

马氏田口系统的量子行为二进制粒子群特征选择优化方法

刘久富¹, 郑锐¹, 丁晓彬¹, 刘海阳², 杨忠¹, 王志胜¹

(1.南京航空航天大学自动化学院, 江苏南京 211106; 2.东南大学电子科学与工程学院, 江苏南京 210096)

摘要:针对标准二进制粒子群用于马氏田口系统的特征选择优化时, 存在迭代速度慢, 容易陷入局部最优解等不足, 提出一种改进的基于量子行为二进制粒子群的马氏田口系统变量选择优化方法。首先, 为了规避可能存在的复共线性特性对距离度量结果的影响, 本研究采用Gram-Schmidt正交化法计算马氏距离值, 对系统进行标准化处理, 对各属性向量进行正交化后计算各类别的马氏距离集合, 通过ROC曲线确定系统分类的最佳阈值点, 定义误分类率概念和被选择变量占比最小作为变量筛选标准, 构建多目标的混合规划模型。运用改进的量子行为粒子群算法求解优化组合, 为适应二值化的变量优化问题, 算法基于概率对粒子进行二进制编码, 求取目标函数的适应值, 并完成粒子群的优化迭代过程。采用优化的变量组合, 构建精简的马氏田口系统, 建立度量预测模型, 完成精确判别任务。最后, 以胎心分娩力造影术测量的胎儿健康诊断为例, 对标准二进制粒子群算法和二进制量子粒子群优化算法进行对比验证, 实验结果表明, 本文方法可以有效地提升粒子的迭代速度和寻优精度, 优化后的马氏田口系统的预测准确率明显提高。

关键词:马氏田口系统; 特征选择; 量子行为二进制粒子群; 优化

中图分类号: TP206.2

文献标志码: A

文章编号: 2096-3246(2019)06-0152-07

Feature Selection Optimization for Mahalanobis-Taguchi System with Binary Quantum Behavior Particle Swarm

LIU Jiufu¹, ZHENG Rui¹, DING Xiaobin¹, LIU Haiyang², YANG Zhong¹, WANG Zhisheng¹

(1.College of Automation, Nanjing Univ. of Aeronautics and Astronautics, Nanjing 211106, China;

2.College of Electronic Sci. and Eng., Southeast Univ., Nanjing 210096, China)

Abstract: When the standard binary particle swarm is used for the feature selection optimization of Mahalanobis-Taguchi system, the computational speed is slow and the selected features combination of Mahalanobis-Taguchi system is prone to falling into the local optimal solution. To address these problems, a feature selection optimization method of Mahalanobis-Taguchi system based on an improved binary quantum behavior particle swarm was proposed. Firstly, in order to avoid the influence of complex collinearity for the distance metric, the Gram-Schmidt orthogonalization method was used to calculate the Mahalanobis distance value. Through the ROC curve, the optimal threshold point for the system classification was determined. The misclassification rate and the selected variables rate were defined and a multi-objective optimization model was built. Then, an improved quantum behavior particle swarm optimization algorithm was presented to solve the optimization model, which performs binary coding on the particle based on probability. Through the optimized features combination, a new Mahalanobis-Taguchi prediction system was established. Finally, the fetal health diagnosis was carried out. The experimental results showed that the improved quantum behavior particle swarm optimization algorithm could effectively enhance the iterative speed, and the optimized Mahalanobis-Taguchi system had the better prediction accuracy.

Key words: Mahalanobis-Taguchi system; variable selection; binary quantum behavior particle swarm; optimization

马氏田口系统(Mahalanobis-Taguchi system, MTS)是多特征变量数据集的模式识别与分类、诊断和预

测的新方法^[1-5], 传统马氏田口系统在系统特征选择优化过程中, 通过正交表计算信噪比优选变量组合,

收稿日期: 2019-01-16

基金项目: 国家自然科学基金项目(61473144)

作者简介: 刘久富(1970—)男, 副研究员, 博士。研究方向: 机器学习; 模式识别; 数据挖掘等。E-mail: liujiufu2@126.com

网络出版时间: 2019-10-31 14:42:51

网络出版地址: <http://kns.cnki.net/kcms/detail/51.1773.TB.20191031.1323.004.html>

但实际应用中正交表优化方法效率低,筛选后的实际系统分类性能也未必达到预期效果。

曾江辉等^[6]提出采用稳健性设计原理和模糊数学对MTS的基准空间进行优化,没有涉及研究各个特征变量的影响与特征变量的优化选择。Iquebal等^[7]引入粗糙集建立优化模型,用信噪比与正交阵列方法,优化选择特征变量,但随着特征变量数目增大,计算速度会大大降低。牛俊磊^[8]综合考虑分类精度、望大特性信噪比和降维程度等目标,构建了多目标优化模型,并用基于遗传算法的全局准则、全方位优化算法求解以筛选特征变量和确定阈值,该方法可导致局部寻优结果。

