

复杂模糊分类系统的协同进化设计方法

张永, 吴晓蓓, 向峥嵘, 胡维礼

(南京理工大学自动化学院, 江苏南京 210094)

摘要: 提出一种基于协同进化算法的复杂模糊分类系统的设计方法. 该方法由以下3步组成: 1) 利用Simba算法进行特征变量选择; 2) 采用模糊聚类算法辨识初始的模糊模型; 3) 利用协同进化算法对所获得的初始模糊模型进行结构和参数的优化. 协同进化算法由三类种群组成: 规则数种群, 规则前件种群和隶属函数种群; 其适应度函数同时考虑模型的精确性和解释性, 采用三类种群合作计算的策略. 利用该方法对多个典型问题进行分类, 仿真结果验证了方法的有效性.

关键词: 模糊分类系统; 特征变量选择; 协同进化算法; 解释性
中图分类号: TP273 **文献标识码:** A

Design of complex fuzzy classification system based on cooperative coevolutionary algorithm

ZHANG Yong, WU Xiao-bei, XIANG Zheng-rong, HU Wei-li

(College of Automation, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China)

Abstract: A novel approach to construct complex fuzzy classification system based on cooperative coevolutionary(Co-evolution) algorithm is proposed in this paper. The approach is composed of three phases: 1) feature selection is accomplished by the Simba algorithm; 2) the initial fuzzy system is identified using the fuzzy clustering algorithm; 3) the structure and parameters of the fuzzy system are optimized by the Co-evolution algorithm. The Co-evolution algorithm owns three species including the number of fuzzy rules species, the premise structure species and the parameters species. Considering both precision and interpretability, the fitness function is calculated on the cooperation of individuals from the three species. The proposed approach had been applied to several benchmark problems, the results showed its validity.

Key words: fuzzy classification systems; feature selection; fuzzy clustering; co-evolution algorithm; interpretability

1 引言(Introduction)

模糊集合理论自1965年由Zadeh提出后, 得到了迅速的发展, 并在仿真、分类、数据挖掘、模式识别、预测及控制等方面得到广泛的应用. 早期的模糊建模主要针对简单系统, 采用总结专家经验的方式进行, 因此得到的模糊模型容易被人们所理解. 对于复杂系统, 由于难以获得完备的专家知识, 而数据相对容易获得, 因此近年来基于数据的精确性和解释性模糊建模成为研究的热点^[1~8].

目前关于模糊模型解释性的研究主要有3条途径: 一是以精确性为目标, 采用模糊聚类/决策树等方法进行模糊建模, 然后对得到的模型进行简化, 提高模糊模型的解释性^[1,2]; 二是采用模糊聚类/栅格等方法构造初始的模糊模型, 将初始的模糊模型编

码为染色体的形式, 以解释性的主要因素(如模糊规则数目)和精确性为目标, 采用多目标遗传算法优化模糊模型^[3~7]; 三是将模糊系统等价为模糊神经网络, 采用神经网络学习算法, 实现模糊模型精确性与解释性的折中^[8].

遗传算法在众多方法中, 因其易于理解, 鲁棒性强, 具有在大范围内搜索全局最优解的能力而得到广泛的应用. 文献[9]利用遗传算法同时优化模糊模型结构和隶属函数的参数, 提高了模型的精确性. 文献[5]利用分层遗传算法优化模糊模型结构和隶属函数参数, 但只涉及了低维的问题. 对于高维复杂系统, 利用遗传算法同时优化模糊模型结构和隶属函数参数的文献较少^[7]. 文献[7]提出了一种智能遗传算法, 将变量选择与规则选择、规则参数等编码

为染色体, 以正确分类的样本数目、变量与规则数目为目标函数, 构造了解释性和精确性的模糊分类系统. 但该方法对模糊分类系统的所有相关参数采用二进制编码, 使得染色体编码过长, 搜索空间大; 同时其语义词与隶属函数之间不是严格的一一对应关系, 存在多个隶属函数同属一个语义词的可能.

遗传算法在处理高维、复杂问题时, 必然面对系统中大量可调参数的优化, 目前多使用匹茨堡型编码方式, 该编码方式往往使得染色体编码过长, 遗传算法的性能大大降低. 为了解决遗传算法的上述问题, 文献[10]提出了协同进化的方法, 该方法把复杂的问题分解为较为简单的子问题, 每个子问题相互适应, 分别进化, 合作产生问题的完整解. 文献[11]发展并提出协同进化算法的一般模式.

本文的主要工作是: 利用Simba算法^[12]对高维分类系统进行预处理, 实现了系统的降维; 根据问题的分类数预先给定一个冗余的初始聚类数, 采用模糊聚类算法构造初始模糊模型; 利用协同进化算法对所获得的初始的模糊模型进行结构和参数的同时优化, 保证了优化后模型的精确性; 在协同进化算法中利用基于相似性的模型简化方法对模型进行约简, 保证了优化后模型的解释性. 利用该方法研究Wisconsin Breast Cancer, Wine, Iris, Glass, Ionosphere和Pima等典型分类问题, 仿真结果验证了方法的有效性.

