

第6章

線形ファジィクラスタリングによる楕円型ファジィモデル

5章では、多次元空間に分布しているデータから線形部分構造にもとづくファジィクラスタリング手法の提案を行った。その結果は3次元上までのデータであれば、散布図によって簡単に確認することが可能である。しかしより高次元上のデータでは、多次元空間上の線形関係を把握する必要があり、その結果を認識することは非常に困難である。本章では、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の提案を行う。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。

6.1 はじめに

ファジィモデリング [22] 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。本章では、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、構造を探求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。この章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6.2節で従来のファジィモデルについていくつかを簡単に紹介する。次に6.3節において、楕円型ファジィモデルを提案する。最後に6.4節において簡単な実際の問題として、水質予測モデルへの応用例を紹介する。

6.2 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 [48]。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 [49] と、構造の非線形性に注目したルール型回帰モデル [22] 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 [22]。

$$\text{ルール } R_i \begin{cases} \text{if } z_1 \text{ is } A_{i1}, z_2 \text{ is } A_{i2}, \text{ and } \dots \\ \text{then } y = c_{i0} + \sum_j c_{ij} x_j \end{cases} \quad (6.1)$$

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

$$\hat{y} = \frac{\sum_{i=1}^c w_i \hat{y}_i}{\sum_{i=1}^c w_i}, \quad w_i = \prod_j \mu_{ij}(z_j^*) \quad (6.2)$$

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 [21]。

$$\text{ルール } R_i \left\{ \begin{array}{l} \text{if } z_1 \text{ is } A_{i1}, z_2 \text{ is } A_{i2}, \text{ and } \dots \\ \text{then } y \text{ is } B_i \end{array} \right. \quad (6.3)$$

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

$$\hat{\mu}(y) = \frac{\sum_{i=1}^c w_i \mu_i(y)}{\sum_{i=1}^c w_i}, \quad w_i = \prod_j \mu_{ij}(z_j^*) \quad (6.4)$$

というメンバ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シップ関数を用いたファジィモデルが提案されている [50]。

$$\text{ルール } R_i \left\{ \begin{array}{l} \text{if } z \text{ is } F_i, \\ \text{then } y = c_{i0} + \sum_j c_{ij} x_j \end{array} \right. \quad (6.5)$$

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。

$$\hat{y} = \frac{\sum_{i=1}^c f_i(z^*) \hat{y}_i}{\sum_{i=1}^c f_i(z^*)} \quad (6.6)$$

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スプな値をとる。このモデルではメンバ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。

6.3 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

$$f_i(\mathbf{z}) = \exp \left\{ - \sum_{j=1}^p \frac{1}{2\lambda_{ij}} \left[(\mathbf{z} - \mathbf{v}_i)^\top \mathbf{e}_{ij} \right]^2 \right\} \quad (6.7)$$

と定める。ただし、 \mathbf{z} は R^p 上の点である。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、回帰式によるクリスプ出力ではなく、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

$$f_{ij}(x_j | \mathbf{x}^*) = \sup_{\mathbf{a} \in \mathbf{R}^{\#\{Y\}-1}} f_i(x_j, \mathbf{a}, \mathbf{x}^*) \quad (6.8)$$

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

$$\hat{x}_{ij}(\mathbf{x}^*) = \arg \sup_{x_j} f_{ij}(x_j | \mathbf{x}^*) \quad (6.9)$$

と定義し、そのメンバシップ値

$$w_i(\mathbf{x}^*) = \sup_{x_j} f_{ij}(x_j | \mathbf{x}^*) \quad (6.10)$$

を用いて、

$$\hat{x}_j(\mathbf{x}^*) = \frac{\sum_{i=1}^c w_i(\mathbf{x}^*) \hat{x}_{ij}(\mathbf{x}^*)}{\sum_{i=1}^c w_i(\mathbf{x}^*)} \quad (6.11)$$

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。

6.4 水質予測モデル

超楕円体クラスタリング法 [21] 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 [21] 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 $c = 3$, $q = 1.5$, $l = 0.4$, $\varepsilon = 10^{-4}$ と設定し、図 6.1 のように、文献 [21] 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 6.1 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。

