DOI：10.13718/j.cnki.xdzk.2014.11.028

基于 PCA-RBF 神经网络的成都市建设用地需求预测①

谢　汀[1,2]，　伍　文[3]，　邓良基[1,2]，　高雪松[1,2]，
李启权[1]，　徐安琪[1,2]，　谢　鑫[1]

1. 四川农业大学 资源环境学院，成都 611130；2. 四川农业大学 生态环境研究所，成都 611130；
3. 成都市国土规划地籍事务中心，成都 610072

摘要：为建立“突变”增长条件下建设用地需求预测模型以利于区域土地资源的可持续利用和土地利用总体规划的定期修编，本文提出了基于主元分析(PCA)与径向基函数(RBF)神经网络模型相结合的建设用地需求预测方法. 首先，利用灰色关联分析筛选出主要驱动因子作为输入数据；然后，利用主元分析消除其相关性，并达到降维目的；最后，以 PCA 结果为输入建立建设用地需求的 RBF 神经网络预测模型. 实例研究表明，建设用地需求 PCA-RBF 神经网络预测模型具有较好的预测效果，平均绝对误差(*MAE*)、误差均方根(*RMSE*)、平均相对误差(*MRE*)、分年度残差和突变点残差都较常规的多元线性回归(MLR)模型更小，更适宜在复杂非线性条件下应用.

关　键　词：建设用地；预测；主元分析；径向基神经网络模型；多元线性回归模型

中图分类号：　　**文献标志码：**A　　**文章编号：**1673－9868(2014)11－0183－08

我国已经历 20 多年的快速发展，经济的发展和工业化进程的进一步推进，会对地区的土地景观造成剧烈的冲击[1]. 其主要表现为大量耕地转变为建设用地和其他城市用地. 据统计，我国城市扩展占用耕地的比例很大，平均达到 70.0%左右，其中西部高达 80.9%[2]. 然而，这种由城市化带来的建设用地的扩展导致的土地利用方式的转变会引起一系列的问题，如生态和粮食安全、城市外部成本的增加等，最终又会影响到经济发展和居民生活质量的提升[3]. 因此，对建设用地需求开展科学合理的预测对于促进经济、社会和环境的和谐发展具有重要意义.

目前，用于建设用地需求预测的方法主要有时间趋势法[4]、马尔科夫模型[5]、GM(1，1)模型[6-7]等. 这些方法在实际应用中大多能得到比较满意的结果，但仍存在不足之处，如未考虑自然因素、社会经济因素和政策因素对建设用地需求的影响且仅适用于历年建设用地规模没有太大波动的情况. 近年来，为满足建设用地预测的实际需要，进一步提高预测精度，许多学者引入了神经网络方法，并取得了一定成果[8-10]. 在这些研究中，BP 神经网络得到了广泛应用[8-9]. 然而，BP 神经网络本身具有学习速度较慢、需要调节的参数多等缺点[11]. 同时，在这些研究中，往往简单地把所有因素或人为地将部分主要因素作为神经网络的输入而忽略了各因素之间可能存在的相关性，最终造成信息的大量丢失或重叠，从而影响预测精度[12]. 此外，当前，大多学者都是基于平稳数据开展预测研究，专门针对具有“突变”特征数据的研究并不多见. 鉴于此，本文以成都市为研究区域，尝试采用主元分析(PCA)与径向基函数(RBF)神经网络模型相结合的方法

① 收稿日期：2013－07－20
基金项目：国土资源部公益性行业科研专项项目(201211050)；国家支撑计划资助项目(2012BAD14B18).
作者简介：谢　汀(1989－)，男，硕士研究生，主要从事土地规划与管理研究.
通信作者：邓良基，教授.

对建设用地需求进行预测. 另外，为进一步说明 PCA-RBF 神经网络的优点，本文同时采用常用的多元线性回归(MLR)模型进行了预测，并就预测精度进行了综合对比，以期为区域土地资源可持续利用和土地利用总体规划的定期修编提供决策依据.

1 研究区概况与研究方法

1.1 研究区概况

成都市位于四川省中部，四川盆地西部，介于 102°54′－104°53′E 和 30°05′－31°26′N 之间，全市东西长 192 km，南北宽 166 km，总面积 22 390 km^2. 全市地势差异显著，西北高，东南低，西部属于四川盆地边缘地区，以深丘和山地为主，海拔大多在 450～720 m 之间. 成都市位于川西北高原向四川盆地过渡的交接地带，地处亚热带湿润地区，地形地貌复杂，自然生态环境多样，生物资源十分丰富.

