとつをLLGMNの教師なし学習を行なう学習用データ  $N_l$ 、もう一方をLLGMNによる分割の妥当性を判定する検証用データ  $N_v$  とする ( $N_a = N_l + N_v, N_l \cap N_v = \{\emptyset\}$ )。まず、学習用データ  $N_l = A$  としてLLGMNの教師なし学習を行なう。3.2で述べる方法を用いた学習したLLGMNで検証用データ  $N_v$  を分類し、エントロピー  $H(\mathbf{x})$  を求める。そして、この  $H(\mathbf{x})$  の平均値  $H_E$  を分岐生成の判定基準として用いる。

$$H_E = \frac{1}{|N_v|} \sum_{\mathbf{x}^{(n)} \in N_v} H(\mathbf{x}^{(n)}) \quad (13)$$

$$H(\mathbf{x}^{(n)}) = - \sum_{c=1}^C {}^{(3)}O_c^{(n)} \log {}^{(3)}O_c^{(n)} \quad (14)$$

$|N_v|$  は  $N_v$  に属するデータの数を表わす。対象ノードに分類された検証用データのエントロピーの平均値  $H_E$  が、指定した閾値  $H_T$  より高ければ、検証用データに対して十分な汎化性が得られていないと考え、そのノードでの分割を停止し、対応するノードに分類されたデータはひとつのクラスとする。このように、LLGMNを用いることによって算出できる検証用データのエントロピーを分岐生成の停止判定基準として利用することにより、汎化性が低下する分類に対しては学習データ特有の分布構造と考え、分類を停止することで過剰な分割を避けることができる。

また、検証用データが分割したサブクラスのうち、片方のクラスのみ分類された場合は学習データに外れ値が存在したと考え、対応するノードでの分割を停止する。さらに、終端ノードにおいて学習データが1つになった場合はさらなる分割は不可能なため、分割を停止し、停止したノードに分類されたデータをひとつのクラスとする。

### 3.4 枝刈り

本手法では学習用データに外れ値などが存在する場合、そのデータをひとつのクラスとした木構造を構築してしまう。特に、木の深さが大きくなってくると、学習用データの数が少なくなり、外れ値の影響が相対的に高くなってしまふ。そこで、過剰な分割を避けるため、木構造の構築後に不必要なノードの枝刈りを行なう。

ノードの枝刈りにおいて、木のどの部分の削除を行なうかを判定するために、本手法では各終端ノードに分類された学習用データの総数を利用する。各終端ノードに分類された学習用データの数と全学習データ数との比が閾値  $\alpha_T$  より少なければ、対応する終端ノードを削除していく。このように、構築した木構造から枝刈りを行なうことによって、学習用データ内に存在する外れ値に対応したクラスの追加を防ぐことが可能である。

## 4. 実験

提案法の有効性の検証を行なうため、人工データの分類を試み、ほかの教師なし学習手法との比較を行なった。また、障害者の生体信号の識別問題において、発声ができない場合や切断患者などの場合には、計測した生体信号と意図する動作間の客観的な対応関係が得られないため、教師あり学習を行なうパターン識別では識別精度が低下してしまう。そこで、筋電位信号の分類、識別実験を行ない、教師信号の信頼性が低い問題において、教師あり学習に対する提案手法による分類の有効性を示す。

### 4.1 人工データ分類による検証

Fig. 2に実験に用いた人工データを示す。データは2次元空間上に分布する5つのクラスから構成され、各クラスのデータは正規分布から生成される。各クラスのパラメータはTable 1に示す。このような分布から生成されたデータを属するクラスに関する教師信号を利用せずに分類を行ない、分類結果を教師信号を利用して評価を行なった。学習

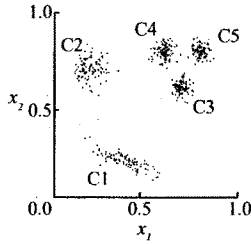


Fig. 2 Examples of artificial data, the parameters of which are shown in Table 1

Table 1 Parameters of the artificial data used in experiments

class	$\mu_{x_1}$	$\mu_{x_2}$	$\delta_{x_1}$	$\delta_{x_2}$	$\delta_{x_1 x_2}$
C1	0.4	0.2	0.08	0.04	-0.8
C2	0.2	0.7	0.05	0.05	0
C3	0.7	0.6	0.03	0.03	0
C4	0.6	0.8	0.03	0.03	0
C5	0.8	0.8	0.03	0.03	0

