JAIST Repository
https://dspace.jaist.ac.jp/

Title	ファジィモデル同定のための前件部変数選択を考慮したファジィクラスタリング
Author(s)	領家，美奈；飯沼，剛志；中森，義輝；田村，坦之
Citation	日本ファジィ学会誌，12(1)：105-113
Issue Date	2000-02-15
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/7956
Rights	Copyright (C) 2000 日本知能情報ファジィ学会．領家　美奈，飯沼　剛志，中森　義輝，田村　坦之，日本ファジィ学会誌，12(1)，2000，105-113．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	

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Japan Society for Fuzzy Theory and Systems

日本ファジィ学会誌 Vol.12, No.1, pp.105-113(2000)

論 文

ファジィモデル同定のための前件部変数選択を考慮したファジィクラスタリング†

領家 美奈*1 飯沼 剛志*2 中森 義輝*1 田村 坦之*2

本論文では，ファジィモデルにおける前件部変数の選択とファジィルールの個数を考慮したファジィモデル構築法を提案する。本手法は，多数の変数により構成されている大規模データ集合に対して有効である。またファジィモデリングにおけるデータ分割問題と後件部変数回帰係数同定を，後件部変数が仮定されたとき，同時に同定するファジィクラスタリング手法を提案する。ファジィクラスタリングの規範は後件部変数で張られる空間では線形関係を，一方，前件部変数で張られる空間ではデータ分布の連続性を重視するものである。

1. はじめに

多入力系に適したファジィモデリングの方法として，高木・菅野のファジィモデル[1]があるが，その同定は困難であり，発見的方法や人間の視察を利用した対話型アプローチによる同定法が一般的である。対話型アプローチは，モデラーが対象に対する知見を与えることができるが，与えられる多次元観測データが大規模化するにつれ，モデラーにかかる負担は大きくなる。

本論文では，説明変数があらかじめ仮定されたとき，ファジィモデリングにおける，データ分割問題と回帰係数同定問題，ならびに前件部変数選択問題を同時に扱うクラスタリング手法を提案し，適切なルール数を考慮しながらモデラーの負担を軽減化するファジィモデル同定法を提案する。クラスタ数の設定方法に関する研究が[2]により行なわれているが，その評価はクラスタリング終了後にクラスタの体積と密度の指標に基づきクラスタ数を選択するというものである。本研究では，ファジィモデルを構築するという目的から，最終的にもっとも精度のよいモデルを導くルール数を選択することで，適切なルール数の決定を考慮する。

2. 高木・菅野のファジィモデル

対象とする高木・菅野のファジィモデル[1]は次のようなルールの集合から成る。

$$\text{Rule } R^i \begin{cases} \text{if } z_1 \text{ is } A_1^i,\ z_2 \text{ is } A_2^i,\ldots,z_l \text{ is } A_l^i \\ \text{then } y^i = c_0^i + \sum_{j=1}^{s} c_j^i \cdot x_j \end{cases} \tag{1}$$

ここで，$z_1,\cdots,z_l$は前件部変数，$A_1^i,\cdots,A_l^i$はファジィ部分集合，x_jは後件部変数と呼ばれる。モデルの出力値は，次式により求められる。

$$y = \frac{\sum_{i=1}^{c} w^i \cdot y^i}{\sum_{i=1}^{c} w^i},\quad w^i = \prod_{p=1}^{l} A_p^i(z_{p*}) \tag{2}$$

ここで，w^iはルールR^iの適合度を表し，前件部変数の入力値z_{p*}に対するメンバシップ値の積で与えられる。

ファジィモデル同定問題は、次の3つの部分問題にわけることができる。各ルールを構成する部分データ集合の同定問題，精度の良い回帰モデル同定のための後件部変数選択問題と後件部回帰係数同定問題を含む後件部同定問題，ルールを特徴づける前件部変数選択問題とメンバシップ関数同定問題を含む前件部同定問題である。

3. ファジィモデル同定法

まず，提案するファジィモデル同定法の大まかな流れについて説明する。

本手法は大きく前件部変数選択を考慮したクラスタリングを行なう第1段階とメンバシップ関数のチューニングを行なう第2段階にわけることができる。各段階の入出力関係を図1に示す。次に，各段階について詳細に述べる。

3.1 第1段階

第1段階では，前件部変数選択を考慮したデータ分割と回帰係数を同時同定するファジィクラスタリングを行う。まず準備として，目的変数，後件部変数，前件

† Fuzzy Clustering with Selection of Premise Variables for Identification of Fuzzy Models

Mina RYOKE, Tsuyoshi IINUMA, Yoshiteru NAKAMORI and Hiroyuki TAMURA
*1 北陸先端科学技術大学院大学 知識科学研究科
Graduate School of Knowledge Science, Japan Advanced Institute of Science and Technology
*2 大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