Reséndiz等^[9]在MTS系统优化中采用二进制蚁群算法取代正交表筛选方法,优化速度较慢。Pal等^[10]将二进制粒子群优化(binary particle swarm optimization, BPSO)应用到MTS模型降维中,计算速度有较大的提高,但优化结果容易趋于局部最优。传统的粒子群优化(particle swarm optimization, PSO)算法的粒子收敛以轨道形式实现,同时粒子速度会受到限制,使得搜索范围是一个有限的区域,很难覆盖整个解空间,这就导致了PSO算法不能保证以概率1收敛到全局。而量子行为粒子群优化(quantum behaved particle swarm optimization, QPSO)理论将粒子引入到量子空间中,量子束缚态的粒子能够依赖一定的概率密度到达空间中的任意点,可行解空间能够达到全局,且不会发散到无穷远处。

本文提出将MTS模型与量子行为粒子群算法相结合,运用改进的二进制量子行为粒子群算法求解MTS多目标优化问题,实现特征变量的约简,提高优化的计算效率与计算精度。本文首先融合Gram-Schmidt正交化法和ROC曲线法,引入误分类率概念,构建多目标的混合0-1降维模型。然后引入全局搜索能力更强、收敛性更好的量子行为粒子群对马氏田口系统进行变量优化,并对量子行为粒子群进行二进制编码的改进,以适应目标函数的求解,通过迭代计算完成特征变量的优选。

1 马氏田口系统

马氏田口系统以马氏距离(Mahalanobis distance, MD)、信噪比和正交表为工具完成系统的诊断和预测,具体操作步骤如下^[11]:

1) 构建马氏基准空间

需要从收集的原始样本数据中确定正常和异常的样本,选择合适的特征变量。选取正常组别的 n 个样本,特征变量数为 k ,先依据式(1)对向量组的元素进行标准化,再根据式(2)使用逆矩阵法计算各正常

样本的马氏距离,以构建马氏基准空间。

$$z_{ij} = \frac{1}{s_j}(x_{ij} - \bar{x}_j), i = 1, 2, \dots, n, j = 1, 2, \dots, k \quad (1)$$

$$MD_i = d_i^2 = \frac{1}{k} \mathbf{Z}_i^T \mathbf{S}^{-1} \mathbf{Z}_i, i = 1, 2, \dots, n \quad (2)$$

式中: $\mathbf{Z}_i = [z_{i1}, z_{i2}, \dots, z_{ik}]$, $i = 1, 2, \dots, n$ 为第 i 个样本的归一化向量; \bar{x}_j 表示第 j 维变量的所有样本的数据均值; s_j 表示第 j 维变量的所有样本的数据标准差; \mathbf{S}^{-1} 表示样本总体的关联矩阵的逆矩阵。

2) 验证基准空间有效性

一般在基准空间的基础上,仍沿用之前正常样本的均值、标准差和关联矩阵,对训练数据集计算马氏距离,通常采用的训练数据是已经预分类为异常的样本数据。如果异常样本的马氏距离相较于基准的马氏距离皆偏大,能够很好地区分两种类别的数据,则说明该基准空间有效。

3) 基准空间的优化

MTS引入正交表和信噪比进行系统特征属性的优化约简,采用二水平正交表,根据特征的数量选择合适的规格,按照每一行选定的变量依次进行实验,建立基准空间,依据式(3)计算异常样本的信噪比值^[5]:

$$SN_i = -10 \log \left[\frac{1}{n} \sum_{i=1}^n (1/MD_i^2) \right] \quad (3)$$

对于变量 x_j ,计算正交表中所有选中了该变量的信噪比均值,记作 SN_j^+ ;再计算该变量未被选中的信噪比均值,记作 SN_j^- ;用 $SN_j^+ - SN_j^-$ 表示该变量增益值。若变量的增益值为正,则保留;否则,剔除。用新筛选的变量重新构建基准空间,同样需要对基准空间检验其有效性,计算正常样本和异常样本的信噪比,验证优化后的信噪比相较于筛选前是否更高。

4) 确定阈值并预测分类

确定分类阈值 T ,使用优化后的测度空间,计算各未知样本的 MD ,未知样本依据阈值实现系统的诊断和预测。对于二分类问题,即把样本分为正常类和异常类两种,如果样本计算得到马氏距离小于阈值 T ,则被归类为正常类;反之,被归类为异常类。

2 降维模型

2.1 基于Gram-Schmidt正交化法的马氏距离

马氏距离是MTS中用于度量样本间相似程度的测量尺度^[2,12-13]。考虑到变量之间会存在多重共线性而导致样本间的关联矩阵无法求逆,本文通过Gram-Schmidt正交化法(Gram-Schmidt orthogonalization process, GSP)求取马氏距离,可以使得样本的各属性间完全正交,以避免多重共线性对距离度量的影响。

