

# 基于粗糙集和支持向量机的标准农田地力等级评价

赖红松, 吴次芳

(浙江大学 公共管理学院, 杭州 310029)

**摘要:** 标准农田是耕地的精华,是确保国家粮食安全的关键。科学评价标准农田地力等级对标准农田培肥和土壤改良有着重要意义。将粗糙集(Rough Set, RS)理论和支持向量机(Support Vector Machine, SVM)相结合,提出了基于RS和SVM的标准农田地力等级评价方法,同时,利用遗传算法的并行搜索结构和模拟退火的概率突跳特性,提出了GASA优化SVM参数算法。该方法首先在确定标准农田地力等级评价指标的基础上,利用地力调查样本数据及传统的指数和法评价结果构建RS决策表,应用RS穷尽算法对决策表进行约简,剔除冗余的评价指标,然后用约简后的评价指标作为SVM的输入,运用GASA优化SVM参数算法对SVM进行训练,建立标准农田地力等级的RS-SVM评价模型。应用该方法对温州市鹿城区标准农田地力等级进行评价,与未用RS约简的SVM模型和BP神经网络模型评价结果进行对比,SVM模型和BP神经网络模型的输入指标数均为15个,其评价正确率分别为100%和90%;RS-SVM模型的输入指标数为14个,其评价正确率分别为100%,结果表明,该方法通过RS约简评价指标后,SVM评价精度并没有降低,但降低了SVM输入向量维数和计算复杂度,提高了训练效率;SVM用于标准农田地力等级评价,具有比BP神经网络更高的评价精度,可有效用于标准农田地力等级评价,为耕地地力评价提供了新方法。

**关键词:** 耕地地力评价; 标准农田; 粗糙集; 支持向量机; 鹿城区

**中图分类号:** S158; TP18

**文献标志码:** A

**文章编号:** 1000-3037(2011)12-2141-14

标准农田是通过土地整理、农业综合开发等方法建成的田成方、渠相通、路相连,林成网、灌得进、排得出的稳产农田,是耕地中的精华,是现代农业发展的基础。标准农田地力等级评价是加强标准农田质量建设的基础,是制订标准农田保护、培肥、改良、利用规划的主要科学依据,是推进标准农田土壤改良,确保粮食生产能力的重要举措。随着计算机和信息技术的发展,近年来,国内外利用地理信息系统技术,采用数值法进行耕地地力评价研究较多,如指数和法<sup>[1-7]</sup>和人工神经网络法<sup>[8]</sup>等。传统的指数和法是在确定评价因子及其权重和隶属度(评分值)的基础上,采用加法模型计算评价单元的综合地力指数,然后根据综合地力指数分布确定分级方案,划分地力等级。该方法步骤复杂,工作量大,且易受人为主观因素影响。人工神经网络处理非线性关系的能力虽然较强,但存在网络结构难以确定、局部最优和泛化能力差等一些难以克服的缺陷<sup>[9-10]</sup>。支持向量机(Support Vector Machine, SVM)是于20世纪90年代中期由Vladimir N. Vapnik等提出的,基于统计学习理论的一种新的机器学习方法。它通过寻求结构风险最小化来实现经验风险最小化,较好地解决了神经网络的固有问题,已被成功应用于模式识别、回归分析、时序预测等领域<sup>[11-13]</sup>。和其他学习算法一

收稿日期: 2011-03-15; 修订日期: 2011-08-15。

第一作者简介: 赖红松(1967-),男,江西赣县人,博士后,主要从事土地利用规划、土地资源评价和土地资源管理等研究。E-mail: maplhs@163.com

样, SVM 的性能依赖于学习机的参数,但是,目前 SVM 参数的选择主要依靠经验和试算,还没有指导 SVM 参数选择的好方法。粗糙集(Rough Set, RS)理论是由波兰数学家 Z. Pawlak 于 1982 年提出的一种处理不完整性 and 不确定性的数学工具,它不需要预先给定关于数据的任何附加信息,在保留关键信息的前提下对数据进行化简并求得知识的最小表达,识别并评估数据之间的依赖关系,从经验数据中获取最小规则,从而在机器学习、决策分析、过程控制等领域得到了广泛应用<sup>[14-15]</sup>。本文将 RS 理论和 SVM 相结合,提出基于 RS 和 SVM 标准农田地力等级评价方法,同时针对 SVM 参数的选择问题,利用遗传算法(Genetic Algorithms, GA)<sup>[16]</sup>的并行搜索结构和模拟退火(Simulated Annealing, SA)<sup>[17]</sup>的概率突跳特性,提出 GASA 优化 SVM 参数算法,实现 SVM 参数的自动优化选择。该方法首先在建立标准农田地力等级评价指标体系的基础上,利用地力调查样本数据及传统的指数和法评价结果构建 RS 决策表,用 RS 属性约简方法进行属性约简,剔除冗余的评价指标,然后将约简后的指标作为 SVM 的输入,运用 GASA 优化 SVM 参数算法训练 SVM,建立 RS-SVM 评价模型,进行标准农田地力等级评价。尝试将该方法用于温州市鹿城区标准农田地力等级评价,并将其结果与未用 RS 约简的 SVM 和 BP 神经网络方法的评价结果进行比较分析,验证该方法的可行性和有效性,以期为耕地地力评价提供一种新方法。

## 1 粗糙集属性约简

属性约简是粗糙集理论的核心内容之一。它是在保持信息系统的分类或决策能力不变的前提下,删除不相关或不重要的冗余属性。信息系统表示为一个决策表  $S = (U, C \cup D)$ , 其中  $U$  是有限对象集,  $C$  是条件属性集,  $D$  是决策属性集,属性约简是在保持  $C$  与  $D$  之间的依赖度  $\gamma(C, D)$  不变的前提下,删除  $C$  中的冗余属性。子集  $P \subseteq C$ , 若  $\gamma(P, D) = \gamma(C, D)$ , 且  $\forall P' \subset P, \gamma(P', D) \neq \gamma(P, D)$ , 称  $P$  是  $C$  的一个  $D$  约简。

约简通常不惟一,已经证明寻找属性的所有约简或最小约简都是 NP-hard 问题。目前,粗糙集属性约简方法主要有穷尽算法<sup>[18]</sup>、Johnson 贪婪算法<sup>[19]</sup>、属性重要性启发式算法<sup>[20]</sup>、互信息约简算法<sup>[21]</sup>、遗传算法<sup>[22-23]</sup>、复合系统约简<sup>[24]</sup>、扩展法则<sup>[25]</sup>、动态约简<sup>[26]</sup>、二维约简算法<sup>[27]</sup>、概念格<sup>[28]</sup>等。穷尽算法是通过构造分辨矩阵导出分辨函数,并应用吸收律对分辨函数进行化简,使之成为最小析取范式(其中每个主蕴涵均为约简),进而求得数据属性集的约简。穷尽算法虽然算法复杂度高,只适用于较小的数据集,但可以求出所有的约简。本文选用 Rosetta 数据分析软件穷尽算法约简标准农田地力等级评价指标。

