

文章编号: 1000-5013(2012)04-0446-05

# 厦门市工程估价的 RBF 神经网络预测模型

叶青

(华侨大学 土木工程学院, 福建 厦门 361021)

**摘要:** 选取 55 个厦门市典型工程造价指标, 利用 SPSS 统计分析软件对工程特征和训练样本进行相关性分析、归类合并, 得出 11 个工程特征作为平米造价的主要影响因素. 以径向基函数(RBF)神经网络原理为基础, 建立工程造价估算模型, 通过试验, 选择  $\text{net} = \text{newrb}(\mathbf{P}, \mathbf{T}, 0.01, 1.0)$  建立 RBF 网络, 用  $y = \text{sim}(\text{net1}, \mathbf{P})$  对样本进行训练测试. 实证分析结果显示: 该模型具有计算快捷简便的优势, 估算误差在允许范围内, 可用于实际工程造价的辅助估算.

**关键词:** 工程估价; 预测模型; 径向基函数; 人工神经网络; 厦门市

**中图分类号:** TU 723.3; TP 183

**文献标志码:** A

人工神经网络由于具有良好的自学能力和容错能力, 成为工程造价预测主要方法之一. 基于神经网络的工程造价估算方法的主要限制, 在于工程特征和训练样本的选取上. 为确保估算模型的准确度, 要求做到选取的工程特征能够反映工程本质, 便于从众多样本中区分开, 同时选取的训练样本要和当前工程类似, 才可能使估算模型为问题求解提供支持. 然而这两个方面的选取工作, 目前主要是凭借经验来完成, 缺乏理论的指导, 容易造成个别输出目标值偏离实际值. 1988 年, Moody 等<sup>[1-2]</sup> 首先将 RBF 应用于人工神经网络设计, 构成了 RBF 神经网络. BP 神经网络与 RBF 神经网络的主要功能都是实现非线性映射, 但在对生物神经系统的模拟、网络结构与表达能力、训练算法、逼近方式等方面存在区别, 使得 RBF 神经网络在表达能力、学习速度、泛化性能上具有更大的优越性<sup>[3-4]</sup>. 径向基函数为局部分布的、中心径向对称的、非负衰减的非线性函数, 它的两个参数——基中心和基宽度, 将确定对输入产生显著响应的范围. RBF 网络对非线性函数进行逼近是切实可行的, 在实际应用中为使用径向基函数设计神经网络提供了理论基础<sup>[5-6]</sup>. 本文以径向基函数神经网络原理为基础, 建立工程造价估算模型, 并基于 SPSS 统计分析软件对工程特征和训练样本进行选取.

## 1 基于 SPSS 的工程造价影响因素分析

### 1.1 数据筛选

根据厦门市造价指标的工程特征描述, 提取 19 个因素进行案例的汇总, 共收集 2007 年 3 月至 2009 年 11 月的 70 多个案例, 但因为有些案例数据不全, 有的只有地下工程, 没有地上工程, 有的缺少桩基础的造价, 最后选取 58 个案例汇编形成案例库. 在进行因子的相关性分析之前, 对所选变量的分布情况进行检验, 在资料数据集合上, 基于不同建筑物类型的进行了数据筛选. 其目的在于鉴别基于建筑物不同类型的变量分布情况的重要偏离量. 上述过程借助 SPSS 软件的箱型图来进行, 发现项目编号为 20, 33, 41 的项目被标出为异常, 经分析决定, 剔除这些样本数据. 此时样本容量为 55 个, 这 55 个数据全部用来进行相关性分析.

