

# 粗糙集和支持向量机的表具识别算法研究

唐 亮<sup>1,2</sup>, 仲元昌<sup>2</sup>, 沈甲甲<sup>2</sup>, 马天智<sup>2</sup>

TANG Liang<sup>1,2</sup>, ZHONG Yuanchang<sup>2</sup>, SHEN Jiajia<sup>2</sup>, MA Tianzhi<sup>2</sup>

1. 西南计算机有限责任公司, 重庆 400060

2. 重庆大学 通信工程学院, 重庆 400044

1. Southwest Computer CO., LTD., Chongqing 400060, China

2. College of Communication Engineering, Chongqing University, Chongqing 400044, China

TANG Liang, ZHONG Yuanchang, SHEN Jiajia, et al. Digital recognition based on rough set and support vector machine optimized by improved quantum-behaved particle swarm optimization. Computer Engineering and Applications, 2018, 54(2): 253-258.

**Abstract:** In order to restrain particles from trapping in local optimum, this paper presents a new algorithm, which is based on the combination of rough sets and Support Vector Machine (SVM) optimized by Improved Quantum-behaved Particle Swarm Optimization (IQPSO), the artificial bee colony algorithm and immune algorithm are introduced to speed up the search space and improve the convergence speed. Firstly, the rough set optimized by IQPSO is used to reduce the attributes of the feature vector, and then the parameters of SVM are optimized by IQPSO. Finally, the simulation experiment shows that the improved algorithm can effectively reduce the number of decision attribute, improve rough set attribute reduction ability, obtain the better parameters of SVM, this algorithm has a fast convergence speed and high recognition accuracy.

**Key words:** gas meter recognition; quantum-behaved particle swarm optimization; rough set; attribute reduction; support vector machine

**摘 要:** 针对量子粒子群算法具有陷入局部值缺点, 提出了一种基于改进量子粒子群算法优化的粗糙集和支持向量机相结合的表具识别算法, 引入人工蜂群算法和免疫算法, 来提高算法搜索空间、收敛速度。首先通过改进量子粒子群算法优化的粗糙集对得到的特征向量进行属性约简, 然后经过改进量子粒子群算法优化支持向量机参数。最后通过实验仿真表明, 改进的算法能有效地减少决策属性的个数, 提高了粗糙集属性约简能力, 优化了支持向量机的参数, 算法收敛速度快, 识别准确率高。

**关键词:** 表具识别; 量子粒子群算法; 粗糙集; 属性约简; 支持向量机

**文献标志码:** A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.1701-0271

## 1 引言

物联网技术已成为当今科技发展最有影响力的新技术之一, 无线远程智能抄表技术就是其典型应用。家用燃气表、水表和电表的三表合一的“表联网”的时代正在悄然兴起。在大规模的数据统计中, 表具数字的识别是整个智能“表联网”的一个重要环节, 因此成为多年来研究的一个热点<sup>[1]</sup>。

由波兰科学家 Pawlak 提出的粗糙集理论是一种处理模糊性和不精确性知识的数学工具, 已经在机器学习、数据挖掘、模式识别等众多领域得到了广泛的应用<sup>[2]</sup>。粗糙集理论作为一种强大的数据分析处理理论, 在数据预处理过程中可以对特征更准确地提取。利用粗糙集的属性约简特性, 对数据集进行降维操作。目前研究更多的是包含属性相对最小的属性约简, 文献[3]

**基金项目:** 国家创新基金项目 (No.10C26215113031); 中央高校基本科研业务费专项项目 (No.106112016CDJXZ168815, No.106112017CDJZRPY0101); 重庆市科技创新专项 (No.cstc2017shmsA1034)。

**作者简介:** 唐亮 (1975—), 男, 硕士, 主要从事计算机网络、数字通信、物联网等领域的研究与开发; 仲元昌 (1965—), 通讯作者, 男, 博士, 教授, 研究方向: 无线传感器网络, 物联网及大数据处理, E-mail: zyc@cqu.edu.cn。