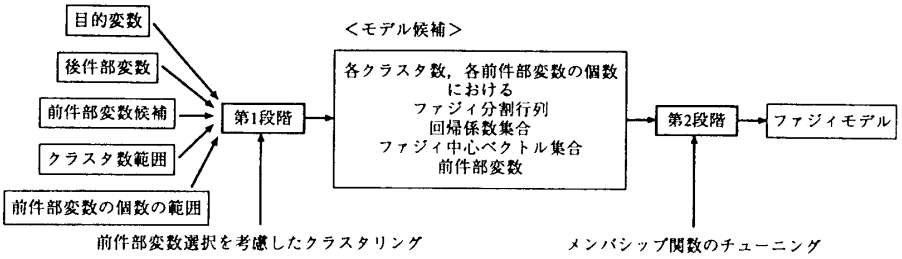


図1 各段階の入出力関係

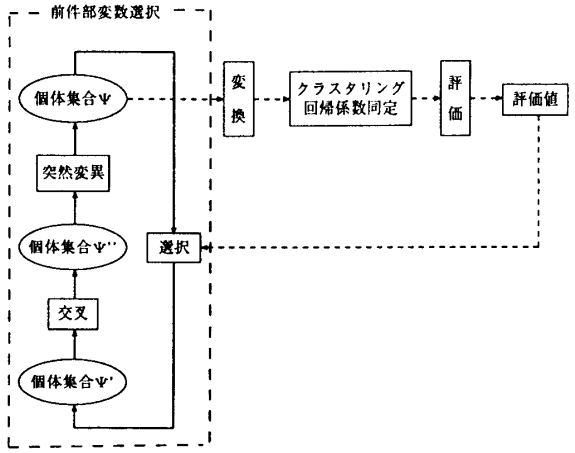


図2 第1段階の流れ

部変数候補，ルール数の範囲，前件部変数の個数の範囲を与えておく。そして，各ルール数ごとに，前件部変数候補集合の部分集合を用いて，ファジィクラスタリングを行い，データ分割と回帰係数同定を行う。

次に前件部変数選択について述べる。本来なら，前件部変数候補集合から全組合せについて調べるべきだが，大規模な多次元観測データに対応できないため，前件部変数の組を記号とする個体による遺伝的アルゴリズム[3]により行う。それぞれの個体について，クラスタリングを実行し，クラスタリング結果と評価値を得る。評価値はクラスタリングで用いる規範値であり，それが個体の適応度になる。

第1段階では，各ルール数ごとに，図2に示すアルゴリズムを実行することになる。このアルゴリズムが終了すると，前件部変数の個数ごとに，評価値最良の個体をモデル候補とする。よって，第1段階では，各クラスタ数，各前件部変数の個数ごとに，1つのモデル候補が選出される。

次に提案するファジィクラスタリングについて述べる。ファジィクラスタリングにおいては，ファジィ分割行列 $U=(u_{ik})$ により，データ k がクラスタ i に属す度合を表す。ファジィ分割行列 U の要素は，c をクラスタ数，n をデータ数とすると，以下の条件を満たすことが要求されている。

$$\begin{aligned}&(\mathrm{i})\ 0\le u_{ik}\le 1,\quad i=1,2,\cdots,c,\\ &\qquad\qquad k=1,2,\cdots,n\\ &(\mathrm{ii})\ 0<\sum_{k=1}^{n}u_{ik}<n,\quad i=1,2,\cdots,c\\ &(\mathrm{iii})\ \sum_{i=1}^{c}u_{ik}=1,\quad k=1,2,\cdots,n\end{aligned}\tag{3}$$

問題を一般的に定式化するためにここでは目的変数が複数の場合を扱う。目的変数を $y_1,y_2,\cdots,y_t$，後件部変数を $x_1,x_2,\cdots,x_s$，前件部変数候補を $z_1,z_2,\cdots,z_p$ とする。また，k 番目の前件部変数候補データベクトルを $\boldsymbol{z}_k=(z_{k1},z_{k2},\cdots,z_{kp})$，後件部変数データベクトルを $\boldsymbol{x}_k=(x_{k1},x_{k2},\cdots,x_{ks})$，目的変数データベクトルを $\boldsymbol{y}_k=(y_{k1},y_{k2},\cdots,y_{kt})$ とする。

クラスタリングで用いられる変数は，目的変数と後件部変数，および前件部変数候補の和集合からなる。クラスタリングに先立ち，各クラスタリング変数を平均0，分散 $s_i^2(i=1,2,\cdots,r\ ;\ r\le(t+s+p))$ に尺度変換する。ここで，s_i^2 は重視度パラメータ[4]であり，この値を大きくすることは，その変数を重視することに対応するので，デザインパラメータの一種と考えることができる。ただし，線形回帰モデルの最終的な係数は元のスケールに逆変換する。

ここで，h 番目の目的変数に対する回帰モデルを，(4)式のように定義する。

$$\begin{aligned}y_{kh}&=f_{ih}(\boldsymbol{x}_k\ ;\ \boldsymbol{\beta}_{ih})+\varepsilon_{kh}\\ &=\beta_{ih0}+\beta_{ih1}x_{k1}+\cdots+\beta_{ihs}x_{ks}+\varepsilon_{kh}\\ &\qquad i=1,2,\cdots,c\end{aligned}\tag{4}$$

ここで，$\boldsymbol{\beta}_{ih}$ は，h 番目の目的変数の i 番目のルールに対する回帰パラメータで，ε_{kh} は残差項である。回帰係数集合を Ω としておく。

クラスタリング規範は次式で表される。

$$J(U,\Omega,\overline{Z},B,\gamma_1,\cdots,\gamma_c)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^mL_{ik}(\Omega,\overline{Z},B)-\eta(B)\sum_{i=1}^{c}\gamma_i\|\overline{\boldsymbol{z}}_i\|_B^2\tag{5}$$

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

ここで，$m(>1)$はスムージングパラメータであり，この値が大きいほど分割結果のあいまい度は大きくなる。また，Ωは回帰係数集合，$\overline{Z}=(\overline{z}_1,\cdots,\overline{z}_c)$は前件部候補変数空間におけるクラスタ中心ベクトル集合，Bは前件部変数行列，$\gamma_1,\cdots,\gamma_c$は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である。はじめに第1項について述べた後で，第2項について記述する。