2 预备知识(Preliminaries)

2.1 模糊分类系统(Fuzzy classification system)

考虑 n 维 Nc 类 N 样本的分类问题, 其中 $x \in X \subseteq \mathbb{R}^n$, $x = (x_1, x_2, \dots, x_n)$ 为特征变量, $\{g_1, g_2, \dots, g_M\}$ 为输出类别, 则模糊分类系统的典型规则形式如下:

$$\begin{aligned} R_i: & \text{if } x_1 \text{ is } \mu_{i1} \text{ and } x_2 \text{ is } \mu_{i2} \text{ and } \dots \text{ and } x_n \\ & \text{is } \mu_{in}, \text{ then the pattern } (x_1, \dots, x_n) \\ & \text{belongs to } g_i. \end{aligned} \quad (1)$$

其中 $\mu_{i1}, \dots, \mu_{in}$ 为定义在输入论域中的隶属函数, 可以取三角形、高斯型等. 本文采用高斯型函数:

$$\mu_{ij}(x_j) = \exp\left(-\frac{(x_j - v_{ij})^2}{2\sigma_{ij}^2}\right). \quad (2)$$

其中 v_{ij}, σ_{ij} 分别代表函数的中心和方差.

对未知样本 x_k , 模糊分类系统的输出采用“胜者为王”的策略, 即系统的输出为具备最大激励强度的规则所对应的类别输出:

$$x_k \in g_{i^*}, \quad i^* = \arg \max(\beta_i(x_k)). \quad (3)$$

其中 β_i 为第 i 条规则的激励强度:

$$\beta_i(x_k) = \prod_{j=1}^n \mu_{ij}(x_{jk}). \quad (4)$$

2.2 模糊模型的解释性(Interpretability issues in fuzzy model)

与精确性等可以量化的性能指标不同, 模糊模型的解释性, 目前尚无明确的标准和定义, 但是一般认为, 模糊系统的解释性, 与模型结构、特征变量和模糊规则数目、隶属函数特性等密切相关, 现将主要因素陈述如下^[13,14]:

1) 特征变量数目: 人们很难通过高维模糊系统来分析系统行为, 因此为了提高模型的解释性需进行特征变量的选择, 使模糊系统采用尽可能少的特征变量.

2) 模糊规则数目: 模糊系统的规则数目越多, 其解释性越低. 经验认为, 可解释的模糊系统, 其规则数目不超过10个, 这是由人在理解、推理时的思维能力所决定的.

3) 隶属函数的特性: 隶属函数必须是凸的, 隶属函数划分必须是完备的, 即对于任何的特征变量, 在其论域内的任何值, 至少有一个隶属函数相对应, 在形式上表现为隶属函数之间存在位置的交叉. 隶属函数划分必须是可区分的, 即对于同一变量, 隶属函数之间存在明显的位置区别, 以便赋予一定的语义项.

4) 模糊规则库的完整性、一致性和精简性: 模糊规则要完整覆盖特征论域, 对每一有效的特征变量组合, 至少有一条模糊规则被激励, 即完整性. 模糊规则之间必须相容而不能有任何两条规则相互矛盾, 即一致性. 在规则数目尽可能小的前提下, 不能包含冗余规则, 如某规则的前件是另一规则的子集等, 即精简性.

3 模糊分类系统的初始化(Initial fuzzy classification system)

3.1 特征变量的选择(Feature selection)

为解决高维模糊分类系统中的“维数灾难”问题, 通常采用特征变量选择法实现高维系统的降维. 本文采用Simba算法进行变量选择, 该算法利用边界法计算每个特征变量的权值, 权值的大小对应于特征变量区分类别的能力. 预先设定某一阈值, 所有高于该阈值的权值所对应的变量为所要选择的特征变量.

Simba算法的目标函数为

$$e(w) = \sum_{x \in P} \theta_P^w(x). \quad (5)$$

其中 w 为特征变量集合 x 的加权矢量, P 为训练数据集, $\theta_P^w(x)$ 为 x 的边界:

$$\theta_P^w(x) = \frac{1}{2} (\|x - \text{nearmiss}(x)\|_w - \|x - \text{nearhit}(x)\|_w). \quad (6)$$

其中 $\text{nearmiss}(x)$ 和 $\text{nearhit}(x)$ 分别为 x 到 P 中最近的不同类别点和相同类别点的欧氏距离, $\|z\|_w = \sqrt{\sum_i w_i^2 z_i^2}$, 且 $\max(w_i^2) = 1$.