**收稿日期:** 2017-01-18 **修回日期:** 2017-07-11 **文章编号:** 1002-8331(2018)02-0253-06

应用遗传算法的思想来设计最小属性约简算法。遗传约简算法具有搜索范围广的优点,但其收敛速度慢,规则较复杂,如交叉变异等操作。文献[4]提出了求最小属性约简的一种带种子粒子保护的粒子群算法,粒子群算法的优点是收敛速度快,缺陷是容易陷入局部最优解,算法不稳定。文献[5]进一步结合免疫算法的思想,提出一个基于免疫粒子群优化思想的最小属性约简算法。文献[6]提出引入人工蜂群搜索算子的粒子群算法。本文引入免疫算法中免疫机制和人工蜂群搜索算子的量子粒子群算法与粗糙集理论相结合进行属性约简,从而提出了一种新的属性约简算法。

数字识别一般通过特征匹配及特征判别的传统方法进行处理,但这些方法的识别率不高<sup>[7]</sup>。对于数字仪表显示值进行识别,文献[8]介绍了BP神经网络的图像识别研究,文献[9]介绍了基于粒子群优化算法的BP神经网络在图像识别中的应用,目前SVM已经成为模式识别领域的研究热点,因此一些学者也开始把支持向量机技术应用到数字识别中<sup>[10]</sup>。文献[11]介绍了基于粒子群算法优化支持向量机,取得了良好的实验效果。比较遗传算法和粒子群算法优化SVM实验效果,本文在改进的量子粒子群算法的基础上引入增强跳出局部值能力的人工蜂群搜索算子和加快算法收敛免疫算法,优化支持向量机参数。

对无指针式数字仪表(以燃气表为例)表盘的数字识别问题,本文利用粗糙集对数据的约简能力以及支持向量机在数字字符具有良好的识别能力,提出了一种改进量子粒子群(Improvement Quantum-behaved Particle Swarm Optimization, IQPSO)分别优化粗糙集和支持向量机参数的算法。该算法不但能够准确提取数字的关键特征、减小计算的复杂度并且有效解决支持向量机参数的优化问题,提高训练速度以及识别精度。

## 2 粗糙集理论分析

为构建基于改进量子粒子群算法优化的粗糙集的数据字符识别算法,引入以下相关定义。

**定义1** 在粗糙集理论中,将信息表知识表达系统四元组  $S=(U, A, V, f)$  定义为一个信息系统,其中,  $U$  称作论域,即对象的非空有限集合;  $A$  为属性的非空有限集合;  $V = \bigcup_{\alpha \in A} V_{\alpha}$  为属性值的集合( $V_{\alpha}$  为属性  $\alpha$  的值域);  $f: U \times A \rightarrow V$  表示信息系统的映射函数。

**定义2** 在给定信息表知识表达系统对每个样例子集  $\forall X \subseteq U$  和论域  $U$  上的一个等价关系  $R \in IND(K)$ , 根据不可分辨关系  $R$  能确定包含于  $X$  中的最大可定义集称为子集  $X$  关于  $R$  的下近似集:  $\underline{R} = \{x | (\forall x \in U) \wedge ([x]_R \subseteq X)\}$  即:  $\underline{R}(X)$ 。

**定义3** 设  $P \subseteq C$ , 对划分  $U/IND(D)$  的  $P$ -近似精度为:

$$\gamma_P = \frac{\sum_{i=1}^k Card(\underline{P}(X_i))}{Card(U)}$$

其中  $Card(\cdot)$  表示集合的基数。

**定义4** 设  $P \subseteq C, \gamma_P = \gamma_C$ , 且不存在  $R \subset P$ , 使  $\gamma_R = \gamma_C$ , 则称  $P$  为  $C$  一个属性约简。

最小属性约简是一个非线性约束组合问题,问题可以描述为:

$$\begin{cases} \min Card(P) \\ P \subseteq C \\ \gamma_P = \gamma_C \\ \forall q \in P, \gamma_{P \setminus \{q\}} < \gamma_P \end{cases} \text{ s.t.}$$

## 3 改进量子粒子群的粗糙集属性约简算法

由 Kennedy 和 Eberhart 提出的粒子群优化(PSO)算法是近年来发展起来的一种新的群体智能优化方法。文献[12]介绍了一种二进制编码的量子粒子群优化算法,用量子粒子群算法来解决离散空间的优化问题。