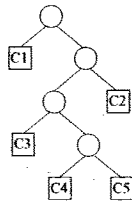


Fig. 3 The tree structure for the five classes data

するデータ数は、各クラスごとに100データを学習用、200データを検証用に用いて実験を行なった。分岐生成の閾値は  $H_T = 0.2$ 、枝狩りの閾値は  $\alpha_T = 0.01$ 、学習時の閾値は  $H_{T'} = 0.001$  とした。LLGMNのパラメータは、コンポーネント数を  $M_c = 1(c = 1, 2)$ 、学習係数は  $\eta = 0.01$ 、学習回数は学習データ追加ごとに100回とした。

Fig. 3 に提案法によるクラスタリング例を示す。図中の○は非終端ノード、□は終端ノードを示し、分類されたクラスに対応する。各終端ノード内の文字は分類された学習データが属するクラスを示す。各試行ごとに重みの初期値を乱数で与え、20試行の分類実験を行なったところ、平均分類精度が98.5%、標準偏差は0.64%となり、高い分類精度が実現された。ここで、分類精度は木構造の構築後に各クラス300データの分類を行ない、正しいクラスに分類される確率とした。

提案法の有効性の検証のため、寺島の方法<sup>8)</sup>とBIRCHとの比較を行なった。寺島の方法ではSOMの自己組織化を行ない、各ユニットに対応するデータ数のヒストグラムからクラス数の推定を行なう。一方、BIRCHは木構造にデータをひとつずつ追加していきながらクラスタリングを変更していき、木構造を構築することでクラス数の推定を行なう。データ数は、各クラスごとに100データを学習用、200データを検証用データとした。

Table 2 に提案法および寺島の方法、BIRCHにより10試行学習を行なった時に各手法において各クラス数が推定された回数を示す。提案手法におけるLLGMNの重みは各試

Table 2 The number of classes estimated by the proposed method, the Terashima's method<sup>8)</sup> and BIRCH

The number of classes	4	5	6	7
The proposed method	0	9	1	0
The Terashima's method ( $N_U = 50$ )	2	7	1	0
The Terashima's method ( $N_U = 40$ )	5	5	0	0
The Terashima's method ( $N_U = 30$ )	7	3	0	0
BIRCH	0	0	3	7

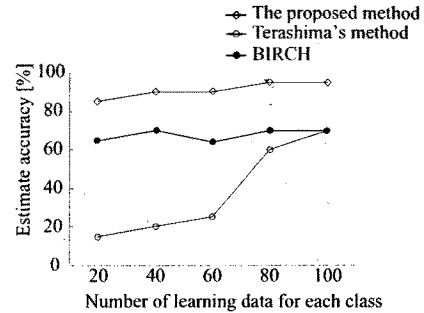


Fig. 4 Effect of the number of learning data on estimate accuracy

行ごとに乱数で与えた。寺島の方法におけるSOMのユニット数はそれぞれ、 $N_U = 30, 40, 50$  とした。

提案法では従来手法に比べて、より正確なクラス数の推定が可能であることがわかる。寺島の方法においてはSOMのユニット数が多い場合は正確なクラス数に近づくが、SOMのユニット数が少ない場合、データの分布を表現する正確なヒストグラムが構成できないため、正確なクラス数が推定できなくなる。また、BIRCHでは各クラスに属するデータの分散和を基準として分割を行なっていくため、分散値が大きいクラス(たとえばクラスC1)を複数のクラスに分類してしまい、推定したクラス数が実際のクラス数よりも大きくなる。このように、従来手法ではユニット数や閾値などのパラメータを問題ごとに適した値に事前に決定する必要があり、この決定法に明確な基準が存在しないため試行錯誤的に調整する必要がある。提案法では、問題に適した数のLLGMNでネットワークを構築でき、かつ、各クラスの分布構造も学習を通じて推定できるので、これらの問題を回避することができる。