算法初始化 $w = (1, 1, \dots, 1)$, 利用梯度下降法计算目标函数 $e(w)$ 的最大值所对应的 w , 其中训练数据集 P 梯度计算如下:

$$\begin{aligned} (\nabla e(w))_i &= \frac{\partial e(w)}{\partial w_i} = \\ &= \frac{1}{2} \sum_{x \in P} \left(\frac{(x_i - \text{nearmiss}(x)_i)^2}{\|x - \text{nearmiss}(x)\|_w} - \frac{(x_i - \text{nearhit}(x)_i)^2}{\|x - \text{nearhit}(x)\|_w} \right) w_i. \end{aligned} \quad (7)$$

3.2 初始的模糊模型的构建(Design of the initial fuzzy model)

为保证协同进化算法从较合理的初始模型开始学习, 本文采用模糊聚类算法构建初始的模糊模型, 从而有效地降低算法的搜索空间.

模糊聚类是基于数据模糊建模最常用的方法之一. 本文采用Gustafson-Kessel (GK)聚类算法离线辨识模糊系统的前件参数. 关于具体算法细节请参考前期工作^[2].

为确定规则的后件类的标号值, 首先定义如下函数:

$$M_{ij} = \frac{\sum_{k=1}^N \mu_{ik} f_j(k)}{\sum_{k=1}^N f_j(k)}. \quad (8)$$

其中 μ_{ik} 是第 k 个数据相对于第 i 个聚类中心的隶属度, 且满足以下条件:

$$\mu_{ik} \in [0, 1]; \sum_{i=1}^C \mu_{ik} = 1; 0 < \sum_{k=1}^N \mu_{ik} < N. \quad (9)$$

$$f_j(k) = \begin{cases} 1, & \text{如果 } x_k \text{ 属于类 } j, \\ 0, & \text{如果 } x_k \text{ 不属于类 } j. \end{cases} \quad (10)$$

对于第 i 条规则, 后件类标值 g_{i^*} 可用下式计算:

$$i^* = \arg(\max(M_{ij})), \quad j = 1, 2, \dots, N_c. \quad (11)$$

4 基于相似性的模糊模型简化(Similarity-driven rule base simplification)

协同进化算法中随机产生的染色体反编码为模糊模型后, 首先利用基于相似性的模糊模型简化方法自动约简模糊系统, 有效地保证模糊模型的解

性, 然后计算模糊模型的精确性指标和解释性指标. 基于相似性的模糊模型简化分为: 模糊集合的相似性分析与融合, 模糊规则的相似性分析与融合两部分内容.

4.1 模糊集合的相似性分析与融合(Similar fuzzy sets merging)

初始的模糊模型和协同进化算法中随机产生的模糊模型, 其模糊集合可能存在冗余, 表现为模糊集合间存在过度的交叉或重叠, 从而难以赋予相应的语义值, 降低了解释性, 因此需要对每个变量的隶属函数进行相似性分析和融合, 从而提高模糊模型的解释性.

对模糊集合 A 和 B , 定义其相似性测度如下^[1]:

$$S(A, B) = \frac{\sum_{k=1}^N [\mu_A(x_k) \wedge \mu_B(x_k)]}{\sum_{k=1}^N [\mu_A(x_k) \vee \mu_B(x_k)]}. \quad (12)$$

其中 \wedge 和 \vee 分别为最小最大算子. S 为定义在 $[0, 1]$ 间的相似性测度, $S = 1$ 表示两个集合完全相等, 而 $S = 0$ 意味着两个集合没有交叉或重叠. 关于具体模糊集合融合过程请参考文献^[2].

4.2 模糊规则的相似性分析与融合(Similar fuzzy rules merging)

在协同进化算法中, 由于模糊集合的相似性分析与融合, 或由于遗传操作的随机性, 可能使得两条模糊规则前件相似或相同. 此时可以通过规则的相似度^[15]来计算规则之间的相似性, 考虑如下两条规则:

$$\begin{aligned} R^i: & \text{ If } x_1 \text{ is } \mu_{i1}(x_1) \text{ and } x_2 \text{ is } \mu_{i2}(x_2) \text{ and } \dots \\ & x_n \text{ is } \mu_{in}(x_n) \\ & \text{ then the pattern } (x_1, \dots, x_n) \\ & \text{ belongs to } g_i \end{aligned}$$

$$\begin{aligned} R^j: & \text{ If } x_1 \text{ is } \mu_{j1}(x_1) \text{ and } x_2 \text{ is } \mu_{j2}(x_2) \text{ and } \dots \\ & x_n \text{ is } \mu_{jn}(x_n) \\ & \text{ then the pattern } (x_1, \dots, x_n) \\ & \text{ belongs to } g_j \end{aligned}$$

则两条规则前件的相似度 S_R 定义如下:

$$S_R(R^i, R^j) = \min_{k=1}^n S(\mu_{ik}, \mu_{jk}). \quad (13)$$

其中 $S(\cdot)$ 的计算如式(12)所示. 当该值为1时, 表明两条规则的前件相同; 当该值较小时, 表明两条规则的前件有很大的不同; 对于任意一组模糊规则, 若其相似度大于给定阈值 $[0.9 - 1]$, 可将该组规则随机删除一条, 实现规则的约简.

