

Title	線形多様体クラスタリングと楕円型ファジィモデル
Author(s)	馬屋原, 一孝; 中森, 義輝
Citation	日本ファジィ学会誌, 10(1): 142-149
Issue Date	1998-02-15
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/7962
Rights	Copyright (C) 1998 日本知能情報ファジィ学会. 馬屋原一孝, 中森義輝, 日本ファジィ学会誌, 10(1), 1998, 142-149. 本文データは学協会の許諾に基づき CiNiiから複製したものである.
Description	

線形多様体クラスタリングと 楕円型ファジィモデル†

馬屋原 一孝*¹ 中森 義輝*²

与えられたデータに基づいて複数の線形多様体を発見するための新しいファジィクラスタリング手法を提案する。主な提案は、それぞれのクラスタ内データの分布に応じた適切な次元の線形多様体を得ることをねらった目的関数の導入である。つぎに、楕円型メンバシップ関数で表される新しいタイプのファジィモデルを提案する。このモデルの構築にはクラスタリング結果のみを用い、構造同定を行う必要はない。このモデルによる対話的シミュレーションを実行すれば、多次元空間におけるデータ分布と変数間関係を理解することが容易になる。

1. はじめに

ファジィモデリング[1]における重要な課題は、システムの部分的な構造をとらえたサブモデル群をいかに発見するかということである。本論文では、ファジィクラスタリング手法を用いた新たなファジィモデリングおよびシミュレーション手法を提案する。

従来のファジィモデリングでは、事前に変数間の構造モデルを仮定する。しかし、複雑で大規模なシステムに対しては、構造を探索するためにデータからモデルを構築してみるという側面がある。また多くの場合、数値データは一様に存在せず、変数の選択や、サブモデルの担当領域の決定が困難である。

こういった状況でモデリングを実行するためには、対象をよく理解する必要がある。本論文で提案する楕円型ファジィモデルは、変数間の関係をデータに忠実に表現しようとするもので、最終モデルというよりは、対象をよりよく理解するための道具である。

論文の構成は以下のとおりである。まず、線形多様体を発見する目的で開発されているファジィクラスタリング手法の概略を記述し、3節において、次元の異なる線形多様体を発見するためのクラスタリング手法を提案する。つぎに、従来のファジィモデルを概観した後、5節において、楕円型ファジィモデルを提案する。最後に、実際問題への応用例を紹介する。

2. ファジィクラスタリング

ファジィモデリングに利用できるファジィクラスタリング手法について簡単に記述する。変数 x_j ($j=1, 2, \dots, m$) の k 番目の標準化データを x_{kj} ($k=1, 2, \dots, n$) とし、 w_k を k 番目のデータベクトルとする。すなわち、

$$w_k = (x_{k1}, x_{k2}, \dots, x_{km})^T, \quad k=1, 2, \dots, n$$

ここで、 $(\cdot)^T$ は転置を意味する。データベクトルの集合 $S = \{w_1, w_2, \dots, w_n\}$ を c 個のクラスタにファジィ分割することを考える。 u_{ik} を w_k のクラスタ i へのメンバシップ値とし、以下の条件を満たすものとする。

- (i) $0 \leq u_{ik} \leq 1, \quad i=1, 2, \dots, c, \quad k=1, 2, \dots, n$
- (ii) $\sum_{k=1}^n u_{ik} > 0, \quad i=1, 2, \dots, c$
- (iii) $\sum_{i=1}^c u_{ik} = 1, \quad k=1, 2, \dots, n$

Fuzzy c -Means (FCM) 法[2]では、メンバシップ行列 $U = (u_{ik})$ とクラスタの中心 $V = \{v_1, v_2, \dots, v_c\}$ を、つぎの評価規範を最小化するように決定する。

$$J_{FCM}(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^q d_{ik}(v_i) \quad (1)$$

ただし、 q は 1 より大きい実数であり、 $d_{ik}(v_i)$ は w_k と v_i の 2 乗距離

$$d_{ik}(v_i) = \|w_k - v_i\|^2 \quad (2)$$

である。

Fuzzy c -Varieties (FCV) 法[3]では、 R^m における c 個の r 次元 ($0 \leq r < m$) 線形多様体

$$V_i^r = \left\{ z \in R^m \mid z = v_i + \sum_{j=1}^r t_{ij} e_{ij}, t_{ij} \in R \right\} \quad (3)$$