### 1.2 因素分析

选用 SPSS 软件的因子分析工具, 输入 19 个因素的数据资料进行相关性分析, 得到 19 个因素的相

**收稿日期:** 2011-11-09

**通信作者:** 叶青(1968-), 女, 教授, 主要从事工程造价和房地产价格评估的研究. E-mail: yeqing@hqu.edu.cn.

**基金项目:** 中央高校基本科研业务费专项资金资助项目(JB-ZR1162); 福建省泉州市科技计划项目(2009Z52); 华侨大学高层次人才科研启动项目(12BS131)

关矩阵,其中装修因素间的相关矩阵如表 1 所示.因平米造价是要预测的因子,与它的相关性越大越好,其他 18 个因素之间的相关性不能太大.混凝土与地上层数有较大的相关性(0.637),故删除.砌筑与内墙面(0.666)、门窗(0.588)、楼地面(0.528)有较大的相关性,内墙面、外墙面、门窗、楼地面、天棚之间也有较大的相关性,根据因素的特征,将这 5 个因素合并为装修因素,而砌筑因素仍保留.建筑面积与地下层数、地上层数、工程类别有较大的相关性,也删除.工期及造价指数与平米造价(0.082,0.105)的相关性较小,考虑这 2 个因素的关系,将它们相乘合并为造价指数及工期 1 个因素.地下层数与地上层数、地上层数与工程用途也有较大的相关性,分别为 0.681,0.532,考虑它们对工程造价的影响,故保留.

剩余 12 个因素的相关矩阵,如表 2 所示.除了地下层数与地上层数(0.681)、地上层数与工程用途(0.532)、砌筑与装修(0.734)有较大的相关性以外,其他的相关性都在合理的范围.最终确定工程用途  $I_1$ 、工程类别  $I_2$ 、结构类型  $I_3$ 、地下层数  $I_4$ 、地上层数  $I_5$ 、基础  $I_6$ 、桩基础  $I_7$ 、砌筑  $I_8$ 、屋面防水保温  $I_9$ 、装修  $I_{10}$ 、造价指数及工期  $I_{11}$  这 11 个因素作为平米造价  $O_1$  的影响因素.

表 1 装修因素间的相关矩阵  
Tab.1 Correlation matrix among the factors of decoration

因素	内墙面	外墙面	门窗	楼地面	天棚
内墙面	1.000	0.404	0.461	0.507	0.475
外墙面	0.404	1.000	0.371	0.215	0.316
门窗	0.461	0.371	1.000	0.292	0.429
楼地面	0.507	0.215	0.292	1.000	0.392
天棚	0.475	0.316	0.429	0.392	1.000

表 2 12 个因素的相关矩阵  
Tab.2 Correlation matrix of 12 factors

因素	$O_1$	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$I_9$	$I_{10}$	$I_{11}$
$O_1$	1.000	0.520	0.074	-0.014	0.081	0.242	0.542	0.453	0.517	0.278	0.670	0.122
$I_1$	0.520	1.000	0.043	-0.051	0.235	0.532	0.173	0.198	0.327	-0.009	0.446	-0.114
$I_2$	0.074	0.043	1.000	-0.223	-0.477	-0.405	-0.290	-0.121	0.396	-0.262	0.368	0.145
$I_3$	-0.014	-0.051	-0.223	1.000	0.149	0.118	0.097	-0.035	-0.241	0.334	-0.190	-0.026
$I_4$	0.081	0.235	-0.477	0.149	1.000	0.681	0.411	0.009	-0.320	0.081	-0.213	-0.056
$I_5$	0.242	0.532	-0.405	0.118	0.681	1.000	0.401	0.102	-0.173	-0.127	-0.075	-0.056
$I_6$	0.542	0.173	-0.290	0.097	0.411	0.401	1.000	0.094	-0.032	0.251	0.148	0.023
$I_7$	0.453	0.198	-0.121	-0.035	0.009	0.102	0.094	1.000	0.249	-0.110	0.285	0.070
$I_8$	0.517	0.327	0.396	-0.241	-0.320	-0.173	-0.032	0.249	1.000	0.081	0.734	-0.003
$I_9$	0.278	-0.009	-0.262	0.334	0.081	-0.127	0.251	-0.110	0.081	1.000	0.074	-0.190
$I_{10}$	0.670	0.446	0.368	-0.190	-0.213	-0.075	0.148	0.285	0.734	0.074	1.000	-0.144
$I_{11}$	0.122	-0.114	0.145	-0.026	-0.056	-0.056	0.023	0.070	-0.003	-0.190	-0.144	1.000

