

基于 VISIT 算法设计模糊分类器的进化方法

白治江, 刘广钟

(上海海事大学信息工程学院, 上海 200135)

摘要:介绍了在没有数据分布先验知识的情况下, 用进化方法直接从训练数据中建立紧致模糊分类系统的方法。使用 VISIT 算法获取每个个体模糊系统, 再用遗传算法从中搜索最优的模糊系统。规则和隶属函数是在进化过程中自动建立和优化的。为了同时有效地评价系统的精度和紧致性, 用一个模糊专家系统作适应度函数。在 2 个基准分类问题上的实验结果表明了新方法的有效性。

关键词: VISIT; 遗传算法; 紧致模糊分类器; 规则抽取

VISIT-based Evolutionary Approach for Fuzzy Classifier Design

BAI Zhijiang, LIU Guangzhong

(Information Engineering College, Shanghai Maritime University, Shanghai 200135)

【Abstract】 An evolutionary approach for designing compact fuzzy classifier directly from data without any a priori knowledge of the data distribution is proposed. The variable input spread inference training (VISIT) algorithm is used to create each individual fuzzy system, and then searches the best one via genetic algorithm. Rules and membership functions are automatically created and optimized in an evolutionary process. In order to effectively evaluate the accuracy and compactness simultaneously, a fuzzy expert system acts as the fitness function. The experiments on two benchmark classification problems show the effectiveness of the new method.

【Key words】 VISIT; Genetic algorithm; Compact fuzzy classifier; Rule extraction

模糊逻辑具有处理不确定性和模糊性的强大能力, 它已经被应用于复杂的分类系统中。模糊逻辑将类别重叠的事实和软决策机制引入分类系统中, 因此一个模式可能以不同的隶属度属于若干类别。此外, 模糊分类器由 if-then 模糊规则集组成, 规则集既能使我们洞察分类器的结构, 又能提高系统的可解释性。

模糊分类器最重要的优势之一是其规则库结构的可解释性, 可解释性与模糊规则的数量有关, 规则较少的(紧致的)模糊系统往往有更好的可解释性。此外, 小的规则库能消除系统的过拟合现象, 增强系统的泛化能力和鲁棒性。已提出若干设计紧致模糊系统的方法, 如选择策略、迭代复杂性归约技术及基于 GA 的多目标优化。但这些方法都需要用某种学习技术先确定模糊系统的初始结构(包括规则的数量, 隶属函数的形状, 隶属函数的数量等), 然后再对初始结构细调。例如, 文献[1]应用 MAC 算法先找出初始结构, 再用快速递归最小二乘算法细调参数; 文献[2]先用模糊聚类技术开发模糊模型的结构, 再用实数编码 GA 和相似性驱动约简相结合的办法细调结构和参数; 文献[3]利用 LVQ 探寻模糊系统的结构, 然后用整数编码 GA 细调规则和隶属函数。

本文提出了在没有任何关于数据分布的先验知识的前提下, 不需要任何初始模型, 从数据中直接建立紧致模糊分类系统的进化方法。

1 模糊分类器的体系结构

模糊分类器中一个规则的前提在 N-维特征空间上定义了一个决策域, 规则结论是一个类别标识。设全部类别的有限集合为 $\{C_1, C_2, \dots, C_M\}$, 则典型的模糊分类规则形式如下:

Rule l: If x_1 is A_1^l and x_2 is A_2^l and ... and x_n is A_n^l , then y is C_m

其中 $l=1, \dots, L$, L 是规则数, $m \in 1, \dots, M$, M 是类别数, n

是特征数, $A_j^l(j=1, 2, \dots, n)$ 是特征变量 x_j 的模糊集。 A_j^l 的隶属函数 $u_{A_j^l}(x_j)$ 可以是三角形, 高斯型或其他任何形状。本文中 使用非对称高斯型隶属函数, 定义为

$$u_{A_j^l}(x_j) = \begin{cases} e^{-\frac{1}{2}(\frac{x_j - m_j^l}{\delta_{j,l}^l})^2}, & x_j < m_j^l \\ e^{-\frac{1}{2}(\frac{x_j - m_j^l}{\delta_{j,l}^r})^2}, & x_j \geq m_j^l \end{cases} \quad (1)$$

