

辐射源识别的模糊联想分类器设计^{*}

方 敏 王宝树

(西安电子科技大学计算机学院 西安 710071)

Design of Fuzzy Associative Classifier for Identification of Radar Radiant Points

FANG Min WANG Bao-Shu

(Institute of Computer Science, Xidian University, Xi'an 710071)

Abstract The fuzzy associative classifier is investigated in this paper. The design methods of the fuzzy associative classifier with genetic algorithm for training are presented. This method trains the weight and back terms to obtain classification rules automatically. Radar radiant points are classified by using of this algorithm, and the simulation results show that the method has higher identification precision than available fuzzy classifiers.

Keywords Classification, Fuzzy associative classifiers, Genetic algorithm, Optimization

1. 前言

Kosk 于 1986 年提出了模糊联想记忆(Fuzzy Associative Memory, FAM)系统,它模拟了人脑通过模糊联想记忆,能够对目标进行划分并对此进行识别判断的方法,对目标进行智能分类识别。它将模糊系统与人工神经网络首次接合起来,形成了模糊神经网络。近年来,模糊联想记忆已在函数估计、模糊控制、图像编码分类识别等方面得到广泛的应用^[1,2]。

在情报侦察数据融合专家系统中,对目标所携带辐射源类型的识别具有十分重要的意义,是进行态势估计和威胁评估的基础。对雷达功能进行分类,目前还没有一个统一的标准。一般情况下接收到的雷达信号分布非常复杂,具有不确定性和模糊性,因而,需要提取适用的雷达分类模糊规则和采用大量的合乎模糊神经网络要求的雷达信号样本对神经网络进行网络学习,以得到合理的分类结果。模糊联想分类器既具有模糊系统的特性,又具有神经网络的自学习能力和并行结构特征^[3],因此,把模糊联想记忆引入到雷达辐射源识别中,以实现目标的智能分类识别。

模糊联想分类器是将分类规则模糊化,并对分类规则的模糊前件、后件以及权值进行学习。常常采用进化优化算法对模糊联想分类器进行训练,获得分类器的分类规则,但求解过程依赖于种群初始值的选择。因此,如何有效划分模式空间和如何减少分类器的训练时间成为研究的重点^[4]。本文提出一种基于进化计算的模糊联想分类器的设计,首先对输入模式空间进行划分,得到每个模糊子集的隶属度,通过遗传算法对分类规则的模糊后件以及权值进行优化学习,提取分类规则集,实现模糊联想分类器。并将该算法应用于雷达体制分类识别问题,通过仿真数据实验表明,基于进化的模糊联想分类器能够很好地提取识别规则,并有效地提高分类精度。

2. 模糊联想分类器的结构与分类规则

模糊联想分类是一类特殊的模糊系统,它定义了一个连续映射:

$$S: F(2^X) \rightarrow F(2^Y) \quad (1)$$

其中 X 是其输入论域, Y 是其输出论域, $F(2^X)$ 是 X 的模糊幂集, $F(2^Y)$ 是 Y 的模糊幂集。作为分类器的模糊联想记忆,输入论域是指待分类的模式集合—模式空间,输出论域为模式类别的集合。若模式空间为 Ω , 其中的模式共有 m 类, 分别标记为 $u_1, u_2, \dots, u_m$, 则有:

$$X = \Omega$$

$$Y = \{u_1, u_2, \dots, u_m\} \quad (2)$$

设 $\omega \in \Omega$ 是任一模式, 则第 i ($i=1, 2, \dots, N$) 条分类规则为: 如果模式 ω 具有属性或特征 A_i , 那么 ω 是 U_i , 即 (A_i, U_i, W_i) , 表示为:

$$\text{if } A_i, \text{ then } (a_{i1}/u_1, a_{i2}/u_2, \dots, a_{im}/u_m)$$

其中 A_i 是输入论域 X 上的模糊子集; $U_i = \{a_{i1}/u_1, a_{i2}/u_2, \dots, a_{im}/u_m\}$ 是输出论域 Y 上的模糊子集, 描述了该模式属于 u_1 的可能性为 a_{i1} , 属于 u_2 的可能性为 a_{i2} , 属于 u_m 的可能性为 a_{im} , a_{ij} 是指类别标记 u_i 相对模糊子集 U_i 的隶属程度, 且有:

$$\sum_{j=1}^m a_{ij} = 1, a_{ij} \geq 0; i=1, 2, \dots, N \quad (3)$$

权值 $W_i > 0$, 它表示规则 i 的强度或可信度。

分类过程为:

