

# 规则库自提取和简化的模糊系统

郭海湘<sup>1</sup>, 刘涛<sup>2</sup>, 诸克军<sup>1</sup>

(1. 中国地质大学管理学院, 武汉 430074; 2. 许昌职业技术学校, 河南许昌 461000)

**[摘要]** 介绍了一种新的规则自提取和简化的模糊模型算法, 该算法由3个步骤组成: 首先, 通过构造一个模糊软划分判别准则对输入输出空间进行模糊划分, 得到最佳分类数, 从而得到了最佳的规则数; 其次, 根据最佳规则数和神经网络来构造初始的模糊模型; 第三, 通过运用模糊相似分析法, 可以得到计算2个模糊集相似度的方程, 从而根据事先确定的阈值来合并相似的模糊集。这样就得到了一个既满足对精度的要求又简单且具有可解释性的模糊系统。最后, 用该算法对我国全要素生产力进行了模糊规则的提取。

**[关键词]** 相似分析法; 模糊模型; 模糊规则; 模糊集; 神经网络

**[中图分类号]** TP181 **[文献标识码]** A **[文章编号]** 1009-1742(2004)10-0052-07

## 1 引言

模糊系统有广泛的应用领域, 例如, 模式识别、数据挖掘、分类、预测、非线性系统模拟和过程控制等方面<sup>[1~3]</sup>。模糊系统的主要优点是方便使用 if-then 规则的形式来表达知识, 具有人脑性质。然而, 通过模糊系统从数据中得到的初始规则库, 由于存在相似或者冗余的模糊集, 导致系统可能不具有很好的可解释性。为了提高模糊系统的可解释性, 已经提出了很多解决方法。其中一些方法注重数字的精确性和语言的可解释性<sup>[4, 5]</sup>, 这些方法的主要缺点是当输入变量增加时, 规则库会以相应的几何级数进行增长<sup>[6]</sup>。另外一些方法注重系统的精确性和简单性<sup>[6, 7]</sup>, 这些方法在保证精度的前提下尽量简化系统用来提高系统的可解释性。近几年已经提出了许多模糊系统的产生和简化的方法。Chao 和 Chen<sup>[6]</sup>提出了一种基于相似分析的模糊规则库的简化方法。Setnes 等<sup>[5]</sup>提出了一种用集理论相似分析方法去减少模型中模糊集的数目, 用来简化规则库。笔者介绍了一种新的模糊规则库自提取和简化的模糊模型算法。该算法分3个步骤: 首

先, 通过构造一个新的判别准则来进行模糊分类, 确定样本数据的最佳分类, 从而确定最佳规则数; 其次, 根据最佳规则数和神经网络来构造初始的模糊系统; 第三, 通过运用模糊相似分析法, 可以得到计算2个模糊集的相似度的方程, 再根据事先确定的阈值来合并相似的模糊集, 这样就得到了一个既满足精度又简单且具有可解释性的模糊系统。最后, 用该系统提取我国全要素生产力的模糊规则。

## 2 规则库的自提取

众所周知, 具有相似几何分布的数据可以划分为一类。模糊分类可以确定样本数据的最佳分类数, 从而确定最佳规则数, 也就得到了初始的模糊系统的结构。用于分类的方法很多, 应用最广泛的方法是 Fuzzy C-mean (FCM)<sup>[8, 9]</sup>, 因为该方法简单且效率高。然而, 运用该方法时分类数  $C$  必须事先确定。一个好的划分类别的判别准则必须既考虑同一类中的紧凑程度又要兼顾类与类之间的分离程度。笔者结合 FCM 算法, 介绍了一种新的划分类别的判别准则, 即

**[收稿日期]** 2003-11-21; **修回日期** 2004-07-15

**[基金项目]** 国家自然科学基金项目资助(70273044)

**[作者简介]** 郭海湘(1978-), 男, 湖南湘乡市人, 中国地质大学硕士研究生

$$V_p(U, C) = \frac{1}{n} \sum_{k=1}^n \max_i(u_{ik}) - \frac{1}{K} \sum_{i=1}^{C-1} \sum_{j=i+1}^C \left[ \frac{1}{n} \sum_{k=1}^n \min(u_{ik}, u_{jk}) \right],$$

$$K = \sum_{i=1}^{C-1} i \quad (1)$$

其中

$$V_i = \frac{\sum_{k=1}^N (u_{ik})^m X_k}{\sum_{k=1}^N (u_{ik})^m}, i = 1, 2, \dots, C \quad (2)$$