其中 m_j^l 是隶属函数的中心, $\delta_{j,l}^l$ 及 $\delta_{j,l}^r$ 表示其左右宽度。如果 $\hat{x}_k = [x_1 \ x_2 \ \dots \ x_n]$ 是一个给定的特征向量, 则规则 R^l 的激励强度用下式计算:

$$u^l(\hat{x}_k) = u_{A_1^l}(x_1) u_{A_2^l}(x_2) \dots u_{A_n^l}(x_n) \quad (2)$$

对给定模式 \hat{x}_k , 若第 l 条规则具有最大激励强度, 则 \hat{x}_k 被分入该规则结论所标识的类别中。模式 \hat{x}_k 的分类误差由下式计算:

$$e_k = \begin{cases} 0, & \hat{x}_k \text{ 正确分类} \\ 1, & \hat{x}_k \text{ 错误分类} \end{cases} \quad (3)$$

2 由 VISIT 算法抽取模糊规则

原始 VISIT 算法中假定所有的论域是完全等同的, 因此开始时选定的参数是标量值 w, a 和 δ 。在实际系统中, 不同特征所作用的论域亦不相同, 本文对 VISIT 算法稍作修正, 即使用 3 个向量参数 W, A 和 Σ 代替标量参数 w, a 和 δ 。W,

基金项目: 上海市重点学科建设基金资助项目(T0602); 上海海事大学重点学科建设基金资助项目

作者简介: 白治江(1962—), 男, 讲师、博士, 主研方向: 人工智能; 刘广钟, 教授、博士

收稿日期: 2006-05-12 **E-mail:** cen.shmtu.edu.cn

A 和 Σ 选择不同的值, 将产生不同的模糊系统, 找出优良 VISIT 分类器的问题就归结为找出参数 W, A 和 Σ 的优良值问题。

VISIT 算法开始之前, 先选定各个论域上创建第一个隶属函数的初始宽度向量 $\Sigma=[\delta_1, \delta_2, \dots, \delta_n]$, 还要选定各个论域上隶属函数的 α -截集值向量 $A=[a_1, a_2, \dots, a_n]$ 以及确定相邻隶属函数之间重叠程度的向量 $W=[w_1, w_2, \dots, w_n]$ 。下面给出修正 VISIT 算法:

(1) 设 $i=1$, 取第一个训练模式 $tp^1=(x_1, x_2, \dots, x_n, y)^1$, 其中 y^1 是一个实数, 表示训练模式 tp^1 的类别。在每个输入论域上形成一个模糊集 A_j^1 , 其隶属函数是中心在 $m_{j,0}=x_j^1$ 和宽度为 δ_j 的对称高斯型函数 $u_{A_j^1}(x_j^1)$ 。在输出论域上形成一个模糊集 B^1 , 其隶属函数是支集为 y^1 的单值隶属函数。

(2) 增加如下规则

If x_1 is A_1^1 and x_2 is A_2^1 and \dots and x_n is A_n^1 , then y is B^1 。

(3) 若再无训练模式, 则训练结束。否则, 置 $i=i+1$, 取下一个训练模式 $tp^i=(x_1, x_2, \dots, x_n, y)^i$ 。

(4) 用当前分类器评价 tp^i , 若分类正确, 转到(3); 否则转到(5)。

(5) 在第 $j(j=1, \dots, n)$ 个输入论域上, 在 x_j^i 点隶属函数值最大的模糊集记为 A_j^{\max} , 其 a_j -截集记为 $Cut_{a_j}(A_j^{\max})$ 。若 $x_j^i \notin Cut_{a_j}(A_j^{\max})$, 则在第 j 个论域上创建一个新的模糊集 A_j^{new} , 其隶属函数的中心为 $m_j^{\text{new}}=x_j^i$, 左右宽度分别为(6)中定义的 δ_j^L 及 δ_j^R 。在输出论域上, 若模式 tp^i 属于新类别, 则增加一个新模糊集 B^i , 其隶属函数是支集为 y^i 的单值隶属函数。否则, 类别 B^i 已经存在。