つぎに、学習用データ数を変更した際のクラス数の推定精度をFig. 4に示す。学習用データ数は各クラスごとに20, 40, 60, 80, 100とした。また、検証用データ数はそれぞれ、学習用データ数の2倍とした。縦軸は各学習用データ数ごとに20試行の学習を行なった中で、推定クラス数が真のクラス数と等しかった回数の割合である。ここで、寺島の方法のSOMのユニット数は $N_U = 30, 40, 50$ とした。Fig. 4から、寺島の方法ではデータ数が少なくなるにつれて精度が大きく低下しているが、提案法では、データ数が少なくなった場合でもほぼ正確なクラス数での分類が行なえていることがわかる。ネットワーク構造にSOMに採用している寺島の方法では、SOMの各ユニットに分類されるデータ数から

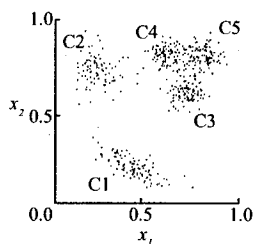


Fig. 5 Examples of artificial data, the parameters of which are shown in Table 3

Table 3 Parameters of the artificial data used in experiments

class	$\mu_{x_1}$	$\mu_{x_2}$	$\delta_{x_1}$	$\delta_{x_2}$	$\delta_{y_{x_2}}$
C1	0.4	0.2	0.08	0.1	-0.8
C2	0.2	0.7	0.08	0.08	0
C3	0.7	0.6	0.05	0.05	0
C4	0.6	0.8	0.05	0.05	0
C5	0.8	0.8	0.05	0.05	0

Table 4 The number of classes estimated by the proposed method, the Terashima's method<sup>8)</sup> and BIRCH

The number of classes	3	4	5	6	7
The proposed method	1	2	6	1	0
The Terashima's method ( $N_U = 50$ )	9	1	0	0	0
The Terashima's method ( $N_U = 40$ )	9	1	0	0	0
The Terashima's method ( $N_U = 30$ )	10	0	0	0	0
BIRCH	0	0	2	2	6

クラス数を推定するが、分散が大きいデータ (たとえばクラス C1) では、各ユニットにおいて分類されるデータが少なくなる場合があり、ヒストグラムではひとつのクラスと識別されなくなり、正確なクラス数よりも少なく推定してしまう。一方、提案手法では、LLGMNによりデータの分散値に関しても学習可能であるため、分散値の大きいクラスに対しても正確にひとつのクラスとして推定している。また、BIRCHでは、データ数が少なくなった場合でも推定精度は低下しないが、全体的な推定精度において提案法の有効性が確認できる。

つぎに、人工データの分散値を変更した場合のクラス数推定精度を示す。Fig. 5 に実験に用いた人工データを、Table 3 に各クラスのパラメータを示す。データ数は各クラスごとに 100 データを学習用に、200 データを検証用に用いて実験を行なった。Table 4 に提案手法および従来手法によって 10 試行学習を行なったときの推定したクラス数を示す。寺島の方法では、重なり合ったクラスをひとつのクラスとして分類するため、少ないクラス数での推定が行なわれている。また、BIRCHでは、先程の結果と同様に分散和パラメータの値に応じて異なるクラス数の推定が行なわれている。一方、提案法では各クラスのデータに重なりがある場合でも高い割合で正しいクラス数が推定できている。また、5 クラスに分類されたときの平均識別率が 89.5%、標準偏差 3.271% となり、前述した Fig. 2 のデータの識別精度よりは低下するものの、重なりがある場合にも高い精度が得られて

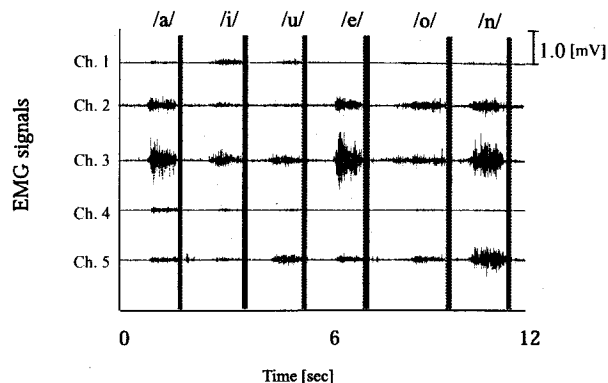


Fig. 6 Examples of raw EMG signals measured from cervical spine injury patient's

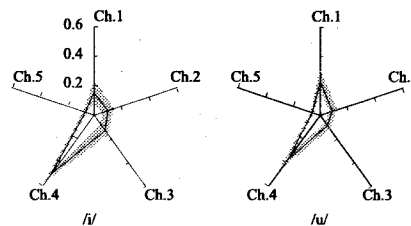


Fig. 7 Radar charts of EMG patterns of /i/ and /u/. The line indicates the mean values of each channel, and the regions of shade imply  $\pm 1$  S.D.