$$\mu_{ik} = \left( \sum_{j=1}^C (d_{ik}/d_{jk})^{2/(m-1)} \right)^{-1} \quad (3)$$

$$d_{jk}^2 = \|X_k - V_j\|^2 = (X_k - V_j)^T (X_k - V_j) \quad (4)$$

可以看出，聚类判别准则  $V_p$  由 2 部分组成，第一部分反映了同一类中的紧凑程度。第  $k$  个样本  $X_k$  越接近模糊类中心，最大隶属度  $\max_i(u_{ik})$  就越接近 1。因此，对每一个样本  $X_k$  来说，模糊集  $\max_i(u_{ik})$  被认为是一个好的分类指标，这个值越大，代表同一类的紧凑程度就越好。另一方面，式 (1) 的第二部分体现了类与类之间的分离程度。这里，用 2 个模糊集的交集来评价类  $V_i$  和  $V_j$  的分离程度。事实上，如果  $X_k$  接近类中心  $V_i$ ， $\min(u_{ik}, u_{jk})$  接近于零，结果类  $V_i$  和  $V_j$  很明显被分开。另一方面，如果  $\min(u_{ik}, u_{jk})$  接近于  $1/C$ ， $X_k$  属于所有类的隶属度相等，此时划分的类最模糊。这个新的判别准则  $V_p$  既考虑了同一类的紧凑程度，又考虑了类与类之间的分离程度。这样相对于最大值  $V_p$  的聚类数  $C$  就是一个最佳聚类数。

结合 FCM 算法和新的判别准则  $V_p$ ，可以得到最佳类别  $C$ ，具体步骤如下：

Step 1 选取最大的聚类数  $C_{\max}$  ( $C_{\max} \leq n^{1/2}$ )，最大迭代步数  $T$ ，指数  $m$  一般取 2，精度  $\epsilon > 0$ ；

Step 2 选取  $C$  ( $C = 2, 3, \dots, C_{\max}$ )；初始化类中心位置  $V_0(v_{10}, v_{20}, \dots, v_{C0},)$ ；

Step 3 运用式 (2) 计算类中心，然后由  $V_i$  和式 (3)、式 (4)，计算新的软分类矩阵  $U = [u_{ij}]$ ；

Step 4 对于预先给定的小整数  $\epsilon$  (如  $10^{-2}$ ,  $10^{-3}$ )，如果  $\max_j [|u_{ij}^* - u_{ij}|] < \epsilon$ ，则  $u_{ij}^*$  及相应得到的  $v_i^*$  即是所求结果，否则再返回到 Step 3 其中  $\epsilon$  越小，结果越精确；

Step 5 通过式 (1) 计算判别准则  $V_p$  的值；如果  $C < C_{\max}$ ，返回 Step 2。否则，终止程序，最佳的聚类数  $C = C_m$ ，这里  $C_m$  满足下面条件：

$$V_p(U, C_m) = \max \{V_p(U, C)\},$$

$$(C = 2, 3, \dots, C_{\max}) \quad (5)$$

有效聚类后，规则数和类中心  $V_i = (v_{i1}, v_{i2}, \dots, v_{is}, v_{i(s+1)})$  都得到了，这里  $i = 1, 2, \dots, C_{\text{opt}}$

