

## 灰色理論による学習ベクトル量子化法への一提案

山口 大輔<sup>†</sup>      李 国棟<sup>†</sup>      水谷 晃三<sup>†</sup>      永井 正武<sup>†</sup>

本論文は自己組織化マップなどの工学手法と同じ命題として、灰色理論を用いた分類ネットワーク (GCN: Grey Classification Network) を提案する。本手法は勝者ノードの決定処理にユークリッド距離型灰色分析を適用する。また、本手法は Kohonen 学習則によらない学習アルゴリズムを提案している。教師なし学習法として、Wong らの提案する灰色クラスター分析を改善したアルゴリズムを提案する。教師あり学習法として、Taylor 逐次近似法を用いた非線形最小二乗法の原理によるベクトル量子化法を提案する。数値例に本手法を適用し、学習に必要なパラメータ、学習の性質および効果について議論する。

### Grey Classification Network Based on Self-Organizing Feature Map

Daisuke YAMAGUCHI<sup>†</sup>   Guo-Dong LI<sup>†</sup>   Kozo MIZUTANI<sup>†</sup>   Masatake NAGAI<sup>†</sup>

This paper proposes the grey classification network (GCN) that succeeded to the concepts of traditional self-organizing feature maps (SOFM). In the SOFM, winner node is decided by Euclidian distance. GCN uses a Euclidian distance type grey relational analysis to decide a winner node. GCN has two learning algorithms that don't depend on the Kohonen's learning rule. One is an unsupervised learning algorithm by grey cluster analysis which is already proposed by authors, and another is a vector quantization method using Taylor sequential approximation. Iterative learning is possible in each algorithm. Numerical example is given for GCN, which discusses efficient learning.

#### 1. はじめに

クラスタリングやパターン認識、および画像圧縮などにおいて、自己組織化マップ [1] や学習ベクトル量子化法は有効な工学的手法である。学習アルゴリズムを持つこれらの手法は、適用事例に応じて柔軟にモデルを変えることができるのが利点である。

灰色理論 [2, 5] は制御、時系列予測モデル GM [15] および灰色関連度分析 (灰色分析と呼ぶ) など、時間軸を持つ離散データの解析手法として研究が始まった。近年ではニューラルネットワークに GM を適用する [3] など、ソフトコンピューティングの分野にも使われるようになってきている。Hu ら [4] は自己組織化マップ (SOFM: Self-organizing Feature Map) に灰色分析を導入した GSOFM (Grey SOFM) を提案し、巡回セールスマン問題に適用している。

筆者ら [5, 6] はパターン認識アルゴリズムとして、以前に灰色判別モデルを提案した。このモデルは、灰色分析で得られる類似度 (灰色関連度と呼ぶ) を用いて判別処理を実現している。しかし、判別基準値の設定方法と灰色分析における距離計算の妥当性など問題点をいくつか抱えている。

本論文は灰色判別モデルの拡張展開モデルとして、灰色理論による学習ベクトル量子化法 GCN (Grey Classification Network) を提案する。GCN は自己組織化マップにおける結合重みベクトルや競合演算、

教師なし/教師あり学習などの概念を引き継ぎ、アルゴリズムの数理拠点として灰色理論を用いる。ただし、Deng [2] の提案した伝統的な灰色分析ではなく筆者らが提案しているユークリッド距離型灰色分析 [7] を用いる。両者の違いについては次節以降に述べる。

GCN の機能として、①ユークリッド距離型灰色分析による勝者ノード決定法、②灰色クラスター分析による教師なし学習法、③Taylor 逐次近似法による教師あり学習法の 3 点を提案する。アルゴリズムの提案後に数値例を適用し、GCN の学習に必要なパラメータ、学習の性質および効果について議論する。

#### 2. 灰色判別モデルとその問題点

##### 2.1 灰色判別モデルのアルゴリズム

灰色判別モデルは、 $n$  個の因子を持つデータ  $x_i$  を灰色分析のための比較数列

$$x_i = \{x_i(1), x_i(2), \dots, x_i(n)\}, \quad i = 1, 2, \dots, m \quad (1)$$

として与える。同様に、 $n$  個の因子を持つ判別対象グループ  $g$  ( $1 \leq g \leq m$ ) の代表値を基準数列

$$s_g = \{s_g(1), s_g(2), \dots, s_g(n)\} \quad (2)$$

とする。灰色判別モデルは基準数列  $s_g$  と比較数列  $x_i$  の局所型灰色関連度を灰色分析で調べ、最も関連

<sup>†</sup> 帝京大学大学院理工学研究科, Graduate School of Science and Engineering, Teikyo University

の高いグループ  $s_g$  に入力データ  $x_i$  を判別する方法である。

基準数列  $s_g$  と各比較数列  $x_i$  との灰色関連度  $\Gamma_{gi}$  は、

$$\Gamma_{gi} = \frac{1}{n} \sum_{j=1}^n \left( \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{gi}(j) + \zeta \Delta_{\max}} \right) \quad (3)$$