模糊分类系统的规则后件为类标, 故不考虑规则

后件的相似性。

5 协同进化算法(Co-evolution algorithm)

应用协同进化算法, 除常规进化算法遇到的问题外, 还需解决以下问题: 1) 求解问题适当的分解; 2) 子种群染色体之间的数据约束关系以及相互合作关系; 3) 多种群中适应度函数的计算策略。

5.1 模型的分解和各种群染色体编码(Decomposition of fuzzy model and coding of individuals from each species)

式(1)所表示的模糊分类系统, 在保证合理解决系统的精确性和解释性的前提下, 分解的种群越少, 则协同进化算法的计算越简单, 搜索空间和时间越少。本文将模糊模型分解为以下3类种群, 种群A描述模糊模型被激励的规则数; 种群B描述模型的规则前件; 种群C描述模型隶属函数的参数; 3类种群分别采用不同编码方式。

5.1.1 模型的规则数种群A(Number of fuzzy rules)

模型的规则数种群主要用来优选模型的规则, 采用二进制编码方式。其编码形式如下:

$$\begin{matrix} R_1 & \cdots & R_i & \cdots & R_c \end{matrix}$$

其中 R_i 取值为 $\{1, 0\}$, 当 R_i 取值为1时, 表示第 i 条规则被激励; 当 R_i 取值为0时, 表示第 i 条规则被屏蔽。根据问题的分类数预先给定一个冗余的初始聚类数 c , 即初始的模糊模型规则数, 本文取经验值 $c = 3N_c$ 。

给定种群大小为 L , 第1条染色体取值为全“1”, 表示选择全部规则, 其余的 $L - 1$ 条染色体, 在二进制空间内随机均匀生成, 从而形成初始规则数种群。

5.1.2 模型的规则前件种群B(Antecedent of fuzzy rules)

模型规则前件种群采用二进制编码。其中二进制染色体反编码后的表现型为:

$$\begin{matrix} \{ \text{规则1} \} & \{ \text{规则}i \} & \{ \text{规则}c \} \\ \beta_{11} \cdots \beta_{1n} \cdots \beta_{ij} \cdots \beta_{c1} \cdots \beta_{cn} \end{matrix}$$

其中 β_{ij} 取值为 $\{0, 1, 2, \dots, c\}$, 其中 c 为模糊规则数, 当 $\beta_{ij} = 0$ 时, 表示第 i 条规则中的第 j 个变量为变量无关项(don't care)^[6], 对应一条不完整规则, 当 $\beta_{ij} \neq 0$ 时, 表示第 i 条规则中的第 j 个变量所对应的第 β_{ij} 个模糊集合被激励。

给定种群大小为 L , 第1条染色体编码对应于模糊聚类产生的初始模型的前件, 其余的 $L - 1$ 个染色体, 在二进制空间内随机均匀生成。

5.1.3 模型的隶属函数参数种群C(All parameters of the membership functions)

模糊模型隶属函数的参数种群, 采用实数编码方式。实数编码染色体比二进制编码的染色体长度短,

编码方式简洁自然, 减轻了遗传算法的计算负担, 提高了运算效率, 能够更好地保持种群多样性。

模糊聚类产生的初始模型的前件参数, 可直接编码产生第1条染色体, 待编码的参数为隶属函数的中心 v_{ij} 和方差 σ_{ij} , 因此每条染色体共有 $2 * c * n$ 个实数, 此时的 n 为经过特征变量选择后的变量的维数, 第1条染色体其编码为

$$H_1 = (v_{11}, \dots, v_{cn}, \sigma_{11}, \dots, \sigma_{cn}). \quad (14)$$

给定种群大小为 L , 染色体表示为 $H_p (p = 1, 2, \dots, L)$, 给定搜索空间 $[H^{\min}, H^{\max}]$:

$$H^{\min} = (v_{11}^{\min}, \dots, v_{cn}^{\min}, \sigma_{11}^{\min}, \dots, \sigma_{cn}^{\min}). \quad (15)$$

$$H^{\max} = (v_{11}^{\max}, \dots, v_{cn}^{\max}, \sigma_{11}^{\max}, \dots, \sigma_{cn}^{\max}). \quad (16)$$

其中 $v_{ij}^{\max}, v_{ij}^{\min}, \sigma_{ij}^{\min}, \sigma_{ij}^{\max}$ 为对应隶属函数的中心和方差的最大最小约束值。其余的 $L - 1$ 个染色体, 以染色体 H_1 为中心, 在搜索空间内随机均匀生成, 从而形成初始参数种群。

5.2 子种群染色体之间的合作关系(Cooperation of individuals from each species)