$L_{ik}(\Omega,\overline{Z},B)$は次式によって定義される。

$$L_{ik}(\Omega,\overline{Z},B)=\|\boldsymbol{y}_k-\boldsymbol{f}_i(\boldsymbol{x}_k,\Omega)\|^2_{R^i}+\eta(B)\|\overline{\boldsymbol{z}}_i-\boldsymbol{z}_k\|^2_B \tag{6}$$

ここで，第1項はi番目のルールによる目的変数のk番目のデータy_kに対する回帰残差の平方を表す。第2項は前件部変数空間におけるクラスタ中心$\overline{\boldsymbol{z}}_i$とデータ点$\boldsymbol{z}_k$との2乗距離を表している。$B\in\boldsymbol{R}^{p\times p}$は前件部変数行列で，各要素は以下の値をとる。

$$b_{ij}=\begin{cases}1 & i=j\text{であり，変数}i\text{が前件部変数}\\ 0 & \text{その他}\end{cases}$$
$$i=1,2,\cdots,p,\ j=1,2,\cdots,p \tag{7}$$

ここで前件部変数行列の対角成分を対角成分ベクトル$\boldsymbol{b}=(b_{11},b_{22},\cdots,b_{pp})^\top$と定義する。また，$\eta(B)$は，前件部変数の数$\mathrm{tr}(B)$を用いると，$\eta(B)=\mathrm{tr}(B)^{-1}$で表されるパラメータである。すなわち，前件部変数空間の次元の数で正規化していることになる。

一方，(5)式の第2項$\|\overline{\boldsymbol{z}}_i\|^2_B(=\overline{\boldsymbol{z}}_iB\,\overline{\boldsymbol{z}}_i^\top)$は前件部変数空間におけるクラスタ中心と原点との2乗距離である。全変数のデータは正規化されているので，原点は全データの重心を表す。全データの重心との距離により，他のクラスタとの距離の値を近似的に表している。

次に，(5)式の各パラメータの同定について述べる。

(a) Uの同定

ファジィ分割行列Uの最適化は，(5)式の局所的最小のための必要条件によって得られる。(3)式の(iii)の制約条件をもとに，ラグランジュ乗数法を用いると，必要条件は，

$$u_{ik}=\left[\sum_{j=1}^{c}\left(\frac{L_{ik}}{L_{jk}}\right)^{\frac{1}{m-1}}\right]^{-2} \tag{8}$$

となり，ファジィ分割行列Uの更新式に用いられる。ただし，$L_{ik}=0$となるようなiが存在するときu_{ik}は次式で与えられる。

$$u_{ik}=\begin{cases}1/|I_k|, & i\in I_k\\ 0, & i\notin I_k\end{cases} \tag{9}$$

ここで，$I_k=\{i\,|\,i=1,2,\cdots,c,\ L_{ik}=0\}$

(b) Ωの同定

h番目の目的変数に対するi番目のルールによる回帰モデルは，(4)式で表される。これを行列の形で書くと，

$$\boldsymbol{\theta}_h=X\boldsymbol{\beta}_{ih}+\boldsymbol{\varepsilon}_{ih} \tag{10}$$
$$=\begin{pmatrix}1 & \boldsymbol{x}_1\\ 1 & \boldsymbol{x}_2\\ \vdots & \vdots\\ 1 & \boldsymbol{x}_n\end{pmatrix}\boldsymbol{\beta}_{ih}+\boldsymbol{\varepsilon}_{ih} \tag{11}$$

と書くことができる。ただし，$\boldsymbol{\theta}_h\in\boldsymbol{R}^n$はデータ$y_{kh}$を第$k$要素とする列ベクトル，$X\in\boldsymbol{R}^{n\times(s+1)}$はデータベクトル$\boldsymbol{x}_k$を持つ行列(ただし，第1列はすべて1)，$\boldsymbol{\varepsilon}_{ih}$は残差ベクトルである。

回帰係数ベクトルの最適化は，

$$\boldsymbol{\varepsilon}_{ih}\Gamma_i\boldsymbol{\varepsilon}_{ih}\rightarrow\min \tag{12}$$

の解によって得られる。ただし，$\Gamma_i\in\boldsymbol{R}^{n\times n}$は$(u_{ik})^m$を第$(i,i)$要素とする対角行列である。(12)式の解は，

$$\boldsymbol{\beta}_{ih}=(X^\top\Gamma_iX)^{-1}X^\top\Gamma_i\boldsymbol{\theta}_h \tag{13}$$

となり，回帰係数ベクトル集合Ωの更新式に使われる。

(c) $\overline{Z}$の同定

クラスタ中心ベクトル集合$\overline{Z}$の最適化は，(5)式の局所的最小のための必要条件によって得られる。すなわち，

$$\frac{\partial J}{\partial\overline{\boldsymbol{z}}_i}=0\Leftrightarrow\overline{\boldsymbol{z}}_i=\frac{\sum_{k=1}^{n}(u_{ik})^m\overline{\boldsymbol{z}}_k}{\sum_{k=1}^{n}(u_{ik})^m-\gamma_i} \tag{14}$$

となり，クラスタ中心ベクトル集合$\overline{Z}$の更新式に使われる。

(d) γ_iの同定

γ_iはクラスタリング過程において適応的に変化し，クラスタリング規範における第3項のクラスタ中心間距離項の重視度合を決めるパラメータである。γ_iが大きくなるにつれて，(14)式より，クラスタ中心が原点から遠ざかっていき，前件部変数空間でいっそうクラスタの分離度合が増す傾向が強まることになる。同定の方針を以下に述べる。