ただし、

$$\left. \begin{aligned} \Delta_{\min} &= \min_{\forall i} \min_{\forall j} \{ \Delta_{gi}(j) \} \\ \Delta_{\max} &= \max_{\forall i} \max_{\forall j} \{ \Delta_{gi}(j) \} \\ \Delta_{gi}(j) &= |s_g(j) - x_i(j)| \\ \zeta &= 0.5, j = 1, 2, \dots, n \end{aligned} \right\} \quad (4)$$

である。全ての基準数列  $s_g$  について局部型灰色関連度を計算すると、

$$R = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \cdots & \Gamma_{1m} \\ \Gamma_{21} & \Gamma_{22} & \cdots & \Gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{g1} & \Gamma_{g2} & \cdots & \Gamma_{gm} \end{bmatrix} \quad (5)$$

となる行列  $R$  に整理することができる。各入力データ  $x_i$  は行列  $R$  を使用して、

$$s_g = \left\{ x_i \mid \max_{\forall g} \{ \Gamma_{gi} \} \right\}, \quad \text{for all } i \quad (6)$$

となるグループ  $s_g$  に判別する。

## 2.2 灰色判別モデルの問題点

灰色判別モデルの構成は図 1 に示すことができる。文献 [6] では図 1 に示すような表形式の構成図を提案した。ところが、自己組織化マップやニューラルネットワークで使われるネットワーク形式の図を採用すると、灰色判別モデルは図 2 に示すような構成図とみることもできる。比較数列  $x_i$  を入力ベクトル、基準数列  $s_g$  を結合重みベクトルであるとすれば、灰色判別処理は勝者ノードの決定処理を灰色分析により実現していると解釈することができる。

伝統的な自己組織化マップでは結合重みベクトルの値を競合学習によって求めている。しかし、灰色判別モデルは基準数列  $s_g$  の値を分析者が主観的に決めている。文献 [6] に示した適用例では基準数列として各グループに属する入力データの算術平均を使用し、ある程度の判別精度を得ている。

しかし、現状よりもさらに精度向上を望むには、分析者自らが全ての基準数列を再設定しなければならない。分析者の作業負担の軽減と判別精度向上のため、本論文では灰色理論に基づく教師なし学習法および教師あり学習法を提案する。

筆者らは伝統的な灰色分析についても議論を行っている [7]。筆者らは式 (4) の計算方法について議論している。伝統的な灰色分析では式 (4) により距離計算を行っている。式 (4) は 2 つの数列  $x_i, x_j$  間のスカラーによる距離である。しかし、数列  $x_i$  が  $n$  個の因子を持っているとき、 $x_i$  はスカラーではな

	$x_1$	$x_2$	$\cdots$	$x_m$
$s_1$	$\Gamma_{11}$	$\Gamma_{12}$	$\cdots$	$\Gamma_{1m}$
$s_2$	$\Gamma_{21}$	$\Gamma_{22}$	$\cdots$	$\Gamma_{2m}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$s_g$	$\Gamma_{g1}$	$\Gamma_{g2}$	$\cdots$	$\Gamma_{gm}$

図 1 灰色判別モデルの表的解釈

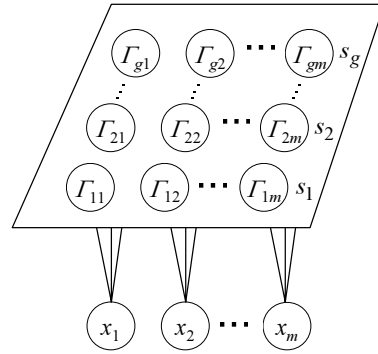


図 2 灰色判別モデルのネットワーク的解釈

くベクトルとなる。ベクトルによる距離、すなわちユークリッド距離にて距離計算を行えば、灰色関連度をより正確に計算することができる。本論文では筆者らが提案 [7] しているユークリッド距離型灰色分析を用いた勝者ノード決定法を提案する。

## 3. 灰色理論による学習ベクトル量子化法の一提案

### 3.1 ネットワーク構成

本論文で提案するネットワークは図 3 に示す構成となっている。ネットワークに入力するベクトルを式 (1) の定義から、

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}], \quad i = 1, 2, \dots, m \quad (7)$$

と再定義する。結合重みベクトルは、式 (2) の定義から、

$$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jn}], \quad j = 1, 2, \dots, g \leq m \quad (8)$$

と再定義する。 $\mathbf{x}_i, \mathbf{w}_j$  はともに  $n$  個の因子を持つ。各因子の値を座標とすれば、 $\mathbf{x}_i, \mathbf{w}_j$  は  $n$  次元空間に配置することができる。図 3 は  $n = 2$  としたときのネットワーク構成である。

