

基于 FSA-LSSVR 模型的安徽省耕地变化预测

赵艳玲, 何厅厅, 刘亚萍, 石娟娟, 冉艳艳, 倪巍, 吴国伟

(中国矿业大学(北京)土地复垦与生态重建研究所, 北京 100083)

摘 要:针对耕地变化内部规律及其外部驱动因子的特点,利用鱼群算法优化最小二乘支持向量机回归(FSA-LSSVR)模型,探讨耕地变化预测模型,为耕地保护政策制定提供参考依据。结果表明:鱼群算法的全局搜索能力能使支持向量机算法有效地收敛到参数 γ 和 σ 的全局最优解;FSA-LSSVR 模型的预测精度指标远高于多元线性、GM(1,1) 和 BP 神经网络模型,且优于 FSA-SVM,求解速度明显优于 SVM。FSA-LSSVR 模型可以解决 SVM 内部参数难以确定的问题,适用于多因素参与的高维非线性的耕地变化预测,而且速度快、精度高,具有推广价值。

关键词:土地变化;驱动因子;鱼群算法;最小二乘支持向量机

中图分类号:F301.24

文献标识码:A

文章编号:1005-3409(2014)03-0136-05

Prediction of Cultivated Land Change of Anhui Province Based on FSA-LSSVR Model

ZHAO Yan-ling, HE Ting-ting, LIU Ya-ping, SHI Juan-juan, RAN Yan-yan, NI Wei, WU Guo-wei

(Institute of Land Reclamation and Ecological Reconstruction,

China University of Mining & Technology (Beijing), Beijing 100083, China)

Abstract: With respect to accordance with the internal law of cultivated land change and the characteristics of external driving factors, the change predictive model of cultivated land was developed by using the method of FSA(Fish Swarm Algorithm) optimize Least Square Support Vector Regression (FSA-LSSVR) model. It could provide reference for cultivated land protection policy. The results indicated that the global search capability of FSA could make the support vector machine algorithm effectively converge to global optimal solution of the parameters γ and σ . Furthermore, the FSA-LSSVR model prediction accuracy index was much higher than the multiple, GM (1,1) and BP neural network model, and it was superior to FSA-SVM, the speed of processing data is obviously better than the SVM. Therefore, it concluded that the FSA-LSSVR model could solve the problem of SVM internal parameters which were difficult to be determined. It was applicable to many factors involved in high-dimensional nonlinear prediction of cultivated land change. Moreover, it was high speed, high precision and worthy of promotion.

Key words: land change; driving factors; Fish Swarm Algorithm; least square support sector regression

耕地变化预测是土地利用规划中的重要内容之一^[1],可为区域耕地保护政策制定提供重要依据。目前对于耕地变化预测方法的研究较多,有单一方法如回归预测、灰色预测、BP 神经网络预测^[2-5]等,其中前两种方法是时间序列数据的纯数学处理,缺乏对社会和经济方面变化的综合考虑;BP 神经网络易陷入局部最优。也有组合方法如灰色—马儿科夫模型^[6]、遗传算法最小二乘支持向量机模型^[7]等。前者仍是对时间序列数据的数学处理,后者中的支持向量机

(Support Vector Machine, SVM)是基于统计学习理论的 VC 维理论和结构风险最小化原则的新型学习机器^[8],该方法具有小样本学习、学习效率高与推广性好的特点,能有效避免“过学习”和“维数灾难”等问题,在模式识别^[9-10]和回归估计^[11-12]等研究领域都取得了很好效果。由于 SVM 模型最终转化为求解一个二次凸规划问题,当训练数据量大和维数高时, SVM 解算速度较慢。Suykens 等^[13]在标准 SVM 的目标函数上增加了误差平方和项,提出最小二乘支持

收稿日期:2014-02-19

修回日期:2014-03-19

资助项目:教育部新世纪优秀人才支持计划(NCET-12-0964);中央高校基本科研业务费专项资金(2009QD05)

作者简介:赵艳玲(1976—),女,河北献县人,副教授,博士,主要从事土地整治工程、土地复垦与生态重建方面的研究。E-mail:zhaoyl7677@163.com

向量机回归算法(Least Square Support Vector Machine Regression, LSSVR)。它将线性不等式约束转化为线性等式约束,使得 SVM 的二次凸规划问题等价于线性方程组求解,极大地提高了 SVM 的求解效率,也降低了 SVM 的学习难度^[14-16]。将遗传算法与最小二乘支持向量机算法结合,虽然可提高支持向量机模型的预测精度,但是遗传算法同样存在优化过程中容易陷入局部最优的缺陷。