† Linear Varieties Clustering and Elliptic Type Fuzzy Models

Kazutaka Umayahara and Yoshiteru Nakamori

*¹ 筑波大学 電子・情報工学系

Institute of Information Sciences and Electronics, University of Tsukuba

*² 中南大学 理学部応用数学科

Department of Applied Mathematics, Konan University

を考える。ここでは、 $e_{ij}(j=1,2,\dots,r)$ をファジィ散布行列

$$S_i = \sum_{k=1}^n (u_{ik})^q (\mathbf{w}_k - \mathbf{v}_i) (\mathbf{w}_k - \mathbf{v}_i)^T \quad (4)$$

の固有値 $\lambda_{i1} \geq \lambda_{i2} \geq \dots \geq \lambda_{im}$ に対応する固有ベクトルの最初の r 個を正規化したものとする。FCV 法では、評価規範として

$$J_{FCV}^r(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^q D_{ik}^r(\mathbf{v}_i) \quad (5)$$

を用いる。ただし、 $D_{ik}^r(\mathbf{v}_i)$ は \mathbf{w}_k と V_i^r の 2 乗距離

$$D_{ik}^r(\mathbf{v}_i) = \|\mathbf{w}_k - \mathbf{v}_i\|^2 - \sum_{j=1}^r |\langle \mathbf{w}_k - \mathbf{v}_i, \mathbf{e}_{ij} \rangle|^2 \quad (6)$$

である。ここで、 $\langle \cdot, \cdot \rangle$ は内積を表す。

Adaptive Fuzzy c-Elliptotypes(AFC)法[4]では、 $d_{ik}(\mathbf{v}_i)$ と $D_{ik}^r(\mathbf{v}_i)$ をクラスタリング過程において適応的に変化するパラメータ

$$\alpha_i = 1 - \frac{\lambda_{im}}{\lambda_{i1}}, \quad i=1,2,\dots,c \quad (7)$$

を用いて加重平均する評価規範

$$J_{AFC}^r(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^q [(1 - \alpha_i) d_{ik}(\mathbf{v}_i) + \alpha_i D_{ik}^r(\mathbf{v}_i)] \quad (8)$$

を用いる。これにより、 r 次元の異なった形状のクラスタを発見できる可能性が高くなる。ただし、文献[4]では 2 次元平面における直線の同定を扱っており、ここでは拡張して定式化している。

3. 次元の異なる線形多様体発見法

AFC 法では、データと r 次元線形多様体との 2 乗距離と、データとファジィ重心(0 次元の線形多様体)との 2 乗距離を、データ分布を考慮して加重平均した規範を用いている。本論文では、この考え方をさらに推し進め、データとすべての次元の線形多様体との 2 乗距離をデータ分布を考慮して加重平均する規範を導入する。すなわち、つぎのような最小化すべき評価規範を考える。

$$J_{FVD}^r(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^q E_{ik}(\mathbf{v}_i) \quad (9)$$

ただし、

$$(a) \quad E_{ik}(\mathbf{v}_i) = \sum_{r=0}^{m-1} \beta_i^r \frac{G_{ik}^r(\mathbf{v}_i)}{m-r}$$

$$(b) \quad G_{ik}^r(\mathbf{v}_i) = \|\mathbf{w}_k - \mathbf{v}_i\|^2 - \sum_{j=1}^r |\langle \mathbf{w}_k - \mathbf{v}_i, \mathbf{e}_{ij} \rangle|^2$$

$r=0$ のときは $G_{ik}^0(\mathbf{v}_i) = \|\mathbf{w}_k - \mathbf{v}_i\|^2$

$$(c) \quad \beta_i^r = \frac{\gamma_i^r}{\sum_{r=0}^{m-1} \gamma_i^r}$$

$$(d) \quad \gamma_i^r = (\lambda_{ir} - \lambda_{i,r+1})^l$$

$r=0$ のときは $\gamma_i^0 = (\lambda_{im})^l$

まず、(a) (b) について説明する。 $G_{ik}^0 (= d_{ik})$ と特定の $G_{ik}^r (= D_{ik}^r)$ の 2 つだけでなく、すべての次元の線形多様体 $V_i^r (r=0,1,\dots,m-1)$ を考え、データ \mathbf{w}_k と V_i^r との 2 乗距離 G_{ik}^r に $1/(m-r)$ の重みづけをおこない、さらに以下で説明するクラスタの形状に応じた重み β_i^r をかけて加算する。これが $E_{ik}(\mathbf{v}_i)$ である。

G_{ik}^r は m 次元空間における r 次元の線形多様体 V_i^r の直交補空間における 2 乗距離である。 G_{ik}^r を直交補空間の次元数 $m-r$ で割ることにより、次元の増加による G_{ik}^r の自然増を押さえ、平等化を図っている。