勝者ノードの決定法は、灰色判別モデルのアルゴリズムを基にした方法を提案する。基準数列を  $\mathbf{x}_i$ 、比較数列を  $\mathbf{w}_j$  としたときの灰色関連度を  $\gamma_{ij}$  とす

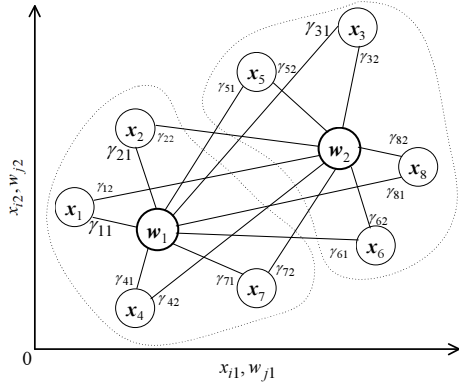


図 3 本提案ネットワークの構成

ると,

$$R = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1s} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{ms} \end{bmatrix} \quad (9)$$

となる行列  $R$  が得られる。ただし、灰色関連度  $\gamma_{ij}$  はユークリッド距離型灰色関連度

$$\gamma_{ij} = \frac{\max_{\forall i} \{\Delta_{ij}\} - \Delta_{ij}}{\max_{\forall i} \{\Delta_{ij}\} - \min_{\forall i} \{\Delta_{ij}\}} \quad (10)$$

$$\Delta_{ij} = \sqrt[\zeta]{\sum_{k=1}^n |x_{ik} - w_{jk}|^{\zeta}} \quad (11)$$

により計算する ( $\zeta \geq 1$ )。  $\zeta$  は  $\Delta_{ij}$  を増幅させる値であり、ノルムの成立条件 [7] を満たす必要がある。

勝者ノード  $w_j^*$  は入力ベクトル  $x_i$  と最も灰色関連度の値が高い,

$$w_j^* = \left\{ x_i \mid \max_{\forall j} \{\gamma_{ij}\} = 1 \right\} \quad \text{for all } i \quad (12)$$

とする。以下では、入力ベクトル  $x_i$  が勝者ノード  $w_j^*$  のグループに属している場合、  $x_i \in w_j^*$  と表記する。

上記の処理は、  $n$  次元空間において入力ベクトル  $x_i$  と最も距離の近い結合重みベクトル  $w_j$  を勝者ノードとしている。これは伝統的な勝者ノード決定法と同じ原理である。ただし、灰色関連度は距離の近いベクトルを 1、距離の遠いベクトルが 0 に近づくように正規化している。

灰色判別モデルでは、  $w_j$  を基準数列、  $x_i$  を比較数列として局部型灰色分析の計算を行った。本提案手法は、入力ベクトル  $x_i$  を基準数列、各結合重みベクトル  $w_j$  を比較数列として局部型灰色分析を行っている。なぜならば、筆者らが提案しているユークリッド型灰色関連度は、接近性という性質により灰色関連度の最大値は 1 となる [7]。提案手法は、勝者ノードの値を必ず 1 とすることで処理結果をよりわかりやすくしている。

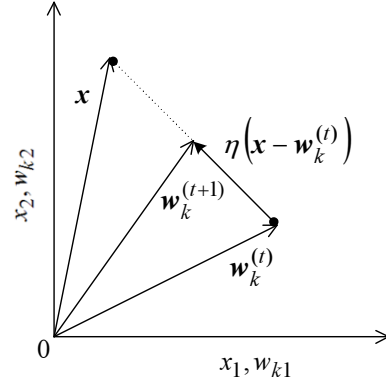


図 4 Kohonen 学習則による  $w_{ik}$  の学習方法

### 3.2 灰色クラスター分析による教師なし学習法の提案

伝統的な自己組織化マップにおける教師なし学習は、Kohonen 学習則 [1] を用いる。Kohonen 学習則は勝者ノードとその近傍を求めた上で、結合重みベクトル  $w_{ik} (k = 1, 2, \dots, n)$  を

$$w_{ik}^{(t+1)} = w_{ik}^{(t)} + \eta (x_k - w_{ik}^{(t)}) \quad (13)$$

により更新し、入力ベクトル  $x_i$  に近づける手続きである。式 (13) の手続きは図 4 のように表すことができる。図 4 により、  $\eta(x_k - w_{ik}^{(t)})$  は結合重みベクトルの移動量を表している。式 (13) において  $t$  は現在の学習回数、  $\eta$  は学習率を表している。学習の効果は最大学習回数、学習率、近傍の大きさ、および  $w_{ik}$  の初期値の与え方によって左右される。

Hu らによって提案されている GSOFM は Kohonen 学習則に灰色分析を適用している。Hu らによる結合重みベクトルの計算式 [4] は、

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta^{(t)} \cdot [\xi_{ij}^{(t)}]^k \cdot (x_j^{(t)} - w_{ij}^{(t)}) \quad (14)$$