(6) 在每个输入论域上新增隶属函数两边的宽度初始化为 δ_j 。如果新函数左边和(或)右边有近邻函数, 则新函数相应侧边的宽度及其左和(或)右近邻函数的宽度要重新计算。新函数的右宽度和右近邻函数的左宽度重新计算为

$$\delta_j^R = \frac{1}{w_j} |m_j^{\text{new}} - m_j^{\text{RN}}| \quad (4)$$

新函数的左宽度和左近邻函数的右宽度重新计算为

$$\delta_j^L = \frac{1}{w_j} |m_j^{\text{new}} - m_j^{\text{LN}}| \quad (5)$$

其中, m_j^{RN} 表示 m_j^{new} 右侧的近邻函数中心, m_j^{LN} 表示 m_j^{new} 左侧的近邻函数中心。

(7) 若由(6)形成了新函数, 则考虑如下候选规则, 其中规则前提中的模糊集按(5)中的方式获得:

If x_1 is A_1^{\max} and x_2 is A_2^{\max} and \dots and x_n is A_n^{\max} , then y is B^i 。

若规则库中没有与该规则不一致的规则(即相同的前提, 不同的结论), 则将该规则加入到库中, 否则丢弃 tp^i 并转入(3)。

(8) 若还有训练模式, 则转入(3); 否则训练过程结束。

3 通过进化算法抽取模糊规则

考虑训练数据集 $X = \{\hat{x}_k = (x_1, x_2, \dots, x_n, y)^k\}$, 对每个 k , 点 \hat{x}_k 是空间 $R^n \times R$ 中的一个元素, VISIT 算法可看作是数据集 X 到模糊系统定义的一个函数映射 $F(\cdot)$, 即

$$FIS = \underset{W, A, \Sigma, M}{F}(X) \quad (6)$$

此处引入另一个参数向量 $M=[m_{1,0}, m_{2,0}, \dots, m_{n,0}]$ 组成初

始隶属函数的中心, 以减弱 VISIT 对模式输入次序的依赖性。由 VISIT 构建优良模糊分类器等价于最优化参数 W, A, Σ 和 M , 其性能由指标 J (诸如下述的 J_1 和 J_2) 评价。由 VISIT 从数据中生成模糊系统的问题可表达为多目标优化问题:

$$\min_{W, A, \Sigma, M} (J) \quad (7)$$

$$s.t. \quad w^{\min} < w_j < w^{\max} \quad (8)$$

$$0 < a_j < 1 \quad (9)$$

$$\delta_j^{\min} < \delta_j < \delta_j^{\max} \quad (10)$$

$$x_j^{\min} < m_{j,0} < x_j^{\max} \quad (11)$$

其中, $[w^{\min}, w^{\max}]$ 定义可接受的重叠范围, $[\delta_j^{\min}, \delta_j^{\max}]$ 定义第 j 个特征的初始隶属函数宽度范围, $[x_j^{\min}, x_j^{\max}]$ 是第 j 个论域。 J 是评价模糊系统的性能指标, 通常从两个方面评价模糊系统: (1) 精度: 模糊系统应当精确地描述数据的特征; (2) 可解释性: 模糊规则库应当足够小使得系统易于理解。可解释性是模糊分类器区别于其他分类器重要特征之一, 模糊系统好的解释性往往意味着系统的鲁棒性。模糊系统的可解释性与隶属函数数量、规则的数量和前提长度呈负相关关系。

上述优化问题中无法满足目标函数 J 关于 W, A, Σ 和 M 可微, 因此 GA 优化成为最适合的选择。本文所用的 GA 算法描述如下:

(1) 设 $t=0$, 初始群体 P_t 包含 N 个个体 (N 是偶数)。

(2) 估计 P_t 中 N 个个体的适应度 f_i^t , $i=1, \dots, N$ (适应度下文定义)。

(3) 从群体 P_t 中选择 $N/2$ 个最适应个体作变异运算, 生成新的 $N/2$ 个个体 P_t' 。

(4) 从群体 P_t 中随机选择 $N/4$ 对个体进行交叉运算生成另外 $N/2$ 个新个体 P_t'' 。

(5) 从现有群体 P_t, P_t' 和 P_t'' 中选 N 个最适应个体, 组成下一代群体 P_{t+1} 。

(6) 设 $t=t+1$, 转到(3)。重复该过程直到 $t > t_{\text{end}}$, 其中 t_{end} 是终止时间。

3.1 遗传算子

将群体中的个体编码成向量形式的染色体 $P_t(i)=[w_1^t(i), \dots, w_n^t(i), a_1^t(i), \dots, a_n^t(i), \delta_1^t(i), \dots, \delta_n^t(i), m_{1,0}^t(i), \dots, m_{n,0}^t(i)] \in R^{4n+1}$ 。向量 $P_t(i)$ 在遗传算法中表示一个个体, 其中 t 表示进化的代数, i 表示群体中的一个个体。本文使用的两类遗传算子为: 整体算术交叉和高斯变异。第 k 对被选个体 $P_t(i)$ 和 $P_t(j)$ 用整体算术交叉算子作交叉运算, 所得的两个后代为

