

基于自适应进化相关向量机的耕地面积预测模型

罗亦泳^{1,2}, 张豪^{3*}, 张立亭¹

(1. 东华理工大学测绘工程学院, 南昌 330013; 2. 武汉大学测绘学院, 武汉 430079;
3. 浙江工业大学建筑工程学院, 杭州 310032)

摘要:为解决耕地面积预测模型建立过程中的非线性、稀疏化及结果可靠性评价等问题, 该文将相关向量机与差分进化优化算法进行融合及改进, 提出基于自适应进化相关向量机的耕地面积预测模型。该文以黄石市为例, 建立基于自适应进化相关向量机的短期、中期耕地预测模型, 并与多元线性回归、BP神经网络、支持向量机算法在精度、计算效率及可靠性方面进行对比分析。试验数据表明, 自适应进化相关向量机的预测精度大约是其余3种方法的2倍以上; 模型的计算效率是多元线性回归方法的2倍, 比BP神经网络、支持向量机高出2个数量级; 测试数据的实际耕地面积均在自适应进化相关向量机估计的95%置信度的置信区间内, 并且由后验差比、小误差概率判定模型等级属于“好”。基于以上数据, 证实该模型是一种精度高、计算快、可靠性强的耕地预测新方法。

关键词: 土地利用; 算法; 支持向量机; 耕地面积

doi: 10.11975/j.issn.1002-6819.2015.09.039

中图分类号: TP301.6; F301.21

文献标志码: A

文章编号: 1002-6819(2015)-09-0257-08

罗亦泳, 张豪, 张立亭. 基于自适应进化相关向量机的耕地面积预测模型[J]. 农业工程学报, 2015, 31(9): 257—264.

Luo Yiyong, Zhang Hao, Zhang Liting. Prediction model for cultivated land area based on self-adaptive differential evolution and relevance vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(9): 257—264. (in Chinese with English abstract)

0 引言

耕地是土地资源利用方式的重要组成部分, 保持一定数量的耕地是人类赖以生存和社会发展的物质基础。然而, 由于中国社会、经济快速发展, 人口数量持续增长, 工业化和城镇化快速推进, 耕地数量不断减小, 人地矛盾日益尖锐。中国一直处于人多地少状态, 改革开放初期至2011年耕地面积减少1100万 hm^2 , 2011年末人均耕地面积仅为0.091 hm^2 , 不足世界平均水平的40%, 如果耕地得不到有效保护, 将严重威胁到中国的粮食安全, 影响社会稳定及经济可持续发展^[1-2]。因此, 耕地面积变化及驱动力分析研究引起国内外学者广泛关注, 如何构建耕地面积与驱动力之间复杂的函数模型, 预测耕地面积变化情况成为当今土地利用变化热点研究问题。建立合理的耕地面积变化预测模型, 准确掌握未来耕地变化趋势, 可为土地利用总体规划及耕地管理政策制定提供可靠理论依据, 有助于合理利用耕地资源、控制耕地面积进一步减少, 对保障粮食生产安全, 促进社会和谐发展有着重要现实意义。

当前耕地面积变化预测模型研究取得丰富成果, 传

统的耕地预测模型主要包括线性回归模型、时间序列分析、灰色理论、马尔科夫及模糊数学等。岳云华等建立基于灰色理论的耕地预测模型^[3]; 刘耀林等^[4]学者将灰色理论、模糊数学与马尔柯夫链结合, 较大程度克服马尔科夫链在耕地预测中的缺点。但这些模型未充分综合考虑社会、自然与经济等复杂影响因素对耕地变化的影响, 导致传统耕地预测模型对耕地变化这一复杂非线性过程拟合精度不高。随着神经网络及元胞自动机等人工智能算法的提出, 曹银贵等^[5]、邓羽等^[6]基于经验风险的神经网络模型建立耕地预测方法, 该算法较好地拟合了耕地变化与驱动力之间函数关系, 精度优于传统方法。但是神经网络存在容易陷入局部最优、结构设计复杂等缺点, 影响该算法的使用精度。Arsanjani等^[7]、Yang等^[8]基于元胞自动机模型预测耕地变化情况, 在时空分布变化取得很好效果, 但在宏观数量预测精度有待于提高。针对神经网络算法的缺点, 张豪等^[9-10]、王霞等^[11]将基于结构风险最小原则的支持向量机(SVM)算法应用于耕地变化预测, 该算法预测精度较神经网络有较大提高。但是SVM算法存在待确定参数较多、核函数限制苛刻等问题, 影响该模型在耕地面积预测的实际应用效果。综上所述, 当前耕地预测模型研究热点主要集中于人工智能、机器学习(如SVM)等方法, 其耕地预测精度较传统方法有较大提高。这些方法研究重点集中在预测精度提高方面, 对算法稀疏化及预测结果的不确定性分析研究很少, 研究稀疏化和具有概率意义的耕地预测模型对提高模型精度、计算效率及结果可靠性具有重要意义。

收稿日期: 2014-11-26 修订日期: 2015-03-29

基金项目: 国家自然科学基金项目(41204003); 国家自然科学基金项目(41374007); 江西省数字国土重点实验室开放研究基金(DLLJ201411)

作者简介: 罗亦泳, 男, 江西南昌人, 博士生, 讲师, 主要从事土地信息技术研究。南昌 东华理工大学, 测绘学院, 330013。Email: ecityyluo82@163.com
*通信作者: 张豪, 男, 江苏南京人, 教授, 博士, 主要从事土地信息技术研究。杭州 浙江工业大学, 建筑工程学院, 310032。Email: zhanghao@zjut.edu.cn