## 2 工程估价模型的建立<sup>[7]</sup>

### 2.1 MATLAB 神经网络工具

径向基函数网络的训练可以用两个函数来实现:newrbe()或 newrb(). 函数 solverbe()用来设计径向基函数网络时,可以达到零误差,其调用格式为 net=newrbe ( $\boldsymbol{P},\boldsymbol{T}$ ). 其中: $\boldsymbol{P}$  和  $\boldsymbol{T}$  分别为输入样本向量、输出目标向量. 函数 newrbe()每次生成一个径向基函数网络的神经元,并如此反复不断地增加神经元个数,即可达到误差指标或最大的训练步数. 用 newrb()函数可以求出合适的网络权值和阈值,使得所设计的径向基函数网络精确地输出目标向量  $\boldsymbol{T}$ . 文中选用函数 newrb()对径向基函数神经网络模型进行训练,在 MATLAB 环境中运用 net=newrb( $\boldsymbol{P},\boldsymbol{T}$ ,goal,sp,mn,df)建立网络. 其中:goal 表示目标误差;sp 表示分布系数;mn 表示神经元个数上限值;df 表示训练过程显示的频率. 参数集合未被指明时,自动取缺省值.

### 2.2 模型结构

1) 数据库模块. 基于径向基函数神经网络的投资估算模型的建立,需要通过大量的样本进行训练.

这些收集到的原始样本数据,内容包括已建项目的具体工程的工程特征、造价及主要材料用量.训练样本数量越多,样本的类似结构越多,样本的时间跨度越小,建立模型进行测算的精确度就越高.

2) 神经元计算模型.采用径向基函数神经网络,径向基函数神经网络的学习、模拟过程可直接使用 MATLAB 语言提供的神经网络工具箱来实现.

3) 径向基函数神经网络知识库与自学习模块.模型设计的核心是神经网络知识库的构建及自学习功能的设计.径向基函数神经网络知识库模块主要包括造价资料集、反映知识规则的权重集和阈值,权重和阈值由 MATLAB 神经网络工具箱中的函数 newrb()完成.

4) 人-机对话功能模块. a) 输入模块.神经网络输入(即工程特征向量)对模块的准确度起着至关重要的作用,选择不当可能导致模型的失败.在确定网络输入变量之后,需进行尺度变换和预处理,把输入数据归一化到能使网络所有权值都在一个不太大的范围之内,以减轻网络训练时的难度. b) 输入处理模块.将神经网络的输入矩阵( $N$  列  $P$  行)数值转化为 $[-1,1]$ 之间的数,以适应神经网络的处理要求. c) 解释模块和输出模块.将神经网络的输出值转化为实际值,得出单方造价、主要材料用量输出,供用户使用.

2.3 模型的训练和测试

建立工程造价估算样本库,将样本集随机地分为训练样本和测试样本.在对训练样本进行充分学习后,即可确定神经网络的结构参数,通过测试样本检测模型的估算功能.在估算模型投入应用之前,要对模型的泛化能力进行验证.泛化能力是表征网络学以致用的程度,是评定网络学习好坏的重要指标.训练样本数目太少或不具有代表性,工程特征向量选取不足以反映工程主要成本的构成,由不能避免的不可量化、不可预测引起的内在噪声等均可能导致一个网络的泛化能力差<sup>[7]</sup>.

2.4 工程特征向量的量化处理

将影响工程平米造价的 11 个主要因素  $I_1 \sim I_{11}$  作为神经网络的输入向量,平米造价  $O_1$  作为神经网络的输出向量.将收集到的 55 个样本分为两组,任选 50 个样本作为训练样本,其余 5 个样本作为测试样本.将神经网络输入和输出向量进行量化,量化标准和量化结果分别如表 3、4 所示<sup>[3]</sup>.