1996 年以来，成都市经济进入了快速发展期，建设用地规模随着经济的发展而迅速扩大，到 2011 年，其建设用地规模净增长了 10.65×10^4 hm^2，年均增长 7 100 hm^2，其中在 2008 年后，由于“5.12 汶川地震”和分类标准差异等原因，出现了一次建设用地的“突变”增长(图 1). 不难想象，随着成都市天府新区建设的进行和“五大兴市战略”的逐步实施，建设用地规模将进一步壮大，大量的农用地尤其是耕地将被占用，人地矛盾将日益凸显. 由此可见，适时开展建设用地需求预测研究将是成都市的必然选择.

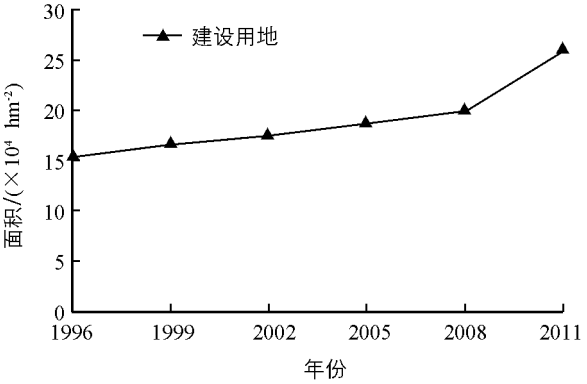

图 1 1996－2011 年成都市建设用地面积变化

1.2 数据来源

本文的建设用地面积数据主要来自变更调查数据(1996－2008,2010,2011)和第二次全国土地调查数据(2009)；社会经济数据主要来源于《中国城市统计年鉴》(1997－2012)，《四川统计年鉴》(1997－2012)，《成都统计年鉴》(1997－2012)等.

1.3 研究方法

1.3.1 灰色关联分析

考虑到建设用地规模与社会经济因素二者之间关系的复杂性，本文采用能较全面分析较多因素相互作用的灰色关联度模型[13]，定量筛选出影响建设用地规模的主要社会经济驱指标. 为便于横向对比，采用 z-score 方法对原始数据进行无量纲化处理.

$$X'_i=\frac{X_i-\overline{X_i}}{\sigma X_i}\qquad i=1,2,3,\cdots,m \tag{1}$$

$$Y'_j=\frac{Y_j-\overline{Y_j}}{\sigma Y_j}\qquad j=1,2,3,\cdots,s \tag{2}$$

式中：X'_i,Y'_i 为无量纲化数据；$\overline{X_i}$,$\overline{Y_i}$ 分别为 X_i 和 Y_j 的期望值；σX_i,σY_i 分别为 X_i 和 Y_i 的标准差.

在此基础上计算指标间的灰色关联系数.

$$l_{ij(k)}=\frac{\Delta_{\min}+\rho\Delta_{\max}}{\Delta_{ij(k)}+\rho\Delta_{\max}} \tag{3}$$

式中：$l_{ij(k)}$ 为 k 时刻 $X'_i(k)$ 与 $Y'_j(k)$ 两比较序列的绝对差即关联系数；$\Delta_{\max}$ 和 $\Delta_{\min}$ 分别表示所有比较序列各个时刻绝对差中的最大值与最小值. 因为比较序列相交，故一般取 $\Delta_{\min}=0$；ρ 称为分辨系数，其意义是削弱最大绝对差数值太大引起的失真，提高关联系数之间的差异显著性，$\rho\in(0,1)$，一般情况下取 0.5.

将灰色关联系数按样本数 N 求平均值，得到关联度矩阵.

$$r_{ij}=\frac{1}{N}\sum_{k=1}^{N}l_{ij(k)}\qquad N=1,2,3,\cdots,n \tag{4}$$

关联度取值范围在(0，1]间内，r_{ij} 值越大，关联性越大；当 $r_{ij}=1$，说明 $X'_i(k)$ 与 $Y'_j(k)$ 间的变化

规律完全相同.

1.3.2 多元回归模型

多元回归分析是依据多个自变量的最优组合建立回归方程来预测因变量的多元统计分析方法，其模型如下[8]：

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_pX_p+\varepsilon \tag{5}$$

式中：Y 为因变量；$X_1,X_2,\cdots,X_p$ 为自变量；$\beta_0,\beta_1,\beta_2,\cdots,\beta_p$ 是总体回归参数；ε 是随机扰动项.

多元回归分析的检验包括相关系数检验、标准误差检验、F 检验、t 检验、DW 检验和多重共线性检验.