1. クラスタに含まれるデータ数が少ないほど，(5)式の第2項の影響が大きくなるので，γ_iの値を小さくとる。
2. 他のクラスタとの中心間距離が大きくなるにつれ，γ_iの値を小さく設定することで，必要以上にクラスタ中心が原点から遠ざかるのを避ける。

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

3. γ_iの値はクラスタに含まれるデータ数の値より，十分小さくしなければならない。

3. の理由はクラスタ中心ベクトルの更新式(14)による制約である。もし，γ_iの値が$\sum_{k=1}^{n}(u_{ik})^m$の値より大きければ，(14)式の分母が負となり，クラスタ中心が原点に対して対称の象限に移動することになってしまう。また，$\sum_{k=1}^{n}(u_{ik})^m$の値に限りなく近づくと，(14)式の分母の値が限りなく0に近くなり，クラスタ中心が原点のはるか遠方に行くことになる。従って，γ_iの同定に方針3.が必要となる。

以上から，γ_iは次式で与える。

$$\gamma_i=\min\left\{\frac{\sum_{k=1}^{n}(u_{ik})^m}{\eta\sum_{j\neq i}^{c}\|\bar{z}_i-\bar{z}_j\|_B^2},\ \frac{\sum_{k=1}^{n}(u_{ik})^m}{\xi}\right\} \tag{15}$$

(e) B の同定

次に，前件部変数行列 B の同定について述べる。前件部変数行列 B の同定は，クラスタリング規範(5)式の値を適応度とした遺伝的アルゴリズムによる。1つの前件部変数の組は，個体の遺伝子表現として，2つの値{0,1}の対立遺伝子からなる p 文字長の記号列(個体)として表される。一様乱数を用いてランダムに P 個の初期個体を生成し，設定された世代数まで，(5)式の値を適応度として遺伝的演算を繰り返す。計算が終了すると，前の世代にもさかのぼって，各個体を前件部変数の個数によって分ける。次に，設定した前件部変数の個数の範囲内で，各前件部変数の個数ごとに，評価値最良の個体を選ぶ。選んだ最良の個体が，その前件部変数の個数での，最適な前件部変数の組を表していることになる。

まとめると，この規範は3項構成で，各項はそれぞれ次のような目的を担っている。

1. 目的変数空間の各ルールの回帰残差の最小化
2. 前件部変数空間の各クラスタのファジィ散布度の最小化
3. 前件部変数空間の各クラスタのクラスタ中心間距離の最大化

ファジィモデルにおける前件部変数としてふさわしい変数は，各々のルールが最もよく力を発揮できる変数，いいかえれば，クラスタがよく分離しているような変数である。クラスタが分離していると思われるのは，次の2つの条件を満たすことである。

・それぞれのクラスタのファジィ散布度が小さい
・それぞれのクラスタ中心間距離が離れている

よって本手法では，前件部変数空間におけるクラスタ中心間距離を考慮することにより，回帰残差が小さく，前件部変数空間におけるファジィ散布度が小さく，クラスタ中心間距離が大きくなるようなファジィ分割ならびに前件部変数空間を発見する。

第1段階のアルゴリズムを以下に述べる。

Step 1 スムージングパラメータ m，クラスタ数の範囲 $c_{\min}$, $c_{\max}$，前件部変数の個数の範囲 $l_{\min}$, $l_{\max}$，γ_iの更新式内のパラメータ ξ，そしてクラスタリングの収束判定値 ε を与える。遺伝的アルゴリズムの各パラメータ(個体数 P,世代数 Φ,選択数 P_s,突然変異率 p_m)を設定する。各変数データを重視度パラメータに基づき，尺度変換する。

Step 2 クラスタ数 $c=c_{\min}$とする。

Step 3 世代数 $\phi=0$とする。初期個体集合 $\Psi(0)=\{\Psi_{00},\cdots,\Psi_{0(P-1)}\}$を与える。

Step 4 個体番号 $\rho=0$とする。

Step 5 各個体 $\Psi_{\phi\rho}$に対し，クラスタリングを実行し，評価値 $J_{\phi\rho}$を得る。

Step 6 $\rho<P-1$であれば，$\rho=\rho+1$として，Step5に戻る。$\rho=P-1$であれば，Step7に進む。

Step 7 $\phi<\Phi-1$であれば，P 個の個体に対し遺伝的操作。次世代個体集合 $\Psi(\phi+1)$を得る。$\phi=\phi+1$として，Step4に戻る。$\phi=\Phi-1$であれば，各前件部変数の個数ごとに最良の個体のコードを前件部変数とし，そのときのクラスタリング結果を採用する。

Step 8 $c<c_{\max}$であれば $c=c+1$として Step3に戻る。$c=c_{\max}$であれば終了。

Step 5のクラスタリングのアルゴリズムを以下に述べる。

Step 5-1 ステップ数 $r=0$とする。$U,\gamma_1,\cdots,\gamma_c$の初期値を与える。

Step 5-2 ファジィ中心ベクトルを計算する。

$$\bar{z}_i^{(r)}=\frac{\sum_{k=1}^{n}(u_{ik}^{(r)m}z_k)}{\sum_{k=1}^{n}(u_{ik}^{(r)})^m-\gamma_i^{(r)})},\quad k=1,2,\cdots,c \tag{16}$$

Step 5-3 回帰モデルの係数を次式により求める。

$$\boldsymbol{\beta}_i^{(r)}=(X^{\top}\Gamma_iX)^{-1}X^{\top}\Gamma_i\boldsymbol{\theta}_h \tag{17}$$