## 2 支持向量分类机

### 2.1 标准 SVM 分类算法

支持向量分类的基本思想是对于给定的样本集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ , 其中  $x_i \in R^n, y_i \in \{1, -1\}$ , 通过一个非线性映射  $\phi(\cdot)$  将  $x_i$  映射到一个高维特征空间(Hilbert 空间),并在这个特征空间中构建一个能将训练样本集中的两类点完全正确地分开,而且要使分类间隔  $2/\|w\|$  最大的最优分类超平面,即  $w \cdot \phi(x) + b = 0$ 。标准 SVM 通过求解式(1)优化问题来确定  $w$  和  $b$ :

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s. t.} \quad & y_i [w \cdot \phi(x_i) + b] \geq 1 - \xi_i \end{aligned}$$

$$\xi_i \geq 0, \quad (i = 1, \dots, l) \quad (1)$$

式中:  $\xi_i$  为松弛变量,  $C$  为惩罚参数。

式(1)的对偶形式为

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(x_i, x_j) \alpha_i \alpha_j \\ \text{s. t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C \quad (i = 1, \dots, l) \end{aligned} \quad (2)$$

式中,  $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$  为核函数。由于局部性核函数学习能力强、泛化性能较弱, 而全局性核函数泛化性能强、学习能力较弱, 为获得学习能力和泛化能力较强的核函数, 本文选用局部性核函数高斯径向基核和全局性核函数 4 阶多项式核组合构成的混合核函数, 即  $K(x_i, x_j) = r(x_i \cdot x_j + 1)^4 + (1 - r) \exp(-\|x_i - x_j\|^2 / \sigma^2)$ , 其中  $\sigma$  为高斯径向基核宽度,  $r(0 \leq r \leq 1)$  为调节高斯径向基核和多项式核作用大小系数。求解式(2)得最优解为  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$ , 计算  $b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j)$ ,  $0 < \alpha_j^* < C$ , 从而求得决策函数  $f(x) = \text{sgn} \left[ \sum_{i=1}^l y_i \alpha_i^* K(x_i, x) + b^* \right]$ 。

## 2.2 类别补偿 SVM

在实际应用中, 训练样本集中的正负两类点个数往往不均衡, 这时用标准 SVM 分类算法进行分类, 对正类点集和负类点集应用相同的惩罚参数  $C$  将导致样本数量多的类别分类误差小, 样本数量少的类别分类误差大<sup>[29]</sup>。为使正负两类之间得到平衡的误差率, 令  $C_+ = \frac{l_+}{l} C$ ,  $C_- = \frac{l_-}{l} C$ ,  $l_+$  和  $l_-$  分别为正类和负类的样本个数, 此时, 标准 SVM 优化问题即式(1)变为

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C_+ \sum_{y=1} \xi_i + C_- \sum_{y=-1} \xi_i \\ \text{s. t.} \quad & y_i [w \cdot \phi(x_i) + b] \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad (i = 1, \dots, l) \end{aligned} \quad (3)$$

其对偶形式为

$$\begin{aligned} \max \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j K(x_i, x_j) \alpha_i \alpha_j \\ \text{s. t.} \quad & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C_+, \quad y_i = 1 \\ & 0 \leq \alpha_i \leq C_-, \quad y_i = -1 \end{aligned} \quad (4)$$

## 2.3 SVM 多类分类算法

SVM 原型是处理两类问题的, 而标准农田地力等级评价属于多类分类问题, 因此, 需采用 SVM 多类分类方法。目前, SVM 多类分类方法主要有一类对一类法<sup>[30]</sup>、一类对余类法<sup>[31]</sup>、决策二叉树法<sup>[32]</sup>、决策导向无环图法<sup>[33]</sup>和纠错输出编码法<sup>[34]</sup>、M-ary 法<sup>[35]</sup>等。一类对余类法对于  $M$  类分类问题只需求解  $M$  个两类分类机, 训练时间较短, 可用于大规模数据。但是, 其缺点是当类别数较大时, 某一类的训练样本将大大少于其他类训练样本的总和, 这

种训练样本间的不均衡将影响分类精度。处理这种训练样本间不平衡,可采用类别补偿 SVM 解决。本文采用一类对余类法,其算法步骤如下:

(1) 给定  $M$  类分类问题的训练集  $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$ , 其中  $x_i \in R^n, y_i \in \{1, \dots, M\}$ 。

(2) 对  $j = 1, \dots, M$  进行如下运算:把第  $j$  类看作正类,把其余的  $M - 1$  类看作负类,构造并求解式(4)得最优解为  $\alpha^* = (\alpha_1^*, \dots, \alpha_l^*)$ , 计算  $b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i, x_j), 0 < \alpha_j^* < C$ , 求得决策函数  $f^j(x) = \text{sgn}(g^j(x))$ , 其中  $g^j(x) = \sum_{i=1}^l y_i \alpha_i^* K(x_i, x) + b^*$ 。

(3) 判定输入  $x$  属于第  $J$  类,其中  $J$  是  $g^1(x), \dots, g^M(x)$  中最大者的上标。

## 2.4 SVM 参数的 GASA 优化算法

由 2.1 节可知,以高斯径向基核和 4 阶多项式核组合构成的混合核函数为核函数的 SVM 分类器有 3 个可调参数:核函数参数  $\sigma, r$  和惩罚参数  $C$ 。 $\sigma$  主要影响样本数据在高维特征空间中分布的复杂程度, $r$  决定高斯径向基核和多项式核在混合核函数中的比重,而惩罚参数  $C$  的作用是在确定的特征空间中调节学习机的置信范围和经验风险的比例核。因此,它们的选择将对 SVM 的泛化能力和学习效率产生较大的影响。目前,SVM 参数选择方法主要有试凑法(穷举法)<sup>[36]</sup>、 $k$  折交叉验证法<sup>[11,37-38]</sup>、梯度下降法<sup>[39]</sup>、网格搜索算法<sup>[40]</sup>、遗传算法<sup>[41-42]</sup>、粒子群算法<sup>[43]</sup>、嵌套均匀设计法<sup>[44]</sup>等。GA 是模拟生物进化过程中的自然选择和遗传变异的一种全局性并行随机搜索的优化算法。SA 是基于 Monte Carlo 迭代求解的一种全局概率型搜索算法。GA 虽有较强的全局搜索性能,但在实际应用中易早熟收敛,而改善其性能需增大种群数目或采用高级进化策略。本文利用 GA 的并行搜索结构和模拟退火 SA 的概率突跳特性,提出 GASA 优化 SVM 参数算法,对 SVM 核函数参数  $\sigma, r$  和惩罚参数  $C$  三个参数进行优化,具体步骤如下:

(1) 确定优化 SVM 参数  $\sigma, r, C$  取值范围,并用浮点数编码表示该参数向量  $V$ 。

(2) 确定 GA 和 SA 参数,初始化种群  $p_0$ ,并令  $k=0$ 。

(3) 用不同样本具有不同惩罚参数  $C$  值的改进的序列最小优化算法<sup>[13]</sup>求解式(4)最优解,计算个体适应度  $f = \frac{l}{\sum_{i=1}^l [y_i - f(x_i)]^2}$ 。