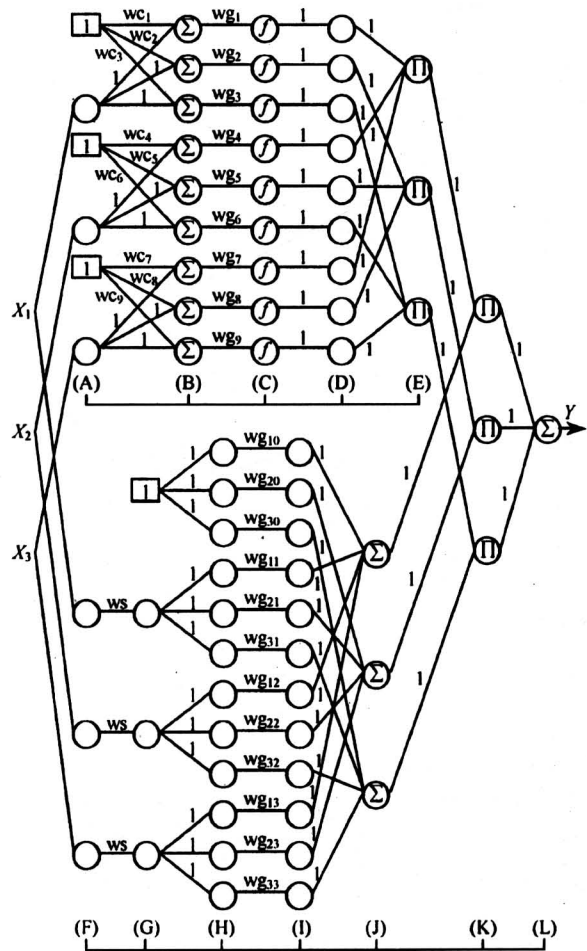


图 1 模糊模型的结构图

Fig.1 The structure of fuzzy system

### 3 相似分析法介绍

参数调整后，精简的模糊系统还没有最后确定，得到的模糊系统可能会由于隶属函数的高度重合而表现为冗余性。为了得到高效率 and 透明的模糊系统，就必须剔除冗余性，使得模糊系统尽可能的简单。一些相似分析法已经被提出用来简化模糊系统<sup>[6, 7, 10]</sup>。笔者介绍了一种新的基于梯形的相似

测度方法，可以应用到任何对称的模糊集。假设  $A$  和  $B$  是 2 个模糊集，模糊集  $A$  和  $B$  的相似度定义为

$$S_{AB} = |A \cap B| / |A \cup B| = |A \cap B| / (|A| + |B| - |A \cap B|) \quad (6)$$

其中  $|*|$  代表模糊集的大小。很明显，计算 2 个模糊集的相似度就是计算 2 个模糊集的交集和并集的比。对高斯和钟型函数来说，因为方程的非线性形状，所以计算交集的大小有困难。为了使得计算式 (6) 变得可行，Chao 等<sup>[6]</sup>和 Lin 和 Lee<sup>[11]</sup>提出了相似分析方法，采用三角函数来计算 2 个高斯函数的相似性。然而，用三角函数很难产生高斯和钟型函数，而用梯形能够很好的模拟径向基函数（包括高斯和钟型函数）如图 2 所示。这里  $a, b$  和  $d, e$  代表梯形的边界点， $c$  代表中心点。三角隶属函数能够被看作是梯形隶属函数的一个特殊的例子，当顶端宽度  $w_t=0$  时，也就是  $d=e=c$  时。因此，就能够用梯形函数计算 2 个对称隶属函数的相似度，对称的隶属函数包括三角函数、高斯函数和钟型函数。问题是怎样得到梯形函数使得能够模拟任何径向基函数。

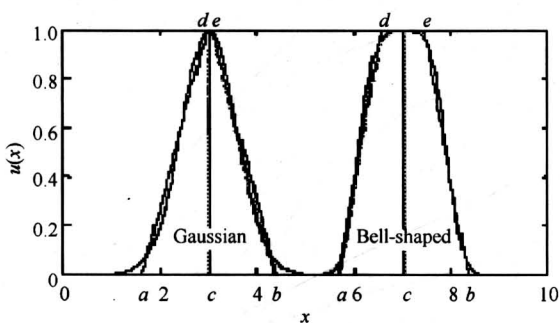


图 2 模拟隶属函数

Fig.2 Membership function approximation

### 3.1 由径向基隶属函数来生成梯形隶属函数

为了确定梯形函数  $T(x) = \max \{ \min \{ (x - a)/(d - a), 1, (b - x)/(b - e) \}, 0 \}$ ，能够模拟高斯函数  $G(x) = \exp \{ - [(x - c)/\sigma]^2 \}$ 。模糊集合的  $\alpha$  割集定义如下：

模糊集  $A$  的  $\alpha$  割集，用  $A_\alpha$  来表示，由论域  $X$  上的所有元素  $x$  所组成的水平集。这个水平集满足隶属度大于等于  $\alpha$ ，如下

$$A_\alpha = \{ x \in X \mid \mu(x) \geq \alpha \} \quad (7)$$

这里  $\alpha$  的取值范围是  $0 < \alpha \leq 1$ 。假设一个用高斯函数  $G(x)$  来代表模糊集  $A$ ，能够用梯形函数  $T(x)$