つぎに、(c) (d) で定義されている β_i^r について説明する。 β_i^r は (d) で定義されている γ_i^r に比例するように決める。 γ_i^r の値が他と比べて大きいということは、1 番目から r 番目までの固有値にある程度の値があり、 $r+1$ 番目以降の固有値が r 番目の固有値よりかなり小さいことを意味している。このとき、評価規範の中で、データと r 次元の線形多様体との 2 乗距離に大きな重みを与える。これにより、 r 次元の線形多様体を発見する可能性を高くする。ただし、 β_i^0 については、最小固有値で定義する。

なお、(d) における l は線形度パラメータで、0 以上の実数である。 l を大きくすることは線形性を強調することを意味する。

クラスタリングはパラメータ c, q, l を与え、評価規範を最小化するように U, V を決定していくプロセスである。2 種類のパラメータを同時に決定することは困難であるから、他のパラメータが与えられたとき、残りのパラメータを最適性の必要条件から決定する。必要条件の導出は FCM 法の場合よりやや複雑であるが、結局 FCM 法の場合と同様な条件式が得られる。

クラスタリングのアルゴリズムを以下に記述する。

Step 1. $t=0$ とし、パラメータ c, q, l, ϵ を定める。

メンバシップ値の初期値 $U^{(0)} = (u_{ik}^{(0)})$ を乱数により与える。

Step 2. クラスタ i のファジィ重心を計算する。

$$\mathbf{v}_i^{(t)} = \frac{\sum_{k=1}^n (u_{ik}^{(t)})^q \mathbf{w}_k}{\sum_{k=1}^n (u_{ik}^{(t)})^q}, \quad i=1,2,\dots,c$$

Step 3. クラスタ i のファジィ散布行列

$$S_i^{(t)} = \sum_{k=1}^n (u_{ik}^{(t)})^q (\mathbf{w}_k - \mathbf{v}_i^{(t)}) (\mathbf{w}_k - \mathbf{v}_i^{(t)})^T$$

の固有値 $\{\lambda^{(g)}\}$ と固有ベクトル $\{e^{(g)}\}$ を求める。ただし、 $\|e^{(g)}\|=1$ と正規化する。

Step 4. 次式により、データ w_k のクラス i へのメンバーシップ値を更新する。

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^c \left(\frac{E_{ik}^{(g)}}{E_{jk}^{(g)}} \right)^{\frac{1}{q-1}} \right]^{-1}$$

$E_{ik}^{(g)}=0$ となる i が存在する場合は次式による。

$$u_{ik}^{(t+1)} = \begin{cases} \frac{1}{\#\{j | E_{jk}^{(g)}=0\}}, & E_{ik}^{(g)}=0 \\ 0, & E_{ik}^{(g)}>0 \end{cases}$$

ただし、 $\#\{\cdot\}$ は集合の要素数を表す。

Step 5. 収束判定条件

$$\max_{i,k} \{|u_{ik}^{(t+1)} - u_{ik}^{(g)}|\} < \varepsilon$$

を満たせば終了。そうでなければ、 $t=t+1$ として

Step 2 へ戻る。

以上の手法を Fuzzy c-Varieties of Different Dimensionalities (FVD) 法と呼ぶことにする。

4. ファジィモデル

ファジィモデリングとは、現実の問題に内包されている種々の「あいまいさ」を反映させるようなモデル構築手法の総称である[5]。回帰モデルの発展形としては、データのあいまい性に注目した可能性線形回帰モデル[6]と、構造の非線形性に注目したルール型回帰モデル[1]が提案されている。後者は一般に「ファジィモデル」と呼ばれている。

次節において提案する楕円型ファジィモデルは後者のファジィモデルに近いが、分布を予測するという点では可能性回帰モデルの特徴を持つものである。

ファジィモデルはつぎのようなルールの集合からなる非線型モデルである[1]。

$$\text{ルール } R_i \begin{cases} \text{if } z_1 \text{ is } A_{i1}, z_2 \text{ is } A_{i2}, \text{ and } \dots \\ \text{then } y = c_{i0} + \sum_j c_{ij} x_j \end{cases} \quad (10)$$

A_{ij} はファジィ部分集合で、メンバーシップ関数を $\mu_{ij}(z_j)$ とする。ファジィモデルによる目的変数 y の予測値は

$$\hat{y} = \frac{\sum_{i=1}^c w_i \hat{y}_i}{\sum_{i=1}^c w_i}, \quad w_i = \prod_j \mu_{ij}(z_j^*) \quad (11)$$

と与えられる。ここに、 \hat{y}_i はルール R_i の出力値である。 w_i はルール R_i の適合度であり、すべての前件部変数の入力値 z_j^* に対応するメンバーシップ値の積で与えられる。