在协同进化算法中, 3类种群的染色体经过合并, 然后反编码为模糊规则的前件。本节以5.1节3类种群为例, 说明子种群染色体间的合作关系。

假设初始规则数 $c = 6$ (类别数为2), 经特征变量选择的特征变量维数 $n=5$, 在3类种群分别任取一条染色体, 合并后为

1	0	1	1	0	1
10341	02306	10341	01235	16545	11605
$v_{11} \cdots v_{15} \cdots v_{61} \cdots v_{65} \cdots \sigma_{11} \cdots \sigma_{15} \cdots \sigma_{61} \cdots \sigma_{65}$					

第1行是模型规则数染色体编码; 第2行(灰色)是模型规则前件染色体二进制编码的表现型; 第3行是隶属函数参数编码, 则上述编码表示的模型规则前件为:

$$\begin{aligned} R_1 : & \text{if } x_1 \text{ is } \mu_{11}(v_{11}, \sigma_{11}) \text{ and } x_3 \text{ is } \mu_{13}(v_{33}, \sigma_{33}) \\ & \text{and } x_4 \text{ is } \mu_{14}(v_{44}, \sigma_{44}) \text{ and } x_5 \text{ is } \mu_{15}(v_{15}, \sigma_{15}) \\ R_3 : & \text{if } x_1 \text{ is } \mu_{31}(v_{11}, \sigma_{11}) \text{ and } x_3 \text{ is } \mu_{33}(v_{33}, \sigma_{33}) \\ & \text{and } x_4 \text{ is } \mu_{34}(v_{44}, \sigma_{44}) \text{ and } x_5 \text{ is } \mu_{35}(v_{15}, \sigma_{15}) \\ R_4 : & \text{if } x_2 \text{ is } \mu_{42}(v_{12}, \sigma_{12}) \text{ and } x_3 \text{ is } \mu_{43}(v_{23}, \sigma_{23}) \\ & \text{and } x_4 \text{ is } \mu_{44}(v_{34}, \sigma_{34}) \text{ and } x_5 \text{ is } \mu_{45}(v_{55}, \sigma_{55}) \\ R_6 : & \text{if } x_1 \text{ is } \mu_{61}(v_{11}, \sigma_{11}) \text{ and } x_2 \text{ is } \mu_{62}(v_{12}, \sigma_{12}) \\ & \text{and } x_3 \text{ is } \mu_{63}(v_{63}, \sigma_{63}) \text{ and } x_5 \text{ is } \mu_{65}(v_{55}, \sigma_{55}) \end{aligned}$$

6条规则中 R_2, R_5 被屏蔽, 其他规则均被激励, 被激励的规则均为不完整规则, R_1, R_3 中的 x_2 , R_4 中 x_1 和 R_6 中 x_4 为变量无关项; R_1 和 R_3 形式上不

同,但其模糊集合参数完全相同,则在基于相似性的模糊规则融合过程中被随机删除1条,所以以上编码实际只表示含有 R_3 , R_4 和 R_6 的3条规则的模糊模型,相关的模糊集合数为9个,若此模糊模型的精确性指标能达到要求,则该3条染色体合并代表的是解释性较好,精确性较高的模糊模型。

通过上述步骤可以获得模糊规则的前件,其后件可以通过式(8)和式(11)计算得到,从而获得完整的模糊模型。

5.3 多种群合作策略(Cooperative strategy of individuals from each species)

5.3.1 适应度函数(Fitness function)

本文的目标是构建精确性和解释性较好折中的模糊模型,其中精确性指标是指模型对样本的错误划分数目,解释性指标是指模糊规则数目和模糊集合数目,表示为

$$f_1(P) = N_{\text{err}}, f_2(P) = R, f_3(P) = F_S. \quad (17)$$

其中 N_{err} 为模型对样本的错误划分数目, R 为模糊规则数目, F_S 为模糊集合数目, P 代表3类种群染色体合并所表示的模糊模型。3个指标函数均为越小越好,采用分量加权求和法,适应度函数为

$$\text{Fit} = \omega_1 * f_1(P) + \omega_2 * f_2(P) + \omega_3 * f_3(P). \quad (18)$$

加权因子 $\omega_1, \omega_2, \omega_3$ 为正实数,在计算适应度函数时预先根据经验设定。

5.3.2 个体适应度的多种群合作计算策略(Fitness evaluation of individuals)

目前多种群个体的合作策略,一般采用文献[11]的方法,利用分量加权求和法将多目标转换为单目标,选择最优个体和部分随机个体作为种群代表,某个种群的个体与其他种群的种群代表合作,生成多个模糊模型,然后计算每个模型的适应度值,将其中最优秀的适应度值作为该种群个体的适应度值。

为直观起见,本节以5.1节的三种群(种群A、种群B和种群C)为例,具体说明个体适应度的多种群合作计算策略,其中每个种群个体数目为40,种群代表数目为3。

如图1所示,种群A的第1条染色体表示最优个体,被选作种群代表,另外2个种群代表在其余39条染色体中随机选择。同理完成种群B, C的代表选择。

如图2所示,种群A的某个体,与种群B, C的代表合作构造模糊模型。种群B和种群C的代表不采用遍历组合法,而是采用位置匹配法,因此该个体构造了3个模糊模型,分别计算3个模糊模型对应的适应度函数值,将最小的适应度函数值作为该个体的适应度值。同理完成种群A其余个体和种群B, C中所有个体的适应度函数计算。