いることがわかる。

最後に、各手法における学習時間の比較結果を示す。提案法の学習時間は寺島の方法 4.23 倍、BIRCH の 2.84 倍となった。この結果、従来法と比較して提案法は識別率が高い反面、学習時間の長さが問題となるため、アルゴリズムの改善など学習時間の短縮が必要であると考えられる。

#### 4.2 筋電位信号

提案法の応用として、言語障害を持つ頸椎損傷患者の発声時に計測した筋電位信号 (以下 EMG 信号) の分類を行なった<sup>13)</sup>。この患者は声を出すことはできないが、口の形を変えることはできる。そこで、頸部および表情筋の計 5Ch. ( $L = 5$ ) から EMG 信号を計測した。測定部位は Ch. 1:咬筋, Ch. 2:顎二腹筋, Ch. 3:下唇下制筋, Ch. 4:大頬骨筋, Ch. 5:口角下制筋である。5 対の電極から測定した EMG 信号を A/D 変換 (サンプリング周波数 1.0[KHz]) し、チャンネルごとに全波整流した後、2 次パワースフィルタ (カットオフ周波数 1.0[Hz]) により平滑化する。この時系列信号を  $E_l(t)$  ( $l = 1, \dots, L$ ) とする。また、時刻  $t$  における EMG 信号より筋力  $F_{EMG}(t)$  を推定し、閾値  $F_T$  を用いて発声時と非発声時を判断する<sup>14)</sup>。分類は、全チャンネルの和が 1 になるように時系列信号を正規化した特徴パターンベクトル  $\mathbf{x}(t)$  を用いる。

$$x_l(t) = \frac{E_l(t) - E_l^{st}}{\sum_{l=1}^L (E_l(t) - E_l^{st})} \quad (15)$$

ここで、 $E_l^{st}$  は安静時における  $E_l(t)$  の時間平均とする。取得した EMG 信号の一例を Fig. 6 に示す。整流、平滑化、正規化を行なった“い”、“う”に属するデータの平均値およ

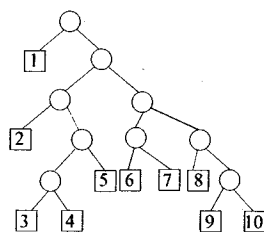


Fig. 8 The constructed tree for the EMG data

Table 5 Clustering results

Node number	/a/	/i/	/u/	/e/	/o/	/n/
1	34	1	0	0	0	0
2	0	23	4	0	0	0
3	2	2	2	0	0	0
4	2	12	8	0	0	0
5	0	0	21	0	0	0
6	0	0	0	0	0	34
7	0	0	0	9	0	0
8	0	0	0	26	0	1
9	2	2	5	5	4	3
10	0	0	0	0	36	3

び標準偏差を Fig. 7 に示す。実線は平均値、影の幅は標準偏差である。図より“い”，“う”に対応する EMG パターンは類似性が高く、特徴の違いがあいまいであるため、教師信号の信頼性が低いと考えられる。

分類には各音ごとに 40 データを学習用，80 データを検証用を用いてクラスタリングを行なった。発声，非発声の判断を行なう閾値は  $F_T = 0.2$ ，分岐生成時のエントロピーの閾値は  $H_T = 0.2$ ，枝符りの閾値は  $\alpha_T = 0.01$ ，学習時の閾値は  $H_{T'} = 0.001$  とした。LLGMN のコンポーネント数は  $M_c = 1(c = 1, 2)$ ，学習係数は  $\eta = 0.01$ ，学習回数は学習データ追加ごとに 500 回とした。