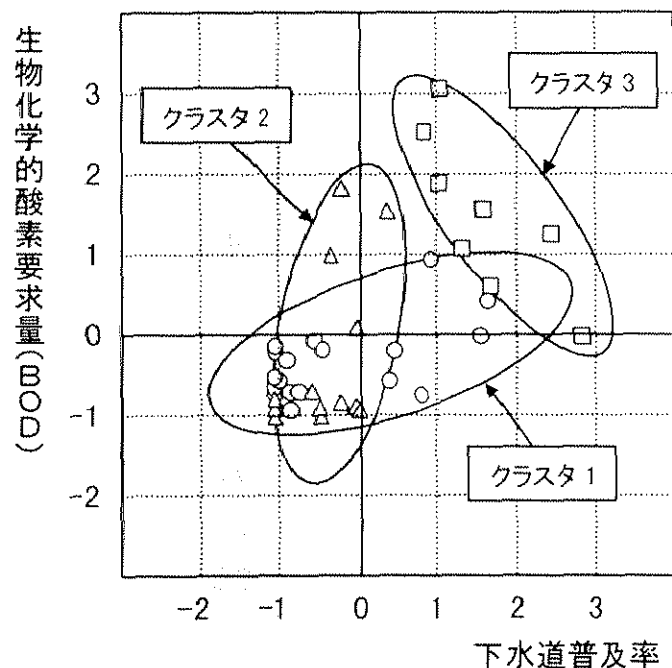


図 6.1: FVD 法によるクラスタリング結果

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シップ関数の射影は図6.2のようになる。図6.2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

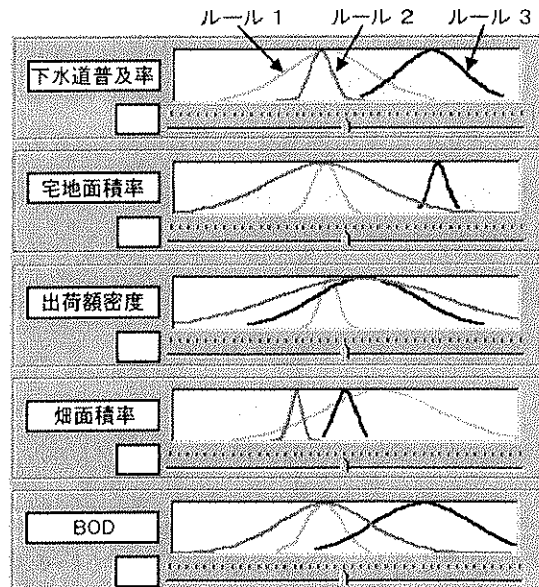


図 6.2: 3つのルールからなる楕円型ファジィモデル

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6.3から図6.8がそれである。データは標準化されているので、図6.3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6.3から図6.8において、下水道普及率以外については(6.8)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6.3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6.3から図6.8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6.2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に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する様子が観察できる。

図6.4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6.5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6.1からも理解できる。

図6.6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6.7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6.8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。

