

杨 静,师学义,荣联伟. PCA-BP 模型在城乡建设用地规模预测中的应用[J]. 江苏农业科学,2017,45(1):246-249.
doi:10.15889/j.issn.1002-1302.2017.01.068

PCA-BP 模型在城乡建设用地规模预测中的应用

杨 静,师学义,荣联伟

[中国地质大学(北京)土地科学技术学院,北京 100083]

摘要:为探索提高 BP 神经网络在建设用地区域预测中的精度,提出主成分分析(PCA)和 BP 神经网络相结合的建设用地规模预测模型,并利用 PCA-BP 模型对山西省晋城市建设用地规模进行预测。首先,利用 Pearson 相关系数法分析筛选出影响建设用地规模的主要驱动因子;其次,利用主成分分析法(PCA)消除其相关性,并达到降维的目的;最后,以 PCA 结果作为输入层建立建设用地规模的 BP 神经网络模型,并利用 BP 模型进行训练预测,得到最终预测值。预测结果表明,PCA-BP 模型的平均绝对误差 MAE、平均相对误差 MRE、均方误差 MSE 比传统 BP 神经网络模型小,平均预测精度 R 更高,具有较高的拟合度及可行性,在预测效率和预测精度方面都有进一步的改善。该方法的提出能为今后科学合理预测建设用地规模提供一种新思路,同时为土地利用总体规划修编提供重要决策基础。

关键词:PCA-BP;建设用地规模;预测;山西晋城

中图分类号:F301.23 **文献标志码:**A **文章编号:**1002-1302(2017)01-0246-03

随着我国经济高速发展、人口不断增长、城镇化进程加快和产业结构不断调整,我国城乡建设用地出现“双扩张”的趋势,耕地保护与建设用地规模控制之间的矛盾日益凸显。在此背景下,土地利用规划修编要求各地根据国家产业发展政策和当地社会经济发展需求,科学测算若干年内各类建设用地需求,保障必要的建设用地、坚持集约用地、严格控制建设用地规模^[1]。因此,对建设用地规模开展科学合理的预测对于促进经济、社会和生态的和谐发展具有重要意义。目前建设用地预测方法研究较为广泛,建设用地需求预测的方法有定额指标法、分项预测法、C-D 生产函数模型、多元线性回归模型、双因素模型、灰色预测模型和 BP 神经网络模型等^[2-7]。这些方法在实际应用中能够取得满意的效果,但仍存在不足之处:如对非线性数据的效果预测不佳,未能全面地考虑自然因素、社会经济因素对建设用地的影响。近年来,鉴于 BP 神经网络具有较强的学习能力和非线性并行处理推理能力^[8-12],BP 神经网络模型在建设用地区域预测中已经得到了广泛的应用。但传统 BP 神经网络预测方法虽然准确性、稳定性较高,但可能会出现局部极值致使预测结果不精确,初始权值随机性导致训练次数多、收敛速度慢等缺点。同时,将影响建设用地规模的所有因素或部分主要因素都代入,很有可能会造成信息重叠或大量信息缺失。针对上述问题,一些专家学者提出了将主成分分析法(PCA)和 BP 神经网络相结合的方法^[13-15],已延伸到各行各业,但在建设用地规模预测方面应用较少。本研究在分析山西省晋城市建设用地规模变化驱动因子的基础上,尝试采用主成分分析法(PCA)与

BP 神经网络模型相结合的方法构建预测建设用地规模模型。另外,为进一步说明 PCA-BP 神经网络的优势性,同时用传统 BP 神经网络模型进行预测,就预测精度等方面进行对比,以期为今后进行建设用地规模预测和土地利用规划修编提供科学依据。

1 数据与方法

1.1 数据来源

数据来源于《1997—2011 年晋城市统计年鉴》《2008—2010 年晋城市国民经济和社会发展统计公报》。根据研究区的自然条件和社会经济条件,选取相关影响因子进行研究。