1.3.3 RBF 神经网络模型

是一种具有单隐层的 3 层前馈型神经网络模型，能以任意精度逼近任意连续函数. 与广泛应用的 BP 神经网络模型相比，在学习速度、参数设置等问题上具有明显优点[14]，其模型结构如图 2 所示.

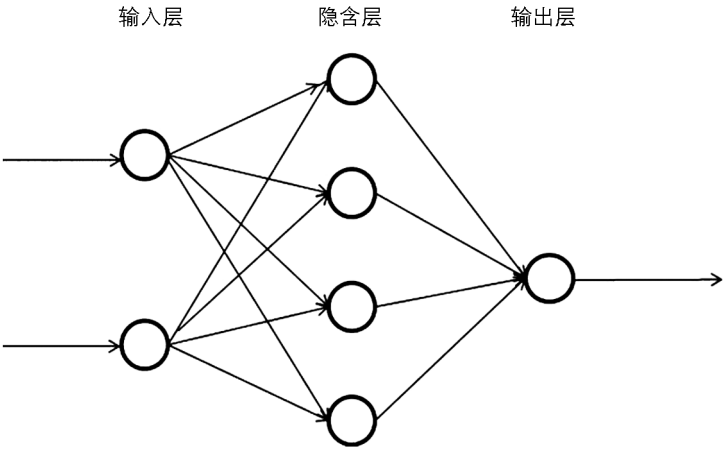

图 2 RBF 神经网络模型结构

RBFNN 的输入和输出的函数关系为：

$$y_i(x)=\sum_{j=1}^{m}w_if_i(x)+b_j \qquad i=1,2,\cdots,n \tag{6}$$

式中：$x=(x_1,x_2,\cdots,x_n)^T\in\mathbf{R}^n$ 为输入向量；y_i 为 i 个输出单元的输出值，m 为中心的个数；w_{ij} 为第 j 个神经元到第 i 个输出单元的权值；b_j 为偏置值；本文隐含传递函数采用高斯非线性基函数，表达式如下[15]：

$$\theta_i(x)=\exp\left(-\frac{\|x-c_j\|}{2\sigma_j^2}\right) \tag{7}$$

式中：$c_j\in\mathbf{R}^n(1\leqslant j\leqslant m)$ 是 RBF 的中心；σ 为宽度参数，调节 RBF 神经元的敏感程度.

1.3.4 预测精度评价

采用平均绝对误差(MAE)、误差均方根($RMSE$)和平均相对误差(MRE)三项指标来评价模型预测精度，其计算公式如下[16]：

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|\hat{Z}_i-Z_i| \tag{8}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Z}_i-Z_i)^2} \tag{9}$$

$$MRE=\frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{Z}_i-Z_i|}{Z_i}\times 100\% \tag{10}$$

式中：$\hat{Z}_i$ 为 MLR 模型或 PCA-RBF 神经网络模型预测值；Z_i 为实际面积；n 为参加训练和检测的总样本数；显然 MAE,$RMSE$ 或 MRE 越小，则误差越小，模型预测精度越高.

2 结果与分析

2.1 指标选择与数据预处理

指标的选择主要结合研究区社会经济发展现状和相关研究成果[17-18]，具体包括 GDP(x_1)、第一产业所占 GDP 比重(x_2)、第二产业所占 GDP 比重(x_3)、第三产业所占 GDP 比重(x_4)、第一产业增加值(x_5)、第二产业增加值(x_6)、第三产业增加值(x_7)、人均 GDP(x_8)、固定资产投资额(x_9)、外商直接投资额(x_{10})、财政收入(x_{11})、财政支出(x_{12})、房地产开发投资额(x_{13})、城镇居民家庭人均可支配收入(x_{14})、城镇居民人均消费性支出(x_{15})、农民人均纯收入(x_{16})、农民人均生活消费支出(x_{17})、社会消费零售总额

(x_{18})、总人口数(x_{19})、非农业人口数(x_{20})、农业人口数(x_{21})、第一产业从业人员数(x_{22})、第二产业从业人员数(x_{23})、第三产业从业人员数(x_{24})、旅客周转量(x_{25})、城镇化率(x_{26})共计26个指标.

经过前期反复试验，最后决定将社会经济数据和建设用地面积数据转换成年变化率的形式，以达到降低数据波动的目的，并对异常变化年份数据(1998和2009年)进行剔除，最终整理得到表1.