(4) 对  $p_k$  中的各个体进行 SA 搜索。首先,用 SA 状态产生函数产生新个体  $V_i(k) = V_i(k-1) + \eta\xi$ ,其中  $\eta$  为扰动幅度参数, $\xi$  为随机扰动变量;然后,计算新、旧个体的适应度之差  $\Delta C$ ;计算接受概率  $P_r = \min[1, \exp(-\Delta C/t_k)]$ ,若  $P_r > \text{random}[0, 1]$ ,用新个体取代旧个体,否则,旧个体不变。

(5) 对  $p_k$  中个体按适应度由大到小进行排序,然后以  $srate$  淘汰率淘汰掉适应度值低的个体,被淘汰掉的个体数为  $n(n = \text{int}(popsize \times srate))$ ,再将未被淘汰掉的优良个体保留到下一代种群  $p_{k+1}$  中。

(6)  $n$  次从  $p_k$  中随机选择两个个体以交叉概率  $p_c = \begin{cases} 0.9 - \frac{0.3(f-f_{avg})}{f_{\max} - f_{\min}}, & f \geq f_{avg} \\ 0.9, & f < f_{avg} \end{cases}$ , 进行

混合交叉操作<sup>[45]</sup>,其中  $f = \max(fits(mate1), fits(mate2))$ ,  $f_{avg} = \frac{\sum_{i=1}^{popsize} fits(x_i)}{popsize}$ , 取适应度值大的新个体填补到  $p_{k+1}$  中,以保持种群规模不变。

$$(7) \text{ 以变异概率 } p_m = \begin{cases} 0.1 - \frac{0.099(f - f_{avg})}{f_{max} - f_{min}}, & f \geq f_{avg} \\ 0.1, & f < f_{avg} \end{cases} \text{ 对步骤(6)产生的新个体进行}$$

高斯变异操作<sup>[45]</sup>,至此产生种群  $p_{k+1}$ 。

(8) 用退温函数  $t_{k+1} = vt_k$  进行 SA 退温。

(9) 若已达到进化最大代数  $maxgen$ ,则终止算法并输出结果;否则令  $k = k + 1$ ,转步骤(4)。

### 3 基于 RS 和 SVM 标准农田地力等级评价方法

首先分析标准农田地力影响因素,选取评价指标,确定评价单元和调查采样点,进行采样点调查、土壤样品采集和分析测试,制定评价指标作用分值标准,利用地力调查样本数据和指数和法评价结果,构建 RS 决策表,应用 RS 属性约简方法进行属性约简,剔除冗余的评价指标,然后将约简后的指标作为 SVM 的输入,用样本训练建立 RS-SVM 评价模型,对评价单元的地力等级进行评价。具体步骤如下:

(1) 选取标准农田地力等级评价指标。

(2) 确定评价单元和调查采样点,采样点调查,土壤样品采集和分析测试。

(3) 制定评价指标作用分值标准。

(4) 应用评价指标作用分值标准,对地力调查样本原始数据进行离散化,利用离散化和指数和法评价结果,构建 RS 决策表。

(5) 运用 Rosetta 数据分析软件穷尽算法对 RS 决策表进行属性约简,剔除冗余的评价指标。

(6) 利用约简后的评价指标数据和指数和法评价结果,构造 SVM 样本集,并将它分成训练样本集和测试样本集两部分。

(7) 运用 GASA 优化 SVM 参数算法用训练样本集对 SVM 进行训练,建立 RS-SVM 评价模型,并用测试样本集测试其性能。

(8) 利用 RS-SVM 评价模型对待评价单元的地力等级进行评价。

## 4 应用实例

### 4.1 研究区概况

鹿城区(120°42'~120°47'E,27°58'~28°09'N)地处温州市中部,瓯江下游内陆南岸,东接龙湾区、乐清市,南临瓯海区,西连丽水市青田县,北隔瓯江与永嘉县相望,辖 16 个街道、1 个镇,是温州市的政治、经济和文化中心。全区东西长 41.43 km,南北宽 20.65 km,呈东南—西北向狭长带状,土地总面积 294.38 km<sup>2</sup>。地势由西北向东南倾斜,海拔介于 2~858 m,地貌类型主要为低山丘陵、平原滩涂、岛屿等。属亚热带海洋季风气候,日照充足,气温温和适中,雨量充沛,年均日照数 1 850 h,年均气温 18 ℃,最冷月(1 月)平均气温 7.6 ℃,最热月(7 月)平均气温 27 ℃,极端最高温 41.3 ℃,极端最低温 -4.5 ℃,年无霜期

280 d, 年均降水量 1 150 mm, 年均蒸发量 894 mm。全区有红壤、粗骨土、潮土和水稻土 4 个土类, 主要农作物为水稻、小麦、番薯、柑橘、甘蔗等。截止 2008 年底, 全区耕地面积 2 937.33 hm<sup>2</sup>, 占土地总面积的 9.98%, 总人口 69.98 × 10<sup>4</sup> 人, 人均耕地 0.004 2 hm<sup>2</sup>, 地区生产总值 (GDP) 286.93 × 10<sup>8</sup> 元, 财政总收入 31.96 × 10<sup>8</sup> 元, 城镇居民人均可支配收入 26 703 元, 农民人均纯收入 11 296 元。

## 4.2 标准农田等级评价因子的确定

根据标准农田立地条件、土壤剖面性状、理化性状等特点, 综合分析影响标准农田地力因素, 选取地貌类型  $c_1$ 、坡度  $c_2$ 、冬季地下水位  $c_3$ 、地表砾石度  $c_4$ 、剖面构型  $c_5$ 、耕层厚度  $c_6$ 、耕层质地  $c_7$ 、容重  $c_8$ 、pH 值  $c_9$ 、阳离子交换量  $c_{10}$ 、水溶性盐总量  $c_{11}$ 、有机质  $c_{12}$ 、有效磷  $c_{13}$ 、速效钾  $c_{14}$  和排涝 (或抗旱) 能力  $c_{15}$  十五项指标, 构建标准农田等级评价指标体系。

## 4.3 评价单元的确定

将标准农田地力调查土壤采样点作为评价单元, 按照《浙江省标准农田地力调查与分等定级技术规范》(以下简称《规范》) 的要求, 对鹿城区 246.78 hm<sup>2</sup> 标准农田进行地力调查取样, 取土样共 20 个, 其中多年种植农田 18 个, 溪滩地 2 个。

## 4.4 指数和法评价结果数据

温州市农业局 2008 年利用鹿城区标准农田地力调查土样数据, 根据《规范》确定的各评价指标生产能力分值 (表 1) 和采用专家经验法确定的各指标权重 (表 2), 应用指数和法计算各土样的综合地力指数 (公式为  $IFI = \sum (F_i \times w_i)$ , 式中:  $IFI$  为标准农田综合地力指数,  $F_i$ 、 $w_i$  分别为第  $i$  个评价指标的生产能力评分值和权重), 依据《规范》确定的标准农田地力等级划分标准 (表 3), 评定了鹿城区标准农田地力调查各土样的地力等级 (表 4)。由表 4 可知, 鹿城区 246.78 hm<sup>2</sup> 标准农田中 2 级田 85.46 hm<sup>2</sup>, 占 34.63%; 3 级田 119.66 hm<sup>2</sup>, 占 48.49%; 4 级田 25.33 hm<sup>2</sup>, 占 10.26%; 5 级田 16.33 hm<sup>2</sup>, 占 6.62%。