Step 5-4 適応型パラメータ $\gamma_i(i=1,\cdots,c)$を次式により更新する。

$$\gamma_i^{(r)}=\min\left\{\frac{\sum_{k=1}^{n}(u_{ik}^{(r)})^m}{\eta\sum_{j\neq i}^{c}\left\|\bar{z}_i^{(r)}-\bar{z}_j^{(r)}\right\|_B^2},\ \frac{\sum_{k=1}^{n}(u_{ik}^{(r)})^m}{\xi}\right\} \tag{18}$$

Step 5-5 分割行列 U を次式により更新する。

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

$$u_{ik}^{(r+1)}=\left[\sum_{h=1}^{c}\left(\frac{L_{ik}^{(r)}}{L_{ih}^{(r)}}\right)^{\frac{1}{m-1}}\right]^{-1} \tag{19}$$

Step 5-6 終了条件

$$\max_{i,k}\{|u_{ik}^{(r+1)}-u_{ik}^{(r)}|\}<\varepsilon \tag{20}$$

を満たせば終了。満たさなければ $r=r+1$ として，Step5-2に戻る。

3.2 第2段階

第2段階では図1に示す通り，メンバシップ関数のチューニングを行なう。第1段階が終了した段階では，各クラスタ数，各前件部変数の個数における，最良の前件部変数の組とクラスタリング結果が求まっている。その時点では，モデルの候補となるクラスタリング結果の組は，$(c_{\max}-c_{\min}+1)\times(l_{\max}-l_{\min}+1)$ 個ある。第2段階では，まずそれらのモデル候補に対し，メンバシップ関数のチューニング前の目的変数の実データの値とモデル出力値との相関係数を調べる。その中から上位 $\tau(\tau=1,2,\cdots,(c_{\max}-c_{\min}+1)\times(l_{\max}-l_{\min}+1))$ 個のモデル候補に対し，メンバシップ関数のチューニングを遺伝的アルゴリズムにより行い，最終的に最も相関係数が高いモデルを採用する。

クラスタ $i(i=1,\cdots,c)$ の変数軸 $p(p=1,\cdots,l)$ に対するメンバシップ関数を次式で定義する。

$$\mu_{ip}(z_p)=\exp\left\{-\frac{(z_p-\bar{Z}_{ip})^2}{2(t_{ip})^2S_{ip}}\right\} \tag{21}$$

ただし，$\bar{Z}_{ip}$ は先に述べたクラスタ中心である。また，S_{ip} はクラスタ i の変数軸 p に対するファジィ分散で，

$$S_{ip}=\frac{\sum_{s=1}^{n}(u_{is})^m(z_{sp}-\bar{Z}_{ip})^2}{\sum_{s=1}^{n}(u_{is})^m} \tag{22}$$

で表される。また，t_{ip} はチューニングパラメータで，デフォルト値は1である。このメンバシップ関数の形は左右対称のガウス型であり，クラスタの中心点がメンバシップ値の最大値1をとる。クラスタのファジィ分散やチューニングパラメータが大きくなれば，各点でメンバシップ値も大きくなる。チューニングパラメータの同定は，目的変数の実データの値とモデル出力値との2乗誤差平均が最小化するように，遺伝的アルゴリズムにより行う。

第2段階のアルゴリズムを以下に述べる。

Step 1 メンバシップ関数のチューニングを行うモデル候補の数 τ，チューニングパラメータが取り得る値の範囲 $t_{\min}, t_{\max}$ を設定する。遺伝的アルゴリズムの各パラメータ(個体数 P，世代数 Φ，選択数 P_s，交叉率 p_c，突然変異率 p_m)を設定する。

Step 2 $(c_{\max}-c_{\min}+1)\times(l_{\max}-l_{\min}+1)$ 個のモデル候補のそれぞれに対し，メンバシップ関数を構成し，目的関数の相関係数を算出する。

Step 3 相関係数をソーティングし，上位 τ 個のモデルを選択する。

Step 4 τ 個のモデル候補に対し，メンバシップ関数のチューニングの作業を行い，目的変数の相関係数を算出する。

Step 5 相関係数をソーティングし，最も相関係数の高いモデル候補をモデルとして採用し終了。

Step 4のメンバシップ関数のチューニングのアルゴリズムを以下に述べる。

Step 4-1 初期化：ランダムに P 個の個体を生成して，初期個体集合 $\Psi(0)$ を作り，世代 $\phi=0$ とする。

Step 4-2 評価：個体集合 $\Psi(\phi)$ 内の個体について，目的変数の2乗誤差平均 e_j を計算する。

Step 4-3 選択：個体集合 $\Psi(\phi)$ に選択演算子を適用し，$\Psi'(\phi)$ を生成する。

Step 4-4 交叉：$\Psi'(\phi)$ に交叉演算子を適用し，$\Psi''(\phi)$ を生成する。

Step 4-5 突然変異：$\Psi''(\phi)$ に突然変異演算子を適用し，次世代の個体集合 $\Psi(\phi+1)$ を生成する。

Step 4-6 判定：$\phi<\Phi-1$ ならば，$\phi=\phi+1$ としてStep4-2へ。$\phi=\Phi-1$ ならば計算終了。その時点での2乗誤差平均が最も小さい個体を解とする。

4. 数値実験

提案するファジィモデル同定法を世界87カ国の統計データに関する数値例に適用する。ここでは，データブック[5]のデータ(ソビエト連邦崩壊前)を用いて，男性平均寿命(歳，1980～1985年)を目的変数とするファジィモデルを構築する。後件部変数は，目的変数と関連するであろう乳児死亡率(x_1：人/1歳未満1000人，1980～1985年)，1人あたりGNP(x_2：1000US/人，1988年)，工業人口比(x_3：%，1981年)の3変数を選択した。また，前件部変数候補として，ルールの特徴を良く記述するものと期待される表1の21変数を設定した。なお，今回の数値実験では，全変数のデータを平均0，分散1に尺度変換している。