表1 1996—2011年成都市建设用地及社会经济指标变化率 %

年份	1997	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2010	2011
建设用地	2.88	4.39	1.53	2.66	1.51	3.80	2.48	0.68	2.51	1.29	2.37	3.92	2.09
x_1	13.37	8.63	10.71	14.29	12.61	14.54	19.11	16.98	16.67	21.38	17.24	23.29	23.48
x_2	−16.42	−6.25	−9.52	−6.32	−5.62	−2.38	−6.10	0.00	−9.09	1.43	−2.82	−13.56	−5.88
x_3	2.26	−0.89	0.45	0.89	0.66	0.66	1.96	−9.19	3.76	2.49	3.10	0.45	2.46
x_4	2.84	2.52	1.79	0.44	0.44	−0.22	−0.87	9.45	−1.81	−2.45	−2.52	1.21	−1.59
x_5	3.63	0.31	3.76	1.82	5.92	9.21	22.76	5.29	10.15	20.48	11.82	6.47	14.82
x_6	11.52	8.14	9.99	16.15	13.93	16.80	23.47	22.17	18.98	23.32	20.14	23.93	26.36
x_7	17.52	10.86	12.63	15.36	12.81	13.81	15.45	14.89	15.76	19.88	15.51	24.73	21.47
x_8	12.35	7.85	9.81	2.69	38.19	−10.11	17.26	14.64	14.11	19.62	16.22	21.99	1.91
x_9	19.79	12.70	13.55	22.34	20.60	22.90	22.54	37.94	30.11	25.93	25.26	5.79	16.18
x_{10}	33.63	34.60	14.92	18.28	69.87	−4.01	22.32	178.89	27.19	−38.38	97.34	73.75	5.38
x_{11}	18.09	17.19	13.19	22.52	25.23	18.77	27.31	32.94	33.71	103.77	13.61	58.07	12.22
x_{12}	18.99	20.22	14.86	27.39	22.74	19.18	21.36	28.91	39.34	123.68	23.70	68.24	1.97
x_{13}	6.69	24.87	29.35	32.20	19.07	20.70	25.90	46.25	35.80	47.53	0.80	35.25	24.82
x_{14}	47.09	10.11	35.94	−15.76	10.38	7.46	7.81	9.28	−1.07	21.17	14.42	13.25	15.70
x_{15}	0.69	5.76	10.79	5.88	−62.52	176.89	27.47	7.17	−13.82	26.04	14.45	2.43	15.85
x_{16}	18.33	5.78	8.37	5.37	6.26	8.23	11.41	10.14	9.36	15.03	14.87	15.09	20.60
x_{17}	−6.69	−0.64	6.60	6.92	8.29	6.77	8.56	4.07	8.77	19.56	14.19	15.64	21.33
x_{18}	18.13	10.84	10.82	13.23	13.07	9.37	13.50	14.21	14.85	17.48	19.50	18.09	18.35
x_{19}	0.86	0.66	0.98	0.65	0.84	1.54	1.47	2.11	1.97	0.80	1.14	0.83	1.23
x_{20}	2.69	2.71	2.90	2.57	3.08	5.61	17.48	19.88	5.07	4.21	2.77	3.42	8.42
x_{21}	0.01	−0.34	0.01	−0.35	−0.35	−0.71	−7.92	−11.20	−1.15	−2.85	−0.75	−2.37	−8.14
x_{22}	−1.49	−5.28	1.47	−6.44	−3.95	−4.56	−4.60	−5.04	−5.70	−5.00	−3.36	−6.20	−5.36
x_{23}	−0.54	−7.07	−0.66	−2.51	6.86	5.26	8.53	5.39	3.18	6.92	2.33	14.09	8.76
x_{24}	4.49	8.59	−6.57	13.42	4.27	5.70	3.23	7.50	11.58	16.81	6.22	4.84	0.23
x_{25}	−1.76	9.68	6.87	9.23	−7.52	−7.95	98.86	13.21	11.94	11.22	7.61	8.02	−31.23
x_{26}	2.56	2.45	2.79	1.75	1.26	3.92	1.03	2.99	2.81	1.71	1.60	1.02	2.27

应用DPS 9.5版软件中的灰色关联分析模块，对表1中建设用地变化率与初选26个社会经济指标变化率做灰色关联分析，得到结果如表2所示. 结果显示建设用地与社会经济指标变化率灰色关联度排序大小为 x_7，x_{12}，x_{11}，x_{15}，x_4，x_1，x_{10}，x_{21}，x_{25}，x_{26}，x_3，x_9，x_{19}，x_{18}，x_{16}，x_{17}，x_{23}，x_6，x_2，x_8，x_{13}，x_{20}，x_{24}，x_{22}，x_5 和 x_{14}. 本次研究中选取与影响建设用地面积变化率关联度大于0.7的社会经济指标作为主要驱动因子，共有14个.