模拟，如图 3 所示。这样，在高斯函数  $G(x)$  的基础上，为了确定梯形函数  $T(x)$  中的参数  $a, b, d, e$ ，介绍 2 个特殊的  $\alpha$  割集  $A_{\alpha_0}$  和  $A_{\alpha_1}$ ，这里  $\alpha_0=0.05, \alpha_1=0.95$  如图 3 所示。这样，参数  $a, b, d, e$  在  $A_{\alpha_0}$  和  $A_{\alpha_1}$  的基础上能够被确定，可以表示为

$$A_{\alpha_0} = [a, b], A_{\alpha_1} = [d, e].$$

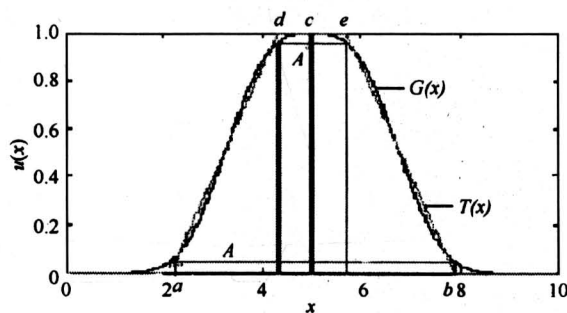


图 3 模糊集的  $\alpha$  割集

Fig.3  $\alpha$  cut of a fuzzy set

得到了边界点  $a, b, d, e$ ，就可以得到顶部宽度  $w_t$  和底部宽度  $w_b$

$$w_t = (e - d)/2, w_b = (b - a)/2.$$

梯形函数的中心为

$$c = (d + e)/2 \text{ 或 } c = (a + b)/2.$$

### 3.2 用相似分析法对模糊集进行分析

基于梯形隶属函数，两个模糊集的相似性的测度分 4 种情况来考虑。用  $A_1, A_2$  来代表模糊集，根据中心  $c_1, c_2$ ，边界点  $a_i, b_i, d_i, e_i (i=1, 2)$ ，其中  $c_i = (e_i + d_i)/2$ 。

假设在下面 4 种情况中  $c_2 > c_1$ ，计算交集面积与并集面积的比例。

第一种情况  $a_1 < a_2, b_1 > b_2, d_1 \leq d_2, e_2 \leq e_1$ 。在这种情况下，模糊集  $A_2$  包含于模糊集  $A_1$  即  $A_2 \subset A_1$ ，如图 4 所示。模糊集  $A_1$  和  $A_2$  的相似性为

$$S = |A_2| / |A_1| = w_2 / w_1,$$

$$\text{即 } S = (b_2 - a_2 + e_2 - d_2) / (b_1 - a_1 + e_1 - d_1) \quad (8)$$

其中  $w_1 = w_{b1} + w_{t1}, w_2 = w_{b2} + w_{t2}$ ，从式 (8) 可以看出  $A_1$  和  $A_2$  的相似度就是  $w_2$  占  $w_1$  的比率。

第二种情况  $|w_{b1} - w_{b2}| \leq c_2 - c_1 \leq w_{b1} + w_{b2}$ 。在这种情况下，有 2 种不同的重叠形式，如图 5 所

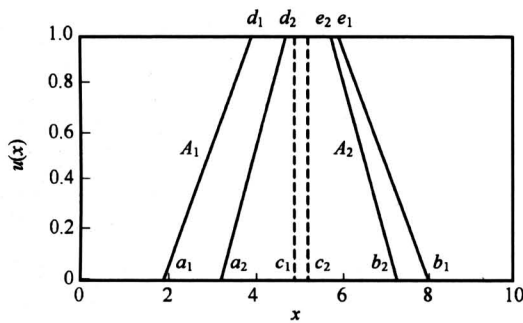


图 4 第一种情况下的两个模糊集合相似性  
Fig.4. Similarity of two fuzzy sets for case 1

示。图 5a 代表  $A_1$  和  $A_2$  的梯形顶端不重合。 $A_1$  和  $A_2$  的相似度为

$$S = h / [2(w_1 + w_2) / (b_1 - a_2) - h] \quad (9)$$

其中  $h = (b_1 - a_2) / (b_1 - e_1 + b_2 - e_2)$ 。图 5b 代表  $A_1$  和  $A_2$  的梯形顶端重叠， $A_1$  和  $A_2$  的相似度为

