

满意特征选择及其应用

张葛祥¹, 金炜东¹, 胡来招²

(1. 西南交通大学 电气工程学院, 四川 成都 610031; 2. 电子对抗国防科技重点实验室, 四川 成都 610036)

摘要: 实际应用中的特征选择是一个满意优化问题. 针对已有特征选择方法较少考虑特征获取代价和特征集维数的自动确定问题, 提出一种满意特征选择方法(SFSM), 将样本分类性能、特征集维数和特征提取复杂性等多种因素综合考虑. 给出特征满意度和特征集满意度定义, 设计出满意度函数, 导出满意特征集评价准则, 详细描述了特征选择算法. 雷达辐射源信号特征选择与识别的实验结果显示, SFSM 在计算效率和选出特征的质量方面明显优于顺序前进法、新特征选择法和多目标遗传算法. 证实了 SFSM 的有效性和实用性.

关键词: 优化; 满意优化; 特征选择; 识别

中图分类号: TP18, O235 **文献标识码:** A

Satisfactory feature selection and its applications

ZHANG Ge-xiang¹, JIN Wei-dong¹, HU Lai-zhao²

(1. School of Electrical Engineering, Southwest Jiaotong University, Chengdu Sichuan 610031, China;

2. National EW Laboratory, Chengdu Sichuan 610036, China)

Abstract: Feature selection is essentially a satisfactory optimization problem in engineering applications. Most of the existing feature selection methods did not consider the cost of feature extraction and automatic decision of the dimension of feature subset. In this paper, a novel approach called satisfactory feature selection method (SFSM) is proposed. SFSM considers compromisingly classification performance of feature samples, the dimension of feature set and the complexity of feature extraction. Feature satisfactory rate and feature set satisfactory rate are defined. Several satisfactory rate functions are designed. Satisfactory feature set evaluation criterion is given in a mathematical way. Satisfactory feature selection algorithm is described in detail. Experimental results of radar emitter signal feature selection and recognition show that SFSM is superior to sequential forward selection using distance criterion, new feature selection method and multi-objective genetic algorithm in computing efficiency and feature qualities. Hence, the validity and applicability of the proposed method are verified.

Key words: optimization; satisfactory optimization; feature selection; recognition

1 引言 (Introduction)

在模式识别、机器学习和数据挖掘等领域中, 特征选择因能降低特征向量维数、减少特征提取代价、简化分类器设计和提高识别率而成为近年来一个充满生机和活力的研究课题, 受到人们的广泛关注^[1~3]. 特征选择的任务是利用模式样本集内部信息, 从一组特征中挑选出性能最优的特征集. 使分类错误概率最小的特征集在理论上是最优的, 但分类错误概率的计算依赖于所采用的特殊分类器和通常未知的类条件分布密度, 且特征选择是一个典型 NP 难解的有约束组合优化问题, 除了穷尽搜索方法外,

难以保证得到的特征集是最优解^[1], 所以, 在实际应用中, 人们总是寻找可计算和实用的类分离判据和性能较好的搜索算法, 以期获得较好的次优解^[2,3]. 由此可知, 特征选择实质上是一个满意优化问题, 得到的解均是满意解. 而且, 已有的特征选择算法较少考虑特征集的维数和特征获取的代价, 造成搜索次优解的过程中需要事先指定特征集的维数和得到的特征集不经济或效率低等问题, 因此, 有必要将多目标满意优化思想引入到特征选择中来综合考虑多种影响因素和评价选出的特征集的质量满意程度.

收稿日期: 2004-04-22; 收修稿日期: 2005-04-19.

基金项目: 国家自然科学基金资助项目(60572143); 国家电子对抗重点实验室基金项目(NEWL51435QT220401); 西南交通大学博士生创新基金资助项目(2003-12); 教育部骨干教师资助计划项目(教技司[2000]65号).

满意优化是针对最优解根本不存在或难以把握的优化问题,或者存在最优解但无法求得或求解代价太大的优化问题而提出来的^[4~8]. 满意优化有着广泛的应用背景,如靳蕃提出神经计算的满意解原理^[4],席裕庚提出复杂工业过程的满意控制^[5],金炜东对满意优化的原理和方法作了系统研究,提出了两类满意优化模型,并用于控制系统参数优化和列车操纵优化中^[6],赵舵等^[7]和张葛祥等^[8]对满意优化展开了进一步研究,分别将它用于模糊控制器设计和多变量控制系统设计中. 这些应用均取得了非常显著的优化效果^[4~8]. 满意优化摒弃了传统的最优概念,强调的是“满意”而不“最优”,它将优化问题的约束和目标融为一体,将性能指标要求的满意设计与参数优化融为一体,具有很大的适用性和灵活性^[4~8].