また，第1段階の設定パラメータとして，$c_{\min}=3$，$c_{\max}=4$，$l_{\min}=2$，$l_{\max}=8$，$m=2.0$，$\xi=2.0$，$\varepsilon=1.0\times10^{-5}$，$\gamma_i$ の初期値を，$\gamma_i=0.0(i=1,\cdots,c)$，GAの各パラメータを，$P=20$，$\Phi=30$，$P_s=10$，$p_m=0.10$ とした。

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

各クラスタ数について，それぞれ遺伝的アルゴリズムにより，前件部変数の探索およびクラスタリングを行い，各前件部変数の個数ごとに，前件部変数の組とクラスタリング結果が求まる。その結果を表2に示す。表2には，各クラスタ数，各前件部変数の個数ごとの前件部変数の組，クラスタリング結果，目的変数の実データの値とモデル出力値の相関係数を示している。

次に，第2段階の設定パラメータとして，$\tau=3$，チューニングパラメータの範囲を，$t_{\min}=0.1$，$t_{\max}=5.0$，GAの各パラメータを，$P=20$，$\Phi=300$，$P_s=10$，$p_c=0.10$，$p_m=0.10$とした。第2段階では，表2のモデル候補について，チューニング前の目的変数の相関係数の上位3組のモデル候補に対し，メンバシップ関数のチューニングを行う。表3にチューニング前の観測データと予測結果の相関係数とチューニング終了後の相関係数を各モデル候補について示す。

チューニングの結果，相関係数の最も高くなった$c=4$，$l=2$のときのモデルが採用される。次に採用されたモデルの詳細な結果を述べる。まず，選択された前件部変数のクラスタリング結果を図3に示す。

選択された前件部変数について検討するために，同じ条件($c=4, l=2$)のもとで，前件部変数候補の全組合せについてクラスタリングを行い評価規範値を求めた結果，GAで選択された変数$\{z_4, z_{12}\}$による評価規範値が，最も良い値を示した。よって本数値例では前件部変数選択において，最適解が求まっているといえる。

クラスタリングによる後件部回帰式の同定結果を以下に示す。ただし，尺度変換データによる回帰モデルの係数を，平均と分散を用いて元のスケールに変換している。

$$\begin{aligned}
R^1&: y=7.649\times10^{1}-2.875\times10^{-1}x_1\\
&\quad +1.013\times10^{-4}x_2-1.004\times10^{-1}x_3\\
R^2&: y=5.758\times10^{1}-1.316\times10^{-1}x_1\\
&\quad +9.540\times10^{-4}x_2+2.948\times10^{-1}x_3\\
R^3&: y=7.523\times10^{1}-2.183\times10^{-1}x_1\\
&\quad -5.060\times10^{-4}x_2+1.124\times10^{-1}x_3\\
R^4&: y=6.862\times10^{1}-1.430\times10^{-1}x_1\\
&\quad +1.119\times10^{-4}x_2+4.267\times10^{-1}x_3 \qquad (23)
\end{aligned}$$

各クラスタに含まれる国家を以下に示す。

クラスタ1：東ドイツ，イギリス，ノルウェー，スウェーデン，チェコスロバキア，スイス，ハンガリー，フランス，ブルガリア，ベルギー，オーストリア，オーストラリア，西ドイツ，デンマーク，カナダ，

表1 前件部変数候補

変数番号	変数
z_1	人口密度 (人/km^2)
z_2	都市人口密度 (%)
z_3	人口増加率 (%)
z_4	出生率 (o/oo)
z_5	死亡率 (o/oo)
z_6	乳児死亡率 (o/oo)
z_7	1人あたり GNP(1000US$/人)
z_8	農業人口比 (%)
z_9	工業人口比 (%)
z_{10}	サービス業人口比 (%)
z_{11}	可耕地面積比 (%)
z_{12}	識字率 (%)
z_{13}	高等教育進学率 (%)
z_{14}	医師数 (人/1万人)
z_{15}	栄養摂取量 (kcal/1d1h)
z_{16}	ラジオ台数 (台/千人)
z_{17}	エネルギー消費 (kg/人)
z_{18}	輸出額 ($/1万人)
z_{19}	輸入額 ($/1万人)
z_{20}	兵員数 (人/1万人)
z_{21}	国防費 ($/1万人)

表2 前件部変数選択ならびに各指標の結果

l	変数	評価値 J	相関係数
クラスタ数 c が 3			
2	z_3, z_4	4.872	0.9581
3	z_3, z_4, z_6	7.982	0.9498
4	z_3, z_4, z_6, z_{12}	9.706	0.9497
5	$z_3, z_4, z_6, z_8, z_{12}$	11.36	0.9455
6	$z_3, z_4, z_6, z_8, z_9, z_{12}$	12.19	0.9487
7	$z_3, z_4, z_6, z_8, z_9, z_{12}, z_{14}$	13.02	0.9514
8	$z_3, z_4, z_6, z_8, z_9, z_{12}, z_{14}, z_{15}$	13.57	0.9512
クラスタ数 c が 4			
2	z_4, z_{12}	3.385	0.9582
3	z_3, z_4, z_{12}	5.164	0.9562
4	z_3, z_4, z_6, z_{12}	6.378	0.9527
5	$z_3, z_4, z_6, z_9, z_{12}$	8.154	0.9521
6	$z_3, z_4, z_6, z_9, z_{12}, z_{15}$	9.124	0.9454
7	$z_3, z_4, z_5, z_6, z_9, z_{12}, z_{15}$	10.25	0.9496
8	$z_3, z_4, z_6, z_8, z_9, z_{12}, z_{14}, z_{15}$	10.29	0.9475