ところで、データの密集状態が存在したり、あるいはデータがほとんど存在しないが将来は起こりうるようなケースでは、後件部として線形回帰モデルを採用できない場合がある。予測モデルを構築するとき後者のケースにしばしば直面する。そのような場合、つぎのように後件部を可能性分布モデルとするファジィモデルが提案されている[7]。

$$\text{ルール } R_i \begin{cases} \text{if } z_1 \text{ is } A_{i1}, z_2 \text{ is } A_{i2}, \text{ and } \dots \\ \text{then } y \text{ is } B_i \end{cases} \quad (12)$$

ただし、 B_i は目的変数空間上のファジィ部分集合で、そのメンバーシップ関数を $\mu_i(y)$ とする。モデルの出力は

$$\hat{\mu}(y) = \frac{\sum_{i=1}^c w_i \mu_i(y)}{\sum_{i=1}^c w_i}, \quad w_i = \prod_j \mu_{ij}(z_j^*) \quad (13)$$

というメンバーシップ関数で与えられる。

さらに、前件部に多次元の楕円型メンバーシップ関数を用いたファジィモデルが提案されている[8]。

$$\text{ルール } R_i \begin{cases} \text{if } z \text{ is } F_i, \\ \text{then } y = c_{i0} + \sum_j c_{ij} x_j \end{cases} \quad (14)$$

ここで、 $z = (z_1, z_2, \dots, z_s)^T$ は前件部変数からなるベクトルで、後件部変数と同じ変数が含まれることもある。 F_i は多次元の楕円型メンバーシップ関数 $f_i(z)$ で表されるファジィ部分集合である。 y の予測値は次式で求める。

$$\hat{y} = \frac{\sum_{i=1}^c f_i(z^*) \hat{y}_i}{\sum_{i=1}^c f_i(z^*)} \quad (15)$$

ここで、 z^* は前件部変数の入力値ベクトルを表す。

(10) のようなルールを用いる高木・菅野型ファジィモデルの入出力はクリस्पな値をとる。このモデルではメンバーシップ関数が変数軸ごとに定義されるため、入力変数間の相関を考慮したシミュレーションを実行するとき注意を要する。この点を改善したものが(14) のようなルールを用いるファジィモデルである。このモデルでは、入力変数値ベクトルの確信度がデータ分布を反映したものとなる。一方、出力についてファジィ化したものが(12) のようなルールを用いるファジィモデルである。このモデルは線形回帰モデルが作れない部分データに対して考えられたものであるが、出力値の確信度を考慮することができる。

次節で提案する新しいファジィモデルは、部分的な線形関係をルール化するという意味ではこれまでのフ

ファジィモデルと同じ発想であるが、部分データに内在する変数間の相関とファジィネスを保存しようとするモデルである。

5. 楕円型ファジィモデル

3節において提案したFVD法によりクラスタリングをおこなうと、クラスタリング結果と同時にクラスタのファジィ重心、ファジィ散布行列の固有値・固有ベクトルという情報が得られる。これらの情報は、データがどこにどのように分布しているかを表わしており、各クラスタを部分モデル、すなわちルールとみなすファジィモデルを考えることができる。

まず、ルール i の多次元メンバシップ関数をクラスタの形状を考慮して

$$f_i(\mathbf{z}) = \exp\left\{-\sum_{j=1}^m \frac{1}{2\lambda_{ij}} [(z - \mathbf{v}_i)^\top \mathbf{e}_{ij}]^2\right\} \quad (16)$$

と定める。ただし、 \mathbf{z} は \mathbf{R}^m の点である。この関数は、クラスタの重心で1となり、クラスタの形状に従ったガウス型分布をしている。

予測したい変数 x_j 軸上のファジィ部分集合 F_{ij} を、メンバシップ関数 $f_i(\mathbf{z})$ を用いて定義することにより、回帰式によるクリस्प出力ではなく、メンバシップ関数によるファジィ出力をおこなうことを考える。このとき、入力変数はファジィモデルの前件部変数のように固定ではなく、全変数中の任意の変数を用いることができる。

いま、入力値を与える変数の集合を X 、入力値ベクトルを \mathbf{x}^* 、入力値を与えない変数の集合を Y 、その要素数を $\#Y$ とする。このとき、変数 $x_j \in Y$ のファジィ部分集合 $F_{ij} = F_{ij}(\mathbf{x}^*)$ のメンバシップ関数を

$$f_{ij}(x_j | \mathbf{x}^*) = \sup_{\mathbf{a} \in \mathbf{R}^{\#Y-1}} f_i(x_j, \mathbf{a}, \mathbf{x}^*) \quad (17)$$

と定めると、 $F_{ij}(\mathbf{x}^*)$ は入力 \mathbf{x}^* に対するルール i の変数 x_j 軸上へのファジィ出力と考えることができる。ただし、 \mathbf{a} は x_j 以外の Y に含まれる変数からなるベクトルである。