Gram-Schmidt正交化处理先要对初始向量组依据式(1)进行标准化处理,再对线性独立向量组计算向量的正交集。GSP的计算过程如下:

$$\begin{cases} \mathbf{U}_1 = \mathbf{Z}_1, \\ \mathbf{U}_2 = \mathbf{Z}_2 - \frac{\mathbf{Z}_2^T \mathbf{U}_1}{\mathbf{U}_1^T \mathbf{U}_1} \mathbf{U}_1, \\ \vdots \\ \mathbf{U}_k = \mathbf{Z}_k - \frac{\mathbf{Z}_k^T \mathbf{U}_1}{\mathbf{U}_1^T \mathbf{U}_1} \mathbf{U}_1 - \dots - \frac{\mathbf{Z}_k^T \mathbf{U}_{k-1}}{\mathbf{U}_{k-1}^T \mathbf{U}_{k-1}} \mathbf{U}_{k-1} \end{cases} \quad (4)$$

式中: $\mathbf{Z}_j = (z_{1j}, z_{2j}, \dots, u_{nj})^T$, $j = 1, 2, \dots, k$ 为 n 个样本中的第 j 维归一化向量; $\mathbf{U}_k = (u_{1k}, u_{2k}, \dots, u_{nk})^T$ 为第 k 维具有相同线性跨度的相互垂直向量。

则第 i 个样本的马氏距离由式(5)给出:

$$MD_i = \frac{1}{k} \left(\frac{u_{i1}^2}{s_1^2} + \frac{u_{i2}^2}{s_2^2} + \dots + \frac{u_{ik}^2}{s_k^2} \right) \quad (5)$$

式中, u_{ik} 为正交向量 \mathbf{U}_k 的元素, s_k 为向量 \mathbf{U}_k 中各元素的标准偏差, k 为变量的总数。

2.2 多目标优化模型

MTS通常用于二分类的问题,即存在正常类和异常类两种判别类型,MTS在变量优化约简的主要目标是希望通过变量的去冗余化,采用最优的变量组合使得分类的错误概率尽可能地降低。

目标1:定义了误分类率的概念作为变量筛选的标准,即分类过程中将正常组的样本错误地归类到异常组,或者将异常组的样本错误地归类到正常组的情况。两种不同类型的误分类结果的严重程度根据所涉及的系统可能有所不同。误分类率的测量是通过目标函数的加权和进行计算的,待优化的变量组合为 $\mathbf{X} = (x_1, x_2, \dots, x_p)$, 其中:

$$x_i = \begin{cases} 1, & \text{如果第 } i \text{ 个变量被选中;} \\ 0, & \text{如果第 } i \text{ 个变量未被选中。} \end{cases}$$

误分类率的表达式如下:

$$\begin{cases} f_1(\mathbf{X}) = \frac{n_1^e}{n_1}, \\ f_2(\mathbf{X}) = \frac{n_2^e}{n_2} \end{cases} \quad (6)$$

式中, n_1^e 表示正常观察值误分类为异常的数目, n_2^e 表示异常观察值误分类为正常的数目, n_1 表示正常组样本总数目, n_2 表示异常组样本总数目。

n_1^e 和 n_2^e 的值通过函数 $\text{Get_}n_1^e(MD^1)$ 和 $\text{Get_}n_2^e(MD^2)$ 获取。函数输入参数 $MD^1 = \{MD_1^1, MD_2^1, \dots, MD_{n_1}^1\}$ 和 $MD^2 = \{MD_1^2, MD_2^2, \dots, MD_{n_2}^2\}$ 分别为使用优化后的变量计算得到的正常组和异常组的马氏距离值; T 为优化前系统预测的阈值,这里采用ROC曲线法确定系统的最佳阈值点。

Get $n_1^e(MD^1)$ 函数伪代码如下:

```
n1e = 0;
for j=1 to n1 do
    if (MDj1 > T) n1e++;
end for
return n1e.
```

Get $n_2^e(MD^2)$ 函数伪代码如下:

```
n2e = 0;
for j=1 to n2 do
    if (MDj2 ≤ T) n2e++;
end for
return n2e.
```

目标2:通过降维使得系统的特征变量数最小。

被选择的特征变量率表示为:

$$f_3 = \frac{p_{\text{sel}}}{p} \quad (7)$$

式中, p_{sel} 表示被选择的特征变量数目。

本文以式(8)~(11)为MTS变量降维的多目标优化模型:

$$\min F(\mathbf{X}) = \omega_1 \frac{n_1^e}{n_1} + \omega_2 \frac{n_2^e}{n_2} + \omega_3 \frac{p_{\text{sel}}}{p} \quad (8)$$

s.t.

$$\sum_{i=1}^p x_i \leq p \quad (9)$$

$$\sum_{i=1}^p x_i = p_{\text{sel}} \quad (10)$$

$$f_1(\mathbf{X}) \leq f_{1\text{max}}, f_2(\mathbf{X}) \leq f_{2\text{max}} \quad (11)$$

式(8)中, ω_1 、 ω_2 、 ω_3 分别为权重。

3 优化求解

优化模型式(8)~(11)是一个混合的二进制整数规划问题,可以通过粒子群算法求解。为了提高计算效率,引入改进的二进制编码的量子行为粒子群算法,用于马氏田口模型的特征变量优化求解。