Fig. 8 に提案法によるクラスタリング例を示す。終端ノード内の数字は，提案法によるクラスタリングのクラス番号である。また，Table 5 に各終端ノードにおいて 6 音に属するデータ数を示す。あるノードに分類されたデータがほぼひとつの発声時のデータの場合，そのノードは対応する発声を示すクラスと考えると，表より“あ”，“お”，“ん”に属するデータの多くはほぼひとつのノードに分類されており，ひとつのクラスとして正確なクラスタリングが行なえていることがわかる。“え”に属するデータの多くはふたつのクラス（ノード番号 7，8）に分類されているが，Fig. 6 の EMG 信号を見ると信号パターンが計測中に大きく変化していることから，データがふたつのクラスに分割されたことは妥当である。また，“い”，“う”に属するデータについては Fig. 7 に見られるように，両音の EMG 信号のパターンが似ているため一部のデータが混在したクラスとして推定されている。

一方，ノード番号 9 に分類されるデータには，6 音のデータがすべて含まれている。Fig. 8 においてノード番号 9 に分類されたデータが Fig. 6 の陰影をつけた部分である。分類されたデータを見ると，すべてのデータが各音の発声終了直前のデータ，つまり推定筋力の閾値  $F_T$  によって非発

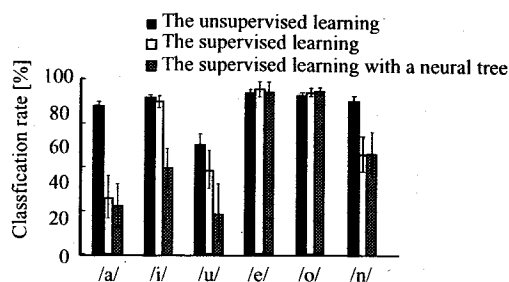


Fig. 9 Classification accuracy

声と判断される寸前のデータである。このようなデータは明確なパターンが見られないあいまいなデータであるため，ひとつのクラスとして分類されたのである。そこで，データ  $x(t)$  がノード番号 9 に分類された場合，1 時刻前のデータ  $x(t-1)$  の識別結果を採用することにより正確な識別が行なえたと考えられる。

つぎに，構築した木構造の識別への利用可能性を検証するため，教師あり学習を行なった LLGMN と識別精度の比較を行なった。比較手法としては教師あり学習を行なった単一の LLGMN と，Neural Tree<sup>15)</sup> の識別子を LLGMN としたものを用いた。識別精度を評価するため，分割されたクラスを学習データの分類結果に基づいてラベル付けを行なう。各クラスに分類されたデータの中で最も多かった音素に該当するクラスが対応しているものと考え，分類されたデータを対応する音として識別する。分類精度は各音ごとに 600 データを選び，正しい音に識別される確率を示したものである。各手法の識別精度を Fig. 9 に示す。各発音の識別率を見ると，図より，全体的にかなり識別率が改善されていることがわかる。全音の識別率は，提案法では 92.6%，標準偏差は 0.86%，単一の LLGMN では識別率は 85% 標準偏差は 2.12%，Neural Tree では 61.4%，標準偏差は 4.3% であった。

以上より，本節で用いた EMG 信号のようにクラス間の差があいまいなデータに関しては，提案法は教師信号を用いた学習と同様の識別精度が得られることがわかる。さらに，あいまいなデータに関しては対象とするクラスとは異なるクラスとしてクラス数の推定，データの分類を行なうため，クラスタリング結果を考慮して識別を行なうことにより，教師あり学習よりも高い精度での識別が期待できる。

## 5. まとめ

本論文では，学習時にクラス数を自動的に決定するため，LLGMN を各非終端ノードでの分類識別子とした階層型クラスタリングを提案した。本手法では，LLGMN の出力から求めたエントロピーによって分類木の分岐生成を判定し，不必要な分岐生成を防ぐことができる。そして，人工データおよび EMG 信号を対象とした検証実験を行ない，提案法の有効性を示した。

今後は， $\alpha_T$  などの閾値の自動決定法，教師無し学習時に



における初期データの選択法や学習則の改善による学習時間の短縮を行なっていく。さらに、さまざまな構造のデータへの適用、パラメータ値の変化によるクラスタリング結果の検証を行ない、その結果に基づいて学習法の修正を検討する予定である。