本文将满意优化思想引入到模式识别中,提出满意特征评价准则来对所提取的特征和挑选出的特征子集进行满意程度的质量评价,并提出满意特征选择方法(SFSM)来从大量特征中选择出具有维数少、复杂度低和分类能力强的满意特征集. 在给出特征满意度和特征集满意度的定义之后,通过设计各个满意度函数,导出满意特征选择的计算模型,并详细描述满意特征的选择算法. 为验证方法的有效性,将SFSM应用来进行雷达辐射源信号特征的筛选,并与基于距离准则的顺序前进法(SFSDC)^[2]、新特征选择法(NFS)^[3]和多目标遗传算法(MOGA)^[9,10]进行比较,结果显示,SFSM在时间复杂性和选出的特征品质方面都优于SFSDC,NFS和MOGA.

2 满意特征评价与选择(Satisfactory feature evaluation and selection)

2.1 评价准则(Evaluation criterion)

衡量特征和特征集的满意程度,首先需要给出评价标准,即满意度. 满意度是满意优化中最基本、最重要的概念,根据不同应用背景和研究方法,已有几种满意度定义被提出^[4~6]. 根据样本的特点,这里给出如下特征满意度和特征集满意度定义:

定义1 对于某一种特征 t ,设有 n 个分类类别,特征满意度定义为

$$\begin{cases} S_i = S(J_i), S_i \in [0,1], \\ J_i = J(J_{ij}), 1 \leq i < j \leq n, \\ J_{ij} = JT(X_i, X_j). \end{cases} \quad (1)$$

式中 $S(J_i)$ 为特征满意度函数,是一单调不减函数,

所得的函数值表示对特征 t 进行评价后所获得的满意度大小, J_i 为对于特征 t 的类别可分离性判据; $J(\cdot)$ 为类别可分离性综合函数,一般可取线性加权函数,max和min函数等. $JT(\cdot)$ 为第 i 类与第 j 类的可分离性准则函数,如距离准则函数、概率准则函数和熵准则函数等, J_{ij} 表示 $JT(\cdot)$ 的函数值, X_i 和 X_j 表示第 i 类与第 j 类的特征样本向量,若样本数为 M ,则 $X_i = [x_{i1}, x_{i2}, \dots, x_{iM}]$, $X_j = [x_{j1}, x_{j2}, \dots, x_{jM}]$.

特征满意度用于对单个特征品质优劣进行评价,满意度值越大,则表示该特征的分类能力越强,反之,则意味着该特征的分类能力越弱. 采用特征满意度对若干特征进行评价后,可采用一些搜索算法(如顺序前进法等)来求得令人满意的解. 但是,采用这些基于排序思想的优化算法得到的特征向量往往不太令人满意,甚至有时会得到质量很差的特征向量,所以,给出特征满意度的定义还不够,还需要给出特征集的满意度定义,以便于挑选质量更好的特征集. 特征集的满意程度不仅与类别可分离性有关,还与该特征集的维数和特征提取的复杂性有关. 所选特征集的维数大小会直接影响到分类器设计和特征提取的代价,而特征复杂度的约束会使挑选出的特征的代价减少. 由此,给出特征集满意度定义如下:

定义2 设有一特征集 $T = [t_1, t_2, \dots, t_z]$, Z 为特征集维数且 $Z \geq 2$,特征集满意度定义如下:

$S_{is} = \varphi(s_d, s_c, s_m)$, $S_{is} \in [0,1]$, (2)
式中 $\varphi(\cdot)$ 为特征集满意度函数, s_d 为特征集维数的满意度, s_c 为特征集复杂性满意度, s_m 为特征集类别可分离性满意度, s_d , s_c 和 s_m 的定义如下:

定义3 设特征集的维数为 d ,且 $d_{\min} \neq d_{\max}$,则该特征集的维数满意度定义为

$$s_d = f(d) = \frac{d_{\max} - d}{d_{\max} - d_{\min}}, s_d \in [0,1], \quad (3)$$

式中 $f(\cdot)$ 为维数满意度函数, $d_{\max}$ 为原始特征集的维数, $d_{\min}$ 为希望选出的特征集的最小维数,若 $d_{\min} = d_{\max}$,则令 $s_d = 1$.