3.1 量子行为粒子群优化

量子行为粒子群优化^[14-15](quantum behaved particle swarm optimization, QPSO)基于量子力学概念,比PSO更容易收敛于全局最优点。量子力学的不确定性原理表明粒子的速度和位置的更新不能同时确定,故在QPSO中每个粒子都有自己的量子行为,相应的量子态函数由波函数 ψ 表示,用 $|\psi|^2$ 表示其位置的概率密度函数。为了增强粒子群的随机性,跳出局部搜索区域, QPSO算法中引入了局部吸引子的概念,如下:

$$p_{ij}(t) = \varphi_j(t)P_{ij\text{best}}(t) + (1 - \varphi_j(t))G_{j\text{best}}(t) \quad (12)$$

$$\varphi_j(t) = \frac{c_1 r_1}{c_1 r_2 + c_2 r_3} \quad (13)$$

式中, $p_{ij}(t)$ 为第*i*个粒子的第*j*维的吸引子, $P_{ij\text{best}}(t)$ 为第*i*个粒子第*j*维特征的局部最优位置, $\varphi_j(t)$ 由(0,1)区间上均匀分布的随机数 r_1 、 r_2 、 r_3 和正常数 c_1 、 c_2 组成。采用蒙特卡罗方法,通过式(14)获得粒子位置的更新:

$$x_{ij}(t+1) = \begin{cases} p_{ij}(t) + \beta |M_{j\text{best}}(t) - x_{ij}(t)| \ln(1.0/u_{ij}), r_6 \geq 0.5; \\ p_{ij}(t) - \beta |M_{j\text{best}}(t) - x_{ij}(t)| \ln(1.0/u_{ij}), \text{其他} \end{cases} \quad (14)$$

式中: $x_{ij}(t+1)$ 为第*i*个粒子的第*j*维的位置; u_{ij} 和 r_6 为(0,1)区间上均匀分布的随机数; β 为收缩-膨胀系数,其控制算法的收敛速度,可通过式(15)计算:

$$\beta = \beta_0 + \frac{(t_{\max} - t) \times (\beta_1 - \beta_0)}{t_{\max}} \quad (15)$$

式中, t 为当前迭代次数, t_{\max} 为最大迭代次数。Sun等^[14]建议将 β 从1.0线性减小到0.5,可以使QPSO算法获得更好的优化效果,一般情况下取 β_1 为1, β_0 为0.5。引入平均最优位置,记为 $M_{\text{best}}(t)$,定义为个体最优位置的平均值,可由式(16)得到:

$$M_{\text{best}}(t) = (M_{1\text{best}}(t), M_{2\text{best}}(t), \dots, M_{n\text{best}}(t)) = \left(\frac{1}{N} \sum_{i=1}^N P_{i1\text{best}}(t), \frac{1}{N} \sum_{i=1}^N P_{i2\text{best}}(t), \dots, \frac{1}{N} \sum_{i=1}^N P_{in\text{best}}(t) \right) \quad (16)$$

3.2 粒子编码原理

由于变量序列都是0、1的二进制字串,而QPSO的粒子是连续的数值串,并不能用于计算粒子的目标函数,所以有必要对粒子的位置数值进行二进制形式的转换。转换方法是把粒子位置映射到Sigmoid函数中,根据粒子的位置数值的大小得到(0,1)区间内的概率值,粒子根据该概率值的大小选择是否更新位置。转换依据公式如下:

$$x_{ij}(t+1) = \begin{cases} 0, \text{sig}(x_{ij}(t)) < \text{rand}(); \\ 1, \text{sig}(x_{ij}(t)) \geq \text{rand}() \end{cases} \quad (17)$$

式中, $\text{rand}()$ 是(0,1)区间上的随机数, Sigmoid函数表示如下:

$$\text{sig}(x) = \frac{1}{1 + e^{-x}}.$$

将连续值转换成二进制字串后的变量序列,能够用于计算目标函数的值,并进行后续有效的更新。

3.3 改进的量子行为二进制粒子群的变量约简算法

提出改进的量子行为二进制粒子群优化(improved binary quantum behaved particle swarm optimization, IBQPSO)的变量约简算法,算法的输入为特征变量序列 $\mathbf{X} = (x_1, x_2, \dots, x_k)$,种群大小为 N ,粒子的维数为 Dim ,粒子的上下界分别为 x_{\max} 和 x_{\min} ,输出为粒子的

全局最优适应度 F_{gbest} 和全局最优粒子序列 $\mathbf{X}_{\text{gbest}}$ 。

首先,完成粒子群的初始化,确定最大迭代次数 $MaxIter$,依据式(17)完成粒子位置的二进制编码。将每个粒子的位置代入目标函数计算每个粒子的适应度,并确定个体最优位置 $\mathbf{P}_{\text{best}}(t)$ 和全局最优位置 $\mathbf{G}_{\text{best}}(t)$ 。接着,进行粒子优化迭代过程,根据式(16)计算粒子的平均最优位置,式(12)计算局部吸引子并完成粒子位置的更新。最后,达到最大迭代次数时停止迭代过程,输出全局最优适应度 F_{gbest} 和粒子的优化序列 $\mathbf{X}_{\text{gbest}}$ 。作者提出的改进的量子行为二进制粒子群算法伪代码如下:

算法 改进的量子行为二进制粒子群算法

输入: 特征变量序列 $\mathbf{X} = (x_1, x_2, \dots, x_k)$, 种群大小 N , 粒子的维数 Dim , 粒子的上下界 x_{\max} 和 x_{\min} 。

输出: 粒子的全局最优适应度 F_{gbest} , 全局最优粒子序列 $\mathbf{X}_{\text{gbest}}$ 。

1 随机初始化粒子群中的每个粒子的当前位置 x_i , 确定最大迭代次数 $MaxIter$;

2 依据式(17)对粒子位置 $x_{ij}(t)$ 进行二进制编码, 计算每个粒子的适应度函数值, 确定个体最优位置 $\mathbf{P}_{\text{best}}(t)$ 和全局最优位置 $\mathbf{G}_{\text{best}}(t)$;

3 repeat

4 根据式(16)计算粒子的平均最优位置;

5 for ($i = 1; i \leq N; i = i + 1$) do

6 for ($j = 1; j \leq Dim; j = j + 1$) do

7 根据式(12)计算局部吸引子 $p_{ij}(t)$;

8 根据式(14)更新粒子的位置 $x_{ij}(t)$;

9 将粒子限制在搜索区间 $[x_{\min}, x_{\max}]$ 中;

10 end for

11 依据式(17)对粒子位置 $x_{ij}(t)$ 进行二进制编码;

12 计算当前粒子的适应度函数值, 并更新粒子的个体最优位置 $\mathbf{P}_{\text{best}}(t)$;

13 将 $\mathbf{P}_{\text{best}}(t)$ 与全局最优位置 $\mathbf{G}_{\text{best}}(t)$ 进行比较, 并更新 $\mathbf{G}_{\text{best}}(t)$;

14 end for

15 until 达到最大迭代次数 $MaxIter$

16 输出全局最优适应度值 F_{gbest} , 粒子的优化序列 $\mathbf{X}_{\text{gbest}}$ 。

4 实例研究

4.1 实验

本文对葡萄牙波尔图大学医学院收集的胎心分娩力造影术(cardiotocography, CTG)测量的胎儿健康诊断数据集进行实例研究, CTG是使用电子胎儿监护仪检测胎儿心率和孕子宫收缩情况的技术, 医疗上可以用于评估胎儿的健康状况。该数据集已由产科专家根据形态学模式和胎儿状态标记了18个特征变量,

这是一个高维的数据集,可能会存在部分属性的冗余,有必要对系统进行降维处理。数据集是已诊断后的样本,诊断结果分为3种类型,包括健康、疑似病症、确诊病例。本文选择健康和确诊病例两类样本用于后续的数据处理,并把健康类别作为正常样本,确诊病例类别作为异常样本,选定569组样本作为实验数据,包括393个正常样本、176个异常样本。

4.2 数据处理与分析

先选取正常样本中前161个样本数据用于建立马氏基准空间,并验证基准空间的有效性。计算表明,基准空间中正常样本的马氏距离均值为0.991,接近于1。计算异常样本的马氏距离值,大多分布在3附近,明显大于正常样本的,故确认基准空间有效。再使用初始特征变量序列,计算正常样本中剩余的232个样本的MD值和异常样本的MD值,如图1所示。

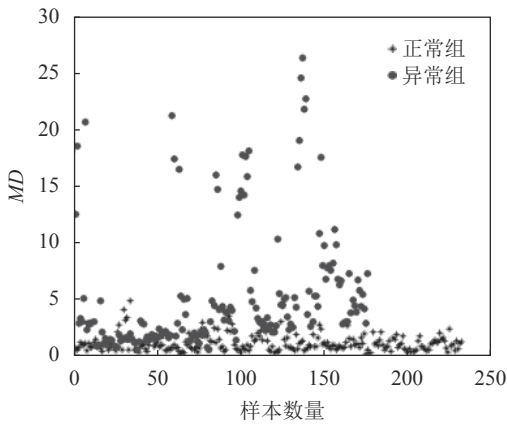


图1 优化前的样本MD分布

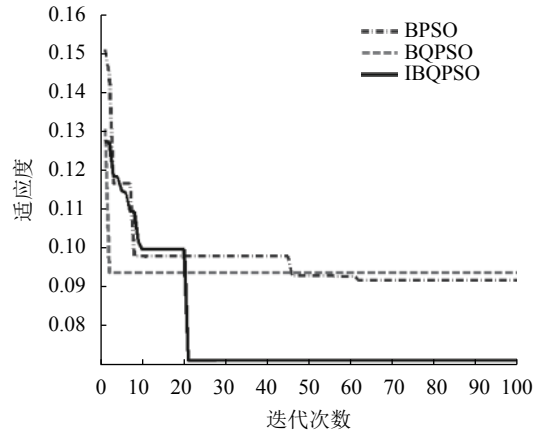
Fig. 1 Distribution of Mahalanobis distance values of samples before optimization