1.2 研究方法

1.2.1 BP 神经网络 BP 神经网络由输入层、隐含层、输出层组成。其中,隐含层可以是 1 层,也可以是多层。前层与后层节点通过权值连接,即拓扑结构为单向传播的多层前向网络。BP 神经网络的学习过程由 2 个部分构成,即正向传播和反向传播。正向传播时,输入信息从输入层经隐含层处理后传入输出层,每一层神经元的状态只影响下一层神经元的状态。如果在输出层得不到希望的输出,则转入反向传播,将误差信号沿原来的神经元连接通路返回。返回过程中,逐一修改各神经元连接的权值。这种过程不断迭代,最后使得信号误差达到合适的范围。

本研究利用 BP 网络预测晋城市城乡建设用地规模,是根据 BP 神经网络的映射原理。假设晋城市城乡建设用地规模(Y)与诸多建设用地驱动因素(X)之间存在映射关系 F ,BP 网络就是在输入和输出过程中建立非线性映射关系,实现 F 值的最优逼近。

1.2.2 PCA-BP 神经网络 PCA 方法以最少的信息丢失为前提,将众多的原有变量综合成较少的几个综合指标,有效地解决了变量信息重叠、多重共线性等问题,简化了数据输入,优化了 BP 网络结构,提高了网络运行速率,但 PCA 无法捕捉影响建设用地规模指标的非线性关系,不能直接用来进行建

收稿日期:2015-11-16

基金项目:国土资源部公益性行业科研专项(编号:201411007)。

作者简介:杨 静(1989—),女,山西晋城人,硕士,主要从事土地资源评价与利用规划研究。E-mail:yang_023675@163.com。

通信作者:师学义,博士,教授,主要从事土地利用规划与土地利用工程研究。E-mail:shixueyi62@163.com。

设用地规模预测。BP 模型则能够很好地捕捉到数据中的非线性关系。因此,将 PCA 与 BP 神经网络相结合,在建设用地区域规模预测中,能够取长补短,充分发挥这 2 种方法的各自优势。

本研究利用 PCA-BP 模型对晋城市建设用地规模进行预测研究。首先,利用 Pearson 相关系数法分析筛选出影响建设用地区域规模的主要驱动因子;其次,利用主成分分析法(PCA)消除其相关性,并达到降维的目的;最后,以 PCA 结果作为输入层建立建设用地区域规模的 BP 神经网络模型,并利用 BP 模型进行训练预测,得到最终预测值。

2 晋城市城乡建设用地规模预测

2.1 驱动因子的选取

科学合理地选取影响建设用地区域规模的驱动因子是预测城乡建设用地规模的前提。影响建设用地区域规模的因子错综复杂,通过专家学者的相关研究,并结合晋城市的自然和社会经济条件,从人口、经济和人民生活水平 3 个方面选取了指标,

具体包括总人口、生产总值、第二产业、第三产业、二三产业所占比重、人均生产总值、耕地面积、粮食总产量、固定资产投资、一般预算支出、财政总收入、城镇居民人均可支配收入、农民人均纯收入、城镇人口、农村人口、城镇化率、非农业人口比重、城乡居民人均收入等 18 个指标。

应用 SPSS 20.0 软件中的相关分析模块,对建设用地区域面积和 18 个指标作二元相关性分析,根据相关分析结果,剔除 7 个相关性不显著的因子,保留相关性显著的 11 个因子作为影响建设用地区域规模的主要驱动因子(表 1),即总人口(X_1)、非农业人口比重(X_2)、生产总值(X_3)、第二产业产值(X_4)、第三产业产值(X_5)、固定资产投资(X_6)、财政总收入(X_7)、一般预算支出(X_8)、城镇居民人均可支配收入(X_9)、城镇化率(X_{10})、耕地面积(X_{11}),选择晋城市城乡建设用地面积为特征因素(Y)。构成影响建设用地区域面积(Y)扩张的驱动模型如下:

$$Y=f(X_1,X_2,X_3,X_4,X_5,X_6,X_7,X_8,X_9,X_{10},X_{11})。$$

表 1 Person 相关系数

因子	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
Y	0.979	0.971	0.869	0.868	0.877	0.825	0.861	0.853	0.899	0.969	0.716

