

基于支持向量机的土地利用变化模拟模型

熊 华^{1,2} 刘耀林^{1,2} 车珊珊¹ 张俊峰¹

(1 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)
(2 武汉大学地理信息系统教育部重点实验室,武汉市珞喻路 129 号,430079)

摘 要:以湖北省为例,选取 5 大类 7 个耕地,利用变化驱动力因子,将 1986~2000 年的数据作为样本训练数据,2001~2004 年的数据作为测试数据,与耕地变化进行基于支持向量机的回归模拟,用遗传算法对参数进行优化,并与 BP、RFB 神经网络模型进行了对比。模拟结果精度分析显示,SVM 模型较 BP 神经网络模型理想,与 RFB 神经网络接近。并运用该模型对湖北省 2010 年的耕地利用变化进行了预测,结果合理。研究表明,SVM 模型有较强的自学习、自适应能力,在土地利用变化模拟中有着广泛的应用前景。

关键词:土地利用变化模拟;支持向量机;神经网络

中图法分类号:P273

土地利用变化是全球变化研究的重要领域^[1],其中,土地利用变化的驱动力研究可以揭示土地利用变化的内在机制,预测未来的变化方向,因而成为土地利用变化研究的核心内容^[2]。耕地变化是土地利用变化的核心类型^[3],在一定区域某时段内,社会经济因素是耕地利用变化的主导力量。目前,土地利用变化驱动力研究多采用因子分析如相关分析、主成分分析方法筛选驱动力因子,在此基础上建立驱动力因子与土地利用变化的回归模拟模型^[4-7]。然而,社会经济驱动因子与耕地利用变化之间往往是复杂的、非线性的反馈关系,建立的线性或指数回归模型并不理想,一些重要的驱动因子难以通过统计学检验而不能进入回归模型^[5-7],这无疑削弱了模型的解释力,导致模型的适应能力差。

支持向量机(SVM)是借助于最优化方法解决机器学习问题的新工具,可以用于分类问题和回归问题的求解,在文本分类、手写识别、图像分类、生物信息学等领域中获得了较好的应用^[8-10]。在国内土地利用领域,杨青生等(2006)应用 SVM 模型进行了空间土地利用变化模拟^[11],而基于驱动力与土地利用变化的 SVM 模拟研究尚未开展。本文重点研究了基于驱动力机制的耕地变化的 SVM 模拟。

1 模型与数据

SVM 回归的基本思想是:通过非线性映射 φ ,把 n 维样本向量从原空间映射到一个高维乃至无穷维的 Hilbert 空间(亦称特征空间),在特征空间中寻找一个线性超平面,使所有样本点离超平面的总偏差 ϵ 最小。当训练样本集不能线性分离(训练样本有重叠现象)时,可以通过“软化”对间隔要求,引入松弛变量 ζ_i 和惩罚参数 C 而转化为可线性分离的情形。本文选用 ϵ 不敏感损失函数, ϵ -SVM 回归机的数学原理见文献[8-10]。

1.1 模型参数的遗传算法优化

SVM 训练首先要确定核函数及参数 γ 、惩罚参数 C 、不敏感损失函数参数 ϵ 。SVM 的推广能力与以上参数的选取有很大关系,对不同的核函数,SVM 模型的计算结果不同,即使对同一核函数,若参数不同,其结果也不同^[12]。常用的核函数有多项式核函数、RBF 核函数、Sigmoid 核函数。本文采用 RBF 核函数。基于 RBF 核函数的 SVM 参数选择,常规方法采用穷举法或网格法^[13],即在模型(SVM、核函数)选择后,首先对惩罚参数 C 、核函数参数 γ 、不敏感损失函数参数 ϵ 赋初始值,开始实验测试,根据测试精度反复调整

参数值,直到得到满意的测试精度为止。

本文对以上三个参数采用遗传算法全局搜索优化^[14]。随机产生初始种群 50 个,以最优目标函数作为适应度函数,交叉概率为 0.9,变异概率为 0.005,最大代数为 500,效率和速度比网格法搜索好,运算时间短。本文采用 C++ 对遗传算法编程,并与 SMO 算法实现的 SVM 软件 libsvm-2.8.1^[13]组合,联立求解。其流程图如图 1 所示。

