

# 基于 PCA 与 NARX 的市政工程造价组合预测

张晓东<sup>1</sup>, 杨圣祥<sup>2</sup>

(1. 江苏开放大学, 江苏 南京 210024; 2. 英国德蒙福特大学, 英国莱斯特)

**摘要:** 针对工程造价影响因素多, 预测困难且周期较长等问题, 通过将主成分分析和 NARX 神经网络相结合, 提出一种新型基于混合算法的市政工程造价预测方法。利用主成分分析对影响市政工程造价的主要因素进行原始数据处理, 消除其相关性, 可有效降低数据冗余, 也降低神经网络运算易出现局部极小点概率。以主成分分析数据作为输入, 单位面积工程造价作为输出, 采用贝叶斯正则化算法构建的 NARX 神经网络模型进行市政工程造价预测。算例结果表明, 基于 PCA 与 NARX 进行市政工程造价快速、准确, 证明预测有效可行。

**关键词:** 工程造价预测; 主成分分析; 神经网络

## Forecasting the Cost of Municipal Engineering Based on PCA and NARX Neural Network

ZHANG Xiao-dong<sup>1</sup> YANG Sheng-xiang<sup>2</sup>

(1. Jiangsu Open University, Nanjing 210024, China; 2. De Montfort University, Leicester LE1 9BH, UK)

**Abstract:** There are many factors influencing the cost of municipal engineering. It is usually difficult and time consuming to predict the cost of municipal engineering. This paper proposes a two-phase method to predict the cost of municipal engineering by combining principal component analysis (PCA) and NARX neural network. In the first phase of the proposed method, PCA is used to analyze the correlations among involved factors based on the original data and extract key factors that influence the cost of municipal engineering. In the second phase, taking the key factors extracted by PCA as input and the unit area engineering cost as the output, a NARX neural network model is constructed based on the Bayesian regularization algorithm to predict the cost of municipal engineering. The experimental results show that the proposed method can predict the cost of municipal engineering fast and accurately, which shows that the proposed prediction method is feasible and effective.

**Key words:** engineering cost forecasting; principal component analysis; NARX neural network

建筑业信息化是指运用计算机、通信、控制、网络、系统集成和信息安全等技术, 改造和提升建筑业技术、生产、管理和服务水平。建筑业信息化要求工程造价信息化, 同时工程造价信息化也是工程造价管理自身发展需求。工程造价信息化指运用信息技术和网络技术形成信息资源, 并将之运用到建设项目的决策与预测中, 提升准确性与效率, 使工程造价管理工作适应发展需要<sup>[1]</sup>。工程造价信息化建设应在确定工程造价信息化目标与相关标准的

基础上, 建设满足各方使用者需求的信息平台<sup>[2]</sup>; 核心工作是着力建设工程造价信息两大数据库, 即造价信息数据库及应用于预测的计算方法数据库。

严格意义的工程造价预测对应设计之前的“工程估算”, 进入到设计阶段的“施工图预算”为广大工程造价人员所熟知与掌握, 相对来说精度较高, 但工作量繁重且效率低下, 又有滞后性。我国传统工程估算的方法有: 单位生产能力投资估算法、生产能力指数估算法、

比例估算法、郎格系数法、工程建筑概算指标估算法。这些方法可以满足工程估算 30% 误差的要求，但是不能满足日益发展的精细化要求。鉴于人工智能的广泛应用，必须建立适合智能计算的科学方法应用数据库，提高预测精度与速度。

### 1. 工程造价预测相关研究

为提高工程造价预测的精度与科学性，相关学者进行了大量研究取得了丰富的成果。在运用线性回归预测方面，秦中伏等<sup>[3]</sup>选取支持向量机构建住宅工程造价预测模型，并通过主成分分析对原始数据进行降噪处理；胡晓娟<sup>[4]</sup>提出采用结构整体最小二乘法估计模型参数解决部分元素受误差影响的问题。在运用灰色预测方面，胡伟勋<sup>[5]</sup>建立工程造价的 GM(1,1)、GM(1,N) 预测模型；王佼，刘艳春<sup>[6]</sup>采用 PSO 优化的 SVR 模型进行工程造价预测，取得了更好的预测精度。在运用模糊数学方面，姜德华等<sup>[7]</sup>较早将模糊数学运用到水电工程造价估算中；于靖等<sup>[8]</sup>根据择近原则原理和已建工程的造价，动态准确地计算出拟建工程的造价。在神经网络预测方面，成果更加丰富，叶青等<sup>[9]</sup>对神经网络进行优化并结合工程实例进行检验；任宏，周其明<sup>[10]</sup>研究运用神经网络估算主要工程量。其他一些研究比较零星，如实例推演估算住宅建筑主要工料消耗<sup>[11]</sup>、时间序列预测工程造价指数<sup>[12]</sup>、运用 Bayes 估算方法，将时间因素和决策者的经验、主观判断等相结合建模预测<sup>[13]</sup>等。国外在运用地理信息系统、联机分析处理及交叉验证方面进行更为前沿的研究。

两种方法同时运用解决单一方法不足方面主要有：模糊理论与神经网络方面，梁喜和刘雨<sup>[14]</sup>利用直觉模糊分析方法找出主要工程特征，作为输入向量构建预测模型；王建平<sup>[15]</sup>针对 BP 网络收敛速度慢，在权值和阈值的修正算法中加入学习率和惯性因子两个参数。杨永明等将灰关联与神经网络相结合对电力工程造价进行预测。潘雨红等<sup>[16]</sup>将遗传算法与人工神经网络模型相结合，将人工智能算法 GA-BP 模型引入公路工程估价估算中。李永明