2.2 PCA-BP 神经网络预测

2.2.1 主成分分析 针对 11 个影响因子标准化后的数据,根据主成分分析法的原理和步骤利用 SPSS 20.0 软件进行主

成分分析,发现 KMO 检验值为 0.745>0.700,说明能够进行主成分分析。根据累积贡献率 $\geq 85\%$,提出前 2 个主成分因子。其载荷系数、特征值、贡献率、累积贡献率见表 2。

表 2 主成分因子的旋转矩阵

主成分因子	载荷系数											特征值	贡献率 (%)	累积贡献率 (%)
	X_1^*	X_2^*	X_3^*	X_4^*	X_5^*	X_6^*	X_7^*	X_8^*	X_9^*	X_{10}^*	X_{11}^*			
Z_1	0.896	0.952	0.994	0.993	0.994	0.977	0.990	0.991	0.997	0.954	0.377	9.635	87.590	87.590
Z_2	-0.309	-0.272	0.049	0.035	0.063	0.177	0.044	0.087	-0.009	-0.255	0.901	1.094	9.947	97.538

由表 2 可知,主成分因子 Z_1 、 Z_2 与 11 个影响因子的标准化数据的线性关系如下:

$$Z_1=0.896X_1^*+0.952X_2^*+0.994X_3^*+0.993X_4^*+0.994X_5^*+0.977X_6^*+0.990X_7^*+0.991X_8^*+0.997X_9^*+0.954X_{10}^*+0.377X_{11}^*;$$

$$Z_2=-0.309X_1^*-0.272X_2^*+0.049X_3^*+0.035X_4^*+0.063X_5^*+0.177X_6^*+0.044X_7^*+0.087X_8^*-0.009X_9^*-0.255X_{10}^*+0.901X_{11}^*。$$

2.2.2 预处理 在 BP 神经网络运行之前,需要对训练样本进行一些前期处理,以便提高计算结果的精度。本研究在 BP 网络输入层样本数据的处理方面,首先通过利用在主成分分析过程中求出的主成分因子 Z_1 、 Z_2 与 11 个影响因子的标准化数据的线性关系,计算出 1996—2013 年的主成分因子 Z_1 、 Z_2 ,并对其进行归一化处理。

2.2.3 模型构建、样本训练与结果预测 通过借鉴 BP 神经网络结构设置经验和系统反复调试,确定输入层、隐含层、输出层的节点数分别为 2、8、1。网络训练权重通过随机方式确定,最大收敛次数为 10 000,输入层至隐含层的传递函数选择 S 形对数函数,隐含层至输出层的传递函数选择线性函数,最大收敛误差设置为 0.000 1,学习速率为 0.1,显示迭代过程间隔为 10,训练方法选择 L-M 优化算法,误差性能(Perform-

ance)通过 MSE 均方差表征,学习方法选择 learngd。应用 Matlab 8.0 软件,编写 BP 神经网络程序代码,将 1996—2007 年的数据作为训练样本,用于网络预测,可反映预测方法对样本数据的拟合能力。并将预处理后的 Z_1 、 Z_2 作为输入数据,将归一化后的建设用地区域总量作为输出期望值。训练结果通过误差性能和误差变化特征曲线进行表征,最后得到建设用地区域总量模拟值(表 3)。同时,选取 2008—2010 年的相关数据作为检测样本,其本身不参与网络训练,可用以检验预测方法的泛化能力,将预处理后 2011—2013 年的主成分因子 Z_1 、 Z_2 作为输入数据,可以预测得到 2011—2013 年晋城市城乡建设用地区域总量的归一化数据,再经反归一化处理,得到建设用地区域总量的预测值(表 3)。

通过相关性分析和变量共线性诊断,进一步确定选择总人口和全社会固定资产投资作为晋城市城乡建设用地区域总量的主要影响因素。将归一化处理后的总人口和全社会固定资产投资作为输入数据,将归一化后的建设用地区域总量作为输出期望值。参数设置同 PCA-BP 网络模型。将 1996—2007 年的相关数据作为训练样本,2008—2010 年的对应数据作为检验数据,检验该网络的泛化能力,将预处理后的 2011—2013 年的主成分因子 Z_1 、 Z_2 作为输入数据,得到晋城市 2011—2013 年的建设用地区域规模(表 4)。