Tipping 基于贝叶斯原理、马尔科夫性质、自动相关决定先验与最大似然等理论提出相关向量机 (relevance vector machine, RVM) [12-13]。相关向量机是一种贝叶斯框架下的稀疏概率模型, 该算法具有稀疏性的同时可获得预测结果的不确定性信息, 并且算法待参数少, 核函数不受 Mercer 条件限制。鉴于相关向量机以上优点, 该算法在解决诸多模式识别和回归估计的实际问题中取得了良好效果, 但在耕地面积预测方面未见相关研究。因此, 针对相关向量机的特点, 本文将该算法应用于耕地预测。并将全局搜索能力强、鲁棒性好与收敛速度快的自适应差分进化算法 [14] 与相关向量机结合, 优化相关向量机待参数, 提出基于自适应进化相关向量机的耕地预测模型。该耕地预测模型在获得预测结果的同时可计算预测变量的均值、方差及置信区间, 并且模型具有很强的稀疏性, 有效提高模型计算效率。以黄石市耕地面积预测为例, 详细分析基于自适应进化相关向量机的耕地预测模型建立过程, 并与多种耕地预测方法进行对比分析, 证明本文方法在模型精度、可靠性及稀疏性方面的优点, 以期对相关向量机进一步研究及应用提供理论与实践参考。

1 自适应进化相关向量机模型

相关向量机待参数较少, 仅包含核函数参数, 但该待参数是否合理对模型预测精度影响较大。因此, 如何优化核函数参数成为相关向量机研究关键问题。鉴于差分进化算法在参数优化方面的优势, 本文对差分进化算法进行改进, 构建自适应差分进化优化算法。利用自适应差分进化算法自适应确定相关向量机待参数, 提出自适应进化相关向量机模型 (SADE-RVM, self-adaptive differential evolution-relevance vector machine), 提高模型的预测精度。

1.1 相关向量机基本理论

对于相关向量机回归模型 (RVM), 设训练样本集 $\{\mathbf{x}_i, t_i\}_{i=1}^N, \mathbf{x}_i \in R^d$ 为输入向量, $t_i \in R$ 为目标值。假设目标 t_i 独立且带有高斯白色噪声 ε_i , t_i 函数模型表示为 [15]:

$$t_i = y(\mathbf{x}_i, \mathbf{w}) + \varepsilon_i \quad (1)$$

式中: $\varepsilon_i \sim N(0, \sigma^2)$, 则 $p(t_i | \mathbf{x}_i) = N(t_i | y(\mathbf{x}_i, \mathbf{w}), \sigma^2)$ 。RVM 输出模型将 $y(\mathbf{x}_i, \mathbf{w})$ 表示为非线性基函数 $\phi_i(\mathbf{x})$ 的加权组合, $\phi_i(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}_i)$ 为基函数 (核函数), 不要求其为正定, 核函数不需要满足 Mercer 条件。则 RVM 模型可表示为:

$$y(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^N w_i \cdot K(\mathbf{x}, \mathbf{x}_i) + w_0 \quad (2)$$

其中权参数向量 $\mathbf{w} = (w_0, w_1, \dots, w_N)$, $K(\mathbf{x}, \mathbf{x}_i)$ 为核函数。由于 t_i 为独立分布, 因此, 数据集 $\mathbf{t} = (t_1, t_2, \dots, t_N)^T$ 的似然估计为:

$$p(\mathbf{t} | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{t} - \mathbf{w}\Phi\|^2\right\} \quad (3)$$

式中: $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_N)$, $\Phi_i = (1, K(\mathbf{x}_1, \mathbf{x}_i), \dots, K(\mathbf{x}_N, \mathbf{x}_i))^T$

($i = 1, \dots, N$)。

如果不对权值进行约束, 直接最大化似然函数式 (3) 估计参数 (\mathbf{w}, σ^2) , 往往导致模型过学习。因此, 根据稀疏贝叶斯原理对每个权参数 w_i 定义零均值高斯先验分布, 则 \mathbf{w} 的先验分布为:

$$p(\mathbf{w} | \boldsymbol{\alpha}) = \prod_{i=1}^N N(w_i | 0, \alpha_i^{-1}) \quad (4)$$

式中: α_i 为 w_i 先验分布超参数, 超参数向量 $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_N)^T$, 每个独立超参数 α_i 控制权参数 w_i 先验分布, 导致 RVM 模型具有稀疏性。

由式 (3)、式 (4) 利用贝叶斯原理计算权参数向量 \mathbf{w} 的后验分布:

$$p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) \cdot p(\mathbf{w} | \boldsymbol{\alpha})}{p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2)} = (2\pi)^{\frac{1+N}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{(\mathbf{w} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{w} - \boldsymbol{\mu})}{2}\right\} = N(\mathbf{w} | \boldsymbol{\mu}, \Sigma) \quad (5)$$

式中: $\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}$, $\boldsymbol{\mu} = \sigma^{-2} \Sigma \Phi^T \mathbf{t}$, $A = \text{diag}(\alpha_0, \dots, \alpha_N)$ 。

由式 (5) 可知, 要确定权参数向量 \mathbf{w} , 要先确定超参数 $\boldsymbol{\alpha}$ 与 σ^2 。由贝叶斯框架可计算超参数似然分布:

$$p(\mathbf{t} | \boldsymbol{\alpha}, \sigma^2) = \int p(\mathbf{t} | \mathbf{w}, \sigma^2) \cdot p(\mathbf{w} | \boldsymbol{\alpha}) \cdot d\mathbf{w} = N(0, C) \quad (6)$$

式中: 协方差 $C = \sigma^2 I + \Phi A^{-1} \Phi^T$ 。

对式 (6) 关于 $\boldsymbol{\alpha}$ 、 σ^2 分别求导数且导数等于 0。根据 MacKay 迭代估计方法构建参数 $\boldsymbol{\alpha}$ 、 σ^2 的迭代算法:

$$\alpha_j^{\text{new}} = \frac{\gamma_j}{u_j}, j = 1, \dots, N \quad (7)$$

式中: u_j 为式 (5) 第 j 个后验平均权值, 其中 $\gamma_j = 1 - \alpha_j M_{jj}$, M_{jj} 为后验权值协方差矩阵 Σ 的第 j 个对角元素, $\boldsymbol{\alpha}$ 、 σ^2 取当前的值。

$$(\sigma^2)^{\text{new}} = \frac{\|\mathbf{t} - \Phi \boldsymbol{\mu}\|^2}{N - \sum_{j=1}^N \gamma_j} \quad (8)$$