## 4.5 粗糙集属性约简

### 4.5.1 数据离散化及 RS 决策表构建

应用评价指标作用分值标准 (表 1), 对各土样原始数据 (表 4) 进行离散化, 利用离散化与指数和法评价结果, 构建 RS 决策表 (表 5)。

### 4.5.2 RS 决策表属性约简

运用由波兰华沙大学和挪威科技大学的科技人员合作开发的 Rosetta 数据分析软件穷尽算法对 RS 决策表进行属性约简, 剔除冗余指标 (水溶性盐总量  $c_{11}$ ) 后, 得到地貌类型  $c_1$ 、坡度  $c_2$ 、冬季地下水位  $c_3$ 、地表砾石度  $c_4$ 、剖面构型  $c_5$ 、耕层厚度  $c_6$ 、耕层质地  $c_7$ 、容重  $c_8$ 、pH 值  $c_9$ 、阳离子交换量  $c_{10}$ 、有机质  $c_{12}$ 、有效磷  $c_{13}$ 、速效钾  $c_{14}$  和排涝 (或抗旱) 能力  $c_{15}$  14 个关键指标。

## 4.6 RS-SVM 评价模型的建立

### 4.6.1 RS-SVM 评价模型样本集

利用 RS 属性约简后的评价指标数据与指数和法评价结果, 构造 RS-SVM 评价模型样本集。将标准农田地力等级为 2 的样本作为正类 (类别标志为 1), 其余样本作为负类 (类别标志为 -1), 构成样本集, 由此构建的 RS-SVM 分类机记为 RS-SVM2, 同理可得 RS-SVM3、RS-SVM4 和 RS-SVM5 的样本集。从 20 个土样中选择 10 个样本 (土样号为 1、2、3、8、9、10、14、16、18、20) 作为 RS-SVM 训练样本 (选择训练样本时各种等级类型应都有, 比较特殊的也应包括在内), 其余 10 个样本 (土样号为 4、5、6、7、11、12、13、15、17、19) 作为测试样本, 用于测试 RS-SVM 模型的性能。

表 1 浙江省标准农田地力等级评价指标生产能力分值

Table 1 Productivity value of productivity evaluation factors of standard cultivated land in Zhejiang Province

地貌类型 $c_1$	水网平原	滨海平原	河谷平原	低丘	高丘
分值	1.0	0.8	0.7	0.5	0.3
坡度 $c_2/(^\circ)$	<3	3~6	6~10	10~15	15~25
分值	1.0	0.8	0.7	0.4	0.1
冬季地下水位 $c_3/\text{cm}$	<20	20~50	50~80	80~100	>100
分值	0.1	0.4	0.7	1.0	0.8
地表砾石度 $c_4/\%$	$\leq 10$	10~25	>25		
分值	1.0	0.5	0.2		
剖面构型 $c_5$	A-Ap-W-C A-[B]-C	A-Ap-P-C A-Ap-Gw-G	A-[B]C-C	A-Ap-C A-Ap-G	A-C
分值	1.0	0.8	0.5	0.3	0.1
耕层厚度 $c_6/\text{cm}$	$\leq 8.0$	8.0~12	12~16	16~20	>20
分值	0.3	0.6	0.8	0.9	1.0
耕层质地 $c_7$	粘壤土	壤土	粘土	砂土	
分值	1.0	0.9	0.7	0.5	
容重 $c_8/(\text{g}/\text{cm}^3)$	0.9~1.1	1.1~1.3, $\leq 0.9$	>1.3		
分值	1.0	0.8	0.5		
pH 值 $c_9$	6.5~7.5	5.5~6.5	7.5~8.5	4.5~5.5	>8.5 $\leq 4.5$
分值	1.0	0.8	0.7	0.4	0.2
阳离子交换量 $c_{10}/(\text{cmol}/\text{kg})$	$\leq 5$	5~10	10~15	15~20	>20
分值	0.1	0.4	0.6	0.9	1.0
水溶性盐总量 $c_{11}/(\text{g}/\text{kg})$	$\leq 1$	1~2	2~3	3~4	4~5 >5
分值	1.0	0.8	0.5	0.3	0.2 0.1
有机质 $c_{12}/(\text{g}/\text{kg})$	>40	30~40	20~30	10~20	$\leq 10$
分值	1.0	0.9	0.8	0.5	0.3
有效磷 $c_{13}/(\text{mg}/\text{kg})$	30~40	20~30	>40, 15~20	10~15	5~10 $\leq 5$
分值	1.0	0.9	0.8	0.7	0.5 0.2
速效钾 $c_{14}/(\text{mg}/\text{kg})$	>150	100~150	80~100	50~80	$\leq 50$
分值	1.0	0.9	0.7	0.5	0.3
排涝(或抗旱)能力 $c_{15}$	一日暴雨一日排出 >70 d	一日暴雨二日排出 50~70 d	一日暴雨三日排出 30~50 d	一日暴雨四日排出 <30 d	
分值	1.0	0.8	0.6	0.4	0.2

表 2 浙江省标准农田地力等级评价指标权重

Table 2 Weight value of productivity evaluation factors of standard cultivated land in Zhejiang Province

评价指标	权重	评价指标	权重	评价指标	权重	评价指标	权重
地貌类型 $c_1$	0.17	地表砾石度 $c_4$	0.06	耕层质地 $c_7$	0.08	阳离子交换量 $c_{10}$	0.08
坡度 $c_2$	0.05	剖面构型 $c_5$	0.03	容重 $c_8$	0.04	水溶性盐总量 $c_{11}$	0.04
冬季地下水位 $c_3$	0.05	耕层厚度 $c_6$	0.07	pH 值 $c_9$	0.06	有机质 $c_{12}$	0.07
						有效磷 $c_{13}$	0.04
						速效钾 $c_{14}$	0.06
						排涝(或抗旱)能力 $c_{15}$	0.10

表3 浙江省标准农田地力等级划分标准

Table 3 Determining standard of productivity grade of standard cultivated land in Zhejiang Province

等级		地力综合指数
一等地	1 级田	$\geq 0.90$
	2 级田	0.80 ~ 0.90
二等地	3 级田	0.70 ~ 0.80
	4 级田	0.60 ~ 0.70
三等地	5 级田	0.50 ~ 0.60
	6 级田	$< 0.50$

表4 温州市鹿城区标准农田地力等级评价指标分值及指数和法评价结果

Table 4 Index value of productivity evaluation of standard cultivated land of Lucheng district of Wenzhou city and its evaluation results of integrated productivity factors method