算法过程如下:设粒子群规模  $M$ , 种群中的第  $i$  ( $1 \leq i \leq M$ ) 个粒子在搜索空间的当前位置定义为  $X_i = \{X_{i,1}, X_{i,2}, \dots, X_{i,D}\}$ , 粒子维数为  $D$ , 表示决策变量的个数。在算法的迭代过程中,记录每个粒子当前搜索到的最优位置定义为  $Pbest_i = [pbest_{i,1}, pbest_{i,2}, \dots, pbest_{i,D}]$ , 所有粒子当前搜索到的最好解是全局最优位置,表示为  $Gbest = [gbest_1, gbest_2, \dots, gbest_D]$ 。粒子的位置通过 0、1 构成的二进制向量表示,同时对吸引子、平均最优位置和粒子更新公式进行了重新定义。

采用个体最优值  $pbest_{id}^t$  和全局最优值  $gbest_{id}^t$  通过平均交叉操作获得局部吸引子  $P_{id}^t$ , 如下:

$$P_{id}^t = \begin{cases} pbest_{id}^t, rand()_{id} > 0.5 \\ gbest_{id}^t, rand()_{id} \leq 0.5 \end{cases} \quad (1)$$

对于平均最优位置  $mbest_d^t$ , 如果超过半数的粒子的个体最优位置  $pbest$  的第  $j$  个比特位都等于 1, 则  $mbest$  的第  $j$  个比特位等于 1, 否则  $mbest$  的第  $j$  个比特位等于 0。如果正好有一半粒子  $pbest$  的该比特位为 1, 则  $mbest$  的该比特位就以概率 0.5 随机等于 1 或 0, 其数学形式如下式:

$$Mbest_{id}^t = \begin{cases} 1, mbest_d^t > 0.5 \\ \text{随机0或1}, mbest_d^t = 0.5 \\ 0, mbest_d^t < 0.5 \end{cases} \quad (2)$$

$$mbest = \frac{1}{M} \left( \sum_{i=1}^M pbest_{id}^t \right) \quad (3)$$

离散量子粒子群算法位置更新公式如下:

$$X_{id}^{t+1} = mutate(P_{id}^t) \quad (4)$$

$$Mutate(y) = \begin{cases} 1 - y, rand() \leq Pr \\ y, rand() > Pr \end{cases} \quad (5)$$

$$Pr = a \times \ln(1/u) \otimes (mbest_d \oplus X_{id}^t) \quad (6)$$

若  $Pr > 1$ , 将  $Pr$  设置为 1。其中  $\otimes$  为与操作,  $\oplus$  为异或操作,  $a$  为收缩扩张因子,  $u$  表示 0 到 1 的随机数。经测试当取:

$$a = (0.95 - 0.55) \times (Maxit - t) / Maxit + 0.55$$

$Maxit$  表示最大迭代次数,  $t$  表示当前迭代次数。即在量子粒子群算法开始阶段, 种群粒子以较大的概率突变, 随着迭代次数的增加, 突变率减小。

人工免疫算法是近年来发展起来的用于模仿自然免疫系统功能的一种智能优化算法。免疫接种和免疫选择是免疫算法中的两个重要算子, 疫苗接种实现如下: 选取疫苗之后, 对粒子群进行一定次数的随机抽取, 即在粒子群中抽取一定数量的粒子, 疫苗注射操作, 使用疫苗的个体信息位来更替原粒子个体相应的信息位, 即用疫苗的粒子位来替换原粒子中相应位上信息, 这样就产生了一个新的粒子个体。每个疫苗个体信息位以概率  $P_{vacc}$  替换原粒子个体的信息位, 本质就是使用疫苗上为 0、1 的有效位来替换原粒子中 0、1 的组合。接种概率如下:

$$P_{vacc} = P_{vacc}^{\max} - (P_{vacc}^{\max} - P_{vacc}^{\min}) \times i / Maxit \quad (7)$$

$P_{vacc}^{\max}$  为最大接种概率, 取 0.9;  $P_{vacc}^{\min}$  为最小接种概率, 取 0.1;  $i$  为当前迭代次数;  $Maxit$  为算法总迭代次数。