と定義している。  $k$  は事前に定めた実数、  $\xi$  はベクトル  $x_i, w_{ij}$  間の局所的な類似度 (灰色関連係数) である ( $0 \leq \xi \leq 1$ )。式 (14) は  $\xi$  の  $k$  乗により近傍を求め、その上で  $w_{ij}$  を学習している。

本提案手法は図 5 に示すイメージにより、Kohonen 学習則によらない教師なし学習法を提案する。勝者ノードの決定には結合重みベクトルと入力ベクトル間のユークリッド距離を計算している。学習とは、結合重みベクトルが入力ベクトルとの距離を最も近くなるように配置する処理だといえる。言い換えれば、結合重みベクトル  $w_j$  は入力ベクトル  $x_i \in w_j^*$  をクラスターとしたときの中心点である。入力ベクトル同士でクラスターを構成し、その中心を見つけることができれば、結合重みベクトルが得られることになる。

本提案手法は、灰色関連度の計算により入力ベクトル同士でクラスターを構成し、その中心を計算す

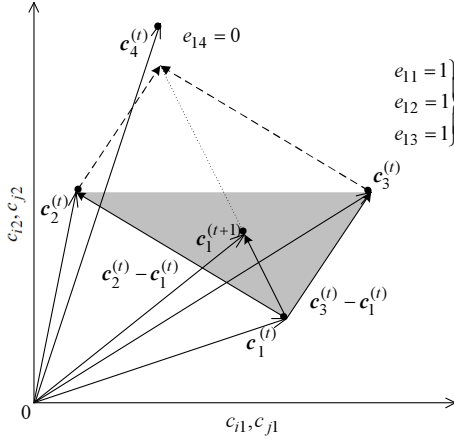


図 5 提案手法による  $c_i$  の学習方法

る教師なし学習法を提案する。灰色関連度によりクラスタ分析を行う方法を灰色クラスタ分析と呼び、既に多数の報告がある [8, 9, 10, 11]。本論文は Wong ら [10] によるクラスタの中心を計算する方法を、 $n$  次元座標空間の計算に拡張展開したアルゴリズムを提案する。

はじめに、クラスタの中心を計算するためのベクトル  $c_i, c_j (i, j = 1, 2, \dots, m)$  を

$$c_i = [c_{i1}, c_{i2}, \dots, c_{in}] \quad (15)$$

で与える。ただし、初期値を

$$c_i^{(t)} = x_i \quad \text{for all } i \quad (16)$$

とする ( $t = 0$ )。

次に、灰色クラスタ分析は仮のクラスタを見つける。全体型灰色分析により全ての  $c_i^{(t)}$  について灰色関連度を計算する。ベクトル  $c_i^{(t)}, c_j^{(t)}$  の全体型関連度を  $\gamma_{ij}$  とすると、全ての組み合わせについて全体型灰色関連度を計算すれば、

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \dots & \gamma_{mm} \end{bmatrix} \quad (17)$$

となる灰色関連マトリクス  $\Gamma$  が得られる。ただし、

$$\gamma_{ij} = 1 - \frac{\Delta_{ij}}{\max_i \max_j \{\Delta_{ij}\}} \quad (18)$$

である。全体型灰色関連度の計算式 (18) は、局所型灰色関連度の計算式 (10) と異なる ( $0 \leq \gamma_{ij} \leq 1$ )。  $\Delta_{ij}$  は式 (11) により計算する。

仮のクラスタは互いに関連が認められるベクトル  $c_i^{(t)}, c_j^{(t)}$  を基に構成する。本提案手法は

$$e_{ij} = \begin{cases} 1 & \gamma_{ij} \geq \psi \\ 0 & \gamma_{ij} < \psi \end{cases} \quad (19)$$

により関連の有無を判定する ( $0 \leq \psi \leq 1$ )。  $\psi$  はクラスタ数の調節係数であり、任意に設定することができる。全てのベクトルについて関連を判定すると、

$$E = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1m} \\ e_{21} & e_{22} & \dots & e_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ e_{m1} & e_{m2} & \dots & e_{mm} \end{bmatrix} \quad (20)$$

となる行列  $E$  が得られる。

図 5 に示すように、  $e_{ij} = 1$ 、すなわち関連が認められる全てのベクトル  $c_i^{(t)}$  で構成されたクラスタは 1 つの  $n$  次元空間と捉えることができる ( $n = 2$  の時は多角形)。本手法はその空間の中心へ移動するための更新値を  $c_i^{(t)}, c_j^{(t)}$  により求める。図 5 により、更新後のベクトル  $c_i^{(t+1)}$  の成分  $c_{ik}^{(t+1)}$  ( $k = 1, 2, \dots, n$ ) は、

$$c_{ik}^{(t+1)} = c_{ik}^{(t)} + \frac{\sum_{j=1}^m e_{ij} (c_{jk}^{(t)} - c_{ik}^{(t)})}{\sum_{j=1}^m e_{ij}} \quad (21)$$