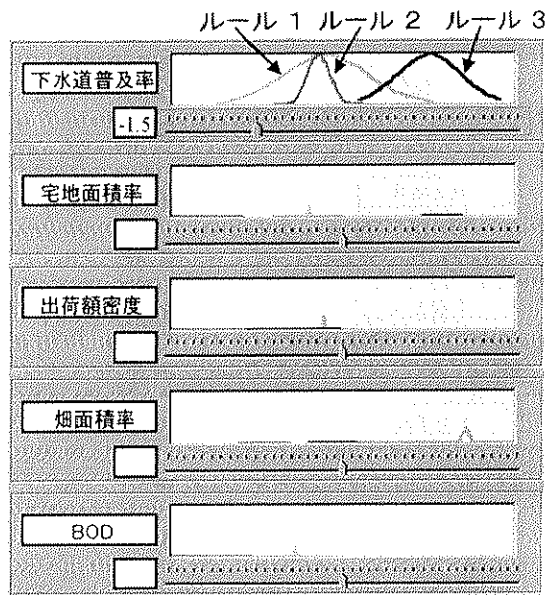


図 6.3: 下水道普及率が -1.5 のときの他の変数の状態

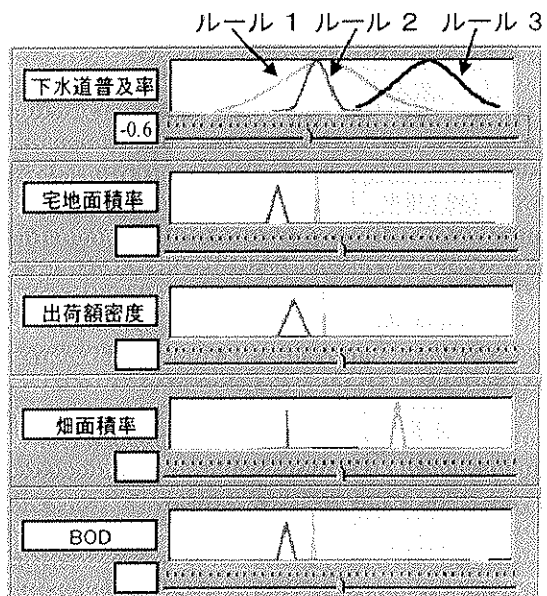


図 6.4: 下水道普及率が -0.6 のときの他の変数の状態

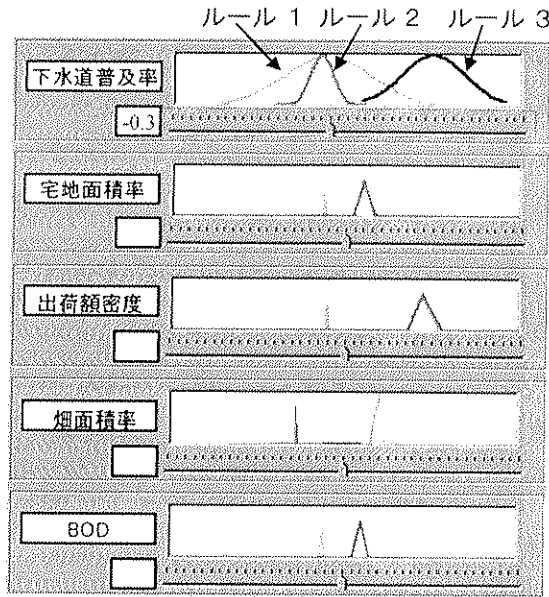


図 6.5: 下水道普及率が -0.3 のときの他の変数の状態

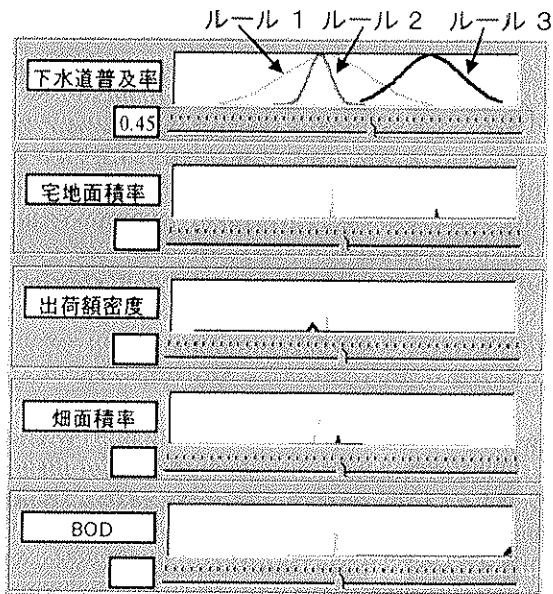


図 6.6: 下水道普及率が 0.45 のときの他の変数の状態

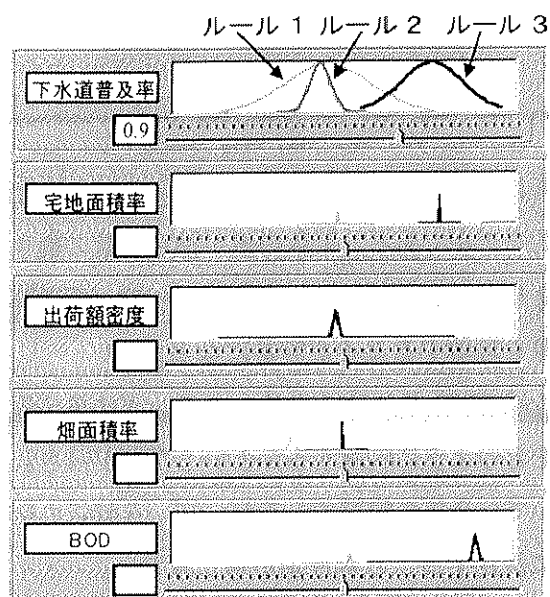


図 6.7: 下水道普及率が 0.9 のときの他の変数の状態

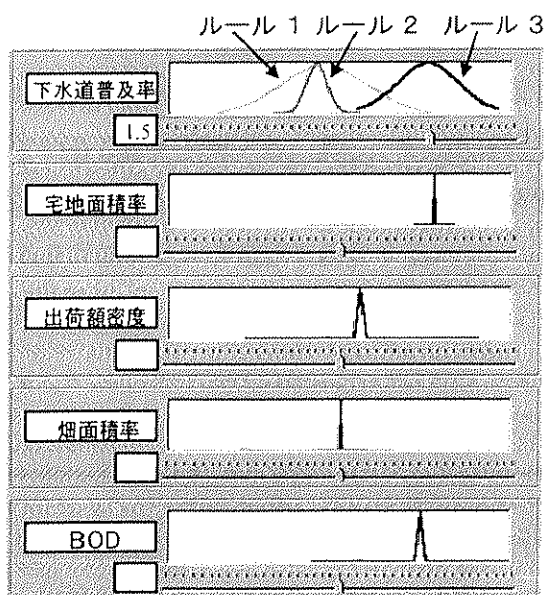


図 6.8: 下水道普及率が 1.5 のときの他の変数の状態

6.5 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もしれない。