表3 数値例の実験結果

c	l	前件部変数	相関係数
4	2	z_4, z_{12}	0.9582 → 0.9709
3	2	z_3, z_4	0.9581 → 0.9640
4	3	z_3, z_4, z_{12}	0.9562 → 0.9692

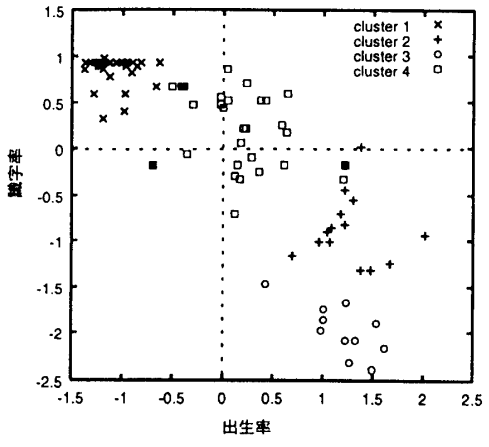


図3 前件部変数空間でのクラスタリング結果

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

ルーマニア，イタリア，日本，アイスランド，アメリカ合衆国，オランダ，ニュージーランド，スペイン，ポーランド，アイルランド，ギリシア，フィンランド，キューバ，ユーゴスラビア，ポルトガル，ソビエト連邦，シンガポール，韓国，イスラエル，アルゼンチン

クラスタ2：ソマリア，マダガスカル，ナイジェリア，イラン，ホンジュラス，ケニア，サウジアラビア，グアテマラ，リビア，アルジェリア，シリア，エジプト，ボリビア，コンゴ

クラスタ3：モザンビーク，アフガニスタン，リベリア，エチオピア，パキスタン，ヨルダン，中央アフリカ，ハイチ，スーダン，ネパール，インド

クラスタ4：マレーシア，コロンビア，ブラジル，パラグアイ，メキシコ，ペルー，フィリピン，ガイアナ，エクアドル，ベネズエラ，ドミニカ共和国，モロッコ，トルコ，エルサルバドル，パナマ，コスタリカ，トリニダードトバゴ，ミャンマー，ジャマイカ，タイ，インドネシア，アラブ首長国連邦，スリランカ，チリ，中華人民共和国，ニカラグア，イラク

クラスタ1に先進資本主義国と(旧)社会主義国が含まれている。クラスタ2，クラスタ3，クラスタ4には発展途上国が含まれているが，クラスタ2には石油輸出国が多く含まれている。また，クラスタ3は南アジア・アフリカ地域の国々，クラスタ4は東南アジア・中南アメリカ地域の国々が多く含まれている。このように，途上国の多様性がクラスタリングに反映された結果となった。

次に，第2段階のメンバシップ関数のチューニング結果について述べる。同定されたチューニングパラメータの値を表4に示す。また，各変数のメンバシップ関数を図4と図5に示す。

目的変数の実データの値とモデル出力値との関係を図6に示す。両者の相関係数は0.9709と高いものになった。

表4 同定されたチューニングパラメータ t の値

クラスタ	前件部変数	t
1	出生率	1.4410
	識字率	0.5071
2	出生率	0.3395
	識字率	0.5071
3	出生率	1.6470
	識字率	1.0050
4	出生率	0.7418
	識字率	1.9090

5. まとめ

本論文では，後件部変数を仮定したとき，前件部変数同定を考慮しながら回帰係数とファジィ分割を同定するファジィクラスタリングを提案した。さらに，提案したファジィクラスタリングを用いて，大規模多次元観測データにも有効な，ファジィモデル同定法を提案した。また，提案した手法を世界の人口関連データに適用し，男性平均寿命を予測した結果，精度が高く，わかりやすいルールを抽出することを示し，本手法の有効性を確認した。今後の課題のひとつには，さらに後件部変数選択問題を考慮した手法が含まれる。また後件部では線形回帰モデルを仮定しているが，線形回帰モデルを後件部に持つよりもファジィ集合を後件部

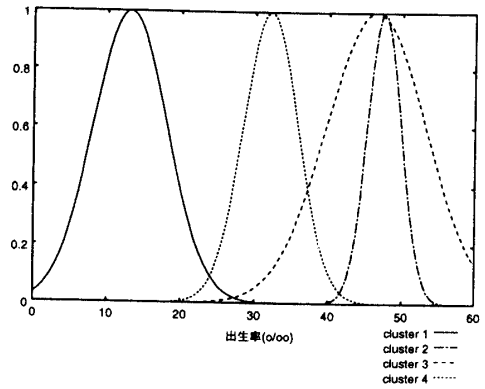


図4 出生率のメンバシップ関数

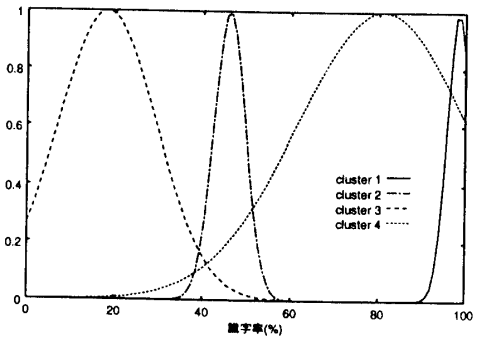


図5 識字率のメンバシップ関数

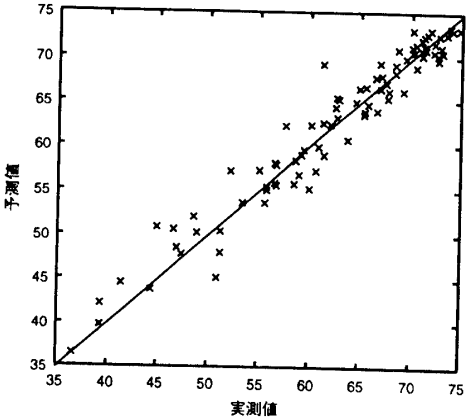


図6 目的変数の実データの値と出力値

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

に持つ方が望ましい場合についても検討する必要がある。本手法では，前件部のファジィ命題により線形関係が成り立つ範囲を制約し，そういった場合に対応していると解釈している。