1) 首先计算输入模式 ω 对于每条规则 $r_i = (A_i, U_i, W_i)$ 前件的适应度值 f_i , 即模式 ω 具有属性 A_i 的程度:

$$f_i = \mu_{A_i}(\omega), i=1, 2, \dots, N$$

规则前件的隶属度可有很多表示形式, 这里每个模糊子集的隶属度函数的形状都取为等腰三角形, 每一维采用相同的模糊划分法。

$$2) \text{ 规则 } r_i \text{ 的激活程度 } H_i \text{ 定义为: } H_i = \prod_{j=1}^T \mu_{r_i}(x_j)$$

3) 计算规则 (A_i, U_i, W_i) 的输出模糊子集:

$$U = \max_{i=1}^N (f_i \cdot U_i \cdot W_i) = \max_{i=1}^N (\mu_{A_i}(\omega) \cdot U_i \cdot W_i) = \{m_{i1}/u_1, m_{i2}/u_2, \dots, m_{im}/u_m\}, i=1, 2, \dots, N \quad (4)$$

采用最大值去模糊法可得分类结果 y 为:

$$y = \begin{cases} u_r, & \text{若 } \exists m_r = \max_{i=1}^N m_i, \text{ 且唯一} \\ \text{拒判}, & \text{若 } \exists m_r = \max_{i=1}^N m_i, \text{ 且不唯一} \end{cases} \quad (5)$$

可以看出模糊联想分类器是通常的模糊分类器的扩展, 分类性能要比采用非模糊量作规则后件的模糊分类器好, 正

^{*} 获国防科技预研基金和图像信息处理与智能控制教育部重点实验室基金 TKLJ0114 资助。

确分类率要比通常的模糊分类器高^[4]。

3. 模糊联想分类规则的获取

模糊联想分类器的关键是模糊分类规则的获取,是产生分类规则集 $(A_i, U_i, W_i) (i=1, 2, \dots, M)$ 。模糊联想分类器的训练问题描述为:设有 n 个取自模式空间 Ω 的 m 个不同模式 $u_1, u_2, \dots, u_m$ 的样本; $A_t = \{A_{t,i}, i=1, 2, \dots, T_t\} t=1, 2, \dots, M$ 是对模式空间 Ω 的 M 种不同的模糊划分,其中 T_t 表示第 t 类模糊划分的模糊子集数。求模糊联想分类器的分类规则集式(6),使得得分错的样本数最小。

$$R = \{(A_{t,i}, U_{t,i}, W_{t,i}) | i=1, 2, \dots, T_t, t=1, 2, \dots, M\} \quad (6)$$

3.1 用遗传算法获得分类规则

设输入论域 X 的模糊划分为 $\{A_{t,i}, i=1, 2, \dots, T_t\}, t=1, 2, \dots, M$, 形成规则集 $R = R_1 \cup R_2 \cup \dots \cup R_M, L = \sum_{i=1}^M |R_i|$, 则它构成了模式空间 Ω 上的一个模糊联想分类器,其中 T_t 表示第 t 种模糊划分的模糊子集数,对于每一种划分建立一套分类规则 R_i , 规则 r_{ij} 表示为:

$$r_{ij}: (A_i, \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im}, W_i) \quad (7)$$

其中 $\alpha_{im}, W_i \in [0, 1]; i=1, 2, \dots, M; j=1, 2, \dots, T_i, |R_i|$ 表示第 i 种划分形成的规则数。 L 为规则总数。

编码方式:每个个体是由 T_i 条规则组成,个体基因串的编码为:

$$(A_1, \alpha_{11}, \alpha_{12}, \dots, \alpha_{1m}, W_1) \dots (A_{T_i}, \alpha_{T_i,1}, \alpha_{T_i,2}, \dots, \alpha_{T_i,m}, W_{T_i}) \quad (8)$$

(1)适应度函数。遗传算法需要根据个体适应度函数的值进行搜索,在每一个学习周期,适应度较低的个体将被淘汰,适应度最高的个体被看作满意解^[5]。对每个训练实例,预测输出与实际输出之间的平均距离越小,该个体的适应度越高,因此,适应度定义为训练样本上的预测输出与实际输出的平均距离:

$$fitness(i) = (\sum_{j=1}^n \frac{1}{[1 + \beta |(O_j - O_{ij}) / O_j|]}) / n \quad (9)$$