免疫选择: 当接种后的粒子个体的适应度函数值比接种前粒子的适应度函数值更好时, 使用接种后的新个体代替当前粒子个体, 否则丢弃接种后的粒子。

为了更好地解决量子粒子群算法过早出现早熟现象, 提高算法的探索能力跳出局部最优, 本文引入了人工蜂群搜索算子。结合粒子群算法, 文献[13]给出了一个新的搜索算子:

$$Z_{i,d} = x_{i,d} + \phi_{i,d}(x_{i,d} - x_{k,d}) + \varphi_{i,j}(p_{g,d} - x_{i,d}) \quad (8)$$

其中  $k \in \{1, 2, \dots, M\}$  和  $d \in \{1, 2, \dots, D\}$ ,  $M$  为粒子数,  $D$  为粒子维数, 并且  $k$  不等于  $i$ ;  $\phi_{i,d}$  是 (0, 1) 之间的随机数,  $\varphi_{i,d}$  是 (0, 1.5) 之间的随机数。

以上搜索算子用于连续数据, 由于求解最小相对属性约简用的是二进制离散值, 结合量子粒子群算法, 使用如下二进制搜索算子:

$$Z_{id}^t = \begin{cases} pbest_{kd}^t, rand()_{id} > \varphi \\ P_{id}^t, rand()_{id} \leq \varphi \end{cases} \quad (9)$$

$$X_{id}^{t+1} = mutate(Z_{id}^t) \quad (10)$$

其中  $k, d$  选取相同,  $pbest_{kd}^t$  为第  $t$  次迭代粒子  $k$  在  $d$  维的个体最优位置,  $P_{id}^t$  是第  $t$  次迭代粒子在  $d$  维上的当前位置。  $\varphi$  叫做偏移概率, 在 (0, 1) 之间取值。首先使用公式(15)计算  $Z_{id}^t$ , 过程产生一个 (0, 1) 随机数, 如果这个数大于  $\varphi$  则  $Z_{id}^t$  取  $pbest_{kd}^t$ , 如果小于等于  $\varphi$  则取

$P_{id}^t$  值。之后将  $Z_{id}^t$  带入位置更新公式, 更新粒子位置  $Z_{id}^{t+1}$ 。这里取  $\varphi = 0.5$ 。如果更新后的粒子适应度比原来的好, 则替换原来的粒子, 否则保留更新前的粒子。

### 3.1 适应度函数的选取

在量子粒子群算法中, 适应度函数是优化的目标函数, 用来评价粒子群中粒子(问题的解)的优劣程度。选取适应度函数如下:

$$F(x) = \begin{cases} (N-m)/N + \gamma_x, \gamma_x \leq \gamma_c \\ (3N-m)/N + \gamma_x, \gamma_x = \gamma_c \end{cases} \quad (11)$$

其中  $Card(C) = N$ , 粒子编码为  $X_i \in \{0, 1\}^N$ ,  $X_i = \{X_{i1}, X_{i2}, \dots, X_{iN}\}$ ,  $m$  为粒子选取属性的个数。

### 3.2 改进量子粒子群约简算法的实现过程

本文提出的改进的量子粒子群算法优化粗糙集求得最小属性约简算法过程如下。

输入信息: 经过连续属性离散化后的决策系统  $S = (U, C \cup D, V, f)$ ; 输出信息: 属性约简的决策系统。

(1) 算法粒子种群, 维数以及参数初始化。

(2) 根据公式(4)更新粒子自身位置。

(3) 根据适应度函数公式(11)计算每个粒子的适应度值, 更新粒子最优个体位置, 即将每个粒子当前适应值与个体最优位置  $pbest_{id}$  的适应度值比较, 若优于个体最优位置  $pbest_{id}$ , 则将更新后的粒子位置赋予个体最优位置  $pbest_{id}$ ; 更新群体最优位置  $Gbest_{id}$ 。