维数满意度反映了设计人员对选出的特征集维数的满意程度. 考虑维数满意度是为了后续优化算法在搜索过程中能自动确定满意特征集的维数,可以解决当前大多特征可分离性评价准则需要人为预先确定所选特征集的维数和需要多次试探不同维数的问题.

定义4 设有一特征集 $T = [t_1, t_2, \dots, t_z]$, Z 为特征集维数,且 $C_{\min} \neq C_{\max}$ 则该特征集的复杂性

满意度定义为

$$s_c = g(C) = \frac{C_{\text{sum}} - C}{C_{\text{sum}} - C_{\text{min}}}, s_c \in [0, 1]. \quad (4)$$

式中若 $C_{\text{min}} = C_{\text{sum}}$, 则 $s_c = 1$, $g(\cdot)$ 为复杂性满意度函数, C_{min} 为原始特征集中复杂性的最小值, C_{sum} 为原始特征集中所有特征复杂性的和, 即

$$C_{\text{sum}} = \sum_{i=1}^L C_i. \quad (5)$$

其中 L 为原始特征集维数, C_i 为第 i 个特征的复杂性. 在式(4)中, C 为特征集 T 的复杂性, 即

$$C = \sum_{i=1}^Z C_i. \quad (6)$$

在模式识别中, 特征提取需要花费代价, 特征不同, 代价也不同, 引入复杂性满意度可克服当前多数特征评价准则没有涉及特征提取的代价问题. 根据实际问题的不同, 特征提取复杂性可用时间、能量等来进行定量计算, 不过, 对于大多数模式识别问题, 如信号识别、图像识别等, 特征的复杂性多表现为获取特征所需的时间.

定义 5 特征集类别可分离性满意度主要是对多种特征组成的多维特征向量 $T = [t_1, t_2, \dots, t_Z]$ (Z 为特征集维数) 的分类能力满意程度的评价, 定义如下:

$$\begin{cases} s_m = \psi(J^m), s_m \in [0, 1], \\ J^m = JM(J_{ij}^m), 1 \leq i < j \leq n, \\ J_{ij}^m = JTM(X_i^m, X_j^m). \end{cases} \quad (7)$$

式中 $\psi(J^m)$ 为特征集类别可分离性满意度函数, 是一单调不减函数, J^m 为特征集类别可分离性判据; $JM(\cdot)$ 为特征集类别可分离性综合函数. $JTM(\cdot)$ 为第 i 类与第 j 类的特征集可分离性准则函数, J_{ij}^m 表示 $JTM(\cdot)$ 的函数值, X_i^m 和 X_j^m 表示第 i 类与第 j 类的特征集样本矩阵, 若样本数为 M , 则:

$$X_i^m = \begin{bmatrix} x_{i1}^{t_1} & x_{i2}^{t_1} & \cdots & x_{iM}^{t_1} \\ x_{i1}^{t_2} & x_{i2}^{t_2} & \cdots & x_{iM}^{t_2} \\ \vdots & \vdots & & \vdots \\ x_{i1}^{t_Z} & x_{i2}^{t_Z} & \cdots & x_{iM}^{t_Z} \end{bmatrix}, \quad (8)$$

$$X_j^m = \begin{bmatrix} x_{j1}^{t_1} & x_{j2}^{t_1} & \cdots & x_{jM}^{t_1} \\ x_{j1}^{t_2} & x_{j2}^{t_2} & \cdots & x_{jM}^{t_2} \\ \vdots & \vdots & & \vdots \\ x_{j1}^{t_Z} & x_{j2}^{t_Z} & \cdots & x_{jM}^{t_Z} \end{bmatrix}. \quad (9)$$

综上所述, 满意特征集评价准则为

$$\begin{cases} S_{ts} = \varphi(s_d, s_c, s_m), \\ s_d = f(d), \\ s_c = g(C), \\ s_m = \psi(J^m), \\ J^m = JM(J_{ij}^m), 1 \leq i < j \leq n, \\ J_{ij}^m = JTM(X_i^m, X_j^m). \end{cases} \quad (10)$$

2.2 特征选择算法 (Feature selection algorithm)

根据满意特征评价准则, 下面逐步描述从一组原始特征集中挑选出满意特征集的步骤:

Step 1 确定待分类类别数 n 和原始特征集的维数 L , 并对于每一种特征的每一类模式, 均产生 M 个特征样本, 于是, 总特征样本数为 $L \times n \times M$.