其中 O_j 为个体在输入实例 j 的期望输出, O_{ij} 为个体 i 在实例 j 的实际输出, n 为训练样本实例数, β 为比例系数。

(2)遗传算法的算子。在这里,使用了遗传算法的 3 种基本算子:选择、交叉和变异^[6]。

初始化:由于对模式空间 Ω 的划分缺乏先验知识,对 A_i 的划分比较困难。当特征空间维数比较大时,模糊前件 A_i 种类数将会很大,所以采用直接从训练样本中获得模式空间的划分,见算法 1。

算法 1:设置种群大小 POP 及模糊空间划分维数的设定值;产生初始化种群。

- 1) 从训练样本取一训练实例,产生划分 A_i ;
- 2) 在一个合理范围,均匀输出 (U_i, W_i) ; 个体中的规则数加 1;
- 3) 每个个体的规则数小于设定值,则转 1); 否则,种群规模加 1;
- 4) 种群规模小于预定设置,则转 1) 产生新个体; 否则,初始化完成;

选择:从当前群体中选择 POP/2 对参与产生下一代个体;当前群体中个体选种概率 P_s 为:

$$P_s(i) = fitness(i) / \sum_{j=1}^{POP} fitness(j) \quad (10)$$

交叉:对选择出的每一对个体进行单点交叉操作,交叉点随机选取,产生 POP 个新个体。这里交叉操作在两个基因的分界处进行。例如有个体 g_i 和 g_j 要进行交叉,设选取的交叉位置,则对 g_i 和 g_j 从选择的交叉点进行交叉,结果为:

$$\text{交叉前: } g_i = r_{i1}r_{i2}r_{i3} | r_{i4}r_{i5}r_{i6}$$

$$g_j = r_{j1}r_{j2}r_{j3} | r_{j4}r_{j5}r_{j6}$$

$$\text{交叉后: } g'_i = r_{i1}r_{i2}r_{i3} | r_{j4}r_{j5}r_{j6}$$

$$g'_j = r_{j1}r_{j2}r_{j3} | r_{i4}r_{i5}r_{i6}$$

变异:对交叉产生的个体串,按(11)式以概率 p_m 进行变异,并对产生个体进行归一化处理,以满足(3)式的约束;

$$\alpha_{ij}(\text{gen}+1) = \alpha_{ij}(\text{gen}) + \text{rnd}(0, 1)$$

$$W_i(\text{gen}+1) = W_i(\text{gen}) + \text{rnd}(0, 1) \quad (11)$$

其中 $\text{rnd}(0, 1)$ 取 $(0, 1)$ 区间的随机数。

终止规则:取繁衍代数是否达到预先设定的代数为终止求解过程的准则。

利用遗传算法进行训练,进行全局搜索,不容易陷入局部极小,并且 GA 算法由于有交叉操作,使得群体能够保持多样性。

4. 实验和分析

由于雷达信号具有模糊性和不确定性^[7,8],因此,通过提取雷达模糊分类规则,利用模糊联想分类器实现雷达功能的分类。选用六类具有代表性的雷达功能类型: {搜索雷达、跟踪雷达、火控雷达、轰炸雷达、导航雷达、气象雷达}。目标特征空间 Ω , 见式(12), 其中 n 为目标总数,并将这些数据的每一维通过平移和比例变换归一化到 $[0, 1]$ 区间。

$$\Omega = \{X_i | X_i = (\text{射频、脉宽、重复频率}), i=1, 2, \dots, n\} \quad (12)$$

问题描述:设特征空间的模糊划分为 $A_i, i=1, 2, \dots, M$, 求解每个 A_i 所对应的规则后件 U_i 及相应的权值 W_i 。

$$U_i = \{\alpha_{i1}/\text{搜索雷达}, \alpha_{i2}/\text{跟踪雷达}, \alpha_{i3}/\text{火控雷达}, \alpha_{i4}/\text{轰炸雷达}, \alpha_{i5}/\text{导航雷达}, \alpha_{i6}/\text{气象雷达}\} \quad (13)$$

设模式空间的划分采用训练样本的特征空间模糊划分方法:

$$\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_N; \Omega_i \cap \Omega_j = \emptyset, i \neq j, N=6 \quad (14)$$