鱼群算法(Fish Swarm Algorithm, FSA)是一种基于动物自治体的优化方法,是集群智能思想的一个具体应用。它能很好地解决非线性函数优化等问题。优化过程中能够很好地跳出局部最优值,并尽可能地搜索到其他的极值,最终搜索到全局极值。将鱼群算法与最小二乘支持向量机回归算法相结合,可避免陷入局部最优,保证了拟合精度,提高了预测能力。

本文以 2002—2011 年《安徽省统计年鉴》中的数据为依据,构建耕地变化的 FSA-LSSVR 预测模型,经代入驱动因子数据验证,该模型速度快、精度高,可应用于耕地变化预测。

1 鱼群算法优化最小二乘法支持向量机回归模型

1.1 最小二乘支持向量机回归模型

最小二乘法支持向量机的主要思想是:通过映射函数把输入数据非线性映射到高维特征空间,然后在高维特征空间中解决回归问题^{P^[17]}。采用最小二乘支持向量机进行非线性函数估计的算法如下^[18-19]:

已知训练样本 $D = \{(x_1, y_1) \mid i = 1, 2, \dots, N\}$, 其中 $x_i \in R^n$ 为输入数据, $y_i \in R$ 是输出类别。在权向量 w 空间(原始空间)中的最小二乘支持向量机的回归函数问题可以描述为如下问题求解:

$$\min_{w, b, e} J(w, b, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (1)$$

$$\text{约束条件: } y_i = w^T \varphi(x_i) + b + e_i, i = 1, 2, \dots, N$$

其中, $\varphi(x_i): R^n \rightarrow R_h^n$ 是核空间映射函数, 权向量 $w \in R_h^n$ (原始空间), 误差变量 $e_i \in R$, b 是偏差量。损失函数 J 是 SSE 误差和规则化量之和, γ 是惩罚因子(常数)。核空间映射函数的目的是从原始空间中抽取特征, 将原始空间中的样本映射为高维特征空间中的一个向量, 以解决原始空间中线性不可分的问题。

由于 w 可能为无限维的, 于是直接计算规划(1)是极其困难的, 因此将这一规划问题转化到其对偶空间中, 构建拉格朗日函数:

$$L(w, b, e; a) = J(w, e) - \sum_{i=1}^N a_i \{w^T \varphi(x_i) + b - y_i + e_i\} \quad (2)$$

式中, $a_i \in R$ 是拉格朗日算子, 在 LSSVR 形式中可以为正, 也可以为负。分别对 w, b, e, a 进行偏微分, 可以得到(2)的最优条件, 如下:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 & w = \sum_{i=1}^N a_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 & \sum_{i=1}^N a_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 & a = r e_i \\ \frac{\partial L}{\partial a_i} = 0 & w^T \varphi(x_i) + b - y_i + e_i = 0 \end{cases} \quad (3)$$

式中, $i = 1, 2, 3, \dots, N$ 。消除 w 和 e , 得矩阵方程:

$$\begin{bmatrix} 0 & I_v^T \\ I_v & \Omega + \frac{1}{\gamma} E \\ & \gamma \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (4)$$

式中, $y = [y_1, y_2, \dots, y_N]^T$, $I = [1, \dots, 1]^T$, $a = [a_1, a_2, \dots, a_N]^T$, E 是 $N \times N$ 维的单位矩阵, $\Omega = \varphi(x_i)^T \varphi(x_l)$, $i, l = 1, 2, \dots, N$ 。根据 mercer 条件^[20], 存在映射函数 φ 和核函数 $K(\cdot, \cdot)$ 使得:

$$K(x_i, x_l) = \varphi(x_i)^T \varphi(x_l) \quad (5)$$

由式(5)进而得到最小二乘法支持向量机回归函数:

$$y(x) = \sum_{i=1}^N a_i K(x, x_i) + b \quad (6)$$

其中 a, b 由式(4)解得。核函数有多种不同形式, 如多项式核函数、高斯核函数、样条核函数、RBF 核函数等^[21]。在本文耕地变化预测模型中, 选择

RBF 核函数。即 $K(x_i, x_l) = \exp(-\frac{\|x_i - x_l\|^2}{2\sigma^2})$,

σ 为核函数参数(常数)。最小二乘支持向量机回归函数的预测精度和收敛速度受参数 (γ, σ) 影响, 本文选用鱼群算法(FSA)对其进行寻优。

1.2 鱼群算法优化最小二乘支持向量机回归算法参数

鱼群算法(Fish Swarm Algorithm, 简称 FSA)是一种模拟鱼群运动规律的优化算法, 是由李晓磊等^[22]在 2002 年提出的一种新型群体智能优化算法, 其基本思想是鱼群向食物浓度较大的水域游动, 鱼群规模最大的地方食物浓度最大。据此, 鱼群算法构造人工鱼个体, 模拟鱼群的觅食、群聚和追尾行为, 通过各个体的局部寻优, 实现全局最优在鱼群中凸现出来的目的。该算法具有良好的克服局部极值的能力。并且算法中只使用目标函数的函数值, 无需目标函数的梯度值等特殊信息, 对问题不需要严格的数学机

理。对搜索空间具有一定的自适应能力,而且算法对初值没有要求,对各参数的选择也不很敏感^[23]。

本文充分利用鱼群算法的全局收敛、初值不敏感、收敛速度快、实时性高等特点^[24],对 LSSVR 模型的惩罚因子 γ 和 RBF 核函数参数 σ 在设定的范围内进行最优或近似最优参数的选择,以使模型的预测精度获得提高,具体的 FSA 和 LSSVR 结合方式如图 1 所示。

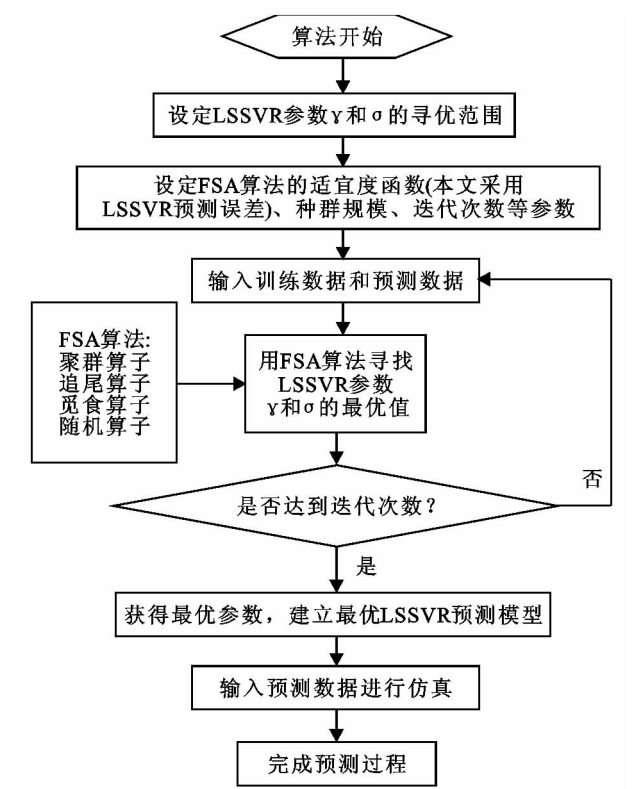


图 1 FSA 优化 LSSVR 算法流程

2 基于 FSA-LSSVR 的耕地变化预测模型的建立

根据已有的研究成果^[25-26],将影响耕地变化的社会经济环境驱动因子划分为自然类(环境变化、自然灾害、气候、地形、地貌等)、人口类(总人口数、农业人口数、第三产业就业人数等)、经济发展类(地区生产总值、第一产业生产总值、人均生产总值、固定资产投资