土样号	标准农田名称	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$	$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$	地力等级	代表面积/hm <sup>2</sup>
1	岙底坦坦畈	0.3	0.1	0.4	0.5	0.1	0.8	0.5	0.8	0.4	0.6	1.0	1.0	0.8	0.5	1.0	5 级	10.93
2	岙底呈岸畈	0.5	0.8	0.4	0.5	0.1	0.8	0.5	0.8	0.2	0.4	1.0	0.9	0.8	0.5	0.2	5 级	5.40
3	上戍西湾畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.4	0.6	1.0	0.9	0.9	0.7	0.2	3 级	10.20
4	上戍西湾畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.4	0.4	1.0	1.0	0.8	0.9	0.2	3 级	13.53
5	临江前盈畈	0.7	0.8	0.4	1.0	0.8	0.9	0.9	0.8	0.4	0.6	1.0	0.9	1.0	0.7	0.4	3 级	5.67
6	临江仁地畈	0.7	0.8	0.4	1.0	0.8	0.9	0.9	0.8	0.4	0.4	1.0	0.8	1.0	0.9	0.2	3 级	14.20
7	临江金岙畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	0.8	0.4	0.1	1.0	0.8	0.8	0.9	0.2	3 级	21.73
8	双潮小旦底段畈	0.7	0.8	0.4	1.0	0.1	0.9	0.5	1.0	0.4	0.1	1.0	0.9	0.5	0.9	0.2	4 级	25.33
9	双潮小旦畈	1.0	0.8	1.0	0.5	0.1	0.9	0.5	1.0	0.4	0.6	1.0	0.9	0.8	0.9	0.2	3 级	0.80
10	藤桥后垵畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.4	0.4	1.0	1.0	0.7	0.9	0.4	3 级	23.53
11	藤桥后垵畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.8	0.4	1.0	1.0	0.7	0.9	0.4	2 级	12.67
12	藤桥垵岸田	1.0	0.8	0.4	0.2	0.8	0.9	0.9	1.0	0.4	0.6	1.0	0.9	0.5	0.9	1.0	3 级	8.47
13	藤桥南岸畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.4	0.6	1.0	1.0	0.9	0.9	0.6	2 级	4.47
14	藤桥北岸畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	1.0	0.8	0.6	1.0	0.9	0.7	0.9	0.6	2 级	3.53
15	藤桥底垵畈	1.0	0.8	0.4	1.0	0.8	0.8	0.9	1.0	0.4	0.6	1.0	0.9	0.8	0.9	0.6	2 级	16.33
16	藤桥下庄畈	1.0	0.8	0.4	1.0	0.8	0.8	0.9	1.0	0.4	0.6	1.0	1.0	0.9	1.0	0.6	2 级	6.27
17	藤桥油岙畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	0.8	0.4	0.6	1.0	1.0	0.8	0.9	0.6	2 级	11.53
18	藤桥潮济畈	0.5	0.8	0.4	0.5	0.5	0.6	0.9	1.0	0.8	0.9	1.0	0.5	0.9	1.0	1.0	3 级	21.53
19	藤桥埭马畈	1.0	0.8	0.4	1.0	1.0	0.8	0.5	0.8	0.8	0.6	1.0	0.9	0.9	1.0	0.6	2 级	7.13
20	藤桥山根畈	1.0	0.8	0.4	1.0	0.8	0.9	0.9	0.5	0.4	0.6	1.0	1.0	1.0	0.9	0.6	2 级	23.53

#### 4.6.2 建立 RS-SVM 评价模型

用 Visual C++6.0 编程实现 GASA 优化 SVM 参数算法。算法参数  $C$ 、 $\sigma$  和  $r$  的取值范围分别为  $[1, 1000]$ 、 $[0.1, 10]$  和  $[0, 1.0]$ ，其他参数选取为  $popsize = 200$ ， $srate = 0.9$ ， $t_0 = 10000$ ， $v = 0.85$ ， $maxgen = 80$ 。运用 GASA 优化 SVM 参数算法用 RS-SVM 训练样本对 SVM 进行训练，得到 RS-SVM2、RS-SVM3、RS-SVM4 和 RS-SVM5 评价模型，其优化参数  $C$ 、 $\sigma$  和  $r$  的优化值和训练误差 MSE 如表 6。



表 5 RS 决策表

Table 5 Decision table of RS

土样号	标准农田名称	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$	$c_{10}$	$c_{11}$	$c_{12}$	$c_{13}$	$c_{14}$	$c_{15}$	$d$
1	岙底坦畈	0	0	0	1	0	1	0	1	1	2	0	3	2	0	3	5
2	岙底呈岸畈	1	1	0	1	0	1	0	1	0	1	0	2	2	0	0	5
3	上戍西湾畈	3	1	0	2	2	2	1	2	1	2	0	2	3	1	0	3
4	上戍西湾畈	3	1	0	2	2	2	1	2	1	1	0	3	2	2	0	3
5	临江前盈畈	2	1	0	2	2	2	1	1	1	2	0	2	4	1	1	3
6	临江仁地畈	2	1	0	2	2	2	1	1	1	1	0	1	4	2	0	3
7	临江金岙畈	3	1	0	2	2	2	1	1	1	0	0	1	2	2	0	3
8	双潮小旦底段畈	2	1	0	2	0	2	0	2	1	0	0	2	0	2	0	4
9	双潮小旦畈	3	1	1	1	0	2	0	2	1	2	0	2	2	2	0	3
10	藤桥后垟畈	3	1	0	2	2	2	1	2	1	1	0	3	1	2	1	3
11	藤桥后垟畈	3	1	0	2	2	2	1	2	2	1	0	3	1	2	1	2
12	藤桥岸田	3	1	0	0	2	2	1	2	1	2	0	2	0	2	3	3
13	藤桥南岸畈	3	1	0	2	2	2	1	2	1	2	0	3	3	2	2	2
14	藤桥北岸畈	3	1	0	2	2	2	1	2	2	2	0	2	1	2	2	2
15	藤桥底垟畈	3	1	0	2	2	1	1	2	1	2	0	2	2	2	2	2
16	藤桥下庄畈	3	1	0	2	2	1	1	2	1	2	0	3	3	3	2	2
17	藤桥油岙畈	3	1	0	2	2	2	1	1	1	2	0	3	2	2	2	2
18	藤桥潮济畈	1	1	0	1	1	0	1	2	2	3	0	0	3	3	3	3
19	藤桥埭马畈	3	1	0	2	3	1	0	1	2	2	0	2	3	3	2	2
20	藤桥山根畈	3	1	0	2	2	2	1	0	1	2	0	3	4	2	2	2

表 6 SVM 模型和 RS-SVM 模型参数优化值

Table 6 Optimizing value of SVM model parameters and RS-SVM model parameters

参数	SVM 模型				RS-SVM 模型			
	SVM2	SVM3	SVM4	SVM5	RS-SVM2	RS-SVM3	RS-SVM4	RS-SVM5
$C$	430.442 7	445.528 4	292.166 6	398.126 5	439.967 2	439.022 7	204.059 7	476.534 8
$\sigma$	5.702 7	7.626 9	5.538 4	6.311 5	7.277 8	6.480 5	4.896 2	7.921 7
$r$	0.000 1	0.000 7	0.000 0	0.000 0	0.000 1	0.000 1	0.000 0	0.000 0
MSE	$9.725 8 \times 10^{-9}$	$5.938 1 \times 10^{-9}$	$1.660 1 \times 10^{-8}$	$1.122 2 \times 10^{-8}$	$6.337 2 \times 10^{-9}$	$7.273 5 \times 10^{-9}$	$1.063 4 \times 10^{-8}$	$1.088 1 \times 10^{-8}$