式中: 分母中 N 指学习样本数据的数量。

RVM 学习过程即通过迭代公式 (7)、式 (8) 不断计算 α_j^{new} 与 $(\sigma^2)^{\text{new}}$, 由式 (5) 更新 Σ 与 $\boldsymbol{\mu}$, 当所有参数都收敛或达到最大训练次数时停止迭代计算, 此时得到 $\boldsymbol{\alpha}$ 、 σ^2 最优估值 $\boldsymbol{\alpha}_{MP}$ 与 σ_{MP}^2 , $\mathbf{w} = \boldsymbol{\mu}_{MP}$ 。在收敛过程中, 许多 α_j 无限大, 对应 w_j 等于 0, 此时删除基函数 $\phi_j(\mathbf{x})$, 非零 w_j 对应学习样本称为相关向量 (RV), 该学习方法显著减少模型基函数数量, 使模型达到稀疏化效果。

假如给定新的输入值 \mathbf{x}_* , 则相应的预测输出的概率分布为:

$$p(\mathbf{t}_* | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = \int p(\mathbf{t} | \mathbf{w}, \sigma_{MP}^2) \cdot p(\mathbf{w} | \mathbf{t}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) \cdot d\mathbf{w} = N(\mathbf{t}_* | \mathbf{y}_*, \sigma_*^2) \quad (9)$$

$$\mathbf{y}_* = \mathbf{u}^T \Phi(\mathbf{x}_*) = \boldsymbol{\mu}_{MP}^T \Phi(\mathbf{x}_*) \quad (10)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \Phi(\mathbf{x}_*)^T \Sigma \Phi(\mathbf{x}_*) \quad (11)$$

式中： y_* 为输入值 x_* 对应输出的预测均值，将其作为预测值。 σ_*^2 表示输出的方差（不确定性），包括数据的估计噪声与模型权重估计过程中的不确定性。

1.2 自适应差分进化算法构建

Price等提出的差分进化算法（differential evolution, DE）是一种基于实数编码的群体智能理论的优化算法，算法原理类似于遗传算法^[16]。通过群体内个体间的合作与竞争产生的群体智能指导优化搜索，所以该算法有更好的全局搜索能力、鲁棒性与收敛速度，较多文献证实DE算法收敛速度和稳定性方面优于遗传算法、粒子群算法和模拟退火算法^[17]。算法通过对种群中的个体进行变异、交叉、选择操作，获得最优个体，确定最优值。鉴于算法的缩比因子 F 和交叉概率 C_r 参数对算法影响较大，本文构建 F 与 C_r 参数自适应的DE算法（self-adaptive differential evolution, SADE）^[18]，算法工作原理如下：

1) 变异操作：对 n 代 m 个体 $v_m(n)$ 进行变异操作，在种群数量范围内随机选取不同 r_1, r_2, r_3 号 $v_{r1}(n), v_{r2}(n), v_{r3}(n)$ 按照式(12)进行变异操作产生 $n+1$ 代个体 $v_m(n+1)$ 。

$$v_m(n+1) = v_{r1}(n) + F(v_{r2}(n) - v_{r3}(n)) \quad (12)$$

2) 交叉操作：对个体 $v_m(n)$ 与个体 $v_m(n+1)$ 的 d 的维数按照式(13)进行交叉，产生个体 $u_m(n+1)$ 。 C_r 交叉概率， k 为在1与最大维数之间随机产生的整数，保证至少有一维数进行交叉。

$$u_{md}(n+1) = \begin{cases} v_{md}(n+1), & \text{rand}(d) < C_r \mid d = k \\ v_m(n), & \text{其他} \end{cases} \quad (13)$$

3) 选择操作：以适应度大小为个体优劣判定依据，对个体 $v_m(n)$ 与 $u_m(n+1)$ 进行选择，优选个体加入下一代种群。

针对 F, C_r 参数对算法性能影响较大这一问题，依据当前个体适应与最大、最小适应度之间关系采取策略式(14)、式(15)对参数进行自适应调整，提高算法可靠性。

$$F_m(n+1) = F_{\min} + (F_{\max} - F_{\min}) \frac{f_m - f_{\min}(n)}{f_{\max}(n) - f_{\min}(n)} \quad (14)$$

$$C_{rd}(n+1) = C_{r\min} + (C_{r\max} - C_{r\min}) \frac{f_m - f_{\min}(n)}{f_{\max}(n) - f_{\min}(n)} \quad (15)$$

式中： F_{\max}, F_{\min} 为最大、最小缩比因子； $C_{r\max}, C_{r\min}$ 为最大、最小交叉概率； $f_{\max}(n), f_{\min}(n)$ 为 n 代最大、最小适应度； f_m 为 m 个体适应度。

1.3 自适应进化相关向量机模型构建

根据SADE与RVM算法特点，以留一交叉验证法建立SADE算法的适应度函数^[19]，自适应确定RVM模型最优参数，提出自适应进化相关向量机模型（SADE-RVM），减少由于人为参数确定不当对RVM算法性能的不利影响，具体模型构建流程如图1所示。

1.4 模型精度评价方法

为对自适应进化相关向量机模型的精度进行合理评价，选取当前使用频率较高的3种耕地预测模型进行对比分析。建立定量与定性的模型精度评价指标，对自适应进化向量机模型精度做出科学、合理的评价。