(4) 疫苗接种, 进行一定次数的随机抽取(本文取 10 次), 从粒子群中抽取一定数量粒子, 按接种概率对选取的粒子进行疫苗接种, 产生新粒子。

(5) 免疫选择, 比较接种前后粒子的适应度值, 若接种后的粒子适应度值更好, 则保留, 否则保留接种前粒子的位置信息。

(6) 根据式(10)计算搜索算子, 根据公式更新自身位置, 如果更新后的粒子适应值比更新前的粒子适应值更优, 则替换原来的粒子。

(7) 根据公式(3)更新种群的平均最优位置  $Mbest$ , 种群中的每个粒子根据公式(1)更新局部吸引子  $P_i$ , 根据公式(6)更新变异概率  $Pr$ 。

(8) 满足终止条件则算法终止, 则返回(2)继续迭代。

## 4 改进量子粒子群算法优化的支持向量机

### 4.1 支持向量机的基本原理

假设给定训练集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (R^n \times Y)^l$ , 其中  $x_i \in R^n$ ,  $y_i \in Y = \{-1, 1\}$ ,  $i = 1, 2, \dots, l$ ;  $x_i$  为特征向量; 通过训练集得到一个实值函数  $g(x)$  构造错误率最小的决策分类函数  $f(x) = \text{sgn}(g(x))$ ; 选取适当的核函数  $K(x, x')$  以及惩罚函数  $C > 0$ , 其中  $K(x, x') = (\varphi(x), \varphi(x'))$ , 在这里非线性映射函数  $\varphi(x)$  为从  $R^n$  到 Hilbert



空间  $H$  的变换,将输入数据映射到高维空间,构造并求出解凸二次规划问题:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(x_i, x_j) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \quad (12)$$

$$\text{s.t.} \sum_{i=1}^l y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l \quad (13)$$

得解为  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$ , 计算  $b^*$ : 选取位于区间  $(0, C)$  中的  $\alpha^*$  的分量  $\alpha_j^*$ , 据此计算:

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j) \quad (14)$$

构造函数  $f(x) = \text{sgn}(g(x))$ , 在这里有:

$$g(x) = \sum_{i=1}^l y_i \alpha_i^* K(x_i, x) + b^* \quad (15)$$

这里是用于二类分类问题,构造分类决策函数,建立最优分类超平面,将样本数据分为两类。数字识别为0到9十个字符的识别,是多分类问题。常见的多分类问题解决办法有:成对分类,一类对余类,SVM决策树分类。本文采用成对分类,径向基内核,主要参数有误差惩罚因子  $C$  和宽度函数  $\sigma$ ,这两个参数是可以调节选择的,且这两个参数的选择直接影响到支持向量机的分类识别性能,本文采用改进的量子粒子群算法来找到参数的最优值。验证性能指标选取  $k$  折交叉验证,总的验证指标为<sup>[14]</sup>:  $\min_{C, \sigma} L = (l_1 + l_2 + \dots + l_k)/k$ , 其中设定惩罚因子和核参数的范围:  $C \in (0, 100], \sigma \in [0, 1000]$ 。适应度函数为:

$$f = (l_1 + l_2 + \dots + l_k)/l \times 100\% \quad (16)$$

其中  $l_i$  为测试样本子集第  $i$  样本集正确分类的样本数目,  $l$  为测试样本集的总数。本文选取  $l=5$ , 即采用5折交叉验证作为参数验证性能。

#### 4.2 基于改进量子粒子群算法优化支持向量机过程

根据该改进量子粒子群算法优化SVM参数优化,如图1,可以得到如下的具体步骤:

(1)初始化粒子群数量以及改进量子粒子群算法参数初始化,初始位置;在编码中,采用二进制编码(将两个参数放在一起进行编码)  $X_t = \{C_{1t}, C_{2t}, \dots, C_{ut}, \sigma_{1t}, \sigma_{2t}, \dots, \sigma_{vt}\}$ , 为算法在  $t$  代某一粒子,  $C_{it}$  为参数  $C$  参数位串,  $\sigma_{it}$  为参数  $\sigma$  位串,其中  $1 \leq i \leq u(v)$ 。本文初始种群取20。