#### 4.7 RS-SVM 评价结果及与单一 SVM 和 BP 神经网络评价结果的比较

利用已训练好的 RS-SVM2、RS-SVM3、RS-SVM4 和 RS-SVM5 评价模型对 RS-SVM 测试样本进行  $g(x)$  值计算,其  $g(x)$  计算值和 RS-SVM 评价结果见表 7。

按 RS-SVM 评价模型样本集构造方法,利用 RS 约简前 15 个评价指标数据与指数和法评价结果,构造 SVM2、SVM3、SVM4 和 SVM5 评价模型样本集。选用与 RS-SVM 评价相同土样号样本作为 SVM 训练样本和测试样本。算法参数  $C$ 、 $\sigma$  和  $r$  的取值范围及  $popsize$ 、 $srate$ 、 $t_0$ 、 $v$ 、 $maxgen$  参数的取值同 RS-SVM 评价模型。运用 GASA 优化 SVM 参数算法用 SVM 训练样本对 SVM 进行训练,得到 SVM2、SVM3、SVM4 和 SVM5 评价模型,其优化参数  $C$ 、 $\sigma$  和  $r$  的优化值和训练误差 MSE 如表 6。利用已训练好的 SVM2、SVM3、SVM4 和 SVM5 评价模型对 SVM 测试样本进行  $g(x)$  值计算,其  $g(x)$  计算值和 SVM 评价结果见表 7。

表 7 SVM 法和 RS-SVM 法评价结果比较

Table 7 Comparison of Evaluation results of SVM method and RS-SVM method

土样号	指数和法	SVM 法					RS-SVM 法				
	等级	SVM2	SVM3	SVM4	SVM5	等级	RS-SVM2	RS-SVM3	RS-SVM4	RS-SVM5	等级
4	3 级	-1.108	1.186	-0.963	-0.981	3 级	-1.110	1.095	-0.966	-0.931	3 级
5	3 级	-0.303	-0.122	-0.907	-0.754	3 级	-0.292	-0.078	-0.916	-0.720	3 级
6	3 级	-0.510	0.114	-0.839	-0.700	3 级	-0.515	0.058	-0.848	-0.648	3 级
7	3 级	-0.632	0.232	-0.788	-0.663	3 级	-0.652	0.141	-0.796	-0.612	3 级
11	2 级	0.075	-0.173	-0.938	-0.947	2 级	0.099	-0.206	-0.946	-0.865	2 级
12	3 级	-0.413	-0.066	-0.792	-0.551	3 级	-0.393	-0.206	-0.775	-0.541	3 级
13	2 级	0.746	-0.721	-1.024	-1.026	2 级	0.753	-0.710	-1.025	-1.020	2 级
15	2 级	0.581	-0.618	-1.011	-0.967	2 级	0.562	-0.619	-1.015	-0.990	2 级
17	2 级	0.653	-0.712	-1.031	-1.005	2 级	0.654	-0.676	-1.038	-1.002	2 级
19	2 级	-0.018	-0.678	-0.843	-0.789	2 级	0.097	-0.437	-0.842	-0.656	2 级

利用 RS 约简前 15 个评价指标数据与指数和法评价结果,以土样指标分值作为 BP 神经网络的输入,以其对应的等级作为 BP 神经网络的输出,即土样等级,若为 2 级,则输出为 (1,0,0,0);若为 3 级,则输出为 (0,1,0,0);若为 4 级,则输出为 (0,0,1,0);若为 5 级,则输出为 (0,0,0,1),构造 BP 神经网络评价模型样本集。选用与 RS-SVM 评价相同土样号样本作为 BP 神经网络的训练样本和测试样本。利用训练样本,选择计算结点的激活函数均为 sigmoid 函数,隐层结点数为 5,即网络结构为 15-5-4,最大训练次数为 7 000 000,动量系数  $\alpha$  为 0.95,学习率  $\eta$  为 0.4 的改进的 3 层 BP 神经网络<sup>[9]</sup>建立 BP 神经网络评价模型。BP 神经网络训练后误差 MSE 为 0.000 362 216 3。以测试样本指标分值作为训练后网络的输入,进行其地力等级评价,结果见表 8。

表 8 BP 神经网络法评价结果

Table 8 Evaluation results of BP networks method

土样号	指数和法	BP 网络法				
	等级	输出值				等级
4	3 级	0.000	1.000	0.000	0.000	3 级
5	3 级	0.001	0.985	0.000	0.001	3 级
6	3 级	0.000	1.000	0.000	0.000	3 级
7	3 级	0.000	1.000	0.001	0.000	3 级
11	2 级	0.302	0.730	0.001	0.000	3 级
12	3 级	0.000	1.000	0.000	0.000	3 级
13	2 级	0.994	0.005	0.000	0.000	2 级
15	2 级	0.985	0.013	0.000	0.000	2 级
17	2 级	1.000	0.000	0.000	0.000	2 级
19	2 级	1.000	0.000	0.000	0.000	2 级

从表 7 可看出,SVM 法、RS-SVM 法与指数和法的评价结果完全相同,SVM 法和 RS-SVM 法的评价正确率均为 100%,表明通过 RS 属性约简后,SVM 评价精度并没有降低,但可以减少使用的评价指标,降低 SVM 计算复杂度,提高训练速度,减少训练时间;SVM 用于标准农田地力等级评价是可行有效的,且评价精度高。从表 8 可看出,BP 神经网络法与指

数和法的评价结果基本一致,惟有 1 个(11 号)土样的评价结果不相同,BP 神经网络法评价正确率为 90%。由此可见,与 BP 神经网络法相比,SVM 法的评价结果更准确可靠,原因是 SVM 克服了 BP 神经网络易陷入局部极小等缺陷,具有比 BP 神经网络更优的性能。

每种方法都有其适用性。综上可知,RS 属性约简方法由于能在保持信息系统分类能力不变的前提下,约去不必要的属性,可用于约简(筛选)评价指标。RS-SVM 法适用于评价指标未筛选的耕地地力评价。SVM 法和 BP 神经网络法均可用于评价指标已筛选的耕地地力评价,但是 SVM 用于耕地地力评价具有比 BP 神经网络更高的精度。指数和法步骤复杂,工作量大,且易受专家的知识、经验等主观因素影响,可用于少量评价单元的耕地地力评价,其评价结果是构建 RS 决策表,应用 RS 属性约简方法对评价指标进行约简(筛选)的依据,也是 SVM 法或 BP 神经网络法的训练样本的来源。

## 5 结论

针对 SVM 参数的选择问题,利用 GA 的并行搜索结构和 SA 的概率突跳特性,提出了 GASA 优化 SVM 参数算法,实现了 SVM 参数的自动优化选择,避免了人为选择 SVM 参数的盲目性。基于小样本的统计学习理论的 SVM 有效克服了人工神经网络过学习、局部极小等缺陷,具有很强的泛化能力,将 SVM 用于标准农田地力等级评价,选用少量典型样本训练建立 SVM 评价模型后,可简便、快捷地给出大规模待定评价单元的地力等级。将 RS 理论和 SVM 相结合,提出了基于 RS 和 SVM 的标准农田地力等级评价方法。实例研究结果表明,通过 RS 约简评价指标后,SVM 评价精度并没有降低,但减少了使用的评价指标,降低了 SVM 输入向量维数和计算复杂度,减少了训练时间;SVM 用于标准农田地力等级评价,具有比 BP 神经网络更高的精度。