となる。Wong ら [10] はベクトルの更新アルゴリズムに

$$c_{ik}^{(t+1)} = \frac{\sum_{j=1}^m e_{ij} c_{jk}^{(t)}}{\sum_{j=1}^m e_{ij}} \quad (22)$$

という式を提案している。図 6 に Wong らの手法の学習イメージを示す。図 5 との違いはベクトルの引き算の有無である。全ての  $c_i^{(t)}$  についてベクトルを学習すれば、  $c_i^{(t)}$  は仮のクラスタの中心に集まる。言い換えれば、結合重みベクトル  $w_j$  はある 1 点に集まる。

更新されたベクトル  $c_i^{(t+1)}$  は式 (17) および式 (18) の再計算により繰り返し学習が可能となる。繰り返し学習を行うと、ベクトル  $c_i^{(t+1)}$  は

$$c_i^{(0)} \neq c_i^{(1)} \neq \dots \neq c_i^{(t)} = c_i^{(t+1)} \quad (23)$$

となり、ある点で固定される。全てのベクトルが固定されたら学習を終了する。このとき、結合重みベクトルは、

$$w_g = \{c_i, c_j | c_i = c_j\} \quad \text{for all } i, j \quad (24)$$

となる。  $g$  ( $1 \leq g \leq m$ ) は本提案手法で得られた重み結合ベクトルの数であり、クラスタの数でもある。

本手法はクラスタ分析であるから、データの分類に使うことができる。ただし、係数  $\psi$  の値によってクラスタ数は可変となる。  $\psi$  の与え次第で、入力ベクトルが 1 つしか存在しないクラスタを作ることが考えられる。孤立した入力ベクトルは判別に適していない可能性がある。



となる。ここで、 $Q_j^{(t)}$  を最小にする更新値  $u_{jk}$  を求める。最小二乗法の条件より、

$$\frac{\partial Q_j^{(t)}}{\partial u_{j1}} = 0, \frac{\partial Q_j^{(t)}}{\partial u_{j2}} = 0, \dots, \frac{\partial Q_j^{(t)}}{\partial u_{jn}} = 0 \quad (30)$$

とする。

式 (30) を展開すると、次式の正規方程式が得られる。

$$\mathbf{a}^T \boldsymbol{\delta}_j^{(t)} = \mathbf{a}^T \mathbf{a} \mathbf{u}_j \quad (31)$$

ただし、

$$\mathbf{a}^T = \begin{bmatrix} \sum_{i=1}^m \frac{\partial \Delta_{ij}^{(t)}}{\partial u_{j1}} \\ \sum_{i=1}^m \frac{\partial \Delta_{ij}^{(t)}}{\partial u_{j2}} \\ \vdots \\ \sum_{i=1}^m \frac{\partial \Delta_{ij}^{(t)}}{\partial u_{jn}} \end{bmatrix}, \mathbf{u}_j = \begin{bmatrix} u_{j1} \\ u_{j2} \\ \vdots \\ u_{jn} \end{bmatrix} \quad (32)$$

である。式 (31) を  $\mathbf{u}_j$  について解くと次式が得られる。

$$\mathbf{u}_j = (\mathbf{a}_j^T \mathbf{a}_j)^{-1} \mathbf{a}_j^T \boldsymbol{\delta}_j^{(t)} \quad (33)$$

ベクトルの更新値  $\mathbf{u}_j$  が大きすぎると中心点にいつまでも移動できない可能性がある。防止策として、更新値を因子数  $n$  で割るなど必要に応じて更新値を小さくする方法を提案する。

$$\mathbf{u}_j = \frac{(\mathbf{a}_j^T \mathbf{a}_j)^{-1} \mathbf{a}_j^T \boldsymbol{\delta}_j^{(t)}}{n} \quad (34)$$

$\mathbf{u}_j$  を展開すると結合重みベクトル  $\mathbf{w}_j$  の繰り返しアルゴリズムが得られる。最終的に、

$$\mathbf{w}_j^{(t+1)} = \mathbf{w}_j^{(t)} + \frac{(\mathbf{a}_j^T \mathbf{a}_j)^{-1} \mathbf{a}_j^T \boldsymbol{\delta}_j^{(t)}}{n} \quad (35)$$

となる繰り返しアルゴリズムが得られる。本アルゴリズムは、 $Q_j^{(t+1)} \leq Q_j^{(t)}$  である間処理を繰り返す。

本教師あり学習法は結合重みベクトルの初期値  $\mathbf{w}_j^{(0)}$  を与えなければならない。Taylor 逐次近似法は与えた初期値によって近似精度が大きく異なることが指摘されている [15]。初期値と近似精度の関係は数値実験で議論する。

## 4. 数値実験

### 4.1 実験概要

本論文で提案した教師なし学習法、教師あり学習法に対して数値例を適用し、学習の性質とその効果について考察する。適用データは Fisher [16] のあやめデータを用いる。このデータは4つの因子を持ち、50 サンプル × 3 グループ = 150 サンプルある。そのうち半分の 75 サンプルをネットワーク構築用を使用し、残りの 75 サンプルを学習効果検証用を使用