表 3 晋城市建设用地规模 PCA-BP 神经网络拟合结果

年份	建设用地规模(hm ²)			相对误差 (%)	备注
	实际值	拟合值	绝对误差		
1996	39 808	40 461	653	1.64	训练样本
1997	42 002	41 692	310	0.74	训练样本
1998	42 265	42 324	59	0.14	训练样本
1999	42 961	42 961	0	0.00	训练样本
2000	44 027	43 973	54	0.12	训练样本
2001	44 483	44 483	0	0.00	训练样本
2002	44 983	44 832	151	0.34	训练样本
2003	45 340	45 179	161	0.36	训练样本
2004	45 555	45 464	91	0.20	训练样本
2005	45 680	45 710	30	0.07	训练样本
2006	46 506	46 448	58	0.12	训练样本
2007	46 922	46 839	83	0.18	训练样本
2008	47 265	47 219	46	0.10	检验样本
2009	47 833	47 771	62	0.13	检验样本
2010	48 277	48 404	127	0.26	检验样本
2011	49 243	49 207	36	0.07	预测样本
2012	49 986	50 101	115	0.23	预测样本
2013	53 074	53 183	109	0.21	预测样本

表 4 晋城市建设用地规模传统 BP 神经网络预测结果

年份	建设用地规模(hm ²)			相对误差 (%)	备注
	实际值	拟合值	绝对误差		
1996	39 808	41 317	1 509	3.79	训练样本
1997	42 002	41 851	151	0.36	训练样本
1998	42 265	42 168	97	0.23	训练样本
1999	42 961	43 063	102	0.24	训练样本
2000	44 027	44 383	356	0.81	训练样本
2001	44 483	44 466	17	0.04	训练样本
2002	44 983	44 942	41	0.09	训练样本
2003	45 340	44 979	361	0.80	训练样本
2004	45 555	45 489	66	0.14	训练样本
2005	45 680	46 027	347	0.76	训练样本
2006	46 506	46 319	187	0.40	训练样本
2007	46 922	46 945	23	0.05	训练样本
2008	47 265	47 476	211	0.45	检验样本
2009	47 833	47 954	121	0.25	检验样本
2010	48 277	47 771	506	1.05	检验样本
2011	49 243	49 007	236	0.48	预测样本
2012	49 986	50 321	335	0.67	预测样本
2013	53 074	53 389	315	0.59	预测样本

3 综合比较分析

从模型的结构来看,本研究传统 BP 神经网络系统的输入层考虑了总人口和社会固定生产投资 2 个指标,虽然这 2 个指标与建设用地规模的相关性都比较高,但考虑因素并不全面;而 PCA-BP 神经网络模型的输入层数据能够涵盖从人口、经济、人民生活水平等 11 个指标因子,在充分保障原始驱动因子信息的同时,可以达到降维及去除因子共线的目的。

从预测精度来看,根据表 3、表 4 对 2 种模型在平均绝对误差 MAE、平均相对误差 MRE、均方误差 MSE、平均预测精度 R 等 4 个方面进行对比发现,传统 BP 神经网络模型预测结果

的平均绝对误差 MAE、平均相对误差 MRE、均方误差 MSE、平均预测精度 R 分别为 276.72、0.62、184 518.28、99.39;而 PCA-BP 神经网络模型预测结果的 4 项指标分别为 119.17、0.27、35 864.06、99.73。显而易见,PCA-BP 神经网络模型预测精度较传统 BP 神经网络模型高(图 1)。