1.2 土地利用变化驱动力因子体系与数据处理

本文综合文献[4-7]对土地利用变化驱动力因子的研究,将影响土地利用变化的社会经济驱动因子分为人口类(总人口、非农业人口及比重)、经济发展类(地方生产总值、固定资产投资、工农业总产值等)、生活水平类(农民人均纯收入、城镇居民可支配收入及二者之比)、农业技术水平类(粮食单产、农业机械总动力、有效灌溉面积、化肥使用量等)及政策类(土地管理政策等)。通过主成分分析,选用贡献率大的重要因子构建影响耕地变化的驱动力指标体系,即总人口 x_1 、非农业人口比重 x_2 、地方生产总值 x_3 、固定资产投资 x_4 、城镇职工人均可支配收入与农村居民人均纯收入比 x_5 、粮食单产 x_6 、土地管理政策 x_7 ,因变量为耕地面积 y 。

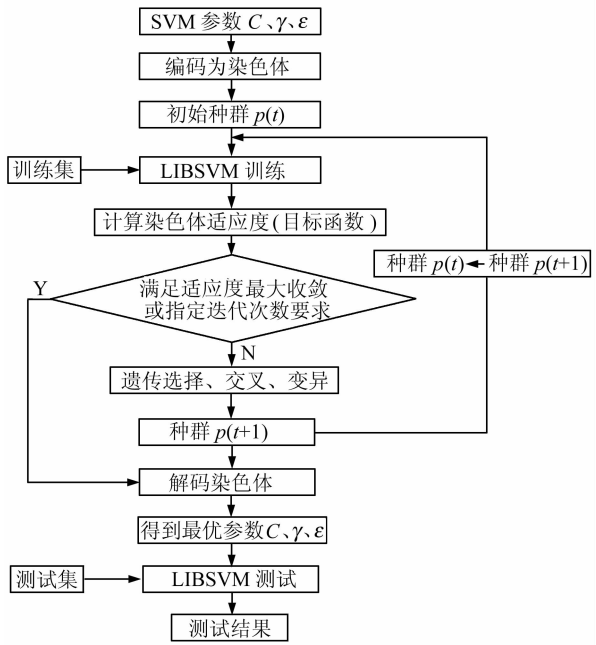


图 1 GA-SVM 模型流程
Fig. 1 Frame of GA-SVM Model

本研究数据来源于 1986~2004 年《湖北省统计年鉴》,为消除量纲的影响,对各指标利用公式 $x_i/\max(x_i)$ 进行标准化处理。处理后的数据如表 1 所示。

表 1 湖北省土地利用驱动力因子数据表(1986~2004 年)

Tab. 1 Data of Land Use Change Driving Force Factors from 1986 to 2004 of Hubei Province

年份/年	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
1986	0.834 4	0.494 4	0.058 3	0.031 2	0.614 2	0.801 1	0.364 8	1.000 0
1987	0.846 3	0.501 4	0.068 3	0.039 2	0.663 5	0.146 1	0.389 9	0.996 2
1988	0.857 0	0.509 6	0.082 6	0.044 9	0.727 7	0.078 7	0.415 0	0.993 1
1989	0.869 2	0.517 4	0.094 6	0.034 6	0.709 1	0.191 0	0.440 1	0.987 6
1990	0.899 1	0.510 1	0.108 7	0.040 4	0.683 2	0.325 8	0.465 2	0.983 7
1991	0.911 1	0.516 9	0.120 5	0.047 1	0.816 0	0.000 0	0.490 2	0.980 0
1992	0.922 3	0.530 6	0.143 6	0.067 4	0.888 0	0.438 2	0.515 3	0.973 8
1993	0.934 5	0.543 2	0.187 9	0.107 3	1.000 0	0.382 0	0.540 4	0.965 9
1994	0.945 3	0.580 1	0.247 8	0.166 0	0.918 4	0.516 9	0.565 5	0.958 1
1995	0.954 1	0.598 9	0.315 4	0.231 3	0.853 6	0.629 2	0.590 6	0.952 2
1996	0.962 8	0.611 4	0.391 8	0.275 5	0.749 6	0.573 0	0.627 4	0.948 1
1997	0.970 7	0.622 6	0.455 1	0.303 3	0.713 9	0.752 8	0.664 1	0.945 0
1998	0.976 4	0.628 3	0.488 6	0.344 6	0.713 5	0.674 2	0.712 1	0.942 1
1999	0.988 9	0.627 6	0.508 9	0.364 5	0.755 1	0.696 6	0.760 1	0.938 4
2000	0.985 1	0.638 7	0.564 1	0.397 9	0.782 1	0.966 3	0.808 1	0.932 8
2001	0.987 5	0.645 7	0.615 0	0.434 3	0.799 6	0.752 8	0.856 1	0.924 9
2002	0.994 4	0.997 2	0.743 0	0.659 6	0.891 5	1.000 0	0.904 0	0.904 3
2003	0.996 9	0.986 3	0.860 0	0.793 4	0.910 4	0.921 3	0.952 0	0.892 9
2004	1.000 0	1.000 0	1.000 0	1.000 0	0.920 8	0.687 6	1.000 0	0.889 0