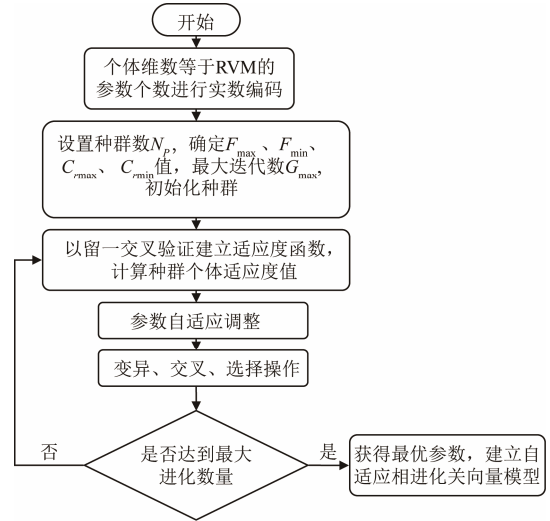


图1 自适应进化相关向量机模型算法建模流程图

Fig.1 SADE - RVM algorithm modeling flowchart

1) 对比模型的建立

选取多元线性回归、BP神经网络、最小二乘支持向量机作为对比模型，分别建立短期、中期耕地预测模型，以Matlab编程实现，具体建模过程如下：①多元线性回归模型（multivariate linear regression, MLR）。以影响耕地面积变化的6种因素为模型输入，耕地面积为输出，以regress函数求解7个模型参数，建立多元线性回归耕地预测模型。②BP神经网络。以影响耕地面积变化的6种因素为模型输入，耕地面积为输出，1个隐含层，5个神经元，传递函数包括{'logsig', 'tansig'}，以trainlm函数为训练函数，建立基于BP神经网络的耕地预测模型。③最小二乘支持向量机（least squares support vector machine, LS-SVM）。以影响耕地面积变化的6种因素为模型输入，耕地面积为输出。选择径向基核函数作为算法核函数，以网格搜索法通过多次试算，确定最小二乘支持向量机的2个参数，建立对应耕地预测模型。

2) 模型精度评价方法

本文选取平均绝对误差（mean absolute error, MAE）、平均相对误差（mean relative error, MRE）、均方根误差（root mean square error, RMSE）等3个精度指标评价各模型精度优劣，具体计算公式如式(16)、式(17)、式(18)。

$$MAE = \frac{1}{v} \sum_{j=1}^v |l_j - l'_j| \quad (16)$$

$$MRE = \frac{1}{v} \sum_{j=1}^v \left| \frac{l_j - l'_j}{l_j} \right| \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^v (l_j - l'_j)^2}{v}} \quad (18)$$

式中： l_j 为真实值； l'_j 为模型预测值； v 为预测值数量。

陈怀录等^[20]、张新长等^[21]依据后验差比 C 及小误差概率 P 对耕地变化预测模型精度进行等级划分，对各耕地变化预测模型进行定性评价，具体如表1所示。

表 1 耕地预测模型精度等级划分

Table 1 Accuracy grade classification of prediction model of cultivated land area

等级 Grade	小误差概率 Small error probability P	后验差比 Posterior error C
好 Good	>0.95	<0.35
合格 Qualified	>0.80	<0.45
勉强 Qualified	>0.70	<0.50
不合格 Unqualified	≤0.70	≥0.65

2 实例分析

本文以黄石市耕地面积变化为例，建立自适应进化耕地面积预测模型，并对模型精度及可靠性进行对比分析，证明本文提出的耕地面积预测方法的可行性及可靠性。

2.1 建模数据源

耕地面积变化受到社会、自然、经济等方面的多种因素影响，杨柳等^[22]通过微粒群和支持向量机对黄石市耕地变化影响因素进行研究，确定黄石市的人口总数 x_1 、城市化水平 x_2 、社会固定资产投资 x_3 、农村居民人均收入 x_4 、经济非农化率 x_5 为耕地面积变化的影响因素。但是这些影响因素中未考虑耕地保护政策这一重要内容，Li 等^[23]、翟文侠和黄贤金^[24]详细地分析耕地保护政策对耕地变化有巨大影响，证实耕地保护政策是耕地变化重要影响因素。因此，本文将耕地保护政策 x_6 作为黄石市耕地面积变化第 6 个影响因素。为了便于建模，利用德尔菲法对耕地保护政策进行量化。通过参考《黄石市统计年鉴》^[25]收集黄石市 1978—2010 年黄石市耕地面积(y)及其影响因素数据，具体数据见表 2。由于各变量单位不同，对建模精度有较大影响。因此，需对原始数据进行归一化处理后才可用于模型建立。

表 2 1978—2010 年黄石市耕地面积及其影响因素数据

Table 2 Data of cultivated land area and its influencing factors in Huangshi city from 1978 to 2010