表2 建设用地与社会经济指标变化率的灰色关联度

指标名称	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
关联度	0.725 2	0.696 7	0.713 8	0.726 3	0.674 1	0.696 8	0.768 3	0.692 5	0.711 1	0.719 0	0.742 0	0.744 4	0.687 8
指标名称	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}
关联度	0.656 7	0.737 4	0.699 7	0.698 4	0.707 2	0.711 0	0.686 9	0.715 0	0.680 2	0.698 4	0.686 7	0.714 7	0.714 3

2.2 多元回归模型预测

考虑到驱动因子之间可能存在多重共线性，为得到具有统计学意义的结果，需要剔除一些变量. 故根据表1，将建设用地面积和主要社会经济驱动因子变化率数据，代入式(5)，并用SPSS17.0软件，回归方法选择逐步回归分析(Stepwise)，分析结果为：

$$y = 4.132 - 0.078x_9 \tag{11}$$

从式(11)和表 3 可以看出，通过逐步回归分析，能够通过统计参数检验的只有固定资产投资(x_9)一个因子，但从前文的灰色关联分析可知，其他各变量变化率与建设用地面积变化率之间也是存在很大关联度的. 据此推论，自变量存在高度的共线性，同时变量之间存在非线性关系，无法用线性关系进行近似处理，以致多元回归模型不能建立[8]. 利用式(11)对 1996－2011 年(除 1998 和 2009 年)成都市建设用地需求变化率进行预测，并最终反算得面积，结果见表 4.

表 3　逐步多元回归分析参数分析

系数	非标准化系数		标准系数	t	Sig.
	B	标准误差			
(常量)	4.132	0.739	—	5.590	0.000
x_9	－0.078	0.033	－0.586	－2.399	0.035

表 4　逐步多元回归模型预测结果

年份	实际面积 /hm²	模型预测值 /hm²	残差值 /hm²	相　对 /%
1997	159 843.15	159 394.06	449.09	0.28
1999	166 856.88	164 862.85	1 994.03	1.20
2000	169 406.16	171 987.40	2 581.24	1.52
2001	173 908.15	173 454.16	453.99	0.26
2002	176 530.23	178 299.83	1 769.60	1.00
2003	183 230.32	180 670.63	2 559.69	1.40
2004	187 774.48	187 580.58	193.90	0.10
2005	189 044.21	189 976.04	931.83	0.49
2006	193 787.35	192 415.37	1 371.98	0.71
2007	196 282.58	197 875.12	1 592.54	0.81
2008	200 940.19	200 524.97	415.22	0.21
2010	256 542.85	255 944.88	597.97	0.23
2011	261 904.74	263 904.91	2 000.17	0.76

2.3　RBF 神经网络模型预测

1) 主元分析

应用 SPSS 19.0 软件中的因子分析功能提取 14 个主要驱动因子变化率的主元，以达到降维和去除因子共线性的作用，共可以提取 12 个主元，累计贡献率为 100%，各主元得分值见表 5.