#### 参 考 文 献

- 1) A. Hiraiwa, K. Shimohara and Y. Tokunaga: EMG pattern analysis and classification by neural network, *IEEE International Conference on Syst., Man and Cybern.*, 1113/1115 (1989)
- 2) 辻, 森, 伊藤: 統計構造を組み込んだニューラルネットによるEMG動作識別法, *電気学会論文誌*, **112-C-8**, 465/473 (1992)
- 3) T. Tsuji, O. Fukuda, H. Ichinobe and M. Kaneko: A log-linearized gaussian mixture network and its application to EEG pattern classification, *IEEE Trans. on System, Man and Cybernetics-Part C: Applications and Reviews*, **29-1**, 60/72 (1999)
- 4) M. R. Anderberg: *Cluster Analysis for Applications*, Academic Press, New York (1974)
- 5) J. H. Ward Jr.: Hierarchical grouping to optimize an objective function, *Journal of the Americal Statistical Association*, **58-301**, 235/244 (1963)
- 6) T. Kohonen: *Self-organization and associative memory, Third Edition*, Springer-Verlag, Berlin (1994)
- 7) J. R. Quinlan: *Induction of decision trees, Machine Learning*, no. 1, 81/106 (1986)
- 8) 寺島, 白石, 山本: 自己組織化特徴マップ上のデータ密度ヒストグラムを用いた教師なしクラスタ分類法, *電子情報通信学会論文誌*, **J79-D-II-7**, 1280/1290 (1996)
- 9) 鈴木, 上田: モジュール競合学習を用いた適応的クラスタリング, *電子情報通信学会論文誌*, **J83-D-II-6**, 1529/1538 (2000)
- 10) T. Zhang, R. Ramakrishnan and M. Livny: An Efficient Data Clustering Method for Very Large Databases, *Proc. 1996 ACM SGMOD int'l Conf. on Management of Data*, 103/114 (1996)
- 11) L. I. Perlovsky and M. M. McManus: Maximum likelihood neural networks for sensor fusion and adaptive classification, *Neural Networks*, **4**, 89/102 (1991)
- 12) Anil K. Jain Robert P. W. Duin and Jianchang Mao: Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern Analysis and Machine Inteligence*, **22-1**, 4/37 (2000)
- 13) 有田, 福田, 大塚, 辻: EMG信号を利用した代用発声システム, 第4回SICEシステムインテグレーション部門講演会, 103/104 (2003)
- 14) 福田, 辻, 金子: EMG信号を利用した手動制御型人間支援マニピュレータ, *日本ロボット学会誌*, **18-3**, 387/394 (2000)
- 15) G. L. Foresti and C. Micheloni: Generalized Neural Trees for Pattern Classification, *IEEE Transactions on Neural Networks*, **13-6**, 1540/1547 (2002)

#### [著 者 紹 介]

#### 岡 本 勝



1979年4月30日生。2004年広島大学大学院工学研究科複雑システム工学専攻博士課程前期修了。現在広島大学大学院工学研究科博士課程後期在学中。ニューラルネット、クラスタリングなどの研究に従事。電子情報通信学会会員

#### ト 楠



1998年中国大連理工大学機械工学科卒業。2001年同大学院機械工学専攻修士課程修了。現在、広島大学大学院工学研究科博士課程後期在学中。ニューラルネット、パターン識別、生体信号解析、ニューロチップなどの研究に従事。IEEE学生会員。

#### 辻 敏 夫 (正会員)



1959年12月25日生。85年広島大学大学院工学研究科博士課程前期修了。同年広島大学工学部助手。94年同助教授を経て、2002年より同大学院工学研究科教授、現在に至る。工学博士。主として、人間とロボットの運動制御、生体信号解析、ニューラルネット、ヒューマン・マシンシステムなどに関する研究に従事。本会学術奨励賞(1986)、論文賞(2002)、バイオメカニズム学会論文賞(1990)、日本義肢装具学会論文賞(2000)、日本医科器械学会論文賞(2003)、IEEE 2003 King-Sun Fu Memorial Best Transactions Paper Award(2004)、日本機械学会ロボティクス・メカトロニクス部門学術業績賞(2004)などを受賞。