年份 Year	人口总数 Total population $x_1/10^4$	城市化水平 Urbanization $x_2/\%$	社会固定资产投资 Social investment in fixed assets $x_3/10^8$ 元	农村居民人均收入 Per capita income of rural residents x_4 /元	经济非农化率 Non-agricultural economy rate $x_5/\%$	耕地保护政策 Land management policies x_6	耕地面积 Cultivated area $y/10^4\text{hm}^2$
1978	166.68	30.00	1.26	97.00	77.30	1	9.29
1979	169.42	30.00	1.46	126.00	75.87	1	9.29
1980	172.41	30.50	1.60	114.00	77.90	1	9.26
1981	176.21	31.00	1.31	139.00	78.64	1	9.23
1982	179.57	32.00	2.03	199.00	79.17	1	9.19
1983	181.63	32.00	2.40	249.00	80.40	1	9.17
1984	184.16	33.00	4.04	318.00	78.27	1	9.12
1985	188.73	34.00	4.67	407.00	77.34	1	9.00
1986	192.13	35.00	5.69	436.00	75.97	3	8.72
1987	195.69	36.00	6.94	427.00	77.07	3	8.54
1988	199.79	37.00	8.07	454.00	79.24	3	8.51
1989	202.81	38.00	6.35	527.00	81.78	3	8.47
1990	211.52	39.00	5.67	590.00	79.57	3	8.43
1991	216.37	40.00	7.18	607.00	83.35	3	8.41
1992	219.65	41.00	13.80	660.00	84.54	3	8.38
1993	226.65	42.00	16.47	737.00	84.67	3	8.29
1994	231.39	43.00	17.89	1001.00	82.46	3	8.18
1995	235.46	44.00	20.20	1270.00	83.00	5	8.16
1996	237.86	45.00	23.84	1594.00	84.69	5	8.12
1997	240.12	46.52	28.62	1771.00	83.89	8	8.08
1998	243.58	47.66	34.61	1853.00	88.25	8	8.06
1999	245.86	49.31	35.83	1962.00	89.28	8	8.10
2000	247.82	49.29	40.86	2065.00	90.08	8	8.02
2001	248.87	49.29	47.35	2140.00	90.73	8	7.85
2002	249.83	49.50	54.45	2230.00	91.18	8	7.56
2003	250.82	49.60	69.23	2335.00	91.39	8	7.32
2004	251.75	49.60	88.30	2626.00	90.52	8	7.43
2005	252.80	49.61	106.46	2810.00	90.69	8	7.67
2006	254.36	49.75	138.05	3182.00	91.73	8	8.19
2007	255.39	50.41	179.85	3742.00	91.61	10	8.74
2008	257.31	50.35	232.67	4374.00	92.19	10	8.71
2009	258.56	50.55	343.05	4811.00	92.08	10	9.07
2010	268.01	53.06	474.06	5524.00	92.23	10	8.98

2.2 耕地面积变化预测模型构建

以耕地面积变化影响因素数据 $X = (x_1, x_2, x_3, x_4, x_5, x_6)$ 为自适应进化相关向量机输入，

耕地面积 y 为模型输出。鉴于文献^[26]对 RVM 模型核函数的比较分析，柯西函数具有较好表现，因此本文选择柯西核函数建立 RVM 模型。通过训练数据集计算模型参

数 α 、 σ^2 和 w ，建立基于自适应进化相关向量机的耕地面积预测模型。将待预测输入 X_* 输入预测模型，计算对应耕地预测值 y_* 及其方差 σ_*^2 。为了检验本文方法对短期（5a）、中期（10a）的耕地面积的预测效果，本文分别建立短期耕地预测模型及中期耕地预测模型，具体建模过程如下：

2.2.1 短期耕地预测模型

以 1978—2005 年的黄石市耕地变化影响因素及耕地面积数据组成训练数据集 $\{X_r, y_r\}_{r=1978}^{2005}$, $X_r \in R^6, y_r \in R$ ，以 2005—2010 年的 $\{X_r, y_r\}_{r=2006}^{2010}$ 为测试数据，用来评价模型对短期耕地面积预测精度。根据 Storn 和 Price 提出的差分进化算法确定参数参考范围，设定模型初始参数 $N_P=80, F_{\max}=0.8, F_{\min}=0.3, C_{r\max}=0.9, C_{r\min}=0.1, G_{\max}=200$ 。自适应进化耕地预测模型通过对训练数据集学习，确定模型参数。模型训练后确定最优柯西核函数核参数为 $\psi=1.302, \sigma_{MP}^2=0.0009, \{X_r, y_r\}_{r=1979, 1982, 1987, 1994, 1998, 2001, 2002}$ 对应的 7 个训练数据为相关向量，相关向量数量与训练数据个数的比例为 1/4，对应的有效权向量 $w=[w_{1979}, w_{1982}, w_{1987}, w_{1994}, w_{1998}, w_{2001}, w_{2002}]$ ，模型常数项 $w_0=bias=0.7720$ 。由以上参数，建立短期耕地预测模型 $y=w^T\Phi(X)+w_0$ 。

2.2.2 中期耕地预测模型

以 1978—2000 年的黄石市耕地变化影响因素及耕地

面积数据组成训练数据集 $\{X_r, y_r\}_{r=1978}^{2000}$, $X_r \in R^6, y_r \in R$ ，以 2010 年的 $\{X_r, y_r\}_{r=2010}$ 为测试数据，评价模型对中期耕地面积的预测精度。模型通过对训练数据进行学习，计算模型参数 $\psi=11.327, \sigma_{MP}^2=0.0077, \{X_r, y_r\}_{r=1981, 1982, 1984, 1988, 1997}$ 对应的 5 个训练数据为相关向量，相关向量数量与训练数据个数的比例为 5/23，对应的有效权向量 $w=[w_{1981}, w_{1982}, w_{1984}, w_{1988}, w_{1997}]$ ，模型常数项 $w_0=bias=0$ 。基于以上参数，建立中期耕地预测模型 $y=w^T\Phi(X)+w_0$ 。

2.3 模型预测结果评价

2.3.1 短期耕地预测结果评价

将测试数据集 $\{X_r, y_r\}_{r=2006}^{2010}$ 中的 $X_{r=2006}^{2010}$ 输入短期耕地预测模型，预测 2006—2010 年耕地面积 $y_{r=2006}^{2010}$ ，对应方差 $\sigma_{r=2006}^{2010}$ 。建立 95%置信度的置信区间计算公式(19)^[27]，其中 $z_{\theta/2}=1.96$ ，并估计 $y_{r=2006}^{2010}$ 对应置信区间。为评价自适应进化相关向量机对短期耕地预测的精度，基于多元线性回归、BP 算法、LS-SVM 建立短期耕地预测模型。各模型具体预测结果见表 3。

$$\{\theta = y_* - \sigma_* z_{\theta/2}, y_* + \sigma_* z_{\theta/2} = \bar{\theta}\} \quad (19)$$