表 5　主元得分值

年份	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9	PC_{10}	PC_{11}	PC_{12}
1997	0.008 9	－0.425 8	－0.539 6	－0.660 3	0.500 9	0.473 5	0.525 4	0.408 2	2.665 6	0.564 6	1.351 4	0.014 7
1999	－0.810 0	－1.141 3	－0.753 4	0.786 3	0.449 9	－0.660 2	－0.326 9	－0.076 9	0.166 6	－1.693 1	－0.338 1	－2.042 7
2000	－0.708 9	－1.114 0	－0.226 8	0.257 0	0.250 0	－0.499 7	－0.563 8	0.969 1	0.130 6	－0.613 2	－1.014 1	2.495 0
2001	0.021 8	－0.635 0	－0.425 6	0.374 0	－0.005 7	0.028 9	－0.168 6	－1.142 2	0.180 3	2.349 3	－1.800 0	－0.476 3
2002	－0.173 8	－0.527 0	－1.393 6	0.449 8	－0.017 0	0.857 1	－0.694 3	－0.436 0	－1.675 4	0.592 3	1.975 9	0.394 6
2003	－0.954 0	－0.873 0	2.392 8	－0.394 4	0.532 9	－0.734 9	1.031 0	－0.402 0	－0.733 7	0.454 5	0.849 8	－0.088 6
2004	0.174 3	0.538 6	0.213 4	0.910 9	－2.781 3	－1.272 5	－0.010 5	－0.012 7	0.480 0	0.036 8	0.537 5	0.074 4
2005	－1.891 4	2.589 2	－0.206 2	0.260 8	0.739 1	0.064 8	－0.066 6	－0.253 4	0.096 6	－0.007 9	－0.192 6	0.140 0
2006	－0.006 9	0.091 4	0.892 4	－0.097 1	－0.665 3	1.856 6	－0.457 4	2.203 2	－0.410 9	0.268 9	－0.529 1	－0.888 1
2007	1.881 2	0.512 2	1.154 1	1.652 4	1.014 8	0.575 1	－0.662 2	－0.814 7	0.420 8	－0.655 7	0.100 0	0.365 7
2008	0.319 9	－0.019 1	－0.459 8	－0.804 8	－0.776 7	1.357 8	2.007 0	－1.144 4	－0.332 3	－1.231 7	－0.725 0	0.421 3
2010	1.496 1	0.679 9	－0.873 1	－0.281 5	0.905 9	－1.511 1	1.251 9	1.374 1	－0.873 5	0.379 4	－0.106 0	－0.240 0
2011	0.642 9	0.323 9	0.225 4	－2.453 1	－0.147 5	－0.535 5	－1.865 1	－0.672 3	－0.114 6	－0.444 2	－0.109 7	－0.170 1

2) RBF 神经网络运算

以提取的 12 个主元得分值作为输入样本，输入层节点数设置为 12，以建设用地变化率作为输出样本，输出层节点数设置为 1，输入输出样本皆进行归一化处理. 选取 2001,2002,2006 和 2008 年数据为检测样本，其本身不参与网络训练，可用以检验预测方法的泛化能力；剩余各年数据为训练样本，用于网络训练，可反映预测方法对样本数据的拟合能力. 另外，为防止人工神经网络方法的过拟合问题，获得最优的网络结构和参数，本文以对样本值预测结果的误差最小为约束条件，将模型的扩展常数初始值设定为 0.05、步长为 0.05、隐层节点数初始值为 2、步长为 1，应用编程搜索确定扩展常数和隐层节点数的最优组合，通过逐步搜索，确定预测建设用地需求的最佳的网络隐层节点数和扩展常数分别为 3 和 0.95，利用该参数完成对研究区建设用地需求的人工神经网络预测(表 6). 以上操作皆利用 Matlab 7.1 软件中提供的函数 newrb 完成.

表 6 RBF 神经网络法预测结果

年份	实际面积 /hm^2	模型预测值 /hm^2	残差值 /hm^2	相对 /%	备注
1997	159 843.15	159 102.19	740.96	0.46	训练样本
1999	166 856.88	166 716.35	140.53	0.08	
2000	169 406.16	171 186.39	1 780.23	1.05	
2003	183 230.32	180 596.12	2 634.20	1.44	
2004	187 774.48	187 498.56	275.92	0.15	
2005	189 044.21	189 012.11	32.10	0.02	
2007	196 282.58	198 161.92	1 879.34	0.96	
2010	256 542.85	256 426.62	116.23	0.05	
2011	261 904.74	262 508.70	603.96	0.23	
2001	173 908.15	173 273.18	634.97	0.37	检验样本
2002	176 530.23	177 876.66	1 346.43	0.76	
2006	193 787.35	193 081.16	706.19	0.36	
2008	200 940.19	200 587.49	352.70	0.18	

2.4 综合对比分析

根据表 4 和表 6，作 2 种模型关于研究区建设用地需求预测结果的误差统计分析表(表 7)，由表可知 PCA-RBF 神经网络模型与 MLR 模型相比，平均绝对误差(*MAE*)、误差均方根(*RMSE*)和平均相对误差(*MRE*)3 项指标值皆相对较小，表明与 MLR 模型相比，PCA-RBF 神经网络模型能够更好地捕捉建设用地需求与多社会经济因子间的复杂关系.