(2)根据公式(4)更新粒子自身位置。

(3)对每个粒子进行解码,计算每一个个体的适应度函数更新粒子最优个体位置  $Pbest_{id}$ , 群体最优位置  $Gbest_{id}$ 。

(4)~(8)步骤和3.2节算法过程一致,此处省略。

算法结束,如果达到预先设定的性能指标后,将最终群体中的最优个体解码即可得到优化后的支持向量

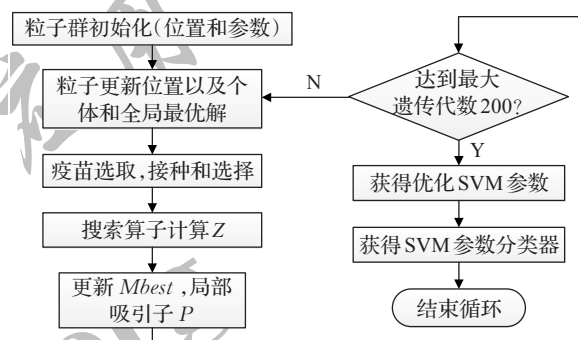


图1 改进量子粒子群算法优化支持向量机参数流程图

#### 5 仿真实验与结果分析

基于改进量子粒子群算法优化的粗糙集和支持向量机的表具图像数字识别模型如图2所示。

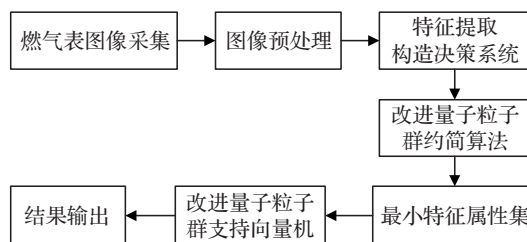


图2 表具图像数字识别系统模型

##### 5.1 图像的预处理

在居民用户实际环境下采集燃气表表盘的图像,由于一些原因使得采集到的图像可能不利于直接去识别,对图像进行增强和去噪及二值化处理,工作改善图像质量,便于后续的识别<sup>[15]</sup>。

选取在不同光照下、不同角度和不同清晰度等情况下的燃气表表盘图像60幅,并分别加0.1、0.2、0.3大小的随机噪声,共产生240幅初始样本。通过倾斜校正以及区域定位分割出目标图像,再进行二值化以及滤波处理,然后再通过垂直以及水平投影相结合并归一化的方法分割,再通过边缘检测的Roberts算子对字符图像进行细化以减少像素个数,如图3。每个燃气表表盘图像提取7个数字字符,从所有字符中随机选取1600个,其中1200个作为训练样本,400个作为测试样本。

##### 5.2 字符特征提取

特征提取是模式识别的第一步,其目标是得出一个长度为  $n$  的向量  $\{t_1, t_2, \dots, t_n\}$  用来表示原始数字字符,特征提取方法的选择是影响识别率至关重要的因素。

本文采取多种特征提取方法结合的方式提取原始特征向量作为神经网络的训练和测试样本,如粗网格特征提取、像素比特特征提取、结构特征提取、笔划密度特征。

##### 5.3 特征向量的属性约简

论域包括总共1600个数字字符,条件属性  $C = \{c_1, c_2, \dots, c_{115}\}$  为粗网格88维、像素比1维、结构特征8维及



密度特征18维共115维特征向量,决策属性  $D=\{0,1,\cdots,9\}$  为对应的10个数字。根据改进量子粒子群优化的粗糙集算法求得最小条件属性  $C$  相对于决策属性  $D$  的相对属性约简为  $\{c_{46},c_{79},c_{82},c_{89},c_{91},c_{94},c_{97},c_{106},c_{112}\}$ ,其中粗网格3维,像素比1维、结构特征3维和密度特征2维共9维特征向量,部分如表1所示。

表1 表具数字特征向量的部分约简结果

样本	$c_{46}$	$c_{79}$	$c_{82}$	$c_{89}$	$c_{91}$	$c_{94}$	$c_{97}$	$c_{106}$	$c_{112}$	数字
1	0.183	4	4	5	0.286	0.357	0.643	0.170	0.091	0
2	0.093	6	3	3	0.214	0.286	0.357	0.071	0.017	1
3	0.171	7	3	3	0.357	0.571	0.214	0.125	0.210	2
4	0.182	5	2	6	0.286	0.429	0.071	0.161	0.120	3
5	0.134	4	3	4	0.571	0.357	0.714	0.089	0.182	4
6	0.179	5	8	5	0.500	0.286	0.571	0.107	0.131	5
7	0.188	3	2	6	0.357	0	0.214	0.223	0.108	6
8	0.122	1	2	4	0	0.286	0.286	0.080	0.222	7
9	0.190	4	5	6	0	0.643	0.429	0.179	0.125	8
10	0.184	8	4	5	0.357	0.429	0.214	0.205	0.125	9