Step 2 确定参数 d_{max} 和 d_{min} . 一般地, d_{max} 取为原始特征集的维数, d_{min} 取为 1. 并计算原始特征集中各个特征的复杂性, 根据定义 4 求得参数 C_{sum} 和 C_{min} .

Step 3 采用 QGA^[11] 得到一组特征集. QGA 是新发展起来的一种基于量子计算原理的概率优化方法, 它以量子计算的一些概念和理论为基础, 用量子位编码来表示染色体, 用量子门作用和量子门更新染色体来完成进化搜索, 具有种群规模小而不影响算法性能、同时兼有“勘探”和“开采”的能力、收敛速度快和全局寻优能力强的特点^[11]. 采用 QGA 获得一组特征集的具体过程为:

1) QGA 的初始化: 确定种群大小 h 和量子位的数目 q , 包含 h 个个体的种群 $P = \{p_1, p_2, \dots, p_h\}$, 其中 $p_j (j = 1, 2, \dots, h)$ 为种群中的第 j 个个体, 描述如下:

$$p_j = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_q \\ \beta_1 & \beta_2 & \cdots & \beta_q \end{bmatrix}. \quad (11)$$

式中 $\alpha_i, \beta_i (i = 1, 2, \dots, q)$ 分别为第 i 个量子位量子基态 $|1\rangle$ 和 $|0\rangle$ 的概率幅, 初始时所有 α_i, β_i 均取为 $1/\sqrt{2}$, 表示在初始搜索时所有状态以相同的概率进行叠加. 与传统遗传算法相比, QGA 的种群 h 可取得很小, 一般取为 20 左右即可^[11], 量子位数目 q 取为与原始特征集的维数相等.

2) 根据 P 中各个体的概率幅构造出量子叠加态的观测态 $R, R = \{a_1, a_2, \dots, a_h\}$, 其中 $a_j (j = 1, 2, \dots, h)$ 为每个个体的观测状态, 即一个由“0”和“1”组成的二进制串, 于是可以得到 h 个二进制串, 每一个二进制串就代表着一个特征集, 其中为“1”的位表示该特征包含在此特征集中, 为“0”的位表

示该特征不包含在此特征集中。

Step 4 根据各特征集的维数分别计算它们的维数满意度,计算公式已由式(3)给出。

Step 5 由各特征集的复杂性分别计算它们的复杂性满意度,计算公式已由式(4)给出。

$$JTM(X_i^m, X_j^m) = 1 - \frac{\int \cdots \int_Z f(x_1, \dots, x_Z) \cdot g(x_1, \dots, x_Z) dx_1 \cdots dx_Z}{\sqrt{\int \cdots \int_Z f^2(x_1, \dots, x_Z) dx_1 \cdots dx_Z} \sqrt{\int \cdots \int_Z g^2(x_1, \dots, x_Z) dx_1 \cdots dx_Z}}. \quad (13)$$

在式(12)中, $w_{ij} (1 \leq i < j \leq n)$ 一般取为相等,当然,若有特殊要求时,也可取不相等。在式(13)中, Z 为特征子集的维数, $f(x_1, \dots, x_Z)$ 和 $g(x_1, \dots, x_Z)$ 分别为特征子集中第 i 类与第 j 类特征样本分布函数,应用场合不同,特征样本的分布函数可能有所不同,但一般可取数学期望为 μ_i, μ_j , 方差为 σ_i, σ_j 的 Gaussian 函数,即 $\mu_i = E(X_i^m), \mu_j = E(X_j^m), \sigma_i = D(X_i^m), \sigma_j = D(X_j^m)$ 。从式(13)可看出,当两类特征样本分布函数在多维特征空间中完全分离时,则该特征集分类能力满意度值达到最大值 1,随着两类特征样本分布函数在高维特征空间中交叠程度的增加,满意度值会逐渐减小,直到两分布函数完全重合时,满意度值变为最小值 0。顺便指出,当采用式(1)对单个特征的品质进行满意度评价时,特征满意度函数 $s(\cdot)$ 和类别可分离性综合函数 $j(\cdot)$ 分别可采用与 $\psi(\cdot)$ 和 $JM(\cdot)$ 相同的函数,而两类可分离性准则函数 $JT(\cdot)$ 可采用 $JTM(\cdot)$ 的特殊形式,即函数 $JTM(\cdot)$ 为一维函数时的情况。