表 7 不同预测模型的统计误差

预测指标	预测方法	总样本点		
		MAE	*RMSE*	*MRE*
建设用地面积	MLR	1 300.866	1 533.858	0.007
	PCA-RBF	864.905	1 160.356	0.005

残差为模型预测值和实际值之间的差，能更直观、更具体地反应模型预测的精度，图 3 为 2 种模型预测残差的对比分析图. 从图中可以看出，PCA-RBF 神经网络模型的残差在多数年度均较 MLR 模型更小，此外，对模型预测能力进行对比的另一个关键点是看模型在突变点上的预测精度. 由图 1 可以发现，2008 年后成都市建设用地面积发生了一次“突变”，在突变点的预测上，PCA-RBF 神经网络模型较 MLR 模型的残差更小，表现出了预测精度优势. 另外，2008 年后建设用地面积发生的“突变”，充分说明了其增长是多因素约束下的非线性增长此时传统的线性回归预测方法的不足就暴露无遗，而神经网络模型特别是 PCA-RBF 神经网络则在处理该问题上却突出地显示出了其技术方法的先进性[19].

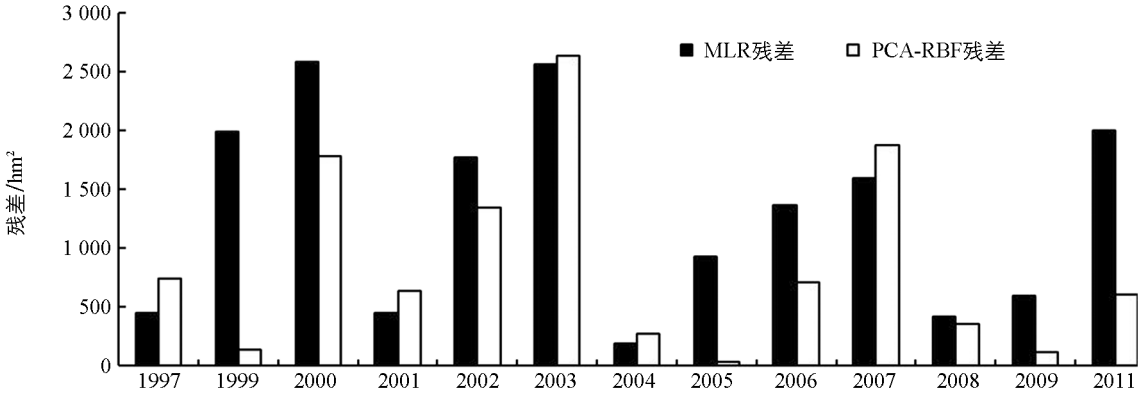

图 3　预测残差对比

3　结论与讨论

1）在成都市建设用地需求预测中，PCA-RBF 神经网络模型预测结果较常规的多元线性回归（MLR）模型相比，平均绝对误差（*MAE*）、误差均方根（*RMSE*）和平均相对误差（*MRE*）分别提高了 50.41%，32.19%和 40.00%且分年度残差和突变点残差更小，精度更高，说明 PCA-RBF 神经网络模型是建设用地需求预测的有效方法之一.

2）以往建设用地需求预测研究，大多是基于平稳数据，而本文的数据在 2008－2009 年出现了一次明显的“突变”性变化，这大大提高了准确预测的难度. 通过前期反复试验发现，如果按照常规研究直接预测面积，预测结果皆不尽人意. 针对此，本文提出了将直接预测面积转换为间接预测面积变化率再反算面积的一种新思路，这为基于“突变”数据条件下开展用地预测研究奠定了基础.

3）前人研究已经发现神经网络模型因其具有良好的非线性映射逼近性能在预测精度方面相比线性回归模型更具优越性；另外，就预测精度而言，神经网络模型中大小依次为：PCA-RBF 神经网络模型，RBF 神经网络模型，BP 神经网络模型；线性回归模型中大小依次为：多元线性回归模型，一元线性回归模型[12，19]. 但在数据具有“突变”特征前提下，上述结果是否依旧适用，还有待研究. 本文首次尝试将神经网络模型中预测精度相对最佳的 PCA-RBF 神经网络模型应用于建设用地需求预测研究，并以常规多元线性回归模型为对照，研究发现基于“突变”数据条件下，PCA-RBF 神经网络模型较之多元线性回归模型在整体和“突变”点预测准度上仍具优越性，有效弥补了相关研究的空白.

4）本文研究仅是建设用地需求预测技术方法层面的一种尝试，虽然基于“突变”数据条件下，PCA-RBF 神经网络模型在成都市建设用地需求预测中获得了较满意的应用效果，但这仅是以成都市为例进行的分析，研究结果是够具有普适性还有待进一步探讨.

参考文献：

[1]　葛春叶，涂建军，李　宽，等. 重庆市建设用地变化及其驱动力分析 [J]. 西南大学学报：自然科学版，2008，30(10)：131－135.