5.4 图像数字识别结果与分析

训练样本的数字向量:运用上述图像预处理的方法,本文选取1 200个作为训练样本,400个作为测试样本,为验证改进算法的有效性,QPSO(参数:种群规模  $N=20$ ),改进QPSO(参数和QPSO相同),未加入粗糙集进行属性约简使用改进型QPSO优化支持向量机的MATLAB仿真如图4~6所示,并得到表2、表3。

经过不同优化算法得到的参数值以及测试样本的准确率如表2所示。

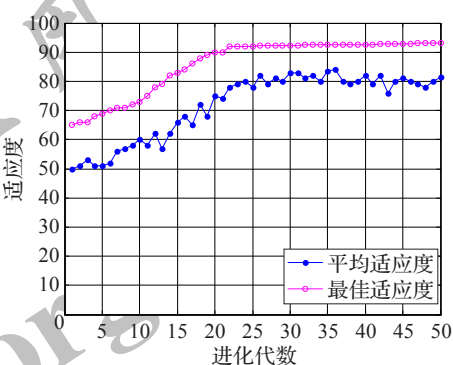


图4 粗糙集和QPSO支持向量机参数适应度曲线

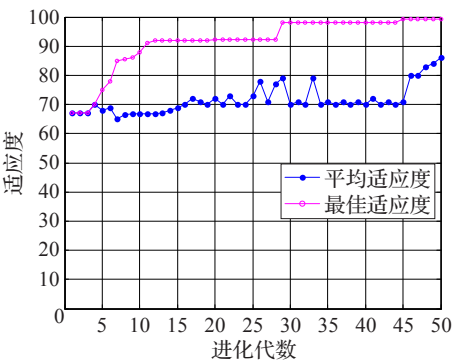


图5 粗糙集和改进QPSO支持向量机参数适应度曲线

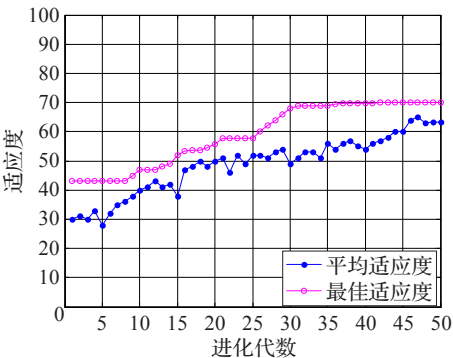


图6 无粗糙集和改进QPSO支持向量机参数适应度曲线

表2 QPSO和改进QPSO分别优化粗糙集识别结果比较

识别结果	QPSO优化粗糙集	改进QPSO法优化粗糙集
迭代次数	34	23
时间/s	47.924 1	23.870 9
识别精度/%	98.695 2	99.537 7

表3 不同优化算法参数及识别结果比较

优化算法	样本数	测试样本	$C$	$\sigma$	识别率/%
QPSO	1 200	400	2.227 8	0.010 1	98.695 2
本文算法	1 200	400	2.213 9	0.008 6	99.537 7

实验结果分析:由以上实验结果可以看出,QPSO算法优化的SVM参数在CV意义下的最高分类识别准确率为93.521 1%。改进型的QPSO在最高分类识别准确率上又得到提高达到97.397 5%。由图4、图5可知,由于人工蜂群搜索算子解决粒子出现趋同现象,使粒子进化中出现多样性,帮助算法快速跳出陷入的局部值;免疫算法使粒子朝着更优的方向进行迭代,起到很好的