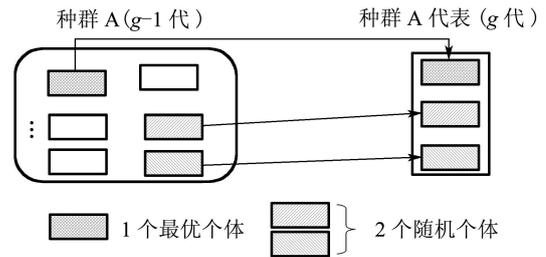


图1 种群A代表的产生

Fig. 1 Representatives of species A

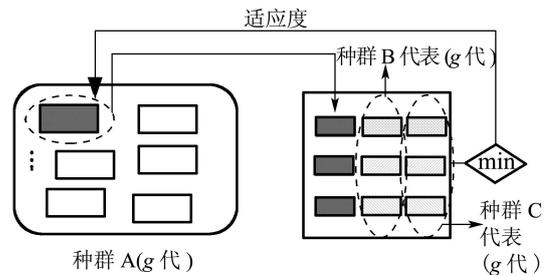


图2 种群A中个体适应度的计算

Fig. 2 Fitness evaluation of individual in species A

5.4 种群的遗传操作(Genetic operators of each species)

协同进化算法的遗传操作主要包括以下3种:选择、交叉和变异。3类种群都选用精英保留和随机遍历抽样法相结合的选择操作。其中种群A和种群B的交叉操作选用单点交叉,变异操作选用二进制变异;由于种群C采用实数编码,为保证种群个体的多样性和算法的有效性,对于交叉和变异操作均给出几种具体的实现方法。交叉操作有:离散交叉,算术交叉,启发式交叉;变异操作有:单点均匀变异,多点均匀变异;在实际的操作过程中,由算法随机选择。

5.5 模糊分类系统的设计流程图(The design flow chart of fuzzy classification system)

图3给出了模糊分类系统的协同进化设计流程图,其中采用Simba算法进行特征变量和模糊聚类构造初始模糊模型的前件只运行一次,而模糊模型后件类标的计算、基于相似性的模型简化、协同进化算法优化模型的结构和参数则经过反复迭代,直至满足中止条件。

6 仿真(Simulation)

为了检验协同进化算法构造精确性和解释性模糊分类系统的性能,本文从UCI机器学习数据库(<http://www.ics.uci.edu/~mllearn/MLrepository.html>)中选择了6类典型分类问题: 1) Wisconsin Breast Cancer(WBC), 2) Wine, 3) Iris, 4) Glass, 5) Ionosphere和6) Pima, 表1中给出了关于这6类问题的数据特性汇总。

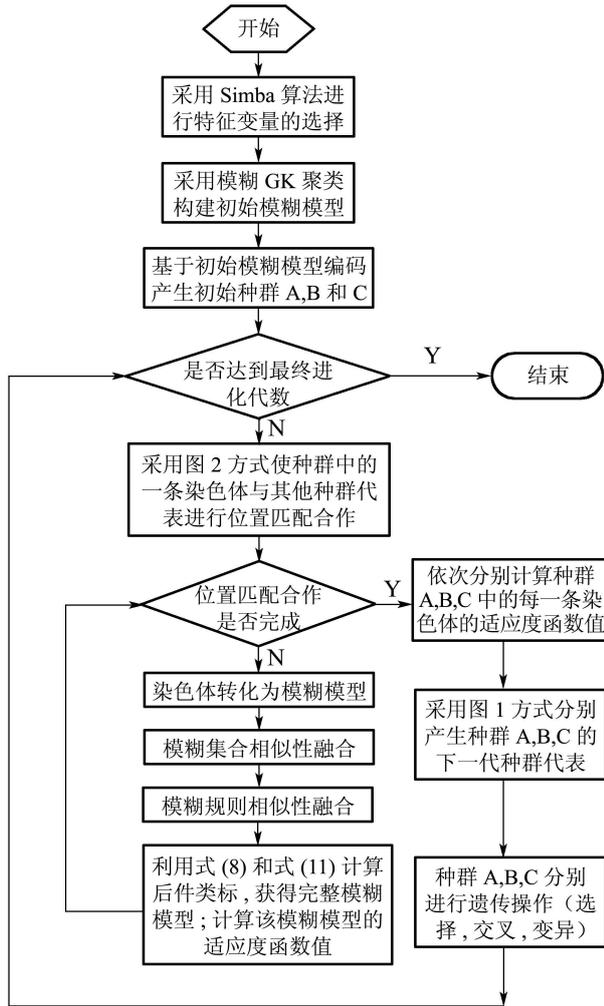


图 3 模糊分类系统的设计流程图

Fig. 3 Design flow chart of fuzzy classification system

表 1 分类问题的数据特性汇总表

Table 1 Characteristics of the six examples

分类问题	样本数	特征变量维数	分类数
WBC	683	9	2
Wine	178	13	3
Iris	150	4	3
Glass	214	9	6
Ionosphere	351	34	2
Pima	768	8	2

仿真程序在MATLAB7.0中运行完成, 算法各参数设置如表2所示. 除WBC分类问题外, 其他分类问题的数据首先进行归一化处理. 对其中的Wine分类问题给出详细的仿真说明, 并将仿真结果和其他文献进行比较; 其他的分类问题给出总结性仿真结果.