$$S = (e_1 - d_2 + b_1 - a_2) / (b_2 - a_1 + e_2 - d_1) \quad (10)$$

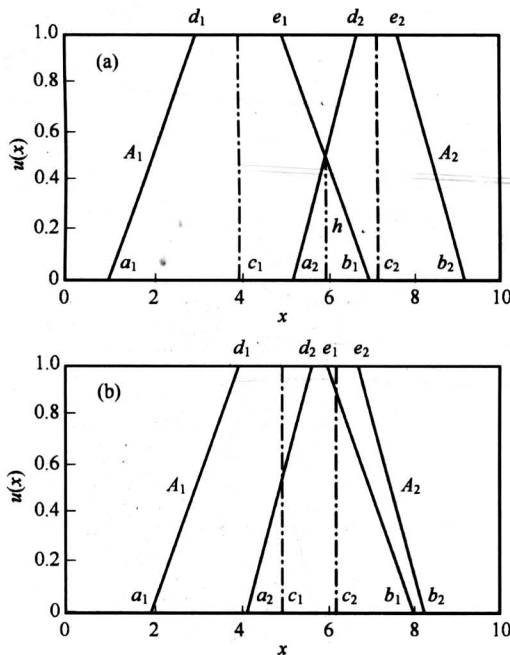


图 5 第二种情况下的两个模糊集合相似性  
Fig.5 Similarity of two fuzzy sets for case 2

第三种情况  $c_1 - c_2 \leq w_{b1} - w_{b2}$ 。在这种情况下应该考虑 2 种不同的重叠情况，如图 6 所示，很容易得到

$$h_1 = (b_1 - a_2) / (b_1 - e_1 + b_2 - e_2)。$$

如果  $w_{b1} > w_{b2}$ ,

$$h_2 = (b_1 - b_2) / ((b_1 - e_1) - (b_2 - e_2)),$$

其中  $l_1 = h_1(b_2 - e_2)$ ,  $l_2 = h_2(b_2 - e_2)$ ,  $l_3 = (b_1 - e_1) - (l_1 + l_2)$ ,  $h_3 = (h_1 + h_2)$ 。

$A_1$  和  $A_2$  的相似度为

$$S = (l_1 h_1 + l_2 h_2 + l_3 h_3) / (2(w_1 + w_2) - (l_1 h_1 + l_2 h_2 + l_3 h_3)) \quad (11)$$

如果重叠部分  $c_1 - c_2 \leq w_{b1} + w_{b2}$ :

对  $w_{b1} > w_{b2}$ ,

$$h = (b_1 - b_2) / ((b_1 - e_1) - (b_2 - e_2)),$$

$$l_1 = h(b_2 - e_2), l_2 = b_2 - e_1 - l_1,$$

$$l_3 = e_1 - d_2, l_4 = b_2 - e_2;$$

对  $w_{b1} \leq w_{b2}$ ,

$$h = (a_2 - a_1) / ((b_1 - e_1) - (b_2 - e_2)),$$

$$l_1 = h(b_1 - e_1), l_2 = d_2 - a_1 - l_1,$$

$$l_3 = e_2 - d_1, l_4 = b_1 - e_1。$$

这样，可以得到  $H = l_1 h + l_2(h + 1) + 2l_3 + l_4$  和相似性

$$S = H / (2(w_1 + w_2) - H) \quad (12)$$

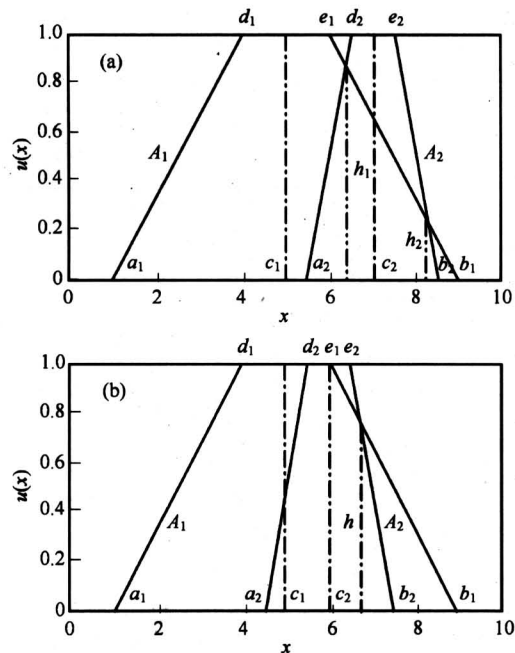


图 6 第三种情况下的两个模糊集合的相似性  
Fig.6 Similarity of two fuzzy sets for case 3

第四种情况  $b_1 \leq b_2$ ，在这种情况下， $A_1$  和  $A_2$  没有交集，如图 7 所示，即  $|A_1 \cap A_2| = 0$ ，则  $S(A_1, A_2) = 0$ 。