ルールの出力はファジィ出力であるが、モデル全体の出力 $\hat{x}_j(\mathbf{x}^*)$ は、各ルールのファジィ出力の中心 $\hat{x}_{ij}(\mathbf{x}^*)$ を、その中心のメンバシップ値 $w_i(\mathbf{x}^*)$ により荷重平均する。すなわち、各ルールのファジィ出力の中心を

$$\hat{x}_{ij}(\mathbf{x}^*) = \arg \sup_{x_j} f_{ij}(x_j | \mathbf{x}^*) \quad (18)$$

と定義し、そのメンバシップ値

$$w_i(\mathbf{x}^*) = \sup_{x_j} f_{ij}(x_j | \mathbf{x}^*) \quad (19)$$

を用いて、

$$\hat{x}_j(\mathbf{x}^*) = \frac{\sum_{i=1}^c w_i(\mathbf{x}^*) \hat{x}_{ij}(\mathbf{x}^*)}{\sum_{i=1}^c w_i(\mathbf{x}^*)} \quad (20)$$

と予測する。

楕円型ファジィモデルでは、各ルールに前件部と後件部の区別がなく、回帰式も用いないため、前件部変数や後件部変数の選択をおこなう必要がない。つまり、入力変数、出力変数という概念のないモデルとなっている。そのため、必要に応じてどの変数の入力値でも指定でき、残りの変数の値を出力することができる。すなわち、指定した入力から出力を予測(シミュレーション)することと、希望の出力から可能な入力を予測(最適化)することを区別しない。

6. 水質予測モデル

超楕円体クラスタリング法[7]によるモデリングの例として紹介されている水質予測モデルの構築に用いたデータを、楕円型ファジィモデルを用いて観察した様子を紹介する。

文献[7]では、関東地方の一級河川地域における21の中小河川の本川合流直前の水質(生物化学的酸素要求量: BOD)を流域特性から予測するモデルを構築している。データは関東地方の一級河川水域の全流域を含む1都7県における1980年度と1985年度のものである。ここで用いる変数は、BOD、下水道普及率、宅地面積率、出荷額密度、畑面積率の5変数である。なお、各変数のデータは平均0、分散1に標準化している。

パラメータを

$$c=3, \quad q=1.5, \quad l=0.4, \quad \epsilon=10^{-4}$$

と設定し、図1のように、文献[7]で発見されている構造と同じような結果を得ることができた。図1は5次元空間を下水道普及率とBODの平面に射影したものである。なお、クラスタ数および他のパラメータの決定は試行錯誤による。

クラスタ i ($i=1, 2, 3$) のファジィ散布行列の固有値の累積寄与率

$$\sigma_{ij} = \frac{\sum_{l=1}^j \lambda_{il}}{\sum_{l=1}^5 \lambda_{il}}, \quad j=1, 2, \dots, 5$$

は以下のものであった。

$$\begin{aligned} \sigma_{11} &= 0.72 & \sigma_{12} &= 0.88 & \sigma_{13} &= 0.97 & \sigma_{14} &= 1 \\ \sigma_{21} &= 0.91 & \sigma_{22} &= 0.96 & \sigma_{23} &= 0.98 & \sigma_{24} &= 1 \\ \sigma_{31} &= 0.57 & \sigma_{32} &= 0.92 & \sigma_{33} &= 0.98 & \sigma_{34} &= 1 \end{aligned}$$

すなわち、3つのクラスタとも、4つの固有値で全変動をとらえており、5次元空間の中の4次元超平面が発見されていると考えられる。したがって、BODに対する線形回帰モデルを、残りの4つの変数を説明変数として構築することができる。ただし、クラスタ2は開発途上の地域のデータと考えられ、すべての変数が高い相関を持っている。したがって、クラスタ2においては説明変数を絞った方が良さそうである。

ここでは、クラスタリング結果を利用して、楕円型ファジィモデルを構築する。各変数軸上へのメンバーシップ関数の射影は図2のようになる。図2の下水道普及率において、もっとも分布の拡がり大きいものがルール1、中央の分布の拡がり小さいものがルール2、右側に位置しているものがルール3に対応する。

下水道普及率の値を小さな値から徐々に大きくし、その他の変数の変化を観察する。図3から図8がそれである。データは標準化されているので、図3において下水道普及率が-1.5というのは、普及率がかなり低いことを意味する。このとき他の変数は明確な値を