基于以上数据，计算各耕地预测模型的 MAE、MRE、RMSE、后验差比 C、小误差概率 P、计算效率、模型等级等评价指标，具体结果见表 4。

表 3 短期耕地面积预测值和实测值

Table 3 Short-term predicted and measured values for cultivated land area

年份 Year	实测值 Measured values/10 ⁴ hm ²	MLR 预测值 Predicted values/10 ⁴ hm ²	BP 预测值 Predicted values/10 ⁴ hm ²	LS-SVM 预测 值 Predicted values/10 ⁴ hm ²	SADE-RVM/10 ⁴ hm ²	
					预测值 Predicted values	置信区间 Confidence interval
2006 年	8.19	7.2095	8.0047	8.2634	8.2212	{8.0864, 8.3560}
2007 年	8.74	6.8415	8.2537	8.7494	8.6414	{8.4506, 8.8322}
2008 年	8.71	6.5225	8.3696	8.6755	8.8369	{8.6169, 9.0569}
2009 年	9.07	5.2009	8.3720	8.5804	8.9677	{8.7257, 9.2098}
2010 年	8.98	3.5213	8.3720	8.5797	9.0186	{8.7663, 9.2708}

表 4 短期和中期耕地面积预测模型评价指标

Table 4 Evaluation index of predicted model for short-term and mid-term cultivated land area

预测模型 Predicted models	短期 Short-term							中期 Mid-term						
	MAE	MRE	RMSE	计算效率 Computational efficiency/s	后验 差比 Posterior error C	小误差 概率 Small error probability P	等级	MAE	MRE	RMSE	计算效率 Computational efficiency/s	后验差比 Posterior error C	小误差 概率 Small error probability P	等级
MLR	2.8789	32.4500	3.2900	4.2209×10 ⁻⁵	0.24	2.15	不合格	2.0286	22.59	2.0286	4.5589×10 ⁻⁵	0.96	0.89	不合格
BP	0.4636	5.2400	0.4987	5.1×10 ⁻³	0.91	0.32	合格	0.9536	10.62	0.9536	5.6×10 ⁻³	0.82	0.42	合格
LS-SVM	0.2014	2.2500	0.2852	6.4×10 ⁻³	0.97	0.23	好	0.4781	5.32	0.4781	7.9×10 ⁻³	0.96	0.21	好
SADE-RVM	0.0795	0.9000	0.0880	2.2724×10 ⁻⁵	1.00	0.07	好	0.2447	2.72	0.2447	9.0052×10 ⁻⁵	1.00	0.12	好

表 3 表明 2006—2010 年耕地面积实际面积均在自适应进化相关向量机估计的 95%置信度置信区间内，证实预测结果可靠。由表 4 可知，对于短期耕地面积预测，自适应进化相关向量机各项精度指标表现最优，模型等

级属于“好”，证实本文提出的耕地面积预测新方法具有很高的短期预测精度。建立自适应进化相关向量机的短期耕地预测模型时，28 组训练数据中只有 7 组为相关向量最终参与预测模型构建，有效简化模型复杂度，具

有很高的稀疏性。因此,该模型在计算效率方面优于其他模型,具有很高的运行速度,这点在表 4 得到证明。基于以上分析数据,自适应进化相关向量机的短期耕地方法计算速度快、精度高、可靠性强。

2.3.2 中期耕地预测结果评价

依据训练数据集 $\{y_r, \mathbf{X}_r\}_{r=1978}^{2000}$ 建立 4 种耕地面积预测模型,并对测试数据集 $\{y_{2010}, \mathbf{X}_{2010}\}$ 进行预测,获得 2010 年黄石市耕地面积预测值。自适应进化相关向量机模型预测结果为 $8.7353 \times 10^4 \text{ hm}^2$, 对应置信区间为 $\{8.6619 \times 10^4, 9.0546 \times 10^4\} \text{ hm}^2$ 。多元线性回归、BP 算法、LS-SVM 模型预测结果分别为 1.10086×10^5 、 8.0264×10^4 、 $8.5019 \times 10^4 \text{ hm}^2$ 。通过各模型预测结果,计算各耕地预测模型的 MAE、MRE、RMSE、后验差比 C 、小误差概率 P 、计算效率、模型等级等评价指标,具体数据如表 4 所示。由表 4 可知,基于自适应进化向量机的中期耕地预测模型的 MAE、MRE、RMSE、后验差比 C 、小误差概率 P 、计算效率、模型等级等评价指标均优于其余 3 种模型,并且 2010 年实际耕地面积介于其置信区间内,证实本文提出的中期耕地预测方法在精度、稀疏性、运算效率及可靠性方面表现突出,优于其余 3 种方法。

3 结 论

将相关向量机、差分进化及留一交叉验证算法优点有机结合提出基于自适应进化相关向量机的短期、中长期耕地面积变化预测新方法,通过预测模型评价分析得到以下结论:

1) 本文提出的自适应进化相关向量机模型有效地解决了核函数参数确定问题,自适应优化核函数参数,可降低由于参数不当对模型精度的影响。

2) 由 MAE、MRE、RMSE 数据可知,对于短期耕地面积预测,自适应进化相关向量机的精度大约是 LS-SVM 方法的 3 倍,是 BP 网络的 6 倍,是多元线性回归的 36 倍;对于中期耕地面积预测,新方法精度大约是 LS-SVM 方法的 2 倍,是 BP 网络的 3.5 倍,是多元线性回归的 8 倍。因此,证实自适应进化相关向量机对短期、中期耕地面积预测有很高精度。

3) 自适应进化相关向量机的短期、中期耕地预测模型中,相关向量数量占训练数据个数比例分别为 1/4、5/23,因此模型具有很好的稀疏性,导致模型的计算效率很高。自适应进化相关向量机计算效率总体来说是多元线性回归的 2 倍,比 BP 网络及 LS-SVM 模型高出 2 个数量级,当对大数据进行建模时,计算效率优势会更加明显。