## 参考文献(References):

- [1] 王瑞燕,赵庚星,李涛. 山东省青州市耕地地力等级评价研究[J]. 土壤, 2004, 36(1): 76-80. [WANG Rui-yan, ZHAO Geng-xing, LI Tao. Grading and assessment of soil productivity of cultivated land in Qingzhou. *Soil*, 2004, 36(1): 76-80.]
- [2] 鲁明星,贺立源,吴礼树,等. 基于 GIS 的华中丘陵区耕地地力评价研究[J]. 农业工程学报, 2006, 22(8): 96-101. [LU Ming-xing, HE Li-yuan, WU Li-shu, et al. Evaluation of the farmland productivity of hilly region of central China based on GIS. *Transactions of the CSAE*, 2006, 22(8): 96-101.]
- [3] 袁秀杰,赵庚星,朱雪欣. 平原和丘陵区耕地地力评价及其指标体系衔接研究[J]. 农业工程学报, 2008, 24(7): 65-71. [YUAN Xiu-jie, ZHAO Geng-xing, ZHU Xue-xin. Linkage of evaluation index system for cultivated land productivity evaluation in plain and hill regions. *Transactions of the CSAE*, 2008, 24(7): 65-71.]
- [4] 方灿华,马友华,钱国平,等. 基于 GIS 的明光市耕地地力评价[J]. 中国农学通报, 2008, 24(12): 308-312. [FANG Can-hua, MA You-hua, QIAN Guo-ping, et al. Evaluation of the farmland productivity of Mingguang city based on GIS. *Chinese Agricultural Science Bulletin*, 2008, 24(12): 308-312.]
- [5] 李贤胜,叶军华,杨平,等. 基于 GIS 的广德县耕地地力定量评价[J]. 土壤, 2009, 41(3): 490-494. [LI Xian-sheng, YE Jun-hua, YANG Ping, et al. Quantitative assessment of cultivated-land capacity in Guangde county based on GIS. *Soil*, 2009, 41(3): 490-494.]
- [6] 刘永文,樊燕,刘洪斌. 丘陵山地耕地地力评价研究[J]. 中国农学通报, 2009, 25(18): 420-425. [LIU Yong-wen, FAN Yan, LIU Hong-bin. Evaluation of cultivated land productivity in hilly area. *Chinese Agricultural Science Bulletin*, 2009, 25(18): 420-425.]
- [7] Dung E J, Sugurnaran R. Development of an agricultural land evaluation and site assessment (LESA) decision support tool using remote sensing and geographic information system [J]. *Journal of Soil and Water Conservation*, 2005, 60(5): 228-235.

- [8] 王瑞燕, 赵庚星, 陈丽丽. 基于 ANN-产量的耕地地力定量评价模型及其应用[J]. 农业工程学报, 2008, 24(1): 113-118. [WANG Rui-yan, ZHAO Geng-xing, CHEN Li-li. Evaluation model of cultivated land productivity using artificial neural network and productivity and its application. *Transactions of the CSAE*, 2008, 24(1): 113-118.]
- [9] 阎平凡, 张长水. 人工神经网络与模拟进化计算[M]. 北京: 清华大学出版社, 2000. [YAN Ping-fan, ZHANG Chang-shui. *Artificial Neural Networks and Simulated Evolutionary Computation*. Beijing: Tsinghua University Press, 2000.]
- [10] 王凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 2001. [WANG Ling. *Intelligent Optimization Algorithms with Application*. Beijing: Tsinghua University Press, 2001.]
- [11] 邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004. [DENG Nai-yang, TIAN Ying-jie. *A New Method for Data Mining—Support Vector Machine*. Beijing: China Science Press, 2004.]
- [12] Nello Cristianini, John Shawe-Taylor. 支持向量机导论[M]. 北京: 机械工业出版社, 2005. [Nello Cristianini, John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Beijing: China Machine Press, 2005.]
- [13] 曹丽娟, 王小明. 金融工程的支持向量机方法[M]. 上海: 上海财经大学出版社, 2007. [CAO Li-juan, WANG Xiao-ming. *Support Vector Machine Based Methods for Financial and Engineering Problems*. Shanghai: Shanghai University of Finance and Economics Press, 2007.]
- [14] 梁吉业, 曲开社, 徐宗本. 信息系统的属性约简[J]. 系统工程理论与实践, 2001(12): 76-80. [LIANG Ji-ye, QU Kai-she, XU Zong-ben. Reduction of attribute in information systems. *Systems Engineering Theory & Practice*, 2001(12): 76-80.]
- [15] 安利平. 基于粗糙集理论的多属性决策分析[M]. 北京: 科学出版社, 2008. [AN Li-ping. *Rough Set Approach to Multi-attribute Decision Analysis*. Beijing: China Science Press, 2008.]
- [16] Goldberg D E. *Genetic Algorithms in Search, Optimization and Machine Learning* [M]. Reading M A: Addison-Wesley, 1989.
- [17] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing [J]. *Science*, 1983, 220(4598): 671-680.
- [18] Skowron A, Rauszer C. The discernibility matrices and functions in information system [C]//Slowinski R. *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht: Kluwer Academic Publishers, 1992: 331-362.
- [19] Nguyen S H, Nguyen H S. Some efficient algorithms for rough set methods [C]//Proceedings of the Fifth International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU'96). Granada, Spain, 1996: 1451-1456.
- [20] Hu Xiaohua. *Knowledge Discovery in Databases; An Attribute-oriented Rough Set Approach* [D]. University of Regina, Canada, 1995.
- [21] 苗夺谦, 胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展, 1999, 36(6): 681-684. [MIAO Duo-qian, HU Gui-rong. A heuristic algorithm for reduction of knowledge. *Journal of Computer Research & Development*, 1999, 36(6): 681-684.]
- [22] Wroblewski J. Finding minimal reducts using genetic algorithms [C]//Wang P P. *Proceedings of the Second Annual Joint Conference on Information Sciences*. Wrightsville, Beach, North Carolina, 1995: 186-189.
- [23] Vinterbo S, Øhrn A. Minimal approximate hitting sets and rule templates [J]. *International Journal of Approximate Reasoning*, 2000, 25(2): 123-143.
- [24] Kryszkiewicz M, Rybinski H. Finding reducts in composed information systems [C]//Ziarko W P. *Proceedings of the Second International Workshop on Rough Sets and Knowledge Discovery (RSKD'93)*. Banff, Alberta, Canada, 1993: 261-273.
- [25] Starzyk J, Nelson D E, Sturtz K. Reduct generation in information system [J]. *Bulletin of International Rough Set Society*, 1999, 3(1): 19-22.
- [26] Bazan G J, Skowron A, Synak P. Dynamic reducts as a tool for extracting laws from decision tables [C]//Polkowski L, Skowron A. *International Symposium on Methodologies for Intelligent Systems*. Charlotte, NC: Springer-Verlag, 1994:

- 346-355.
- [27] Hashemi R R, Jelovsek F R, Razzaghi M. Developmental toxicity risk assessment: A rough sets approach [J]. *Methods of Information in Medicine*, 1993, 32(1): 47-54.
- [28] 张文修, 魏玲, 祁建军. 概念格的属性约简理论与方法[J]. 中国科学 E 辑: 信息科学, 2005, 35(6): 628-639. [ZHANG Wen-xiu, WEI Ling, QI Jian-jun. The attribute reduction theory and method of concept lattice. *Science in China Ser. E Information Sciences*, 2005, 35(6): 628-639.]
- [29] 范昕炜, 杜树新, 吴铁军. 可补偿类别差异的加权支持向量机算法[J]. 中国图象图形学报, 2003, 8(9): 1037-1042. [FAN Xin-wei, DU Shu-xin, WU Tie-jun. Weighted support vector machine based classification algorithm for uneven class size problems. *Journal of Image and Graphics*, 2003, 8(9): 1037-1042.]
- [30] KreBel U H G. Pairwise classification and support vector machines [C]//Schölkopf B, Burges C J C, Smola A J. *Advances in Kernel Methods; Support Vector Learning*. Cambridge, Massachusetts; MIT Press, 1999: 255-268.
- [31] Bottou L, Cortes C, Denker J, et al. A comparison of classifier methods: A case study in handwritten digit recognition [C]//Proceedings of the 12th International Conference on Pattern Recognition. Jerusalem, 1994: 77-87.
- [32] Takahashi F, Abe S. Decision-tree-based multi-class support vector machines [C]//Proceedings of the Ninth International Conference on Neural Information Processing. Singapore, 2002, 1418-1422.
- [33] Platt J C, Cristianini N, Shawe-Taylor J. Large margin DAG's for multi-class classification [C]//Advances in neural information processing systems. Cambridge, Massachusetts: MIT Press, 2000: 547-553.
- [34] Dietterich T G, Bakiri G. Solving multi-class learning problems via error-correcting output codes [J]. *Journal of Artificial Intelligence Research*, 1995(2): 263-286.
- [35] Sebald D J, Buchlew J A. Support vector machines and the multiple hypothesis test problem [J]. *IEEE Transactions on Signal Processing*, 2001, 49(11): 2865-2872.
- [36] 王睿. 关于支持向量机参数选择方法分析[J]. 重庆师范大学学报: 自然科学版, 2007, 24(2): 36-38, 42. [WANG Rui. Method analysis about support vector machine parameter. *Journal of Chongqing Normal University: Natural Science Edition*, 2007, 24(2): 36-38, 42.]
- [37] 邓乃扬, 田英杰. 支持向量机——理论、算法与拓展[M]. 北京: 科学出版社, 2009. [DENG Nai-yang, TIAN Ying-jie. *Support Vector Machine—Theory, Algorithms and Extension*. Beijing: Science Press, 2009.]
- [38] 苏高利, 邓芳萍. 关于支持向量回归机的模型选择[J]. 科技通报, 2006, 22(2): 154-158. [SU Gao-li, DENG Fang-ping. Introduction to model selection of SVM regression. *Bulletin of Science and Technology*, 2006, 22(2): 154-158.]
- [39] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines [J]. *Machine Learning*, 2002, 46(1/3): 131-159.
- [40] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification [R]. Department of Computer Science and information Engineering, National Taiwan University, 2003.
- [41] Zheng Chunhong, Jiao Licheng. Automatic parameters selection for SVM based on GA [C]//Proceedings of the 5th World Congress on Intelligent Control and Automation. Hangzhou, China, 2004: 1869-1872.
- [42] 刘胜, 李妍妍. 自适应 GA-SVM 参数选择算法研究[J]. 哈尔滨工程大学报, 2007, 28(4): 398-402. [LIU Sheng, LI Yan-yan. Parameter selection algorithm for support vector machines based on adaptive genetic algorithm. *Journal of Harbin Engineering University*, 2007, 28(4): 398-402.]
- [43] Lee T F, Cho M Y, Shieh C S, et al. Particle swarm optimization-based SVM for Incipient Fault Classification of Power Transformers [C]//Esposito F, Ras Z W, Malerba D, et al. *Proceedings of the 16th International Symposium on Methodologies for Intelligent Systems, Foundations of Intelligent Systems, Lecture Notes in Computer Science*. Berlin Heidelberg: Springer-Verlag, 2006: 84-90.
- [44] Huang C M, Lee Y J, Lin D K J, et al. Model selection for support vector machines via uniform design [J]. *Computational Statistics & Data Analysis*, 2007, 52(1): 335-346.
- [45] 林丹, 李敏强, 寇纪松. 基于实数编码的遗传算法的收敛性研究[J]. 计算机研究与发展, 2000, 37(11): 1311-1327. [LIN Dan, LI Min-qiang, KOU Ji-song. On the convergence of real-coded genetic algorithms. *Journal Computer Research & Development*, 2000, 37(11): 1311-1327.]

# Productivity Evaluation of Standard Cultivated Land Based on Rough Set and Support Vector Machine

LAI Hong-song, WU Ci-fang

(College of Public Administration, Zhejiang University, Hangzhou 310029, China)

**Abstract:** Standard Cultivated Land (SCL) is the gem of all cultivated land, which is the key of ensuring China's grain security. Scientific evaluating productivity of SCL has a significant sense for improving soil productivity and improving soil of SCL. Integrated Rough Set (RS) theory with Support Vector Machine (SVM), a productivity evaluation method of SCL based on RS and SVM is proposed. Combining the parallel searching structure of genetic algorithms with the probabilistic jumping property of simulated annealing, a GASA-optimized algorithm is put forward for selecting of SVM parameters. In this new evaluating method, on the basis of determining productivity evaluation indexes of SCL, firstly the decision making table is constructed by using the sample data of productivity surveying of SCL and its evaluation results of traditional integrated productivity factors method, and the redundant indexes are removed through the exhaustive reducer of Rosetta software which was jointly developed by Norwegian University of Science and Technology and Warsaw University, then the reduced indexes are used as the input of SVM, finally, the SVM is trained with the training samples through the GASA-optimized algorithm, and the RS-SVM evaluating model of SCL is built. The method is tested on productivity evaluation of SCL of Lucheng District of Wenzhou City, Zhejiang, and its results are compared with that of the SVM without RS reduction method and the BP networks without RS reduction method. The input indexes of the SVM model and the BP networks model are both 15, With correct evaluating rate being 100% and 90% respectively. The input indexes of the RS-SVM model are 14, whose correct evaluating rate is 100%. The results indicate that the suggested SVM method is of the same accuracy as the SVM without RS reduction method, but it can reduce the dimensions of SVM input vectors and the computing complexity of SVM, and raise the SVM training efficiency. The results also show that the method which is used SVM can achieve greater evaluating accuracy than the BP networks without RS reduction method. It is concluded that the suggested SVM method is feasible and effective in evaluating productivity of SCL, and this method has explored a new way for evaluating productivity of cultivated land.

**Key words:** productivity evaluation of cultivated land; standard cultivated land; Rough Set; support vector machine; Lucheng District