由图1可知:总体趋势上正常组的MD值较小,均值为1.18;而异异常组的MD值波动较大,均值达到了5.37;整体看正常组和异常组的MD有一定的区分度,但仍存在较多重叠的部分,进一步分类难度较大。

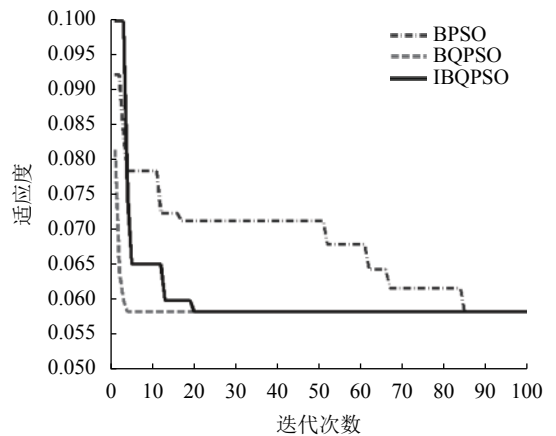
对于多目标优化函数式(8),两种误分类率的权重 ω_1 与 ω_2 对优化结果影响较大,而 ω_3 对优化结果影响较小,故 ω_3 可设定为较小权重比,考虑尝试使用3种配置:1) $\omega_1 = 0.50, \omega_2 = 0.48, \omega_3 = 0.02$;2) $\omega_1 = 0.30, \omega_2 = 0.68, \omega_3 = 0.02$;3) $\omega_1 = 0.80, \omega_2 = 0.18, \omega_3 = 0.02$ 。粒子总数为50,最大迭代次数为100。同时,使用标准二进制粒子群(BPSO)和二进制量子粒子群^[15](binary quantum behaved particle swarm optimization, BQPSO)进行对比实验,两种粒子群的种群数和迭代次数与IBQPSO相同。优化迭代实验连续进行60轮,适应度和收敛时间分别取多次实验后的平均值。

图2(a)、(b)和(c)给出了BPSO、BQPSO、IBQPSO这3种算法的收敛过程。由图2可以看出:

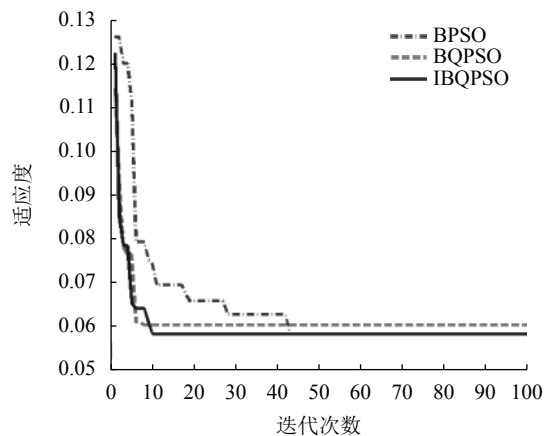
IBQPSO的适应度最优,3种权重组合情况下适应度分别达到了0.023、0.058和0.058。收敛速度方面,BQPSO在3种权重组合下都收敛很快,平均收敛次数基本都是10次以内,但容易陷入局部最优;BPSO收敛最慢,在3种权重组合下都超过了50次迭代才收敛;IBQPSO迭代速度适中,平均收敛次数在20次左右,明显小于BPSO。



(a) $\omega_1=0.50, \omega_2=0.48, \omega_3=0.02$



(b) $\omega_1=0.30, \omega_2=0.68, \omega_3=0.02$



(c) $\omega_1=0.80, \omega_2=0.18, \omega_3=0.02$

图2 算法的收敛过程

Fig. 2 Convergence process for the algorithms

表1列出了应用3种算法后目标函数达到的最优适应度的均值、每种优化算法的迭代次数均值和优化后的特征组合。

由表1可知: 在3种不同权重组合下, BQP SO的空间搜索速度都是最快的, 且平均迭代次数最少, 但是, BQP SO算法的适应度均大于BPSO和IBQP SO算

表 1 目标函数的适应度和最优特征组合

Tab. 1 Fitness value of the objective function and optimal combination of features

权重组合	优化算法	适应度	迭代次数	最优特征组合
$\omega_1=0.50, \omega_2=0.48, \omega_3=0.02$	BPSO	0.057	58	$x_3, x_5, x_6, x_{13}, x_{15}$
	BQP SO	0.063	7	$x_3, x_5, x_6, x_7, x_{15}$
	IBQP SO	0.052	20	$x_3, x_5, x_6, x_{13}, x_{15}$
$\omega_1=0.30, \omega_2=0.68, \omega_3=0.022$	BPSO	0.043	73	x_3, x_6, x_{17}
	BQP SO	0.048	8	$x_3, x_4, x_6, x_{10}, x_{14}, x_{15}$
	IBQP SO	0.042	20	$x_3, x_4, x_6, x_{11}, x_{15}, x_{16}$
$\omega_1=0.80, \omega_2=0.18, \omega_3=0.02$	BPSO	0.042	50	$x_3, x_5, x_6, x_{13}, x_{15}$
	BQP SO	0.045	6	$x_1, x_4, x_5, x_6, x_{13}, x_{15}$
	IBQP SO	0.039	16	$x_3, x_5, x_6, x_{13}, x_{15}$