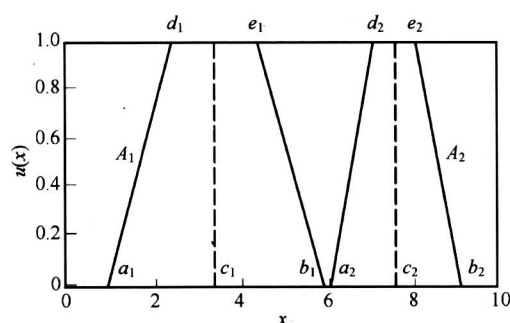


图7 第四种情况下的两个模糊集合相似性

Fig.7 Similarity of two fuzzy sets for case 4

#### 4 对模糊集合进行精简

相似模糊集的合并。针对对称的连续隶属函数，模糊集的相似性可以由式(8)至式(12)计算得出。

如果  $S(A_{ij}, A_{kj}) > \lambda_m$ , 阈值  $\lambda_m \in (0, 1)$ , 则模糊集  $A_{ij}$  和  $A_{kj}$  合并成一个新的模糊集  $A_{pj}$ , 新模糊集  $A_{pj}$  的中心和宽度分别是  $A_{ij}$  和  $A_{kj}$  中心和宽度的算术平均和, 即  $c_{pj} = (c_{ij} + c_{kj})/2$ ,  $w_{pj} = (w_{ij} + w_{kj})/2$ 。阈值  $\lambda_m$  对模糊模型的性能非常明显, 较小的  $\lambda_m$  导致更多的模糊集合并而产生一个更简单的

模糊模型, 但是通常精度不高, 通常  $\lambda_m \in [0.6, 0.85]$  之间。

#### 5 我国全要素生产力的模糊规则提取

设有  $n$  个投入要素, 则产出  $Y$  与投入要素之间的关系为<sup>[12]</sup>

$$Y = F(x_1, x_2, \dots, x_n; t) \quad (13)$$

其中  $t$  为时间。根据著名的 C-D 函数(柯布—道格拉斯)将式(13)写成

$$Y = F(x_1, x_2, \dots, x_n; t) = AK_t^\alpha L_t^\beta N_t^\gamma \quad (14)$$

式中  $K, L, N$  分别代表资本投入、劳动投入、土地(包括环境资源)投入,  $\alpha, \beta, \gamma$  是常数, 并假设  $\alpha + \beta + \gamma = 1$  (即产出的规模效益不变)。从式(14)可得到

$$A = YK_t^{-\alpha} L_t^{-\beta} N_t^{-\gamma} \quad (15)$$

式(15)表示一个地区的技术进步与该地区的产出、资本投入、劳动投入、土地(包括环境资源)投入的一定的依存关系。基于这种关系, 以我国 31 个省、市为样本, 指标为: GDP ( $10^4$  元/人)、固定资产 ( $10^8$  元)、人均耕地 (公顷/人)、人力资本 (以湖北为参照的相对值<sup>[13, 14]</sup>), 见表 1。