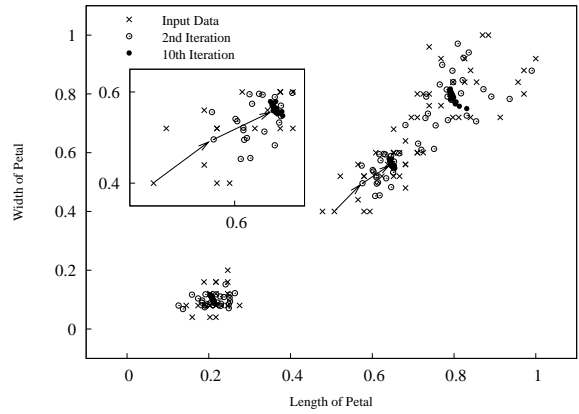


図 8 教師なし学習による  $w_j$  の学習過程

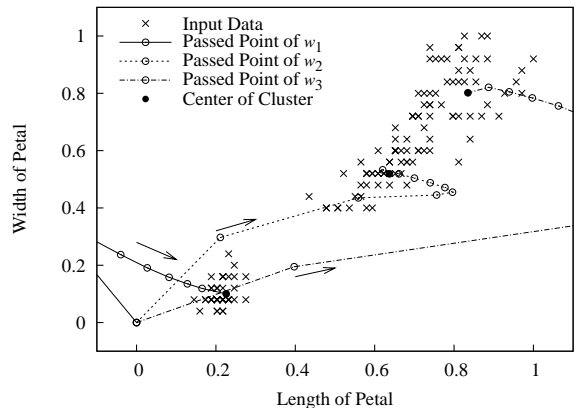


図 9 教師あり学習による  $w_j$  の学習過程

する。本実験は全てのデータを  $[0,1]$  に正規化し、同一環境<sup>‡</sup>下で実験を行っている。

教師なし学習 ( $\zeta = 2, \psi = 0.85$ ) および教師あり学習 ( $\zeta = 2, \mathbf{w}_j^{(0)} = 0$ ) について、結合重みベクトルの学習過程をトレースすると、図 8 および図 9 が得られる。視覚的にわかりやすくするため、4 因子のうち最も変化がわかりやすい花卉の長さ、花卉の幅という 2 つの因子を用いてプロットした図を採用している。図 8 の教師なし学習のほうは、学習が進むにつれてデータ群の中心に集まっているのがわかる。図 9 右の教師あり学習もデータ群の中心に移動しているのが確認できる。

### 4.2 教師なし学習における学習回数と学習結果の関係

教師なし学習は結合重みベクトルの初期値と入力ベクトルを同値とする。ネットワークに対して入力

<sup>‡</sup>以下の環境を用いている。

OS Windows XP Home Edition SP 1  
CPU Pentium 4 M 1.8GHz  
Memory 512MB  
Language Visual C++ 6.0 Professional SP 5

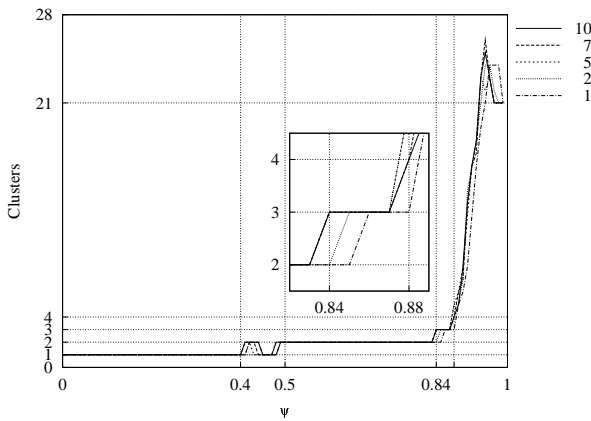


図 10 教師なし学習におけるクラスター数の推移

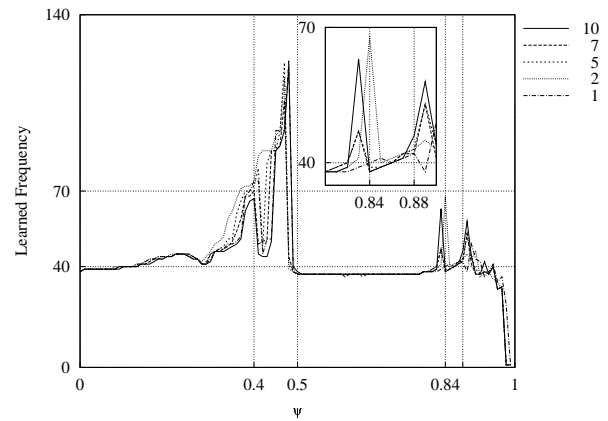


図 11 教師なし学習における学習回数の推移

表 1 パラメータと教師なし学習結果の関係

$\zeta$	$\psi$	学習回数	判別結果
1	0.86	40	72/75 (96%)
2	0.85	40	72/75 (96%)
5	0.84	39	72/75 (96%)
7	0.84	39	72/75 (96%)
10	0.84	38	72/75 (96%)