首先采用Simba算法进行特征变量选择, 13维特征变量所对应的权值如图4所示, 当阈值取0.3时, 所选择的特征变量为 x_1, x_7, x_{10}, x_{13} ; 采用模糊GK聚类构造初始模糊模型, 初始聚类数为9($N_c = 3$).

采用5次折叠交叉验证法, 结果为: 样本平均分类正确率为97.17%; 最差分类正确率为94.44%; 最好分类正确率为100%.其中验证样本分类正确率最好的模糊分类器规则库中有6条规则, 模糊集合的总数为10个, 图5给出该模糊分类系统的隶属函数分布图.

表 2 协同进化算法参数设置

Table 2 Parameter setups of the co-evolution algorithm

参数	值	参数	值
最大进化数	100	种群C精英率	0.025
种群初始个体数	40	种群代表数	3
种群A,B交叉率	1	权值 ω_1	1
种群A,B变异率	0.2	权值 ω_2	0.1
种群A精英率	0.05	权值 ω_3	0.01
种群B精英率	0.1	集合融合阈值	0.4
种群C交叉率	0.85	规则融合阈值	1
种群C变异率	0.1		

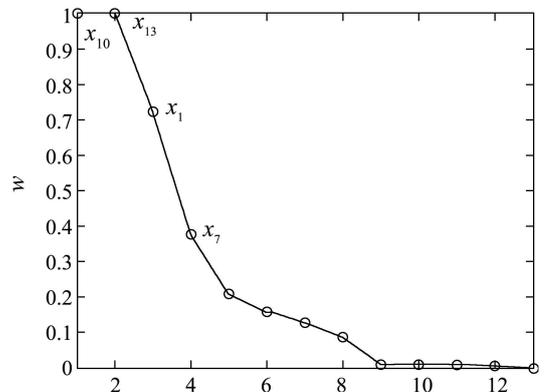


图 4 Wine 问题特征变量所对应的权值分布图

Fig. 4 Wine: weights of feature

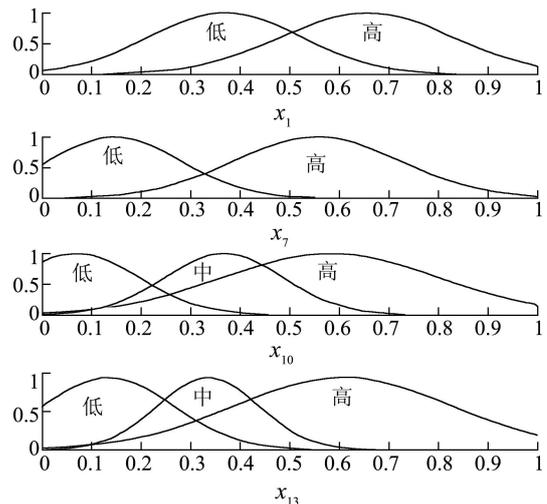


图 5 Wine 模糊分类系统隶属函数分布图

Fig. 5 Wine: membership functions

表3中给出本文方法同其它文献建模方法的性能比较. 文献[6,16]虽然有较高的分类率,但其规则数为60,模型的解释性较差;文献[3]中的结果为最好分类结果,而不是平均值;文献[4]利用决策树初始化模糊模型,然后利用常规遗传算法对初始模糊模型进行优化. 通过以上比较表明本文方法能对分类系统的特征变量进行选择,同时利用较少的模糊分类规则和模糊集合数达到较高的分类率.

本文同时对以下几个典型的分类问题: Iris, Wisconsin Breast Cancer, Glass, Ionosphere和Pima进行了研究,表4中汇总了利用本文方法对这几个典型分类问题的仿真结果.