采用IBQP SO算法优化后的最优变量组合重新构建基准空间, 并计算正常组和异常组样本的MD值, 如图3所示。采用优化后的变量使得正常样本和异常样本的MD值差异更加明显, 重叠部分较优化前减少许多, 系统辨识效率大幅提高。

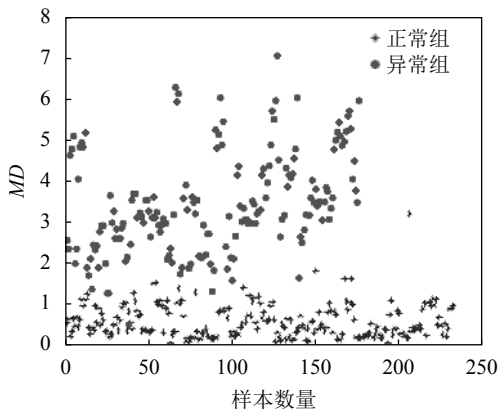


图 3 优化后的样本MD分布

Fig. 3 Distribution of Mahalanobis distance values of samples after optimization

为了证实改进的量子行为粒子群优化 (IBQP SO) 算法优化后的特征变量可以使得度量空间的分类准确性有效提升, 本文结合样本度量结果的真正率和假正率, 得出优化前后的样本ROC曲线, 并找到了优化前后系统的最佳阈值点, 一并绘制于图4中以进行对比。由图4中虚线可以看出, 优化前经ROC法确定的系统的最佳阈值为1.64, 预测准确度仅为81.29%, 可见, 采用样本数据集原始的变量组合用于系统的辨识效果并不理想, 存在一些冗余的对分类结果有

法适应度, 而在3种不同权重组合下IBQP SO算法的适应度是最小的, 因此, 使用IBQP SO算法优化的系统的分类准确性最高。

总体来说, 在优化模型的3种不同权重组合中, 综合考虑迭代速度和收敛精度两个指标, IBQP SO算法的优化性能要好于BPSO和BQP SO算法。

负作用的变量, 故有必要对系统的多元属性进行降维优化处理, 以获取具有较好预测判别性能的精简系统。由图4中实线可以看出, 优化后的样本ROC曲线下的面积达到了0.98, 非常接近于1, 表明优化后的系统分类辨识效果良好。经实验得出当阈值点取值为1.63时, 正常样本的分类精确度达到了94.7%, 分类准确率较优化前有了显著地提升。

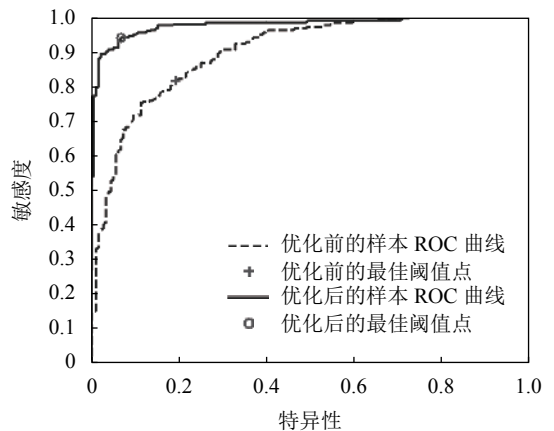


图 4 优化前后的样本ROC曲线

Fig. 4 ROC curves of samples before and after optimization

表2列出了3种不同的优化算法约简系统后, 选取优化后的特征变量重新构建度量模型, 系统预测的准确率情况。从总体平均准确率来看, IBQP SO算法的准确率最高(94%), 均高于BPSO和BQP SO算法的准确率(92%和92.7%), 可以体现IBQP SO算法在对特征变量进行有效约简后, 系统的冗余变量更少, 同时分类结果的判别精度有了较大提高。

表 2 3种算法优化后的系统预测准确率

Tab. 2 System prediction accuracy after optimization of three algorithms

算法	系统预测准确率/%		
	正常样本	异常样本	总体
BPSO	92.9	91.1	92.0
BQPSO	93.8	91.6	92.7
IBQPSO	94.7	93.3	94.0

5 结 论

针对现有的标准二进制粒子群用于马氏田口系统变量降维存在分类成功率低,迭代速度慢的缺陷,本文提出改进的基于量子行为二进制粒子群的马氏田口系统变量选择优化方法,构建了误分类率的多目标降维优化模型,使用IBQPSO粒子群算法进行求解,得到了系统的最优变量组合。在胎儿健康状况诊断实例中应用了IBQPSO算法求解马氏田口系统的最优变量组合,并与BPSO和BQPSO算法的优化结果进行比较,结果表明:与BPSO和IBQPSO算法相比,虽然BQPSO算法的收敛速度较快,但易收敛于局部最优,适应度较大;IBQPSO算法比BPSO算法收敛更快,迭代后的适应度最小,总体来看IBQPSO算法优化效果最好。此外,实验还对优化后的样本集进行了分类的准确性测试,结果表明精简后的系统的预测准确性更好。通过实例分析,采用本文方法后,变量降维的迭代速度和分类成功率显著提高,验证了该方法的有效性。