4) 耕地面积实际值均在自适应进化相关向量机估计的置信区间内,定量地证实模型具有很强的可靠性。通过后验差比、小误差概率分析,自适应进化相关向量机对短期、中期耕地预测结果等级均属于“好”,定性地证实模型具有很好的预测效果。

[参 考 文 献]

- [1] Wang Jing, Chen Yongqi, Shao Xiaomei, et al. Land-use changes and policy dimension driving forces in China: Present, trend and future[J]. Land Use Policy, 2012, 29(4): 737—749.
- [2] 肖思思, 吴春笃, 储金字. 近 15 年中国耕地驱动因素的空间差异分析[J]. 中国土地科学, 2013, 27(7): 55—62.
Xiao Sisi, Wu Chundu, Chu Jinyu. Spatial variation of driving factors on cultivated land change in China for the past 15 years[J]. China Land Sciences, 2013, 27(7): 55—62. (in Chinese with English abstract)
- [3] 岳云华, 冉清红, 孙传敏, 等. 我国耕地数量灰色预测分析与耕地保护[J]. 地理与地理信息科学, 2010, 26(7): 56—59.
Yue Yunhua, Ran Qinghong, Sun Chuanmin, et al. Grey predication of cultivated land change and analysis of farmland protection in China[J]. Geography and Geo-information Science, 2010, 26(7): 56—59. (in Chinese with English abstract)
- [4] 刘耀林, 刘艳芳, 张玉梅. 基于灰色-马尔柯夫链预测模型的耕地需求量预测研究[J]. 武汉大学学报: 信息科学版, 2004, 29(7): 575—580.
Liu Yaolin, Liu Yanfang, Zhang Yumei. Prediction of gross arable land based on Grey-Markov model[J]. Geomatics and Information Science of Wuhan University, 2004, 29(7): 575—580. (in Chinese with English abstract)
- [5] 曹银贵, 王静, 刘爱霞. 基于 BP 神经网络的三峡库区开县耕地面积预测研究[J]. 中国农业资源与区划, 2007, 28(4): 30—34.
Cao Yingui, Wang Jing, Liu Aixia. Studies on forecasting the arable land areas of Kai county in reservoir of three gorges based on BP nerve network[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2007, 28(4): 30—34. (in Chinese with English abstract)
- [6] 邓羽, 刘盛和, 王海军. 基于新陈代谢 BP 神经网络在耕地保有量预测中的应用[J]. 华中师范大学学报: 自然科学版, 2008, 42(3): 494—500.
Deng Yu, Liu Shenghe, Wang Haijun. Study on the prediction of land based on the renewal model of the BP nerve network[J]. Journal of Huazhong Normal University: Nat Sci, 2008, 42(3): 494—500. (in Chinese with English abstract)
- [7] Arsanjani J J, Kainza H, Kainza W. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion[J]. International Journal of Applied Earth Observation and Geoinformation, 2013, 21(3): 265—275.
- [8] Yang Q S, Li X. Cellular automata for simulating land use changes based on support vector machine[J]. Computers & Geosciences, 2008, 34(2008): 592—602.