表 1 我国各地区的相关指标数据\*

Table 1 The index data of each area of China

地区	GDP $\times 10^4$ /元·人 <sup>-1</sup>	人均耕地 /公顷	固定资产 值 $\times 10^8$ /元	人力资本	地区	GDP $\times 10^4$ /元·人 <sup>-1</sup>	人均耕地 /公顷	固定资产 值 $\times 10^8$ /元	人力资本
北京	2.0576	0.02487	8227.30	824.05	湖北	0.7803	0.08284	8257.60	2448.01
天津	1.8328	0.04837	4046.97	553.85	湖南	0.6039	0.05993	6049.00	2292.17
河北	0.8326	0.10275	10867.73	2933.38	广东	1.3681	0.04204	21245.70	4140.11
山西	0.5440	0.14024	3382.22	1089.20	广西	0.4660	0.09206	3994.39	1389.13
内蒙古	0.6503	0.34501	2638.26	926.48	海南	0.6859	0.09574	1549.79	327.02
辽宁	1.2001	0.09954	8519.52	2575.60	重庆	0.5650	0.05112	2516.91	983.77
吉林	0.7553	0.20730	3529.01	1088.67	四川	0.5118	0.08770	8993.94	2410.79
黑龙江	0.9344	0.30892	5414.79	2067.14	贵州	0.2856	0.12907	2186.84	606.95
上海	3.0674	0.01952	12975.97	1188.82	云南	0.4840	0.14979	4358.04	1067.75
江苏	1.2933	0.06882	16511.65	3857.63	西藏	0.5275	0.13787	357.37	55.49
浙江	1.4629	0.04607	14318.02	2357.03	陕西	0.5040	0.14049	3771.24	990.81
安徽	0.5199	0.09437	5133.12	2013.47	甘肃	0.4165	0.19513	2200.20	526.47
福建	1.2365	0.04171	7059.91	1963.60	青海	0.5754	0.13155	831.98	125.26
江西	0.5198	0.07151	3136.84	1499.28	宁夏	0.5300	0.22536	846.54	144.29
山东	1.0439	0.08505	14948.75	4160.06	新疆	0.7918	0.21246	3794.42	623.60
河南	0.5903	0.08488	18354.74	3517.76					

\* 根据 1991 年至 2003 年的中国统计年鉴资料整理

5.1 按科技水平划分最佳类别数

选取指标 GDP、固定资产和人力资本，根据式 (1) 和表 1 的数据，计算结果如表 2 所示。

表 2 各个分类所对应的判别准则  $V_p$  的值

Table 2 The data of criterion  $V_p$  to each cluster

划分结果	C=2	C=3	C=4	C=5
判别准则 $V_p$ 的值	0.772 0	0.774 4	0.722 9	0.772 0

从表 2 可知  $c=3$  时，判别准则  $V_p$  的值取到最大值，所以最优分类结果为  $C_{opt}=3$ ，各类样本如表 3 所示。

表 3 各省市（自治区）科技水平分类表

Table 3 The corresponding area to each cluster by science and technology

类别	地区名称
第一类	北京 天津 上海
第二类	河北 辽宁 黑龙江 江苏 浙江 山东 广东 福建 河南 湖北 湖南 四川
第三类	吉林 内蒙 山西 江西 安徽 云南 重庆 广西 海南 贵州 陕西 西藏 甘肃 青海 宁夏 新疆

5.2 运用简化的模糊系统提取模糊规则

根据式 (12)，选取指标 GDP、人力资本、固定资产和人均耕地面积，再通过模糊系统和简化方法对我国全要素生产力进行模糊规则的提取。我国全要素生产力模糊规则见表 4。

表 4 我国全要素生产力模糊规则

Table 4 The fuzzy rule of the productivity factor of China

规则	结论
Rule 1	if X belongs to 第一类 then $y_1 = 0.826\ 985 + 2.730\ 167x_1 + 1.140\ 115x_2 + 0.598\ 281x_3$
Rule 2	if X belongs to 第二类 then $y_1 = 0.008\ 862 + 0.168\ 800x_1 + 3.276\ 196x_2 + 0.000\ 206x_3$
Rule 3	if X belongs to 第三类 then $y_1 = 0.116\ 152 + 4.139\ 537x_1 + 2.657\ 706x_2 + 0.049\ 274x_3$

6 结论

用笔者介绍的算法可以得到既简单又具解释性的模糊系统，通过实例仿真可知，在第一类中也就

是经济发达地区，人力资本和固定资产对产出有很大的作用；在第二类中也就是东部沿海和中部地区，固定资产对产出有很大的作用；在第三类中也就是西部地区，人力资本和固定资产对产出有很大的作用。

参考文献

[1] M Delgado, A F Gomez-Skarmeta, F Martin. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling [J]. IEEE Trans Fuzzy Systems, 1997, 5 (2): 223~233

[2] Y Fukuyama, M Sugeno. A new method of choosing the number of clusters for the fuzzy c-means method [A]. Proc 5th Fuzzy Syst Symp [C]. Japan, 1989. 247~250

[3] J R Jang. ANFIS: adaptive network-based fuzzy inference system [J]. IEEE Trans Syst Man Cybernet, 1993, 23 (4): 665~684

[4] V de Oliveira. Semantic constraints for membership function optimization [J]. IEEE Trans SMC-A, 1999, 29 (1): 128~138

[5] M Setnes, R Babuska, H Verbruggen. Rule-based modeling: precision and transparency [J]. IEEE Trans SMC-C, 1998, 28(1): 165~170