表 2 教師あり学習における学習結果の関係

初期値	$\zeta$	$Q$	学習回数	判別結果
0	3.2	1.212	11	72/75 (96%)
0.5	6.3	0.971	11	72/75 (96%)
1	4.4	1.097	9	72/75 (96%)
Mean	10	0.898	1	72/75 (96%)
RUSL	10	0.894	3	72/75 (96%)

する係数は $\zeta, \psi$ である。本実験では $\zeta = 1, 2, 5, 7, 10$  および  $0 \leq \psi \leq 1$  としたときに得られる学習結果を比較する。ベクトルの学習は  $1.0 \times 10^{-5}$  単位で終了判定を行う。

図 10 に  $\zeta, \psi$  とクラスター数の関係を示す。同様に、 $\zeta, \psi$  と合計学習回数との関係を図 11 に示す。両図によると、 $\psi$  が同じ値であっても  $\zeta$  が変わると学習回数、学習結果が変わることがわかる。ただし、図 10 において、 $\zeta = 1, 2$  よりも  $\zeta = 5, 7, 10$  の方がより小さい  $\psi$  の値で 3 グループに分類されていることがわかる。

図 11 では、 $0.4 \leq \zeta \leq 0.5$  および  $0.8 \leq \zeta \leq 0.9$  のところで学習回数に大きな変動が起きている。これは、クラスター数の分岐点で学習回数に変動することを表している。

目標クラスター数を 3 としたときの学習結果を表 1 に示す。クラスター数を 3 にするとき、最も学習回数が少ないのは  $\zeta = 10$  のときである。本適用例で最も効率の良い学習結果を得るには、 $\zeta$  を大きい値に設定した上で  $\psi$  の値を調節することを示唆している。ただし、各学習結果を用いて判別処理を行ったところ同じ判別結果となっている。

#### 4.3 教師あり学習における学習回数と学習結果の関係

教師あり学習ではグループ数をあらかじめ 3 と指定し、 $w_1, w_2, w_3$  の初期値を複数通りに定めた場合の学習結果を考察する。本実験

では  $w_j^{(0)} = 0, 1, 0.5, \text{Mean}, \text{RUSL}$  (Result of UnSupervised Learning) を採用する。0 は空間の原点、1 は端点、0.5 は中間点を表し、Mean は各グループに属する入力値の算術平均、RUSL は前節において  $\zeta = 10, \psi = 0.84$  としたときの教師なし学習の結果とする。

図 12 に  $\zeta$  と学習回数との関係を示す。図 13 に  $\zeta$  と残差の式 (29) の関係、すなわち近似精度の関係を示す。図 12 において、Mean, RUSL は安定した学習回数となっている。両者の初期値は既にある程度近似されているため、学習回数が少ないと考えられる。一方、0, 0.5, 1 は  $\zeta$  により学習回数に変動がある。

図 13 においても、Mean, RUSL は初期値が既にある程度近似されているため安定した近似精度が得られている。Mean, RUSL など、初期値に暫定的な近似値を用いる場合は  $\zeta$  を大きい値に設定した方がよいことを示唆している。0, 0.5, 1 は  $\zeta$  と近似精度に大きな変動がある。これは、前述した Taylor 逐次近似法における初期値問題が原因だと考えられる。Mean, RUSL とは異なり、この場合は  $\zeta$  を適切に設定する必要があることを示唆している。

各方法における最も良い近似結果を用いて判別処理を行ったところ、表 2 が得られている。本適用例に関しては教師なし学習と同様、全ての方法について判別結果は同じ結果となっている。