2 结果分析与预测

2.1 模型结果分析

GA-SVM 模型确定的 γ 取值定为(0,1), ϵ 为(0,1), C 为(1,10 000)。求解得到的 SVM 对偶

问题的最优目标函数值为-0.207 132, γ 取值为 0.005 030, C 取值为 52.408 838, ϵ 取值为 0.008 205。

为对比分析,同时将样本数据用 BP 神经网络,RBF 神经网络进行训练和模拟。BP 神经网络输入单元数 7 个,输出单元数 1 个,隐含层数

1,隐含单元 7,隐含层和输出层激发函数用 Sigmoid 函数。对学习率取值采用试错法,以均方误差最小估计模型参数,隐含层学习率取 0.015,输出层学习率取 0.01,动量项取 0.8。RBF 神经网络学习速度取值范围为(1,30)。

将 1986~2000 年数据作为样本训练数据,2001~2004 年数据作为测试(模拟)数据。测试样本的耕地实际值与三种模型模拟得到的耕地预

测值见表 2。为了检验 SVM 模型,与神经网络模型测试结果进行比较,采用残差检验方法,对模型预测值和实际值的误差即绝对误差($\epsilon(t)=y(t)-y_0(t)$)、相对误差($\epsilon(t)/y_0(t)\times 100\%$)进行逐点检验,并作出统计分析。

通过表 2 可以看出,SVM 模型比 BP 神经网络模型模拟的精度高,与 RBF 神经网络模型模拟的结果接近。

表 2 SVM 模型与神经网络模型的测试值与实际值的分析比较

Tab. 2 Comparison Between Prediction Value and Practical Value from 2001 to 2004 with SVM Model and Neural Networks Models

年份	实际值	SVM 模型			BP 神经网络模型			RBF 神经网络模型		
		预测值	绝对误差	相对误差/%	预测值	绝对误差	相对误差/%	预测值	绝对误差	相对误差/%
2001	0.924 9	0.934 2	0.009 3	1.01	0.956 6	0.031 7	3.43	0.928 61	0.003 7	0.40
2002	0.904 3	0.915	0.010 6	1.18	0.960 0	0.055 7	6.16	0.905 80	0.001 5	0.17
2003	0.892 9	0.904 7	0.011 8	1.32	0.960 4	0.067 5	7.55	0.895 97	0.003 1	0.34
2004	0.889 0	0.891	0.002 0	0.23	0.960 5	0.071 5	8.05	0.881 26	-0.007 7	-0.87
平均			0.008 4	0.94		0.056 6	6.30		0.000 1	0.01
相关系数			0.974 1			-0.953 6			0.977 8	
均方误差			0.000 019 4			0.000 320 2			0.000 028 4	
误差标准差			0.004 4			0.017 9			0.005 3	

2.2 预测

2010 年,湖北省耕地利用变化驱动力因子($x_1\sim x_6$)根据《湖北省经济社会发展第十一个五年规划主要目标和措施》确定的目标或历年统计趋势进行推算。标准化后的数据分别为 $x_1=1.027,x_2=1.142,x_3=1.390,x_4=1.460,x_5=1.086,x_6=1.020,x_7=1.100$ 。

利用测试的 SVM 模型对 2010 年耕地利用变化进行预测,预测结果为 0.858 117,反推 2010 年耕地面积为 304.202 万 hm^2 。2010 年耕地面积将比 2004 年减少 15.964 万 hm^2 ,年均减少 2.661 万 hm^2 。变化趋势图如图 2 所示。

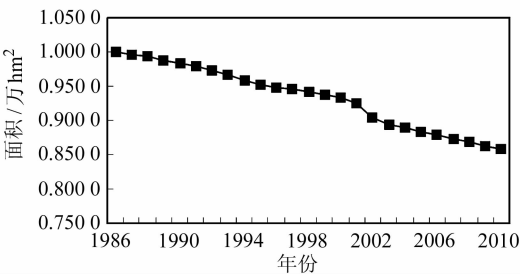


图 2 1986~2010 年耕地变化趋势图

Fig. 2 Trend of Arable Land Change from 1986 to 2010

根据湖北省 2005 年土地利用变化情况进行统计^[15],年内耕地净减少 1.58 万 hm^2 ,预测数据比 2005 年耕地净减少数据大。在未来几年内,总人口持续增长,特别是城市化、工业化水平处于加速