资等)、生活水平类(农民人均纯收入、城镇居民消费水平等)及土地政策类(土地管理^[27]政策、退耕政策、耕地保护政策等)。从众多的耕地变化驱动因子中选择主要驱动因子是耕地变化预测精度的根本。针对安徽省耕地变化的实际情况,根据定性和定量相结合的办法,依据 2002—2011 年《安徽省统计年鉴》,应用主成分分析法、相关分析法、迭代回归分析法^[28]最终确定影响耕地变化驱动因子指标体系,即第一产业生产总值、农民人均纯收入、农民人口数、第三产业人口数、固定资产投资、城市化水平、耕地保护政策,因变量为耕地面积,具体见表 1。其中第一产业总值、农民人均纯收入、农民人口数目、第三产业就业人数、固定资产投资、耕地面积数据直接来源于安徽省统计年鉴,城市化水平通过非农业人数除以人口总数算得,耕地保护政策以数字“1—10”衡量,最严格的耕地保护政策定义为“10”。1998 年实施《土地管理法》首次以立法形式确认了“十分珍惜、合理利用土地和切实保护耕地是我国的基本国策”,保护政策为“1”;2004 年中央 1 号文件《中央关于促进农民增收若干政策的意见》明确提出“各级政府要切实落实最严格的耕地保护制度”,保护政策为“3”;2005 年的《政府工作报告》要求严格保护耕地特别是基本农田,保护政策为“5”;2006 年中央 1 号文件《关于推进社会主义新农村建设的若干意见》在耕地占用税、土地出让金、新增建设用地有偿使用费征缴和使用方面做出了有利于耕地保护的规定,保护政策为“8”;2008 年中央《关于推进农村改革发展若干重大问题的决定》提出“坚持最严格的耕地保护政策,层层落实责任,坚决守住 1.2 亿 hm^2 耕地红线”,保护政策为“10”。

在 Matlab R2010 环境下,应用 SVM 和 LSSVR 工具箱,利用表 1 中 2001—2006 年的数据作为训练数据,分别建立多元线性模型、GM(1,1)、BP 神经网络、FSA-SVM、FSA-LSSVR 耕地变化预测模型,算得 2007—2010 年的耕地面积,并与实际耕地面积比较,对各预测模型进行精度评价和验证,具体的耕地预测结果和精度数据见表 2、表 3 和图 2。

表 1 耕地预测影响因子

耕地预测影响因子	2001 年	2002 年	2003 年	2004 年	2005 年	2006 年	2007 年	2008 年	2009 年	2010 年
第一产业生产总值/亿元	760.77	783.66	749.4	950.5	966.5	1011.03	1200.18	1418.09	1495.45	1729.02
农民人均纯收入/元	2020.04	2118	2127	2499	2640.96	2969.08	3556.27	4202.49	4504.32	5285.17
农民人口数/万人	5068.10	5079.60	5091.49	5118.94	51476.21	51602.29	52083.25	52425.06	52772.52	52762.89
第三产业就业人数/万人	872.9	920.2	872.9	920.2	974	1038	1102.4	1164.3	1273.4	1354.5
固定资产投资/亿元	659.4	791.99	659.4	791.99	1178.52	1606.38	2140.03	3057.86	4461.59	5936.17
耕地保护政策	1	1	1	3	5	8	8	10	10	10
城市化水平/%	19.87	20.24	19.87	20.24	20.57	20.77	20.99	21.70	21.98	22.23
耕地面积/ hm^2	4228860	4177761	4084727	4108856	4092451	4116941	4144170	4144981	4171222	4181295

表 2 各模型耕地预测结果

预测模型	模型内耕地预测						模型外耕地预测			
	2001 年	2002 年	2003 年	2004 年	2005 年	2006 年	2007 年	2008 年	2009 年	2010 年
真实值	4228860	4177761	4084727	4108856	4092451	4116941	4144170	4144981	4171222	4181295
多元线性	4149846	4172634	4129443	4148027	4111402	4128342	4125697	4115665	4185339	4184864
GM(1,1)	4228860	4109242	4115862	4122492	4129133	4135785	4142448	4149121	4155805	4162500
BP 神经网络	4228426	4177781	4085144	4108853	4092455	4116943	4144159	4202061	4179333	4150628
FSA-SVM	4230082	4179119	4084917	4110226	4093399	4117087	4149739	4155918	4190372	4200592
FSA-LSSVR	4228147	4177166	4084004	4108969	4099197	4118062	4140922	4141286	4179202	4189749

表 3 各模型耕地预测精度

预测模型	模型内精度			模型外精度			总精度		
	RMSE	MXMAE	MXARE	RMSE	MXMAE	MXARE	RMSE	MAE	MAE
多元线性	41418	79014	1.8684	18793	29316	0.7073	34213	26386	0.0063560
GM(1,1)	35475	68519	1.6401	12360	18795	0.4495	28569	20889	0.0050488
BP 神经网络	246	434	0.0102	32 651	57080	1.3771	20651	9675	0.0023264
FSA-SVM	1014	1370	0.0333	14914	19297	0.4615	9465	6019	0.0014445
FSA-LSSVR	2833	6746	0.1648	6312	8454	0.2022	4555	3339	0.0008046