表3 Wine分类问题不同建模方法的性能比较
Table 3 Wine: comparison of different results

	规则数	集合数	正确分类率
文献[16]	60	...	100%
文献[3]	3	21	98.31%
文献[6]	60	...	99.4%
文献[4]*	3.6	8.8	91.22%
本文	4.4	8.8	97.17%

加*的结果和本文结果均为5次折叠交叉验证的平均值

表4 其它分类问题的仿真结果
Table 4 Simulation results of other benchmark problems

典型分类问题	模糊规则数	模糊集合总数	相关特征变量数	正确分类率
Iris	3.4	5.6	3	98.67%
WBC	2	5.4	5	97.07%
Glass	9.6	22.6	7	63.08%
Ionosphere	3.6	13.8	10	91.6%
Pima	2.2	4.2	3	77.47%

以上相关数据为5次折叠交叉验证的平均值

7 结论(Conclusion)

本文提出一种基于协同进化算法的复杂模糊分类系统的设计方法. 并论述了与之相关的问题. 本文方法的优点: 1) 首先进行特征变量的选择,减少了以后聚类算法和协同进化算法的计算量; 2) 模糊模型结构和参数的同时优化,保证了优化后模型的精确性; 3) 在协同进化算法中采用基于相似性的模型简化方法约简模糊模型,有效地保证了模糊模型的解释性. 最后利用该方法对Wine等分类问题进行研究. 仿真结果表明该方法能对复杂分类系统的特征变量进行选择,同时利用较少的模糊规则和模糊集合数达到较高的识别率.

参考文献(References):

[1] SETNES M, BABUSKA R, KAYMAK U, LEMKE H R N. Similarity measures in fuzzy rule base simplification[J]. *IEEE Trans on*

Systems, Man and Cybernetics, 1998, 28(3): 376 – 386.

- [2] XING Z Y, JIA L M, ZHANG Yong. A case study of data-driven interpretable fuzzy modeling[J]. *Acta Automatica Sinica*, 2005, 31(6): 815 – 824.
- [3] SETNES M, ROUBOS H. GA-fuzzy modeling and classification: complexity and performance[J]. *IEEE Trans on Fuzzy Systems*, 2000, 8(5): 509 – 522.
- [4] ABONYI J, ROUBOS H, SZEIFERT F. Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision tree initialization[J]. *Int J of Approximate Reasoning*, 2003, 32(1): 1 – 21.
- [5] WANG Han Li, SAM KWONG. Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction[J]. *Fuzzy Sets and Systems*, 2005, 149(1): 149 – 186.
- [6] ISHIBUCHI H, NAKASHIMA T, MURATA T. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems[J]. *IEEE Trans on System, Man and Cybernetics-Part B*, 1999, 29(5): 601 – 618.
- [7] HO S-Y, CHEN H-M. Design of accurate classifiers with a compact fuzzy-rule base using an evolutionary scatter partition of feature space[J]. *IEEE Trans on Systems, Man and Cybernetics-Part B*, 2004, 34(2): 1031 – 1044.
- [8] PAIVA R P, DOURADO A. Interpretability and learning in neuro-fuzzy system[J]. *Fuzzy Sets and Systems*, 2004, 147(1): 17 – 38.
- [9] HOMAIFAR A, McCORMICK E. Simultaneous design of membership functions and rule sets for fuzzy controller using genetic algorithms[J]. *IEEE Trans on Fuzzy System*, 1995, 3(2): 129 – 139.
- [10] PAREDIS J. *Coevolutionary Computation, Artificial Life 2*[M]. Cambridge, MA: MIT Press, 1995: 355 – 375.
- [11] POTTER M A, DE JONG K A. Cooperative coevolution: an architecture for evolving coadapted subcomponents[J]. *Evolutionary Computation*, 2000, 8(1): 1 – 29.
- [12] RAN GILAD-BACHRACH, AMIR NAVOT. Margin based feature selection-theory and algorithms[C]//*Proc of the 21st Int Conf on Machine Learning*. Banff, Canada: ACM Press, 2004.
- [13] CASILLAS J, CORDON O, HERRERA F. *Interpretability improvements to find the balance interpretability-accuracy in fuzzy modeling: an overview*[M]//Chapter of Interpretability Issues in Fuzzy Modeling, Heidelberg, Germany: Springer, 2003: 3 – 22.
- [14] JIN Y. *Advanced Fuzzy Systems Design and Applications*[M]. New York: Physical-verl, 2003.
- [15] JIN Y, SEELEN V W, SENDHOFF B. On generating FC3 fuzzy rule systems from data using evolution strategies[J]. *IEEE Trans on Systems, Man and Cybernetics*, 1999, 29(6): 829 – 845.
- [16] CORCORAN A L, SEN S. Using real-valued genetic algorithms to evolve rule sets for classification[C]//*Proc 1st IEEE Conf on Evolutionary Computation*. Orlando, USA: IEEE Press, 1994: 120 – 124.

作者简介:

张永 (1969—),男,南京理工大学自动化学院博士生,主要研究方向为模糊建模、智能控制等, E-mail: zy69813@gmail.com;

吴晓蓓 (1958—),女,南京理工大学自动化学院教授,博士生导师,主要研究方向为模糊建模、工业过程智能控制与检测等;

向峥嵘 (1969—),男,南京理工大学自动化学院副教授,主要研究方向为非线性控制理论及应用、智能控制与智能系统等;

胡维礼 (1941—),男,南京理工大学自动化学院教授,博士生导师,主要研究方向为智能控制与智能系统、非线性控制理论及应用、高精度数字交流伺服系统等.