参 考 文 献

[1]T. Takagi and M. Sugeno：“Fuzzy Identification of Systems and Its Applications to Modeling and Control” *IEEE Transactions on Systems, Man and Cybernetics*, Vol.SMC-15, No.1, pp.116-132(1985).
[2]I. Gath, A. B. Geva：“Unsupervised Optimal Fuzzy Clustering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.11, No.7, pp.773-781(1989).
[3]玉置久・喜多一・岩本貴司・三宮信夫：“遺伝アルゴリズムーI－GAの基礎”，システム／制御／情報，Vol.39, No.6, pp.295-302(1995).
[4]M. Ryoke, H. Tamura and Y. Nakamori：“Fuzzy Rule Generation by Hyperellipsoidal Clustering”, *Proc. of IIZUKA'96(4th International Conference on Soft Computing)*, Vol.1, pp.86-89(1996).
[5]朝日新聞社：“朝日現代用語1991別冊付録「知恵庫」”179 p.(1991).

(1999年4月21日 受 付)
(1999年8月30日 再受付)

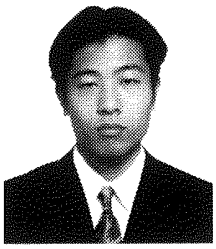
[問い合わせ先]
〒923-1292
石川県能美郡辰口町旭台1-1
北陸先端科学技術大学院大学
知識科学研究科
領家 美奈
TEL ：0761-51-1757
FAX ：0761-51-1149
E-mail：ryoke@jaist.ac.jp

著 者 紹 介

領家 美奈（りょうけ みな）

北陸先端科学技術大学院大学知識科学研究科

1998年 大阪大学大学院基礎工学研究科物理系専攻制御工学分野博士後期課程修了。博士(工学)。1998年4月より北陸先端科学技術大学院大学助手。1993年度 日本ファジィ学会論文賞，1999年度システム制御情報学会奨励賞受賞。日本ファジィ学会，システム制御学会，環境科学会会員。



飯沼 剛志（いいぬま つよし）

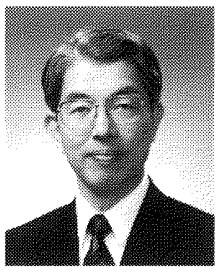
大阪大学大学院基礎工学研究科

1997年3月 大阪大学基礎工学部システム工学科卒業。同年4月 大阪大学基礎工学研究科博士前期課程システム人間系システム科学分野入学。1999年3月 修了。同年4月 関西電力株式会社に入社。現在，大阪北支店勤務。

中森 義輝（なかもり よしてる）

北陸先端科学技術大学院大学知識科学研究科

1979年 京都大学大学院工学研究科数理工学専攻博士課程修了。工学博士。甲南大学理学部応用数学科勤務を経て，1998年4月より北陸先端科学技術大学院大学教授。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。



田村 坦之（たむら ひろゆき）

大阪大学大学院基礎工学研究科

1964年 大阪大学大学院工学研究科修士課程修了。同年～1971年 三菱電機(株)中央研究所勤務，1971年 大阪大学工学部助教授，1987年 同教授，1993年 基礎工学部教授，1997年 大学院基礎工学研究科教授となり現在に至る。その間，1966年～1968年 米国スタンフォード大学大学院 Engineering-Economic Systems 学科に留学，1972年～1973年 英国ケンブリッジ大学客員研究員，1986年～1988年 文部省統計数理研究所併任。大規模複雑系を対象にしたシステム計画論とその生産システム・公共システムへの応用に関する研究に従事。工学博士。1976年，1999年 計測自動制御学会論文賞，1990年 システム制御情報学会樋本記念賞論文賞，1991年 日本OR学会事例研究奨励賞受賞。日本OR学会フェロー，計測自動制御学会，システム制御情報学会，環境経済・政策学会，IEEE，INFORMS，SRA などの会員。

NII-Electronic Library Service

Japan Society for Fuzzy Theory and Systems

Fuzzy Clustering with Selection of Premise Variables for Identification of Fuzzy Models

by

Mina RYOKE, Tsuyoshi IINUMA, Yoshiteru NAKAMORI and Hiroyuki TAMURA

Abstract :

This paper proposes a fuzzy modeling technique taking into account selection of premise variables and the number of fuzzy rules. This technique is effective for a large-scale data set and of a large number of variables. In fuzzy modeling, a fuzzy clustering method is also proposed in order to determine a data partition and consequence parameters simultaneously under the assumed consequence variables. The proposed criterion of the fuzzy clustering has two purposes to detect the linearity in the space of consequence variables and the continuity in the space of premise variables.

Contact Address : **Mina RYOKE**
Graduate School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1-1 *Asahidai, Tatsunokuchi, Ishikawa* 923-1292, *Japan*
TEL : 0761-51-1757
FAX : 0761-51-1149
E-mail : ryoke@jaist.ac.jp

NII-Electronic Library Service