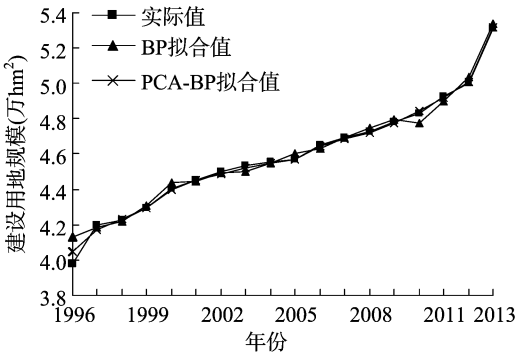


图1 晋城市建设用地规模 2 种模型拟合结果对比

4 结论与讨论

本研究提出将 PCA(主成分分析)和 BP 神经网络模型相结合的建设用地规模预测模型,该模型融合了 PCA(主成分分析)和 BP 神经网络的优点。其中,PCA 既全面保留了影响建设用地规模的人口、社会经济和人民生活水平等因素,又避免了这些驱动因子的共线性情况,将影响建设用地需求量的 11 个因子下降到了 2 个综合分值,起到降维去冗余的作用,提高了计算效率,增强了网络性能。而 BP 神经网络预测建设用地规模具有方便、快捷的特点,可以建立建设用地规模和各驱动因子之间的非线性映射关系,对于解决其复杂内部机制问题具有很好的效果。

在晋城市城乡建设用地规模预测中,PCA-BP 神经网络较传统 BP 神经网络的平均绝对误差 MAE、平均相对误差 MRE、均方误差 MSE 指标值更低,平均预测精度 R 则有了一定程度的提高,也就意味着在预测精度方面有了进一步的改善,能够有效预测建设用地规模。可见,PCA-BP 神经网络模型可以作为今后预测建设用地规模的一种新方法,并为今后进行建设用地规模预测提供一种新思路。

影响建设用地规模的指标体系有待进一步完善。本研究所构建的指标体系可能存在一些不尽完善之处,建设用地扩张受自然、社会经济、政策规划等多方面的影响,特别是地形地貌、城市规划等因素的影响也较为显著,但限于目前的研究水平,这些未能量化的影响因素并未纳入到指标体系中,探讨如何构建更加全面合理的指标体系也是今后研究的重要方向之一。

参考文献:

[1] 杨 韬,朱明仓. 城镇建设用地需求预测研究——基于恩格尔系数与人均建设用地面积模型分析[C]//2006 中国科协年会论文集:下册,2006:159-163.
[2] 姜 海,曲福田. 建设用地需求预测的理论与方法——以江苏省为例[J]. 中国土地科学,2005(2):44-51.
[3] 陈 华,邓传明,狄秀梅. 建设用地需求量预测的探讨[C]//2007 年中国土地学会学术年会论文集,2007:590-597.

晁雷,崔东亮,赵晓光,等.溶解氧对生物转盘技术处理乳制品废水效能的影响[J].江苏农业科学,2017,45(1):249-252.
doi:10.15889/j.issn.1002-1302.2017.01.069

溶解氧对生物转盘技术处理乳制品废水效能的影响

晁雷^{1,2},崔东亮¹,赵晓光^{2,3},尤涛²,徐新阳³,李亚峰¹

(1. 沈阳建筑大学市政与环境工程学院,辽宁沈阳 110168; 2. 辽宁省环境科学研究院,辽宁沈阳 110161;

3. 东北大学资源与土木工程学院,辽宁沈阳 110819)

摘要:在不同的溶解氧浓度条件下,研究接触氧化池生物膜的状态以及乳制品废水污染物的去除效率。结果发现,水力停留时间为 10 h 时,生物转盘转速为 10 r/min,此时生物转盘出水 COD_{Cr}、氨态氮(NH₄⁺-N)含量、总磷(TP)含量分别约为 129、4.67、2.81 mg/L;当接触氧化池水力停留时间 4 h 时,溶解氧浓度在 2.5~3.0 mg/L 之间,COD_{Cr}、NH₄⁺-N 含量、TP 含量出水分别达到 55、4.44、1.42 mg/L。以上结果表明,乳制品废水经过处理后满足《城镇污水处理厂污染物排放标准》一级 B 标准。

关键词:溶解氧;生物转盘;接触氧化;乳制品废水;效能;水力;停留时间

中图分类号: X703 **文献标志码:** A **文章编号:** 1002-1302(2017)01-0249-04

生物膜法与传统的污水处理方法活性污泥法相比,具有对水质水量变动较强的适应性、剩余污泥量少、运行方便等特点,是一种被广泛应用的生物处理技术^[1]。其中,生物转盘技术(rotating biological contactor, RBC)别称浸没式生物滤池,是在生物滤池的基础上发展而来的,因其系统具有设计灵活可靠、安装操作简洁、无须曝气降低能耗等特点,因此成为了既传统同时又在不断研究中的生物膜技术^[2]。近年来,国内外学者对生物转盘的应用范围、脱氮机理、盘片结构、盘片驱动方式作了广泛研究,并取得了大量成果^[3-10]。