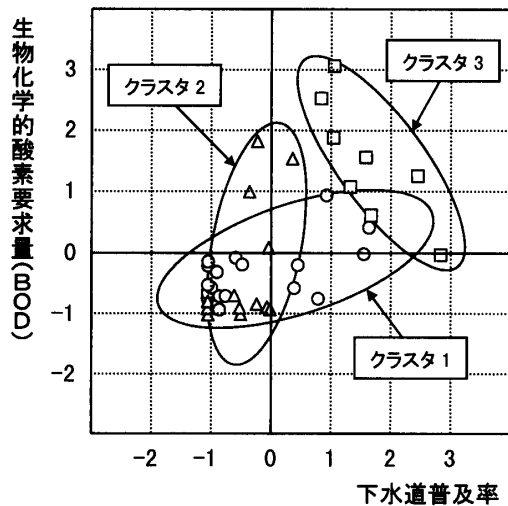


図1 FVD法によるクラスタリング結果

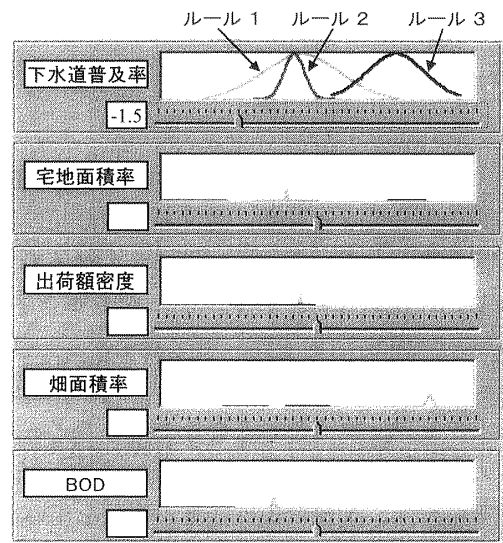


図3 下水道普及率が-1.5のときの他の変数の状態

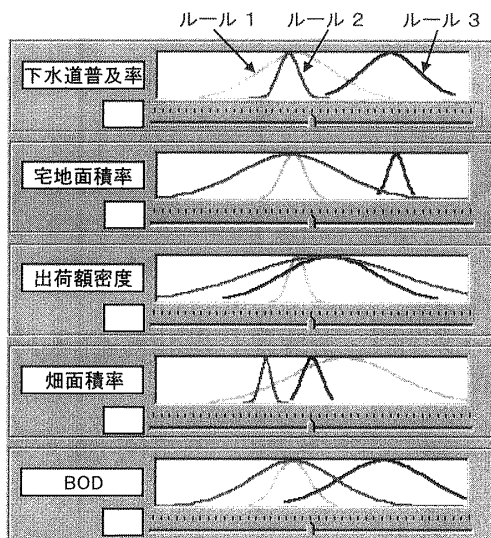


図2 3つのルールからなる楕円型ファジィモデル

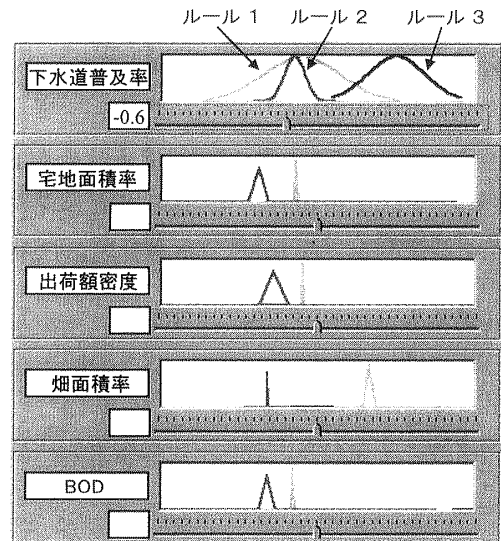


図4 下水道普及率が-0.6のときの他の変数の状態

とらない、つまり、下水道普及率が -1.5 ということは、過去のデータにはほとんどなかったことを示している。

前節の記号との対応については、 $X = \{\text{下水道普及率}\}$, $x^* = \text{下水道普及率の値}$, $Y = \{\text{BOD, 宅地面積率, 出荷額密度, 畑面積率}\}$ である。図3から図8において、下水道普及率以外については(17)式によって計算されるメンバシップ関数が描かれている。たとえば、図3においては、各 $x_j \in Y$ に対して $f_{ij}(x_j | -1.5)$ が描かれている。

なお、図3から図8はコンピュータ画面のハードコピーである。マウスで目盛りの下の針を動かすことにより変数の値が固定される。そのときのみ変数名の下に数字が現れる。また、値が固定される変数については図2のメンバシップ関数がそのまま描かれている。これは、複数の変数の値を固定するときを参照するためである。実際のシミュレーションにおいては、下水道普及率の値をマウスにより連続的に変化させるとき、他の変数のメンバシップ関数が連続的に変化の様子が観察できる。