表 3 工程特征向量的量化标准  
Tab. 3 Quantitative criteria of engineering eigenvectors

工程特征	量化值				
	1	2	3	4	5
工程用途 $I_1$	厂房仓储	综合办公楼	宿舍教学楼	商住楼	
工程类别 $I_2$	1 类	2 类	3 类		
结构类型 $I_3$	框架结构	框剪结构	钢结构等		
地下层数 $I_4$			按实际数据输入		
地上层数 $I_5$			按实际数据输入		
基础 $I_6$	独立基础+带形基础	独立桩承台基础	独立桩承台+满堂基础	满堂基础	
桩基础 $I_7$	预制(预应力)钢筋砼管桩 1	人工挖孔桩 1、沉管灌注砼桩	预应力高强砼管桩 2、人工挖孔桩 2	冲钻孔灌注砼桩	预应力高强砼管桩 3
砌筑 $I_8$	加气混凝土砌块墙厚度小于 120 mm,多孔砖墙 90 mm	加气混凝土砌块墙厚度大于 120 mm,多孔砖墙 190 mm	多孔砖墙大于 190 mm	多孔砖墙+空心砖墙水泥空心砌块	
屋面防水保温 $I_9$	保温玻璃纤维或挤塑聚苯板	高聚物改性沥青防水卷材+挤塑聚苯板保温层	保温层高分子防水卷材+挤塑聚苯板保温层	合成高分子防水涂膜+防水卷材型复合保温夹芯压型板	高分子防水卷材+水泥基聚合物弹性防水涂料+干铺无纺聚酯纤维布一层
装修 $I_{10}$	差	较差	一般	较好	好
造价指数及工期 $I_{11}$	按厦门市造价管理站颁布的各种类型建筑的造价指数的均值乘以工期提前导致的造价增加系数.经实例测算,工期正常取 1,提前 10%~20%取 1.009,提前 20%~30%取 1.015,提前大于 30%取 1.022				
平米造价 $O_1$	按实际数据输入				

2.5 工程特征向量的初始化处理

随机选取编号为 A9,A10,A23,A27,A49 的 5 个案例作为测试样本,剩余 50 个作为训练样本. 根据神经网络的特点,对神经网络的输入、输出值进行归一化处理,使输入向量均匀,易收敛,初始化数据样本,神经网络的输出值通过公式再转化为实际值,这些过程均使用 MATLAB 编程实现.

表 4 训练样本库原始数据  
Tab.4 Data of train samples

影响因素	案例编号				
	A1	A2	...	A57	A58
工程用途 $I_1$	2	2	...	2	1
工程类别 $I_2$	3	3	...	3	3
结构类型 $I_3$	1	1	...	1	1
地下层数 $I_4$	0	1	...	0	0
地上层数 $I_5$	5	9	...	8	4
基础 $I_6$	3	3	...	1	1
桩基础 $I_7$	0	2	...	0	0
砌筑 $I_8$	3	1	...	2	1
屋面防水保温 $I_9$	1	3	...	2	2
装修 $I_{10}$	3	2	...	2	2
造价指数及工期 $I_{11}$	1.132	1.083	...	1.109	1.109
平米造价 $O_1$	1057.79	878.55	...	808.24	925.72

2.6 径向基函数神经网络的训练和测试

采用 MATLAB 神经网络工具箱中的径向基函数神经网络命令进行模型训练与测试,其测试结果如表 5 所示,运行结果如图 1 所示.

根据表 5,选择目标误差 goal=0.01,分布系数 sp=1.0,net=newrb( $\boldsymbol{P},\boldsymbol{T},0.01,1.0$ )建立 RBF 网络,用  $y=sim(net1,\boldsymbol{P})$ 对样本进行训练测试,结果如表 6 所示. 表 6 中: $\eta$ 为相对误差.

由表 6 可知,采用基于 RBF 网络的工程估价模型对测试样本进行造价模拟,输出结果与实际值的相对最大误差的绝对值为 12.33%,在投资估算的误差允许范围内. 说明该模型可用于实际工程造价的辅助估算.

表 5 测试样本在不同分布系数下的相对误差  
Tab.5 Relative error of test samples in the different distribution coefficient %

样本编号	goal=0			goal=0.01		
	sp=0.5	sp=1.0	sp=2.0	sp=0.5	sp=1.0	sp=2.0
A9	0.05	-17.78	12.84	1.95	-0.25	-1.90
A10	-15.10	-275.81	24.26	-10.90	1.92	-5.41
A23	-4.33	-127.86	90.60	-1.63	0.59	-5.23
A27	-12.40	-450.31	174.89	-7.97	3.29	-17.88
A49	11.45	-23.94	23.71	12.48	12.33	7.84