乳制品废水相当于原状牛奶的稀释液,主要含有的污染物是乳蛋白、乳脂、乳糖、含于原乳当中的矿物质以及用于清洗生产设备、管道、容器和车间地面的酸性和碱性洗涤剂

等^[11]。乳制品废水水质主要有以下特点:废水的水质和水量有很大的波动,有机物含量高,良好的可生化性能。废水中生化需氧量(BOD)/化学需氧量(COD)大于 0.5 时,属于可生化性好的有机废水^[12]。本试验采用新式盘片的生物转盘对乳制品废水运行效果进行研究。但目前,我国对于生物转盘和生物接触氧化池 2 个生物膜反应器代表的运行影响因素和运行参数研究尚不全面、深入。因此,本试验通过改变曝气量控制混合液中溶解氧的浓度,研究不同溶解氧条件下接触氧化池生物膜的状态以及污染物的去除效率,旨在获得最佳运行系统和符合实践应用所需要的经济性要求的运行参数。

1 材料与方法

1.1 试验用水

试验采用牛奶配制的模拟乳制品工业废水(表 1)作为研究对象。本试验使用的接种活性污泥取自沈阳北部污水处理厂,主要呈现棕褐色的大块絮团,其污泥沉降比约为 42%,污泥体积指数为 87 mL/g。通过显微镜观察,污泥菌胶团絮体较好,其中丝状菌较少。

收稿日期:2015-11-13

基金项目:水体污染控制与治理科技重大专项(编号:2012ZX07208-003);辽宁省科学技术计划(编号:2014020163);沈阳建筑大学学科培育项目(编号:XKHY-36)。

作者简介:晁雷(1978—),男,辽宁沈阳人,博士,教授,主要从事污水生态处理技术研究。E-mail:chaolei@aliyun.com。

[4]刘胜华,詹长根.基于国民经济和人口发展目标的建设用地需求规模预测研究——以武汉市黄浦区为例[J].中国人口·资源与环境,2005,15(5):47-51.

[5]邱道持,刘力,栗辉,等.城镇建设用地预测方法新探——以重庆市渝北区为例[J].西南师范大学学报(自然科学版),2004,29(1):146-150.

[6]王玉波,唐莹.省域建设用地需求量预测研究——以黑龙江省为例[J].科技管理研究,2009,29(12):161-165.

[7]刘金国,张希.中国城市建设用地需求量预测研究综述[J].国土与自然资源研究,2011(2):16-18.

[8]孙华芬,赵俊三,潘邦龙,等.基于 GIS 和 BP 神经网络技术的建设用地适宜性评价研究[J].国土资源信息化,2008(6):17-19.

[9]王良健,师迎春,林目轩. BP 神经网络结合小波处理在城市扩张

预测中的应用——以长沙市区为例[J].中国土地科学,2008,22(1):39-47.

[10]王增彬,迟恒智.基于 BP 神经网络的济南市建设用地规模预测[J].水土保持研究,2007(5):222-224.

[11]郭杰,欧名豪,刘琼,等.基于 BP 神经网络的南通市建设用地需求预测[J].资源科学,2009,31(8):1355-1361.

[12]段祖亮,张小雷,权晓燕.基于 BP 神经网络模型的新疆建设用地分析[J].中国科学院研究生院学报,2009,26(4):451-457.

[13]朱志洁,张宏伟,韩军,等.基于 PCA-BP 神经网络的煤与瓦斯突出预测研究[J].中国安全科学学报,2013(4):45-50.

[14]蔡红,陈荣耀.基于 PCA-BP 神经网络的股票价格预测研究[J].计算机仿真,2011,28(3):365-368.

[15]陈建宏,郑荣凯,陈浩.基于 PCA 和 BP 神经网络边坡稳定性分析[J].中国安全生产科学技术,2014(5):142-147.