$$P_t''(1) = rP_t(i) + (1-r)P_t(j) \quad (12)$$

$$P_t''(2) = rP_t(j) + (1-r)P_t(i) \quad (13)$$

其中 r 是在 $[0, 1]$ 上均匀分布的随机数。定义如下变异算子从第 t 代中第 i 个个体创建一个新个体:

$$P_t'(i, j) = P_t(i, j) + \gamma_j \rho_j(i) \exp\left(-\frac{f_i^t}{f_{\max}^t}\right) \quad (14)$$

其中 j 表示染色体的一个特定元素, γ_j 是学习系数, 它决定染色体中第 j 个元素从上一代到下一代的最大变化量, $\rho_j(i)$ 是服从 $[-1, 1]$ 均匀分布的随机数, f_i^t 是个体 i 在第 t 代中的适应度, f_{\max}^t 是第 t 代中的最大适应度。变异算子可能产生不可行后代, 必须检查每个后代以确保其满足约束(8)~(11)。如果某个后代不可行, 则将其丢弃, 并将 γ_j 值减半, 再次使

用变异算子直至获得可行后代。

3.2 用模糊专家系统评价个体适应度

模糊系统有两个相互冲突的性能：精度和可解释性。模糊分类器的精度与错分的模式数量有关，错分模式越多精度越低。模糊系统的可解释性与规则数量有关，规则越多可解释性越差。本文用平均分类误差(mean classification error, MCE)度量模糊分类器的精度；用规则数作为度量可解释性的负相关指标。MCE 准则如下：

$$J_1 = \frac{1}{K} \cdot \sum_{k=1}^K e_k \quad (15)$$

其中 e_k 由式(3)定义且 K 是训练集中的模式数量。可解释性由下式度量：

$$J_2 = \beta(L - M) \quad (16)$$

其中 L 是规则数量， M 是训练集中类别总数 (M 也表示模糊系统中规则数量的最小值)，参数 β 是一个因数，它将规则库的大小标准化为 $[0, 1]$ 范围内的一个值。本文取 $\beta=0.01$ 。

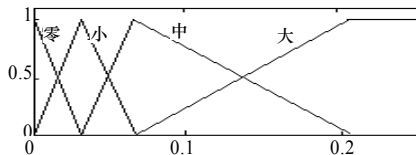
目标函数 J_1 和 J_2 的取值位于 $0 \sim 1$ 之间。一个好的模糊分类器应当有小的 MCE 和少的规则数，但提高精度通常需要增加规则数，因此必须在精度和可解释性之间寻找折衷。本文所设计的模糊系统可形式化为如下二目标优化问题：

$$\min_{W, A, \Sigma, M} (J_1) \text{ and } \min_{W, A, \Sigma, M} (J_2) \quad (17)$$

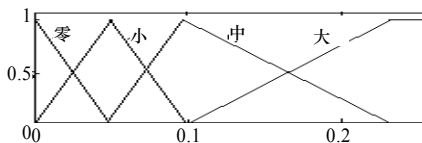
本文用一个模糊专家系统评价个体的两个目标函数，其优点是模糊专家系统结合人类直观知识将两个目标以非线性方式组合在一起。模糊专家系统由两个输入 (J_1 和 J_2) 一个输出 (fit) 和 11 条 TS 规则(零阶)组成。模糊专家系统中规则 k 建议的输出是 fit_k 。表 1 列出模糊专家系统的规则，图 1 表示输入的隶属函数。规则库和隶属函数是我们关于一个好的模糊系统的直觉知识的结果。

表 1 适应度规则

| 规则序号 (k) | 隶属函数 | | |
|-------------|-------|-------|---------|
| | J_1 | J_2 | fit_k |
| 1 | 零 | 零 | 80 |
| 2 | 零 | 小 | 70 |
| 3 | 零 | 中 | 50 |
| 4 | 小 | 零 | 60 |
| 5 | 小 | 小 | 40 |
| 6 | 小 | 中 | 20 |
| 7 | 中 | 零 | 30 |
| 8 | 中 | 小 | 10 |
| 9 | 中 | 中 | 1 |
| 10 | -- | 大 | -20 |
| 11 | 大 | -- | -20 |