对 Ω_i 中的每一个样本,根据式(15)求每一维的最大值和最小值,形成区间 S_{ij} :

$$\begin{cases} \omega_{\max,ij} = \max_{\omega \in \Omega_i} \omega_j & i=1, 2, \dots, 6 \\ \omega_{\min,ij} = \min_{\omega \in \Omega_i} \omega_j & j=1, 2, \dots, 3 \end{cases} \quad (15)$$

$$S_{ij} = [\omega_{\min,ij}, \omega_{\max,ij}] \quad (16)$$

可得一个覆盖第 i 类样本的超立方体:

$$C_i = S_{i,1} \times S_{i,2} \times S_{i,3}; i=1, 2, \dots, N \quad (17)$$

其中“ $\times$ ”为集合的笛卡尔积。可由 C_i 确定一个模糊集 A_i , 其隶属度函数取为:

$$\mu_i(\omega) = \begin{cases} 1; & \text{if } \omega \in C_i \\ \prod_{j=1}^3 \max[1 - \frac{|\omega_j - \omega_{\min,ij}| * \delta}{D_{ij}}, 0]; & \text{if } \omega \notin C_i \end{cases} \quad (18)$$

δ 是控制模糊集 A_i 的支集宽度的参数。令 S'_i 是第 i 类样本与第 k 类样本超立方体的相交部分, S'_i 也构成一个超立方体。

$$S'_i = \begin{cases} 0; & \text{if } S_{ij} \cap S_{kj} = \emptyset, \exists j \in \{1, 2, 3\} \\ (S_{i1} \cap S_{k1}) \times (S_{i2} \cap S_{k2}) \times (S_{i3} \cap S_{k3}); & \text{other} \end{cases}$$

$$i=1,2,3;k=1,2,3, \text{ 且 } k \neq i;$$

$$r=1,2,\dots,15 \quad (19)$$

每个 S_r 也构造出一个模糊集合 A_r , 由 A_i 和 A_r 构成了对特征空间 Ω 的模糊划分。根据上述算法进行雷达体制的分类实验, 算法中参数设定为: $P_m=0.001$, 进化的代数 = 50, $\beta=1$, $\delta=2$ 。用 75 个训练样本 140 个测试样本进行训练, 由雷达功能与雷达的技术特征参数之间的关系, 经训练提取了 18 条模糊分类规则。表 1 给出了利用上述算法与通常的模糊分类器分类效果的比较, 表 2 是基于 GA 训练的模糊联想分类器对 6 类目标的分类结果。由分类结果可见, 这个模糊联想分类器具有较好的分类性能, 比通常的模糊分类器要好, 并且有较高的识别率。

表 1 算法实验结果

每维的模糊划分数	正确分类率 %			
	K=2	K=3	K=4	K=5
基于通常模糊分类器	90.33	92.00	95.12	96.22
基于模糊联想分类器	95.89	97.51	97.95	98.87

表 2 模糊联想分类器雷达辐射源的分类结果(k=3)

雷达种类	识别率 %	拒利率 %	误判率 %
搜索雷达	97.8	0.8	1.4
跟踪雷达	97.3	0.7	2.0
火控雷达	97.4	0.9	1.7
轰炸雷达	97.7	0.9	1.4
导航雷达	96.9	1.0	2.1
气象雷达	98.0	0.2	1.8

(上接第 39 页)

图 3 是根据本体对语音错误的纠正例示, 并在纠正后给出相应的知识^[17]。

结论和进一步的工作 本文从本体论的角度出发探讨了医学知识的获取方法。迄今为止, 我们已建立了一个比较完善的医学本体体系, 并总结出了较多的一些医学知识获取模式以及医学知识获取主体, 形成了一套比较完整的医学知识表示方法和医学知识获取体系。由于医学是一门巨复杂的学科, 要想获取有关医学学科的完全和完整的知识是非常困难的, 这给我们在保证医学知识的完备性方面带来了困难。同时由于在进行形式化本体时由于人的理解以及获取知识的知识源本身都会不可避免地造成获取到的知识在知识库产生不一致, 因此, 我们目前虽然采用了公理来对一致性进行保证。但也很避免避免出现由于本体设计时出现的不一致性, 这就给我们提出了如何分析本体的课题, 比如本体设计的合理性, 本体之间的相似性和等价性, 本体层面上的知识的相似性和等价性。