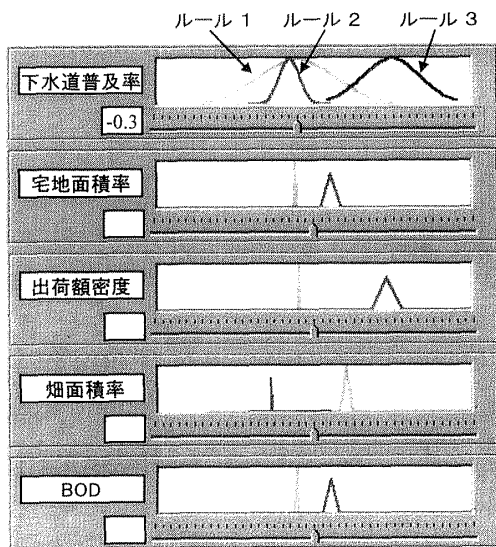


図5 下水道普及率が -0.3 のときの他の変数の状態

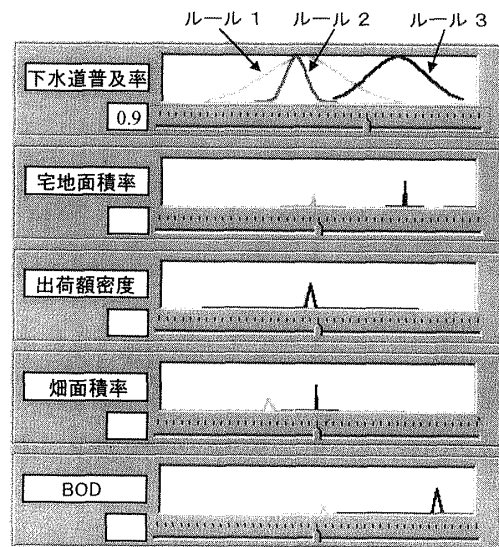


図7 下水道普及率が 0.9 のときの他の変数の状態

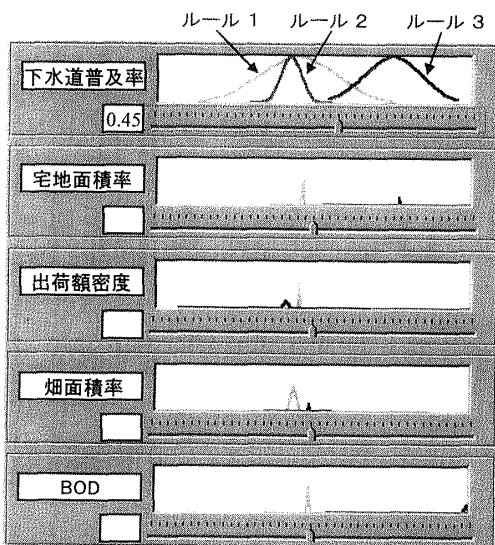


図6 下水道普及率が 0.45 のときの他の変数の状態

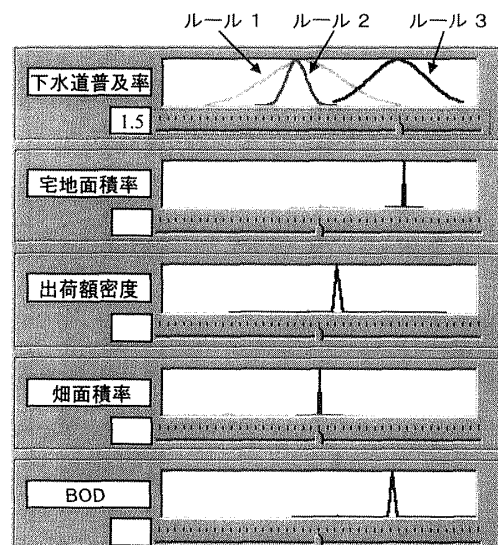


図8 下水道普及率が 1.5 のときの他の変数の状態

図4においては下水道普及率を -0.6 に指定している。この場合、ルール1とルール2が発火していることから、他の4つの変数のとりうる値について2つの可能性があることが読みとれる。下水道普及率を -0.3 とした図5の場合、ルール2において、宅地面積率、出荷額密度、BODの値が大きくなっている。畑面積率は他の変数とは反対の動きをしていることがわかる。ここで、下水道普及率が -0.6 から -0.3 へ変化するとき、BODはクラスタ2のデータで構築したルール2についてのみ大きく変化することは図1からも理解できる。

図6のように下水道普及率を大きくして 0.45 とした場合、BODは非常に大きな値となる可能性があることが理解される。ところが、図7のように、下水道普及率をさらに大きく 0.9 とした場合、BODの値は下がりはじめ、図8のように 1.5 とした場合、BODはさらに小さくなっていく。