(a) J_1



(b) J_2

图 1 模糊专家系统的输入隶属函数

4 实验评价

在 2 个基准数据集上评价新方法的性能，它们是 wine 数据和 iris 数据。这些数据集直接从 <ftp://ftp.ics.uci.edu/pub/machine-learning-database/> 下载。数据被标准化在 $[0, 1]$ 上。

4.1 实验 1: wine 数据

wine 数据包括对 3 个类别的 178 种葡萄酒的化学分析。每种酒有 13 个连续属性(AL, MAC, AC, AA, MA, TP, FL, NFP, PR, CI, HUE, O, P)，3 种类别的模式数量分别为 59, 71, 48。

修正 VISIT 算法用于从 wine 数据中创建一个模糊系统，其参数向量 $\Sigma = [\delta_1, \delta_2, \dots, \delta_{13}]$ ， $A = [a_1, a_2, \dots, a_{13}]$ ， $W = [w_1, w_2, \dots, w_{13}]$ ，和 $M = [m_{1,0}, m_{2,0}, \dots, m_{n,0}]$ 在 (8)~(11) 的范围内作随机初始化。每代种群大小为 200，进化过程经 100 次迭代后停止。

新方法对其中的 7 个特征只创建了一个隶属函数，显然这 7 个特征对所有的规则贡献相同，删除这些冗余的隶属函数之后，系统最终的隶属函数和规则集如图 2 和表 2，所得模糊系统的识别率是 98.9% (2 个模式错分)。新方法在没有任何结构模型假设的前提下，开发了一个紧凑的模糊系统。系统只用了 6 个属性、5 条规则，每个特征的隶属函数不多于 3 个，因此分类规则有很高的可解释性。表 3 是新分类器与一些著名模糊分类器系统在特征数、模糊集数量和识别率方面的比较。

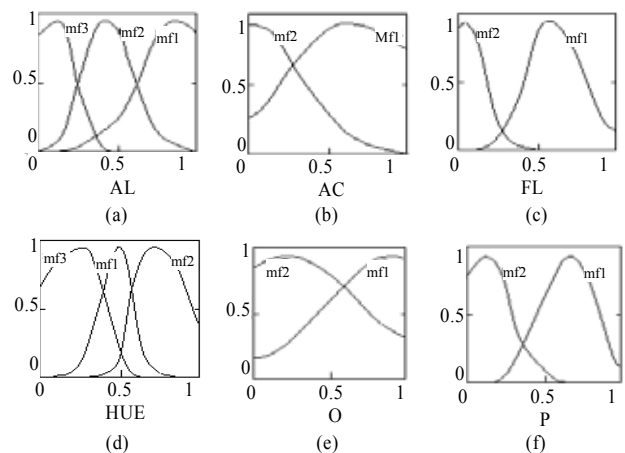


图 2 删除冗余隶属函数之后系统最终的隶属函数 (只有 6 个特征)

表 2 系统的分类规则 (6 个特征)

| Rule | AL | AC | FL | HUE | O | P | Class |
|------|----|----|----|-----|---|---|-------|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 |
| 3 | 2 | 1 | 1 | 2 | 1 | 2 | 2 |
| 4 | 2 | 2 | 1 | 3 | 2 | 2 | 2 |
| 5 | 2 | 1 | 2 | 3 | 2 | 2 | 3 |

表 3 wine 数据上的分类结果比较

| 方法引用 | 特征 | 模糊集 | 规则 | 总识别率(%) |
|-------|----|------------|----|------------------|
| 文献[2] | 9 | 21 | 3 | 98.3 |
| 文献[1] | 13 | 34 | 3 | 99.4 |
| 文献[4] | -- | 9 | 6 | 100 |
| 文献[5] | 5 | 15, 11, 10 | 3 | 98.9, 98.3, 99.4 |
| 本算法 | 6 | 13 | 5 | 98.9 |

4.2 实验 2: Iris 数据

Iris 数据有 4 种属性(SL, SW, PL, PW)，共 150 个数据，分 3 个类别，每种类别包括 50 个样本。新方法使用 Iris 数据中全部 150 个模式设计模糊分类器，参数 W ， A ， Σ 和 M 在 (8)~(11) 的范围内随机初始化。群体中有 200 个个体，100

(下转第 28 页)