引导作用,加快了收敛的速度。由图5、图6可知,经过改进QPSO算法优化的粗糙集进行属性约简后再经过改进QPSO优化SVM参数的支持向量机和没有经过粗糙集进行属性约简直接经过改进QPSO算法优化支持向量机进行识别相比较,前者有更高的识别准确率和更快的收敛速度。经过改进的量子粒子群算法优化粗糙集进行属性约简,去除了无关或者相对不重要的属性,为后来支持向量机的识别进行前期特征向量的数据约简处理,所以有更高识别准确率;由表2可知,在经过QPSO和改进的QPSO分别进行粗糙集进行属性约简后分别通过相同训练识别,可得改进的QPSO优化的粗糙集有更高的识别率,迭代次数为23,准确率达到99.537 7%。由表3可知,几种优化算法的结果比较,改进后的QPSO有更好的优化粗糙集以及支持向量机参数的能力,得到支持向量机惩罚因子 $C$ 为2.213 7和宽度函数 $\sigma$ 为0.008 6,有更高的识别率。所以经过改进QPSO算法优化的粗糙集和支持向量机训练时间最快,能够很快收敛到目标值,识别精度最高。经过多次训练测试,识别率能够达到99.537 7%。

## 6 结束语

本文为解决家用燃气表的摄像直读问题,提出了一种基于改进量子粒子群算法优化的粗糙集和支持向量机相结合的字符识别算法。利用粗糙集的属性约简对字符特征向量属性进行简化得到最小属性约简,有效地减少了图像特征决策属性的个数,优化了支持向量机的参数。仿真实验表明:该算法提高了支持向量机的分类识别效果,具有更高的识别精度,收敛速度快,可以广泛用于无指针式仪表表盘数字识别,可以构造一个手写数字家庭仪表字符识别器,为智能家居和“表联网”的实现开辟了一种新的途径。

## 参考文献:

[1] 高菊,叶桦.一种有效的水表数字图像二次识别算法[J].东南大学学报:自然科学版,2013,43(s1):153-157.

[2] Pawlak Z. Rough sets[J]. International Journal of Computer and Information Sciences, 1982, 11(5): 341-356.

[3] Dai J H, Li Y X. Heuristic genetic algorithm for minimal reduction in decision system based on rough set theory[C]// Proceedings of 2002 International Conference on Machine Learning and Cybernetics, 2002: 833-836.

[4] 叶东毅, 廖建坤. 基于二进制粒子群优化的最小属性约简算法[J]. 模式识别与人工智能, 2007, 20(3): 295-300.

[5] Ye D Y, Chen Z J, Liao J K. A new algorithm for minimum attribute reduction based on binary particle swarm optimization with vaccination[C]// Proceedings of PAKDD07, 2007: 1029-1036.

[6] 高卫峰, 刘三阳, 焦合华, 等. 引入人工蜂群搜索算子的粒子群算法[J]. 控制与决策, 2012, 27(6): 833-838.

[7] Chen X, Liu X, Jia Y. Discriminative structure selection method of Gaussian mixture models with its application to handwritten digit recognition[J]. Neurocomputing, 2011, 74(6): 954-961.

[8] 胡菡静, 王小妮, 王磊, 等. 图像识别技术-神经网络识别方法的分析[J]. 中国新通信, 2016, 18(22).

[9] 李战明, 杨红红. 车牌图像特征提取及改进神经网络的识别算法研究[J]. 现代电子技术, 2016, 39(16): 102-104.

[10] 刘永平, 郭小波. 基于最小二乘支持向量机车牌字符特征识别[J]. 数字技术与应用, 2015(7): 119-120.

[11] 邵信光, 杨慧中, 陈刚. 基于粒子群优化算法的支持向量机参数选择及其应用[J]. 控制理论与应用, 2006, 23(5): 740-743.

[12] 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.

[13] Zhu G, Kwong S. Gbest-guided artificial bee colony algorithm for numerical function optimization[J]. Applied Mathematics and Computation, 2010, 217(7): 3166-3173.

[14] 周辉仁, 郑丕谔, 赵春秀. 基于遗传算法的LS-SVM参数优选及其在经济预测中的应用[J]. 计算机应用, 2007, 27(6): 1418-1429.

[15] Gao Q, Liu J, Zhang H, et al. Enhanced fisher discriminant criterion for image recognition[J]. Pattern Recognition, 2012, 45(10): 3717-3724.