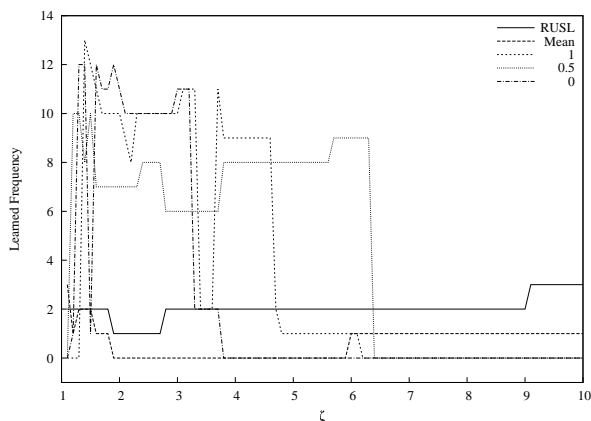


図 12 教師あり学習における学習回数の推移

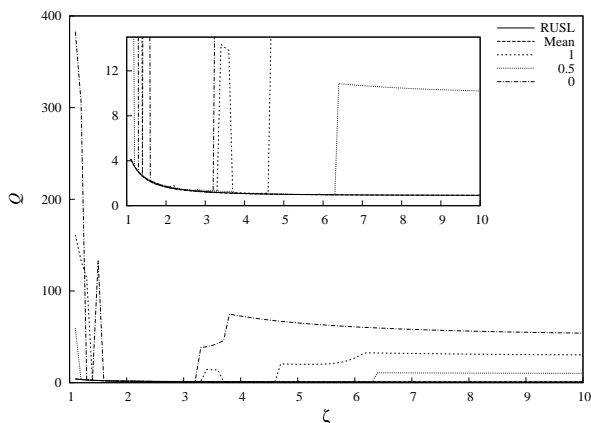


図 13 教師あり学習における近似精度の推移

## 5. おわりに

本論文は灰色理論を用いた分類ネットワーク“GCN”を提案した。GCNの特徴は、

1. 勝者ノードの決定処理にユークリッド距離型灰色分析を適用
2. Wongらのアルゴリズムを拡張したクラスター分析による教師なし学習法
3. Taylor 逐次近似法を用いた、最小二乗法の原理による教師あり学習法
4. 学習回数および学習結果は $\zeta, \psi$ により調節が可能

である。本論文では Kohonen 学習則によらない学習アルゴリズムを提案している。

数値例を適用し、学習の性質について議論した。その結果、教師なし学習では $\zeta, \psi$ の順に調節すること、教師あり学習はベクトルの初期値を適切に選ぶ必要があることなどの仮説を得ることができた。

今後はパターン認識あるいは感性識別をはじめとする実例を通じて、既存の手法と本手法の違いを評価する。

## 参考文献

- [1] T. Kohonen, 自己組織化マップ, シュプリンガー・フェアラーク東京, 1996.
- [2] J.L. Deng, Grey System, China Ocean Press, 1988.
- [3] C.S. Cheng, Y.T. Hsu and C.C. Wu, “Grey Neural Network,” IEICE Trans. Fundamentals, vol.E81-A, no.11, pp.2433–2442, Nov. 1998.
- [4] Y.C. Hu, R.S. Chen, Y.T. Hsu and G.H. Tzang, “Grey Self-organizing Feature Maps,” Neuro Computing, vol.48, no.4, pp.863–877, 2002.
- [5] 永井正武, 山口大輔, 灰色理論と工学応用方法, 共立出版, 2004.
- [6] 山口大輔, 小林俊裕, 水谷晃三, 永井正武, “灰色分析を適用した判別モデルの提案,” 電子情報通信学会技術研究報告, vol.103, no.521, HIP2003-82, pp.31–36, Dec. 2003.
- [7] 山口大輔, 李国棟, 水谷晃三, 永井正武, “ユークリッド距離による灰色関連測度の提案,” 情報処理学会研究報告, FI78-2, Mar. 2005.
- [8] H.C. Hu, T.H. Hung, M.F. Yeh, “Approximated Clustering Via Modified Grey Relational Analysis,” The Journal of Grey System, vol.10, no.2, pp.87–104, 1998.
- [9] C.C. Wong and C.C. Chen, “Data Clustering by Grey Relational Analysis,” The Journal of Grey System, vol.10, no.4, pp.281–288, Dec. 1998.
- [10] C.C. Wong and H.R. Rai, “The Grey Clustering Algorithm,” The 5th National Conf. on Grey System Theory and Applications, pp.277–283, 2000 (in Chinese).
- [11] 山口大輔, 小林俊裕, 水谷晃三, 永井正武, “灰色分析を適用した階層的クラスター分析法の提案,” 情報処理学会研究報告, vol.2004, no.10, AL-93-11, pp.75–92, Jan. 2004.
- [12] 櫻庭祐一, 中本高道, 森泉豊榮, “ファジー理論を用いた学習ベクトル量子化法,” 電子情報通信学会論文誌, vol.J73-D-II, no.11, pp.1863–1871, Nov. 1990.
- [13] N.B. Karayiannis, P.I. Pai, “Fuzzy Algorithms for Learning Vector Quantization,” IEEE Trans. on Neural Network, vol.7, no.5, pp.1196–1211, Sept. 1996.
- [14] M.R. Aaron, “The Use of Least Squares in System Design,” IRE Trans. on Circuit Theory, vol.3, no.4, pp.224–231, Dec. 1956.
- [15] 李国棟, 山口大輔, 水谷晃三, 永井正武, “灰色理論 GM における精度改善への新しい提案と評価,” 情報処理学会研究報告, FI76-24, pp.177–184, Sept. 2004.
- [16] R.A. Fisher, “The use of multiple measurements in taxonomic problems,” Ann. Eugenics, vol.7, no.2, pp.179–188, 1936.