同时, 为了加快医学知识获取的速度, 文本知识自动获取也是我们今后要进行的重要工作。

参 考 文 献

- 1 美国国家医学图书馆. ULMs 项目. <http://www.nlm.nih.gov/research/umls/>
- 2 Bowden P R, Halstead P, Rose T G. Extracting Conceptual Knowledge from Text Using Explicit Relation Markers. In: N. Shadbolt, K. Ohara, G. Schreiber, eds. Advances in Knowledge Acquisition. Lecture Notes In Artificial Intelligence, Springer-Verlag, Berlin, 1996, 1076:147~162
- 3 Hahn U, Schnattinger K, Romacker M. Automatic Knowledge

小结 模糊联想分类器是通常的模糊分类器的扩展, 是将模糊联想记忆系统用于模式分类的形式构成的。它把规则的后件扩展为模糊集, 使得分类界面的变化范围更大, 因而分类性能要比通常的模糊分类器的性能好。本文采用遗传算法作为模糊联想分类器进化训练方法, 进行模糊分类规则的提取, 不但具有对解空间进行全局搜索能力, 同时解过程不依赖于种群初始值的选择。实验结果表明这样的结合可以得到较好的分类效果。遗传算法是通过优化分类规则的边界值来提高分类精度的, 由于它可以对解空间进行全局搜索优化, 优化后的边界值可以是最优的分类问题划分。

参 考 文 献

- 1 Shin F B, Kil D H, Wayland R F. Active impulsive echo discrimination in shallow water by mapping target physics derived features to classifiers[J]. IEEE Journal of Oceanic Engineering 1997, 22(1):66~80
- 2 Ishibuchi H, Hozakik, Tanaka H. Distributed representation of fuzzy rules and its application to pattern recognition. Fuzzy sets and Syst, 1992, 52
- 3 黄建军, 等. 正则模糊联想记忆[J]. 电子学报, 1997, 25(7):68~71
- 4 Simpson P K. Fuzzy min-max neural networks Part 1: Classification[J]. IEEE Transactions on Neural Networks, 1992, 3(5):776~786
- 5 Wolpert D H. Stacked generalization[J]. Neural Networks, 1992, 5(2):241~259
- 6 周明, 孙树栋. 遗传算法原理及应用[M]. 国防工业出版社, 1999
- 7 Li Q, Tufts D W. Principal feature classification [J]. IEEE Transactions on Neural Networks, 1997, 8(1):155~160
- 8 Stewart C, Lu Y C, Larson V. A clustering approach for high resolution radar target classification [J]. Pattern Recognition, 1994, 27(4):503~513
- Acquisition from Medical Texts. Text Knowledge Engineering Lab, 1996
- 4 Lu R Q. New Approaches to Knowledge Acquisition. World Scientific Publishers, 1992
- 5 曹存根. 面向专家的知识获取. 北京: 科学出版社, 1998
- 6 Welty C. The Ontological Nature of Subject Taxonomies. In: Proc. of the First Intl. Conf. (FOIS'98), June 6-8, Trento, Italy. 317~327
- 7 Guarino N. Formal Ontology and Information Systems. In: Proc. of the First Intl. Conf. (FOIS'98), June 6-8, Trento, Italy. 3~15
- 8 《中国大百科全书》之医学卷. 中国大百科全书出版社, 1998
- 9 Chaudhri V K, Farquhar A, et al. The Generic Frame Protocol 2.0:[SRI International Technical Report]. 1997
- 10 Cao Cungen. Extracting and Sharing Medical Knowledge. Journal of Computer Science and Technology, 2002, 3
- 11 Cao Cungen, et al. Progress in the Development of National Knowledge Infrastructure. Journal of Computer Science and Technology, 2002, 17(5)
- 12 张德海, 曹存根, 张宇翔. 国家和城市知识获取与本体论分析. 见: 中国人工智能学会第九届全国学术年会暨中国人工智能学会成立 20 周年庆祝大会, 2001. 366~370
- 13 唐素勤, 曹存根. 智能教学系统: 综述与改进. 见: 中国人工智能学会第九届全国学术年会暨中国人工智能学会成立 20 周年庆祝大会, 2001. 1129~1132
- 14 《疾病和有关健康问题的国际统计分类》第十次修订本. 世界卫生组织
- 15 《医学主题词表》2002. 美国国家医学图书馆
- 16 曹存根. 大百科全书中医学知识的获取. LNCS. 2001, 2101:268~271
- 17 丰强泽, 曹存根, 田雯. 关于知识朗读的思考和结果. CAAI-9, 2001