期,在人口、经济等驱动力作用更强的情况下,耕地减少趋势将比目前快是合理的。预测表明,湖北省未来几年内,耕地保护面临较大的压力。

3 结 语

本文应用支持向量机进行土地利用变化预测模拟,具有如下优点:① SVM 模型应用核函数的展开定理,不需要探求非线性映射的显式表达式,其自学习、自适应等智能化特征使得主要驱动力因子得以考虑,增强了模型模拟的说服力,因而 SVM 模型具有较传统线性回归模型更强的模拟能力;② 运用核函数提供了将数据映射到高维空间来增强线性分类学习的计算能力,与线性模型相比,几乎不增加计算的复杂性,在某种程度上避免了“维数灾难”,计算速度快捷;③ 与普通神经网络相比,SVM 模型的稳定性和模拟精度更高。当然,各区域的数据难免不同,这在一定程度上可能影响模拟结果精度的判断,还需要其他区域的数据进行比较验证。

参 考 文 献

[1] 李秀彬. 全球环境变化研究的核心领域:土地利用/土地覆被变化的国际研究动向[J]. 地理学报,1996,51(6):553-558

[2] 摆万奇,赵士洞. 土地利用变化驱动力系统分析[J]. 资源科学,2001,23(3):39-41

[3] 朱会义,何书金,张明. 环渤海地区土地利用变化的驱动力分析[J]. 地理研究,2001,20(6):669-678

[4] 王良健,刘伟,包浩生. 梧州市土地利用变化的驱动力研究[J]. 经济地理,1999,19(4):75-79

[5] 吴炜,卞正富. 睢宁黄河故道沿线区域土地利用变化特征与人文驱动力研究[J]. 中国土地科学,2005,19(2):29-35

[6] 王晓峰,任志远,黄青. 农牧交错区县域土地利用变化及驱动力分析——以陕北神木县为例[J]. 干旱区地理,2003(4):402-407

[7] 熊华,梁艳,邓鹏,等. 区域土地利用变化及驱动力研究——以南宁市为例[J]. 测绘信息与工程,2006,31(3):50-51

[8] 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

[9] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery,1998,2(2):121-167

[10] Alex J, Smola A J,Olkopf B S. A Tutorial on Support Vector Regression[J]. Statistics and Computing,2004,14:199-222

[11] 杨青生,黎夏. 基于支持向量机的元胞自动机及土地利用变化模拟[J]. 遥感学报,2006,10(6):836-846

[12] 魏宏业,王建华,何巍. 销售量预测的支持向量机建模及参数选择研究[J]. 系统仿真学报,2005,17(1):33-36

[13] Chang C C, Lin C J. LIBSVM: a Library for Support Vector Machines[OL]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001

[14] 王小平,曹立明. 遗传算法——理论、应用与软件实现[M]. 西安:西安交通大学出版社,2002

[15] 湖北省国土资源厅. 湖北省 2005 年度土地利用变化情况[OL]. <http://www.hbtlr.gov.cn>, 2005

第一作者简介:熊华,副教授,博士生。现主要从事土地规划、土地管理研究。
E-mail:harrietxionghua@163.com

Land Use Change Simulation Model Based on Support Vector Machine

XIONG Hua^{1,2} LIU Yaolin^{1,2} CHE Shanshan¹ ZHANG Junfeng¹

(1 School of Resources and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)
(2 Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: We apply support vector machine to simulate the relation between arable land area and datum of driving forces of land use change of Hubei Province from 1986-2004, and compare the test values by GA-SVM with ones by BP and RBF neural networks models. Practice proves the simulation results of land use change by SVM are more effective than those from BP-ANN, and is approximate to the precision of RBF-ANN.

Key words: land use change model; SVM; neural networks

About the first author: XIONG Hua, associate professor, Ph.D candidate, engaged in land plan and land management.
E-mail: harrietxionghua@163.com

(上接第 361 页)

ty, new nodes quantity, time costs and meshes quality as criterions, ETCM and several existing node refinement algorithms were evaluated by simulations. The results show that ETCM has the least time costs, and is more independent of data distribution. Moreover, the performances on other criterions of ETCM are almost the same as those of the best of existing algorithms.

Key words: Delaunay triangulation; feature constrains; node refinement

About the first author: TIAN Fengmin, Ph.D candidate, majors in marine GIS and underwater terrain aided navigation.
E-mail: tianfengmin2000@163.com.cn