[2]　谈明洪，李秀彬，吕昌河. 20 世纪 90 年代中国大中城市建设用地扩张及其对耕地的占用 [J]. 中国科学 D 辑：地球科学，2004，34(12)：1157－1165.

[3]　赵　可，张安录，李　平. 城市建设用地扩张的驱动力——基于省际面板数据的分析 [J]. 自然资源学报，2011，26(8)：1323－1332.

[4]　付　倩. ARIMA 模型在西安卫星城建设预测中的应用 [J]. 中国城市经济，2011，(1)：127－129.

[5]　景亚平，张　鑫，罗　艳. 基于修正组合模型的青海省城市需水量预测 [J]. 自然资源学报，2012，27(6)：1013－1021.

[6]　彭保发，胡曰利，吴远芬，等. 基于灰色系统模型的城乡建设用地规模预测——以常德市鼎城区为例 [J]. 经济地理，2007，27(6)：999－1002.

[7]　陈国建. 刁承泰，黄明星，等. 重庆市区城市建设用地预测研究 [J]. 长江流域资源与环境，2002，11(5)：403－508.

[8]　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.

[9] 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，2007，14(5)：245－246，250.
[10] 基于灰色预测和神经网络的城市建设用地量预测[J]. 测绘信息与工程，2003，28(6)：36－39.
[11] 柴 杰，江青茵，曹志凯. RBF神经网络的函数逼近能力及其算法[J]. 模式识别与人工智能，2002，15(3)：310－316.
[12] 陈昌华，谭 俊，尹健康，等. 基于PCA-RBF神经网络的烟田土壤水分预测[J]. 农业工程学报，2010，26(8)：85－90.
[13] 程叶青，邓吉祥. 吉林省中部粮食主产区城乡系统耦合的关联分析[J]. 经济地理，2012，32(1)：127－145.
[14] ALP M，CIGIZOGLU H K. Suspended Sediment Load Simulationby Two Artificial Neural Network Methods Usinghydrometeorological Data[J]. Environmental Modelling and Software，2007，22：2－13.
[15] 吴 明，姚 尧，贾冯睿，等. 基于径向基函数神经网络方法的城市生态压力预测[J]. 自然资源学报，2013，28(2)：328－335.
[16] 李启权，王昌全，张文江，等. 基于神经网络模型和地统计学方法的土壤养分空间分布预测[J]. 应用生态学报，2013，24(2)：459－466.
[17] 彭文甫，何政伟，周介铭，等. 1996－2002年成都市土地利用变化分析[J]. 四川师范大学学报：自然科学版，2007，30(1)：106－111.
[18] 陈金梁，洪惠坤，陈 辉，等. 城市建设用地扩展研究——以重庆市渝北区为例[J]. 西南师范大学学报：自然科学版，2007，32(2)：140－144.
[19] 张晓瑞，方创琳，王振波，等. 基于RBF神经网络的城市建成区面积预测研究——兼与BP神经网络和线性回归对比分析[J]. 长江流域资源与环境，2013，22(6)：691－697.

Construction Land Demand Forecasting of Chengdu City Based on PCA-RBF Neural Network

XIE Ting[1]， DENG Liang-ji[1]， GAO Xue-song[1]，
LI Qi-quan[1]， WU Wen[2]， XU An-qi[1]， XIE Xin[1]

1. *College of Resources and Environment，Sichuan Agricultural University，Chengdu 611130，China；*
2. *Chengdu Land Planning and Cadastral Centre，Chengdu 610072，China*

Abstract：This paper proposes a construction land forecasting method with the neural network model based on principal component analysis (PCA) and radial basis function (RBF)，aiming at building a forecasting model under a "mutation" growth condition of construction land，so that sustainable utilization of regional land resources and regular modification of general land use planning can be improved. The paper consists of three parts. Firstly，major driving factors were picked，using grey relational analysis，to be the import data. Then，principal component analysis was adopted to eliminate the relevance of the driving factors and to reduce the dimensionality. Finally，PCA results were used as the input to establish an RBF neural network mode to predict construction land demand. The case study shows that PCA-RBF neural network prediction model for construction land demand can give accurate prediction results. The mean absolute error (*MAE*)，root mean square error (*RMSE*)，mean relative error (*MRE*)，sub-annual residuals and point mutations residuals are smaller than those of the conventional multiple linear regression (MLR) model and are more suitable for complex nonlinear conditions.

Key words：construction land；forecast ；principal component analysis；radial basis function neural network model；multiple linear regression model

责任编辑 陈绍兰