Step 7 特征集满意度函数 $\varphi(\cdot)$ 只与 3 种因素 s_d, s_c 和 s_m 有关,因而可直接取为较简单的线性加权函数,即

$$\varphi(s_d, s_c, s_m) = \frac{\omega_d s_d + \omega_c s_c + \omega_m s_m}{\omega_d + \omega_c + \omega_m}. \quad (14)$$

式中 $\omega_d, \omega_c, \omega_m$ 分别为 s_d, s_c, s_m 的加权系数。

Step 8 将 Step 7 中的特征集满意度函数作为 QGA 的适应度函数对种群中的所有个体进行适应度评价。

Step 9 保留最佳个体,并判断是否满足终止条件,若满足,则算法终止,否则,执行下一步。

Step 10 采用文献[11]中的方法计算量子旋转门的旋转角,并将量子门作用于种群中所有个体的概率幅,即将种群 P 中的所有个体的概率幅进行更新。

Step 11 QGA 的进化代数增 1,算法转至

Step 6 根据式(7)计算各特征集的分类能力满意度,在计算过程中,各函数设计如下:

$$\begin{cases} \psi(J^m) = (J^m)^2, \\ JM(J_{ij}^m, 1 \leq i < j \leq n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_{ij} J_{ij}^m. \end{cases} \quad (12)$$

Step 3 继续进行,直到算法结束。

SFSM 的全局收敛性直接由 QGA 的全局收敛性决定,在文献[12]中,采用随机泛函理论对 QGA 的全局收敛性进行了证明,限于篇幅,这里就不再详细讨论。

3 应用实例(Application example)

在我们先前的工作中,已经从 10 种典型雷达辐射源信号中分别提取出了 16 种特征。这 10 种信号分别为: CW, FD, LFM, NLFM, BPSK, QPSK, MPSK, IPFE, FSK 和 CSF。16 种特征分别为信息维数、盒维数、关联维数、两种相像系数、近似熵、Lempel-Ziv 复杂度、范数熵和小波包 3 层分解所得的从低频到高频的 8 个特征。将这 16 种特征组成原始特征集,并分别用标号 1~16 来表示。

在实验前,先对每一种雷达辐射源信号在 5~20dB 的信噪比范围内每隔 5dB 产生 150 个样本,这样,每一种雷达辐射源信号总共有 600 个样本,其中 200 个样本用于特征满意度的评价、特征选择和分类器训练,其余 400 个样本用于分类识别的测试实验。为了便于理解特征满意度的概念和对照分析下面几种方法挑选出的特征集的满意度与特征集所包含的各个特征的满意度之间的关系,先采用特征满意度来分别评价这 16 种特征的质量,得到的从大到小的排序结果如表 1 所示。特征 12 和特征 6 分别为最满意和最不满意的特征。

表 1 原始特征集中各特征的满意度(SR)和排序结果(OR)

Table 1 Satisfactory rates (SR) of each feature of original feature set and ordering results(OR)

特征	1	2	3	4	5	6	7	8
SR	0.798	0.734	0.902	0.922	0.959	0.708	0.887	0.809
OR	13	14	7	6	2	15	8	11
特征	9	10	11	12	13	14	15	16
SR	0.953	0.941	0.853	0.971	0.922	0.844	0.805	0.950
OR	3	5	9	1	6	10	12	4

当采用 SFSM 进行特征选择时,根据前面的分析,参数 $L = 16$, $n = 10$, $M = 200$, $d_{\max} = 16$, $d_{\min} = 1$, $C_{\text{sum}} = 2163.503$, $C_{\min} = 6.203$, $P = 20$, $m = 16$. 满意度指标 s_d , s_e 和 s_m 的权值选择对算法的结果将会产生一定的影响,根据经验^[4-8]和多次反复试验, ω_d , ω_e 和 ω_m 分别取为 0.20, 0.10 和 0.70 比较合适. 算法运行 100 次,所得特征集均由特征 5 和特征 10 组成,特征集满意度为 0.9817. 该特征集并非由表 1 中的满意度值最大的两特征构成,说明采用特征集满意度选出的特征并不是简单的按照单个特征满意度值进行排序选择的,这正体现了特征集满意度还要受维数和复杂性的影响. 为了进行对比,本文还采用 SFSDC^[2], NFS^[3] 和 MOGA^[9,10] 进行特征选择, SFSDC 的距离准则函数为

$$G_i = \sum_j \frac{(m_i - m_j)^2}{\sigma_i^2 + \sigma_j^2}, (i = 1, 2, \dots, U). \quad (15)$$