下一步研究中,考虑改进粒子群优化策略,将混沌量子粒子群优化应用于马氏田口系统中,增加粒子的多样性,进一步增强粒子的全局搜索能力。

参考文献:

- [1] Yang T H, Cheng Y T. The use of Mahalanobis-Taguchi system to improve flip-chip bumping height inspection efficiency[J]. *Microelectronics Reliability*, 2010, 50(3): 407-414.
- [2] Huang C L, Hsu T S, Liu C M. The Mahalanobis-Taguchi system—Neural network algorithm for data-mining in dynamic environments[J]. *Expert Systems with Applications*, 2009, 36(3): 5475-5480.
- [3] Shakyia P, Kulkarni M S, Darpe A K. Bearing diagnosis based on Mahalanobis-Taguchi-Gram-Schmidt method[J]. *Journal of Sound and Vibration*, 2015, 337: 342-362.
- [4] Rizal M, Ghani J A, Nuawi M Z. Cutting tool wear classification and detection using multi-sensor signals and Mahalanobis-Taguchi System[J]. *Wear*, 2017, 376(15): 1759-1765.
- [5] Yazid A M, Rijal J K, Awaluddin M S, et al. Pattern recogni-

tion on remanufacturing automotive component as support decision making using Mahalanobis-Taguchi system[J]. *Procedia CIRP*, 2015, 26: 258-263.

- [6] Zeng Jianghui, Zeng Fengzhang. The measurement scale of Mahalanobis-Taguchi system optimization based on fuzzy robustness discriminant criterion[J]. *Industrial Engineering and Management*, 2008, 13(3): 52-55. [曾江辉, 曾凤章. 基于模糊判别准则的马田系统基准空间优化[J]. *工业工程与管理*, 2008, 13(3): 52-55.]
- [7] Iquebal A S, Pal A, Ceglarek D, et al. Enhancement of Mahalanobis-Taguchi system via rough sets based feature selection[J]. *Expert Systems with Applications*, 2014, 41(17): 8003-8015.
- [8] Niu Junlei. Methods of classification and evaluation using Mahalanobis-Taguchi system based on omni-optimizer algorithm and applications[D]. Nanjing: Nanjing University of Science and Technology, 2012. [牛俊磊. 基于全方位优化算法的马田分类和排序评价方法研究及应用[D]. 南京: 南京理工大学, 2012.]
- [9] Reséndiz E, Moncayo-Martínez L A, Solís G. Binary ant colony optimization applied to variable screening in the Mahalanobis-Taguchi System[J]. *Expert Systems with Applications*, 2013, 40(2): 634-637.
- [10] Pal A, Maiti J. Development of a hybrid methodology for dimensionality reduction in Mahalanobis-Taguchi system using Mahalanobis distance and binary particle swarm optimization[J]. *Expert Systems with Applications*, 2010, 37(2): 1286-1293.
- [11] Jin X H, Chow T W S. Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis-Taguchi system[J]. *Expert Systems with Applications*, 2013, 40(15): 5787-5795.
- [12] Liparas D, Laskaris N, Angelis L. Incorporating resting state dynamics in the analysis of encephalographic responses by means of the Mahalanobis-Taguchi strategy[J]. *Expert Systems with Applications*, 2013, 40(7): 2621-2630.
- [13] Mahalakshmi P, Ganesan K. Mahalanobis Taguchi system based criteria selection for shrimp aquaculture development[J]. *Computers and Electronics in Agriculture*, 2009, 65(2): 192-197.
- [14] Sun Jun, Feng Bin, Xu Wenbo. Particle swarm optimization with particles having quantum behavior[C]//Proceedings of the IEEE 2004 Congress on Evolutionary Computation. *Portland: IEEE*, 2004: 325-331.
- [15] Xi Maolong, Sun Jun, Wu Yong. Quantum-behaved particle swarm optimization with binary encoding[J]. *Control and Decision*, 2010, 25(1): 99-104. [奚茂龙, 孙俊, 吴勇. 一种二进制编码的量子粒子群优化算法[J]. *控制与决策*, 2010, 25(1): 99-104.]

(编辑 赵 婧)

引用格式: Liu Jiu fu, Zheng Rui, Ding Xiaobin, et al. Feature selection optimization for Mahalanobis-Taguchi system with binary quantum behavior particle swarm[J]. *Advanced Engineering Sciences*, 2019, 51(6): 152-158. [刘久富, 郑锐, 丁晓彬, 等. 马氏田口系统的量子行为二进制粒子群特征选择优化方法[J]. *工程科学与技术*, 2019, 51(6): 152-158.]