注:RMES 为均方根误差;MXMAE 为最大绝对误差;MAE 为平均绝对误差;MXARE 为最大相对误差。

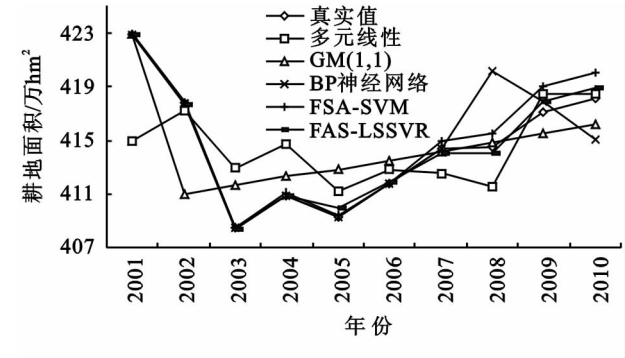


图 2 各预测模型耕地预测值和真实值对比

由表 2、表 3 和图 2 可知:(1) 多元线性耕地预测模型的预测精度较差,其模型内和总精度的均方根误差、最大绝对误差、最大相对误差和平均绝对误差在这几种预测模型中最大,表明多元线性模型不能准确表达非线性耕地变化预测;(2) GM(1,1)耕地预测模型仅考虑耕地变化的内在规律,忽略了其他综合因子对耕地变化的外在影响,所以其预测精度不高;(3) BP神经网络耕地变化预测模型的模型内精度最好,各精度指标均优于其它几种预测模型,但其模型外精度指标均低于其他预测模型,且其网络结构复杂,参数确定困难,易产生过拟合现象^[24];(4) FSA-SVM 耕地变化预测模型和 FSA-LSSVR 耕地变化预测模型的模型外精度和总精度均高于前三个预测模型,模型内精度仅次于 BP 神经网络耕地预测模型;(5) FSA-LSSVR 模型精度略高于 FSA-SVM 模型,因为 LSSVR 模型在标准 SVM 的目标函数上增加了误差平方和项,把线性不等式约束转化为线性等式约束,求解过程转化为解一组线性方程式,避免了耗时的二次凸规划问题求解,改善了 SVM 的求解效率,简

化了 SVM 的学习复杂度,提高了模型的预测精度;(6) 利用本文引用的安徽省 2001—2010 年的耕地数据,FSA 优化 SVM 和 LSSVR 内部参数时,调用 SVM 1 500 次耗时 274.56 s,调用 LSSVR 1 500 次耗时 86.37 s。可见,FSA-LSSVR 的运行速度较快。

3 结论

本文针对耕地变化预测模型的问题,以安徽省耕地变化为例,提出了一种 FSA-LSSVR 耕地变化预测模型。通过对比分析多元线性、GM(1,1)、BP 神经网络、FSA-SVM 和 FSA-LSSVR 模型在耕地变化中的预测能力,得到如下结论:

- (1) 针对 SVM 内部参数难以确定的问题,利用 FSA 的全局搜索能力对惩罚因子 γ 和 RBF 核函数参数 σ 进行最优或近似最优选择,结果证实,FSA 能有效地收敛到参数 γ 和 σ 的全局最优解。
- (2) 耕地变化是一种多因素参与的高维非线性预测系统。多元线性模型和单因素 GM(1,1)模型均不能准确地表达耕地变化趋势。BP 神经网络由于其复杂的网络结构和大量随机的内部参数,使其在耕地变化中的预测精度不高,并且 BP 神经网络缺乏完善的理论依据。
- (3) FSA-LSSVR 模型的预测精度指标远高于多元线性、GM(1,1)和 BP 神经网络模型,且优于 FSA-SVM。

(4) FSA 优化 SVM 和 LSSVR 内部参数时,调用 LSSVR 1 500 次较调用 SVM 1 500 次耗时缩短了近 2/3,运行速度大大提高。

总之,FSA-LSSVR 模型可以解决 SVM 内部参

数难以确定的问题,适用于多因素参与的高维非线性的耕地变化预测,而且速度快、精度高,具有推广价值。

参考文献:

- [1] 杜新波,周伟,司慧娟,等. 青海省 2000—2008 年间耕地变化及驱动力研究[J]. 水土保持研究,2013,20(5):180-86.
- [2] 车明亮,聂宜民,刘登民,等. 区域耕地数量变化预测方法的对比研究[J]. 中国土地科学,2010,24(5):13-18.
- [3] 赵永华,刘晓静,奥勇. 陕西省耕地资源变化及耕地压力指数分析与预测[J]. 农业工程学报,2013,29(11):217-223.
- [4] 赵海英,张明旭. 基于灰色模型的耕地变化预测[J]. 吉林师范大学学报:自然科学版,2007,5(2):66-67.
- [5] 胡喜生,洪伟,吴承祯. 基于 BP 神经网络的福建省耕地预测模型[J]. 福建农林大学学报:自然科学版,2008,37(4):66-67.
- [6] 黄成毅,邓良基,方从刚. 基于灰色—马尔柯夫模型的区域耕地变化预测研究:以四川盆地中部丘陵区为例[J]. 四川师范大学学报:自然科学版,2009,32(6):816-821.
- [7] 张豪,罗亦泳,张立亭,等. 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.
- [8] Yuan S F, Chu F L. Support vector machines-based fault diagnosis for turbo-pump rotor[J]. Mechanical Systems and Signal Processing,2006,20(4):939-952.
- [9] Doumpos M, Zopounidis C, Golfinopoulou V. Additive support vector machines for pattern classification[J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,2007,37(3):540-550.
- [10] Khemchandani R, Chandra S. Twin support vector machines for pattern classification[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on,2007,29(5):905-910.
- [11] Wu Z, Li C, Ng J K Y, et al. Location estimation via support vector regression[J]. Mobile Computing, IEEE Transactions on, 2007,6(3):311-321.
- [12] Hao P Y, Chiang J H. Fuzzy regression analysis by support vector learning approach[J]. Fuzzy Systems, IEEE Transactions on,2008,16(2):428-441.
- [13] Suyken J A K, Vandewalle J. Least squares support vector machine classifiers[J]. Neural Processing Letters,1999,9(3):293-300.
- [14] Van Gestel T, Suykens J A K, Baesens B, et al. Benchmarking least squares support vector machine classifiers[J]. Machine Learning,2004,54(1):5-32.
- [15] Anguita D, Boni A. Digital least squares support vector machines[J]. Neural processing Letters,2003,18(1):65-72.
- [16] Tsujinishi D, Abe S. Fuzzy least squares support vector machines for multiclass problems[J]. Neural Networks,2003,16(5):785-792.
- [17] Vapnik V N. Statistical learning theory[M]. New York:Wiley,1998.
- [18] 朱家元,段宝君,张恒喜. 新型 SVM 对时间序列预测研究[J]. 计算机科学,2003,30(8):124-125.
- [19] 彭珍瑞,孟建军,祝磊,等. 基于支持向量机的铁路客运量的预测[J]. 辽宁工程技术大学学报,2007,26(2):269-272.
- [20] 朱家元,陈开陶,张恒喜. 最小二乘支持向量机算法研究[J]. 计算机科学,2003,30(7):157-159.
- [21] 李波,徐宝松,武金坤,等. 基于最小二乘支持向量机的大坝力学参数反演[J]. 岩土工程学报,2008,30(11):1722-1725.
- [22] 李晓磊,邵之江,钱积新. 一种基于动物自治体的寻优模式:鱼群算法[J]. 系统工程理论与实践,2002,22(11):32-38.
- [23] 周利民. 基于鱼群算法的无线传感器网络覆盖优化研究[D]. 长沙:湖南大学,2010.
- [24] 杨淑霞,韩奇,徐琳茜,等. 鱼群算法与神经网络结合的节能减排效果评价[J]. 中南大学学报:自然科学版,2012,43(4):1538-1544.
- [25] 赵永华,何兴元,胡远满,等. 岷江上游汶川县耕地变化及驱动力研究[J]. 农业工程学报,2006,22(2):94-97.
- [26] 李伟,郝晋珉,冯婷婷,等. 基于计量经济模型的中国耕地数量变化政策与资产因素分析[J]. 农业工程学报,2008,24(6):115-118.
- [27] 刘文智,陈亚恒,李新旺. 基于产能的耕地整理数量质量潜力测算方法研究:以河北省卢龙县为例[J]. 水土保持研究,2010,17(3):227-231.
- [28] 林建平,赵小敏,邓爱珍,等. 城镇建设用地规模影响因素分析及预测:以江西省广丰县为例[J]. 国土资源科技管理,2008,25(2):102-106.