このように、地域の開発の進行に伴い、変数の値が変化していく様子がよく理解できる。実際には、下水道普及率をマウスで少しずつ動かすことにより、他の変数の値が変化していく様子を観察することができる。

7. おわりに

楕円型ファジィモデルは、FVD法によってファジィ分割された部分データの存在している状態を表現する客観的モデルである。変数の値が他の変数の動きに連動して移り変わってゆく様子を観察することを目的としている。楕円型ファジィモデルの挙動を観察することにより、構造の複雑な対象を理解することが容易になる。

楕円型ファジィモデルはFVD法によるクラスタリング結果に依存しており、データ分割が適切であるかどうかをどのように判断するかという問題が残されている。また、適切なパラメータの設定法についても今後の課題であるが、データ解析の宿命として、いくつか試してみても解釈できるものに決定することが最善か

著者紹介



馬屋原 一孝 (うまやはら かずたか)

筑波大学 電子・情報工学系

1991年 大阪府立大学大学院総合科学研究科情報科学専攻修士課程修了。1991年4月より1992年11月まで、青年海外協力隊に参加しマラウイ共和国農業省勤務。1995年4月甲南大学大学院自然科学研究科情報・システム科学専攻博士後期課程入学。1997年4月同課程退学。1997年5月より筑波大学 電子・情報工学系助手、先端学際領域研究センター勤務。日本ファジィ学会会員。

もしれない。

謝辞

本研究の一部は筑波大学先端学際領域研究センターにおける共同研究プロジェクトの一環として行われた。

参考文献

- [1] T. Takagi and M. Sugeno: Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. SMC-15, No. 1, pp. 116-132, 1985.
- [2] J. C. Bezdek: *Fuzzy Mathematics in Pattern Classification*. PhD Thesis, Cornell University, 1973.
- [3] J. C. Bezdek et al.: Detection and characterization of cluster substructure II. fuzzy c-varieties and convex combinations thereof. *SIAM J. Appl. Math.*, Vol. 40, No. 2, pp. 358-372, 1981.
- [4] R. N. Dave: An adaptive fuzzy c-elliptotype clustering algorithm. Proc. of NAFIPS 90: Quarter Century of Fuzziness, Vol. I, pp. 9-12, 1990.
- [5] 田中: ファジィモデリングとその応用。システム制御情報学会編, 朝倉書店, 181p., 1990.
- [6] H. Tanaka, S. Uejima and K. Asai: Linear regression analysis with fuzzy model. *IEEE Trans. on Systems, Man and Cybernetics*, Vol. SMC-12, No. 6, pp. 903-907, 1982.
- [7] 中森義輝: ファジィモデリング。オーム社, pp. 99-126, 1994.
- [8] M. Ryoike, H. Tamura and Y. Nakamori: Fuzzy rule generation by hyperellipsoidal clustering. Proc. of IIZUKA'96, pp. 86-89, 1996.

(1997年6月3日 受付)

(1997年10月30日 再受付)

[問い合わせ先]

〒305-8573

つくば市天王台 1-1-1

筑波大学 電子・情報工学系

馬屋原 一孝

TEL : 0298-53-6553

FAX : 0298-53-6471

E-mail: uma@is.tsukuba.ac.jp



中森 義輝 (なかもり よしてる)

甲南大学 理学部応用数学科

1979年 京都大学大学院工学研究科数理工学専攻修士課程修了。1981年より甲南大学理学部応用数学科勤務。現在教授。1984年10月より1985年11月まで、国際応用システム解析研究所(オーストリア)研究員。1986年4月より環境庁国立環境研究所客員研究員。1992年9月より大連理工大学客員教授兼務。日本ファジィ学会、計測自動制御学会、環境科学会、IEEEなどの会員。

Linear Varieties Clustering and Elliptic Type Fuzzy Models

by

Kazutaka Umayahara and Yoshiteru Nakamori

Abstract :

This paper proposes a new fuzzy clustering technique to detect a number of local linear varieties based on a given data set. The main proposal is the introduction of an objective function that aims at obtaining an appropriate dimensional linear variety corresponding to the data distribution within each cluster. Then, the paper proposes a new type of fuzzy models expressed by elliptic type membership functions. One can develop this kind of models based on the clustering results only, without considering the model structure. An interactive simulation using the developed model helps us understand the data distribution in the multi-dimensional space and the relationships between variables

Contact Address : **Kazutaka Umayahara**

Institute of Information Sciences and Electronics, University of Tsukuba

1-1-1 Ten'nodai, Tsukuba, Ibaraki 305-8573, Japan

TEL : 0298-53-6553

FAX : 0298-53-6471

E-mail : uma@is.tsukuba.ac.jp