式中 m_i, m_j 分别为第 i 类和第 j 类信号的样本均值, σ_i, σ_j 分别为第 i 类和第 j 类信号的样本方差, U 为特征总数. 由于 SFSDC 和 NFS 都需要事先指定特征集的维数,为了与 SFSM 进行识别结果比较, SFSDC 和 NFS 的维数都指定为与采用 SFSM 所获得的特征集维数相同. MOGA 的种群大小取为 100, 交叉和变异概率分别取为 0.8 和 0.007. 算法的最大进化代数设为 500, 4 种特征选择方法消耗的计算时间的 100 次实验的统计值如表 2 所示. 实验是在 Pentium IV (CPU 2.0G, RAM 512MB) 上完成的.

表 2 4 种方法的结果对比

Table 2 Comparison results of 4 methods

方 法	SFSM	SFSDC	NFS	MOGA
计算时间	46.55	24.31	48.29	57.32
特征组成	5, 10	4, 5	6, 7	1, 12
满意度	0.98	0.89	0.94	0.96
BPSK/(%)	0	65.49	51.89	17.89
QPSK/(%)	2.12	43.88	31.87	0.06
MPSK/(%)	5.86	99.98	21.95	0.11
LFM/(%)	0.02	0	2.15	8.06
NLFM/(%)	0	0	29.29	20.00
CW/(%)	0.46	0	0.24	0
FD/(%)	0	0	16.63	0
FSK/(%)	0.12	0.06	36.58	0
IPFE/(%)	0	10.02	1.79	0
CSF/(%)	0	0.12	45.12	40.00
平均误识率/(%)	0.86	21.96	23.75	8.61
训练时间/s	38.33	498.39	129.70	139.48
测试时间/s	1.84	1.84	1.83	1.86

将 SFSDC 和 NFS 的维数都取为 2 并进行特征选择, SFSDC 获得的最佳特征集由特征 4 和 5 组成, NFS 获得的最佳特征集由特征 6 和 7 组成, MOGA 获得的最佳特征集则特征 1 和 12 构成, 各特征集满意度已在表 2 中给出. 分别将 SFSM, SFSDC, NFS 和 MOGA 获得的特征集用来在 5 ~ 20dB 的信噪比范围内进行分类识别测试实验, 实验中所采用的分类器为神经网络分类器, 其结构为 2-15-10, 输入层神经元的数目与特征向量的维数相同, 输出层神经元的数目与待分类信号数相同, 隐含层和输出层分别采用“tansig”和“logsig”传递函数, 训练算法采用弹性反向传播算法, 监督信号取为“0-1”型信号, 输出容差为 0.05, 误差取为 0.001, 最大训练代数取为 10000. 经过 100 次实验所得的分类器平均训练时间、测试时间和 10 种雷达辐射源信号的错误识别率如表 2 所示.

表 2 显示, SFSM 的计算时间比 NFS 和 MOGA 少, 比 SFSDC 多, 但 SFSM 选择出的特征的满意度最高. 从选出特征的质量来看, SFSM 获得的平均误识率仅为 0.86%, 分别比 SFSDC, NFS 和 MOGA 少 21.10, 22.89 和 7.75 个百分点, 且 SFSM 选出的特征集需要的分类器训练时间比 SFSDC, NFS 和 MOGA 少很多, 不过 4 种方法需要的分类器测试时间基本相同. 由此可看出, SFSM 在计算效率和选出特征的品质方面均大大优于 SFSDC, NFS 和 MOGA 3 种方法.

4 结束语 (Conclusions)

在已有的特征选择法中, 较少考虑特征向量的维数和特征提取的代价问题, 这往往会导致搜索算法在寻找较优特征集时需要逐步多次的反复试探和资源浪费或效率低等问题. 本文引入满意特征选择法来从大量特征中挑选出质量、维数和复杂性都令人满意的特征集, 获得了比 SFSDC, NFS 和 MOGA 更高的正确识别率和识别效率. 满意特征评价准则中的各满意度函数可根据不同的应用情况和设计意图进行调整, 具有较大的灵活性, 同时也具有一定的不确定性. 在本文方法中, 量子遗传算法的性能和满意度指标的权值选择对算法的性能有一定影响, 选择高效快速的搜索策略和权值的自适应调整都是需要进一步研究的问题.