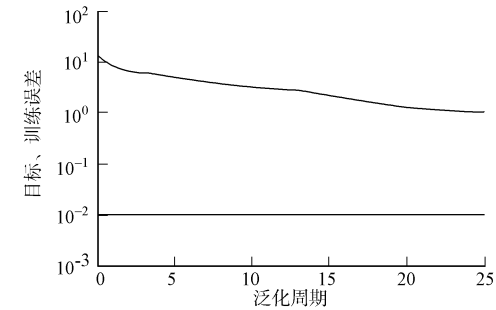


图 1 误差曲线  
Fig.1 Error curve

表 6 测试样本的模拟结果和实际值的比较  
Tab.6 Comparison between the simulation results and actual values in test samples

样本编号	模拟值	实际值	$\eta/\%$
A9	1 083.8	1 086.51	-0.25
A10	1 279.5	1 255.41	1.92
A23	1 138.1	1 131.43	0.59
A27	1 252.4	1 212.49	3.29
A49	1 096.0	975.64	12.33

3 结 论

1) 选取 55 个厦门市典型工程造价指标,基于 SPSS 的相关性分析,归类合并,删除相关性大的因素,得出 11 个工程特征作为平米造价的主要影响因素. 与文献[3,7]相比,主要影响因素的选取更有说服力,解决了由于相关性大可能导致网络的泛化能力差的问题,同时考虑造价指数及工期对造价的影响,增加了模型的准确性和实用性.

2) 通过试验,选择  $\text{net}=\text{newrb}(\mathbf{P},\mathbf{T},0.01,1.0)$  建立 RBF 网络,用  $y=\text{sim}(\text{net1},\mathbf{P})$  对样本进行训练测试,实证分析结果显示该模型具有计算快捷简便的优势,估算误差在允许范围内,可用于实际工程造价的辅助估算.

3) 训练样本的数量和质量直接影响网络泛化能力,采用数量多、质量好的样本数据训练出来的网络,其泛化能力必然比较好. 进一步的研究可收集更多的厦门市典型工程造价指标,通过数据筛选,形成高质量的样本库,提高基于 RBF 神经网络的工程估价模型的适用性.

参考文献:

[1] MOODY J,DARKEN C. Learning with localized receptive fields[C] // Proc Connectionist Models Summer School. San Mateo:Morgan Kaufmann,1998:133-143.

[2] MOODY J,DARKEN C. Fast learning in networks of locally-tuned processing units[J]. Neural Computation,1989,1(2):281-294.

[3] 叶青,王全凤. 基于 BP 神经网络的工程估价模型及其应用[J]. 厦门大学学报:自然科学版,2008,47(6):47-51.

[4] 张新红,雷素娟. 用 QCEA 优化的 RBF 神经网络及其在股市预测的应用[J]. 华侨大学大学学报:自然科学版,2011,32(3):338-342.

[5] CHEN Tian-ping,CHEN Hong. Approximation theory capability to functions of several variables, nonlinear functions, and operators by radial basis functional neural networks[J]. IEEE Trans on Neural Networks,1995,6(4):904-910.

[6] 田景文,高美娟. 人工神经网络算法研究及应用[M]. 北京:北京理工大学出版社,2006.

[7] 段晓牧. 基于 RBF 神经网络的非确定性工程投资估算新方法的研究[D]. 阜新:辽宁工程技术大学,2003.

Prediction Model of the Project Cost Estimation  
Based on RBF Neural Network

YE Qing

(College of Civil Engineering, Huaqiao University, Xiamen 361021, China)

**Abstract:** 55 typical engineering cost indexes in Xiamen city were selected and analyzed, to obtain the correlation of engineering features and training samples by SPSS statistical analysis software. Based on radial basis function (RBF) neural network theory, the project cost estimation model is established. Selecting the  $\text{net}=\text{newrb}(\mathbf{P},\mathbf{T},0.01,1.0)$  to establish RBF network through the test, using  $y=\text{sim}(\text{net1},\mathbf{P})$  to train and test, the calculation results show that the model has the advantage of convenient calculation, the estimation error is small and allowable, and the model is worth in the project prediction estimation.

**Keywords:** project cost estimation; prediction model; radial basis function; artificial neural network; Xiamen city

(责任编辑: 陈志贤      英文审校: 方德平)