- [9] 张豪, 罗亦泳. 基于遗传算法最小二乘支持向量机的耕地变化预测[J]. 农业工程学报, 2009, 25(7): 226—231.
Zhang Hao, Luo Yiyong. Cultivated land change forecast based on genetic algorithm and least squares support vector machines[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2009, 25(7): 226—231. (in Chinese with English abstract)
- [10] 张豪, 罗亦泳. 基于遗传支持向量机的城市扩张非线性组合模型[J]. 地理学报, 2010, 62(6): 656—664.
Zhang Hao, Luo Yiyong. A nonlinear polynomial model for urban expansion incorporating genetic algorithm and support vector machines[J]. Journal of Geographical Sciences, 2010, 62(6): 656—664. (in Chinese with English abstract)
- [11] 王霞, 王占岐, 金贵. 基于核函数支持向量回归机的耕地面积预测[J]. 农业工程学报, 2014, 30(4): 204—211.
Wang Xia, Wang Zhanqi, Jin Gui. Land reserve prediction using different kernel based support vector regression[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2014, 30(4): 204—211. (in Chinese with English abstract)
- [12] Tipping M E. Sparse bayesian learning and the relevance vector machine[J]. Journal of Machine Learning Research 2001, 1(3): 211—244
- [13] Bishop C M, Tipping M E. Variational relevance vector machine[C]. The 16th Conf on Uncertainty in Artificial Intelligence, Morgan Kaufmann, USA, 2000.
- [14] Storn R, Price K. Differential evolution: A simple and efficient heuristic for global optimization over continuous spaces[J]. Journal of Global Optimization, 1997, 11(4): 341—359.
- [15] 孟庆芳, 陈月辉, 冯志全, 等. 基于局域相关向量机回归模型的小尺度网络流量的非线性预测[J]. 物理学报, 2013, 62(15): 1—7.
Meng Qingfang, Chen Yuehui, Feng Zhiquan, et al. Based on local relevance vector machine regression model[J]. Acta Phys Sin, 2013, 62(15): 1—7. (in Chinese with English abstract)
- [16] Price K, Storn R, Lampinen J. Differential evolution: A practical approach to global optimization[M]. Berlin Heidelberg: Springer, 2005.
- [17] 孙昌跃, 刘德顺, 段凯. 基于差分进化算法的钻头波阻辨识研究[J]. 煤炭学报, 2012, 37(2): 350—355.
Sun Changyue, Liu Deshun, Duan Kai. Identifying the characteristic impedance of the drill bit based on differential evolution algorithm[J]. Journal of China Coal Society, 2012, 37(2): 350—355. (in Chinese with English abstract)
- [18] Janez B, Viljem Z, Mirjam S M. Self-adaptive differential evolution algorithm in constrained real parameter optimization[C]. IEEE Congress on Evolutionary Computation, Vancouver Canada, 2006.
- [19] 刘学艺, 李平, 郜传厚. 极限学习机的快速留一交叉验证算法[J]. 上海交通大学学报, 2011, 45(8): 1140—1146.
Liu Xueyi, Li Ping, Gao Chuanhou. Fast leave-one-out cross-validation algorithm for extreme learning machine[J]. Journal of Shanghai Jiao Tong University, 2011, 45(8): 1140—1146. (in Chinese with English abstract)
- [20] 陈怀录, 冯东海. 土地利用规划耕地预测方法对比研究——以甘肃省临夏市为例[J]. 西北师范大学学报: 自然科学版, 2011, 47(1): 99—104.
Chen Huailu, Feng Donghai. The comparative study of the predictive method of the plow in land use planning: Taking Linxia City of Gansu Province for example[J]. Journal of Northwest Normal University: Natural Science, 2011, 47(1): 99—104. (in Chinese with English abstract)
- [21] 张新长, 杨大勇, 潘琼, 等. 城市边缘区耕地变化的发展预测分析研究[J]. 中山大学学报: 自然科学版, 2002, 41(3): 104—109.
Zhang Xinchang, Yang Dayong, Pan Qiong, et al. Study on developing and predicting analysis of cultivated land in urban edge districts[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2002, 41(3): 104—109. (in Chinese with English abstract)
- [22] 杨柳, 刘艳芳. 将微粒群和支持向量机用于耕地驱动因子选择的研究[J]. 武汉大学学报: 信息科学版, 2010, 35(2): 248—252.
Yang Liu, Liu Yanfang. Feature subset selection for driving forces of cultivated land based on PSO-SVM[J]. Geomatics and Information Science of Wuhan University, 2010, 35(2): 248—252. (in Chinese with English abstract)
- [23] Li Wei, Feng Tingting, Hao Jinmin. The evolving concepts of land administration in China: Cultivated land protection perspective[J]. Land Use Policy, 2009, 26(3): 262—272.
- [24] 翟文侠, 黄贤金. 我国耕地保护政策运行效果分析[J]. 中国土地科学, 2003, 17(2): 8—14.
Zhai Wenxia, Huang Xianjin. Analysis on the effect of policies operation of cultivated land protection in China[J]. China Land Science, 2003, 17(2): 8—14. (in Chinese with English abstract)
- [25] 黄石统计局和国家统计局黄石调查队. 黄石统计年鉴 2011[M]. 北京: 中国统计出版社, 2011.
- [26] 楼俊钢, 蒋云良, 申倩, 等. 软件可靠性预测中不同核函数的预测能力评估[J]. 计算机学报, 2013, 36(6): 1303—1311.
Lou Jungang, Jiang Yunliang, Shen Qing, et al. Evaluating the prediction performance kernel based software of different kernel functions in reliability models[J]. Chinese Journal of Computers, 2013, 36(6): 1303—1311. (in Chinese with English abstract)
- [27] Wang F F, Zhang Y R. Relevance vector machine technique for the inverse scattering problem[J]. Chinese Physics B, 2012, 21(5): 19—24.

Prediction model for cultivated land area based on self-adaptive differential evolution and relevance vector machine

Luo Yiyong^{1,2}, Zhang Hao^{3*}, Zhang Liting¹

(1. Faculty of Geomatics, East China University of Technology, Nanchang 330013, China; 2. School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; 3. College of Civil Engineering and Architecture, Zhejiang University of Technology, Hangzhou 310032, China)

Abstract: In order to solve some key problems existing in the previous prediction model for cultivated land area, for instance, the nonlinearity, sparseness and the result reliability, a new prediction model for cultivated land area is proposed through fusing and improving the self-adaptive evolution and relevance vector machine in the paper. By analyzing the characteristics of convergence rate for differential evolution algorithm, the functional relationship among shrinkage ratio factor, crossover probability, maximum fitness and minimum fitness is established. Meanwhile, individual shrinkage ratio factor and crossover probability of the next generation are determined based on the current individual fitness data. And the self-adaptive differential evolution algorithm is also developed in this way which can effectively improve the global convergence ability and the robustness of the algorithm. The current studies have confirmed that the kernel parameters have a greater impact on the prediction accuracy of relevance vector machine. Therefore, in order to improve the accuracy of the model, fitness function is established based on leave one cross validation and the relevance vector machine on the basis of self-adaptive differential evolution is also proposed by optimizing the kernel parameters. As the new model has the advantages of sparsity and nonlinearity, and can output the information of the uncertainty of the results, the new method is used to predict the cultivated land area. In order to prove the excellent properties of the new method, the accuracy of the model is evaluated by choosing 5 kinds of precision indices including mean absolute error (MAE), mean relative error (MRE), root mean square error (RMSE), posterior error and error frequency. The computational efficiency and the reliability of the model are estimated quantitatively by running time and confidence interval. Taking Huangshi City as an example, a short-term and a middle-term prediction model for the cultivated land area are set up on the basis of the self-adaptive evolution and relevance vector machine. And these two established prediction models are also compared with the multivariate linear regression model, back propagation (BP) neural network and least squares support vector machine in terms of accuracy, computational efficiency and reliability. The experimental statistics reveal that the newly established prediction model based on the self-adaptive evolution and relevance vector machine is about 2 times higher than the rest 3 models in accuracy, 2 times as much as multivariate linear regression model and 2 orders of magnitude higher than the BP neural network and least squares support vector machine in computational efficiency; the actual land area of the test data set is all in confidence intervals at the 95% confidence level, which is obtained by prediction model for cultivated land area based on self-adaptive differential evolution and relevance vector machine. All the above data confirms that the model based on the self-adaptive evolution and relevance vector machine is a new approach to the prediction of the cultivated farm land with high accuracy, fast calculation and strong reliability.

Key words: land use; algorithms; support vector machines; cultivated land area