参考文献 (References):

- [1] COVER T M, VANCOMPENHON J M. On the possible ordering in the measurement selection problem [J]. *IEEE Trans on Systems, Man and Cybernetics*, 1977, 7(9): 657-667.

- [2] MOLINA LUIS-CARLOS, BELANCHE LLUFS, NEBOT ANGELA. Feature selection algorithms: a survey and experimental evaluation [C]// *Proc of Int Conf on Data Mining*. Piscataway: IEEE Press, 2002: 306-313.
- [3] 吕铁军, 王河, 肖先赐. 新特征选择方法下的信号调制识别[J]. 电子与信息学报, 2002, 24(5): 661-666.
(LÜ Tiejun, WANG He, XIAO Xianci. Recognition of modulation signal based on a new method of feature selection[J]. *J of Electronics and Information Technology*, 2002, 24(5): 661-666.)
- [4] 靳蕃, 胡飞. 模糊神经计算的满意输出原理[J]. 铁道学报, 1996, 18(2): 102-107.
(JIN Fan, HU Fei. Principles of satisfactory-solution for fuzzy neural computing [J]. *J of the China Railway Society*, 1996, 18(2): 102-107.)
- [5] 席裕庚. 复杂工业过程的满意控制[J]. 信息与控制, 1995, 24(1): 14-20.
(XI Yugeng. Satisfactory control of complex industrial process [J]. *Information and Control*, 1995, 24(1): 14-20.)
- [6] JIN Weidong, LI Chongwei, HU Fei, et al. A study on intelligent computation of methods of optimization operation for train [C]// *Proc of Int Workshop on Autonomous Decentralized System*. Los Alamitos: IEEE Press, 2000: 227-230.
- [7] ZHAO Duo. The application of multi-criterion satisfactory optimization in fuzzy controller design [C]// *Proc of 2nd Int Workshop on Autonomous Decentralized System*. Los Alamitos: IEEE Press, 2002: 162-167.
- [8] 张葛祥, 金炜东, 胡来招. 多变量系统控制器的参数满意优化设计[J]. 控制理论与应用, 2004, 21(3): 362-366.
(ZHANG Gexiang, JIN Weidong, HU Laizhao. Parameters satisfactory optimization of controller in multivariable control system [J]. *Control Theory & Applications*, 2004, 21(3): 362-366.)
- [9] MARSEGUERRA M, ZIO E, PODOFILINI L. Optimal reliability availability of uncertain systems via multi-objective genetic algorithms[J]. *IEEE Trans on Reliability*, 2004, 53(3): 424-434.
- [10] OSMAN M S, ABO-SINNA M A, Mousa A A. An effective genetic algorithm approach to multiobjective routing problems (MORPs)[J]. *Applied Mathematics and Computation*, 2005, 163(2): 769-781.
- [11] 张葛祥, 金炜东, 胡来招. 基于量子遗传算法的特征选择算法[J]. 控制理论与应用, 2005, 22(5): 810-813, 819.
(ZHANG Gexiang, JIN Weidong, HU Laizhao. Feature Selection algorithm based on quantum genetic algorithm [J]. *Control Theory & Applications*, 2005, 22(5): 810-813, 819.)
- [12] ZHANG Gexiang, JIN Weidong, HU Laizhao. Quantum evolutionary algorithm for multi-objective optimization problem [C]// *Proc of IEEE Int Sym on Intelligent Control*. Piscataway: IEEE Press, 2003: 703-708.

作者简介:

张葛祥 (1974—), 男, 西南交通大学电气工程学院教授, 博士, IEEE Associate Member, IRSS 和 CAAI 会员, 研究领域为满意优化、模式识别、信号处理、进化计算、支持向量机等, 出版著作 1 部, 在国内外期刊、会议上发表学术论文 40 余篇, 其中有近 30 篇进入 SCI 或 EI 检索, E-mail: gxzhang@ieee.org;

金炜东 (1959—), 男, 西南交通大学电气工程学院教授, 博士生导师, 博士, 研究领域为优化理论与优化控制、智能信息处理, 系统仿真等;

胡来招 (1945—), 男, 中国电子科技集团第 29 所研究员, 博士生导师, 博士, 研究领域为信号处理, 侦察接收机, 无源定位等。