[6] C T Chao, Y J Chen, C C Teng. Simplification of fuzzy neural systems using similarity analysis [J]. IEEE SMC-B, 1996, 26(2): 344~354

[7] Y Jin. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement [J]. IEEE Trans Fuzzy Systems, 2000, 8 (2): 212~221

[8] J C Bezdek. Cluster validity with fuzzy sets [J]. J Cybernet, 1974, 3 (3): 58~72

[9] J C Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms [M]. Plenum, New York, 1981

[10] Min-You Chen, D A Linkens. Rule-base self-generation and simplification for data-driven fuzzy models [J]. Fuzzy Sets and Systems, 2004, 142: 243~265

[11] C T Lin, C S G Lee. Reinforcement structure/parameter learning for neural-network based fuzzy logic control systems [J]. IEEE Trans Fuzzy Systems, 1994, 2 (1): 46~63

[12] 诸克军, 王小刚, 匡奕军. 基于模糊系统的教育经济贡献率的预测和评价 [J]. 系统工程理论方法应用, 2002, 11(2): 169~172

- [13] Mulligan C, Sala-i-Martin X. A labor-income-based measure of the value of human capital: an application to the states of the United States [J]. *Japan and the World Economy*, 1997, 9 (2): 159~191
- [14] Byeongju Jeong. Measurement of human capital input across countries: a method based on the laborer's income [J]. *Journal of Development Economics*, 2002, 67: 333~349

## Rule-base Self-extraction and Simplification for Fuzzy Systems

Guo Haixiang<sup>1</sup>, Liu Tao<sup>2</sup>, Zhu Kejun<sup>1</sup>

(1. *School of management, Chinese University of Geosciences, Wuhan 430074, China;*

*2. Xuchang Vocational Technical college, Xuchang, Henan 461000, China)*

**[Abstract]** In this paper, a fuzzy model algorithm for a rule-base self-extraction and simplification is introduced. The method consists of three steps: The first step is to classify the out-in space by constructing a fuzzy partition validity index, then the optimal number of clusters and, hence, the optimal number of rules are obtained; The second step is to construct the initial fuzzy system based on the optimal number of rules and neural networks; The third step is to get the function of computing similarity of fuzzy sets by fuzzy similarity analysis method. The similar fuzzy sets are merged to create a common fuzzy set in rule base based on threshold value. Thus a fuzzy system with interpretability and simplicity is obtained. At last, the fuzzy rules of productivity factor of China is extracted by the fuzzy system.

**[Key words]** similarity measures; fuzzy model; fuzzy rules ; fuzzy sets

## 《中国工程科学》2004年第6卷第11期要目预告

- |   |  |
|---|--|
| 生态县 生态市 生态省建设规划编制导则<br>..... 卞有生等         | 改进 LK 算法求解 ..... 田 伟等                    |
| 应该认真查找“豆腐渣”工程设计上的错误<br>..... 张肇伸          | 分布式能量系统浅析 ..... 宋之平                      |
| 机械产品的信息化——面向机械装备<br>的信息技术 ..... 屈梁生等      | 基于V/F的智能多路数据采集系统及<br>串口通信 ..... 贾贵玺等     |
| 基于时间-有色 Petri 网的电梯系统的<br>建模与分析 ..... 宗 群等 | 基于半实物的 S 频段卫星测控仿真系<br>统平台设计 ..... 杨永安等   |
| 提高光流估计性能的渐进性的高斯多<br>维预滤波方法的研究 ..... 付 昀等  | 基于可视化的事故树分析系统研究与<br>开发 ..... 施式亮等        |
| 模糊中心聚类的模式识别学习方法<br>..... 曾黄麟等             | 大庆油田台肇地区低渗透储层裂缝及<br>其开发对策研究 ..... 曾联波等   |
| 时域有限差分算法与遗传算法在平面<br>螺旋电感设计中的应用 ..... 王宏建  | 加压浸出生产金红石及熔盐电解制备<br>海绵钛新工艺探索 ..... 王海北等  |
| 一类高维随机系统的小波分析 ..... 夏学文                   | 人与自然的和谐是人类社会进步的需要<br>..... 苑 菲           |
| 考虑非对称情形的一类拣选问题的                           | 参数自寻优模糊控制器在中央空调温<br>度控制系统中的应用 ..... 吴爱国等 |