等<sup>[17]</sup>将数据挖掘与神经网络相结合，应用数据挖掘技术进行数据预处理结合改进的神经网络模糊系统进行电力工程造价的预测。

### 2. 工程造价预测问题

大数据时代的来临，BIM 的诞生和普及应用，云造价的面世和普及，能够从根本上解决了工程造价行业存在的“数据不连续、数据积累难、协同共享难”等问题，也为工程造价预测提供足够的技术支撑。

#### 2.1 搭建工程造价信息化平台

工程造价信息平台是工程造价信息系统建设的具体呈现，面向不同用户需求的工程造价信息平台应立足使用、结构合理、功能完善（见图 1）。平台会让数据获取方便，也让协同作业成为可能，可进行工程造价预测与决策<sup>[18]</sup>。

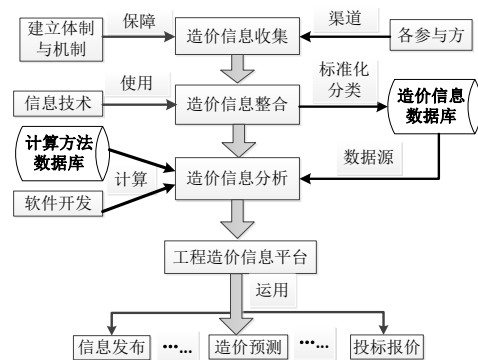


图 1 工程造价预测的技术路线

#### 2.2 工程造价预测

工程造价的费用简单说是分部分项工程量乘以对应单价得出，工程建设规模、工程特征的差异与功能定位是工程的本质区别，决定了工程量的类型与数量不同，同时价格会因为社会发展与市场波动呈现变化。

工程造价预测是根据拟建工程建设规模与工程特征，选取已完工程数据库中的相似工程，运用方法数据库中适合的科学方法，结合专业人员的主观经验和判断能力，对拟建工程造价进行预测。

建设项目的唯一性决定没有完全相同的建设项目，如何从已完工程中选择与拟建工程相似工程成为首先要解决的问题，模型识别

中,特征提取能有效缓解模型识别领域经常出现的维数灾难问题。项目特征是项目区别于其它项目的本质特征,也是影响工程造价的主要要素,这些影响工程造价方面的项目特征之间存在复杂的耦合关系且存在信息冗余,必须进行数据处理,找到问题主要方面。

价格随时间波动变化决定了不能简单从数据库中调用数据直接运用,要考虑数据的历史维度特性,考虑时间序列影响。

本研究探索运用主成分分析(PCA)找到影响工程造价的主要影响因子,即公共因子,消除原始数据的相关性。结合 NARX (Non-linear Auto-Regressive with Exogenous inputs) 动态神经网络可以反映系统历史状态信息,具有记忆功能的特性,将主要影响因子作为神经网络输入,单位面积工程造价作为输出,进行市政工程工程造价预测。

### 3. PCA 与 NARX 的组合预测

#### 3.1 建设项目特征的主成分分析

PCA 是模型识别中常用的面向模式分类的特征提取方法,其主要思想是通过线性空间变换求取主元变量,将高维数据空间投影到低维主元空间,且保留原始数据的大部分方差信息,使问题得到最佳综合<sup>[19]</sup>。

设数据矩阵  $X \in R^{n \times m}$  ( $n$  为样本数,  $m$  为变量数), PCA 模型可表示为

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T \quad (1)$$

式中:  $t_i$  为得分向量(主元);  $p_i$  为负荷向量。

主成分分析步骤为: 1、构成变量采样的原始数据  $X_{n \times p}$  并标准化处理; 2、计算相关系数矩阵  $R$  并得出特征值与特征向量; 3、以特征值  $\geq 1$ , 累计贡献率  $\geq 80\%$  选取主成分个数; 4、计算主成分得分。

#### 3.2 NARX 神经网络

神经网络有很好的自适应性,预测性能随着样本的增加而加强,其实质是通过大量样本学习训练来自我调整连接权值,能模拟任意的非线性输入和输出。神经网络预测方法的缺点在于训练效果对网络初始权值的敏感依赖性,网络初始化不够理想往往会使训练陷入局部

最优<sup>[20]</sup>。

NARX 网络称为有外部输入的非线性自回归模型,是一种有记忆功能的神经网络,可以看作有时延输入的神经网络加上输出到输入的时延反馈连接<sup>[21]</sup>,对非线性数据有着更好适应能力,其模型表达式见图 2,非常适合工程造价复杂、非线性时间序列预测要求的特征。图 2 中参数  $d$  为时延阶数,  $m$  为隐含层神经元个数,  $w$  为权值向量,  $b$  为偏置值,  $f$  为神经网络隐含层激活函数。其模型表达式为:

$$Y(t) = f[y(t-1), y(t-2), \dots, y(t-y_n), u(t-1), u(t-2), \dots, u(t-u_n)] \quad (2)$$

可以看出下一个  $y(t)$  取决于上一个  $y(t)$  和上一个  $u(t)$ ,  $f(\cdot)$  表示过程函数。

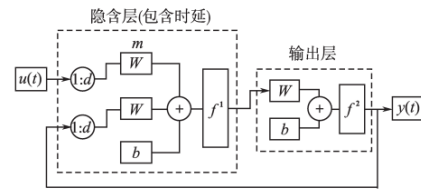


图 2 NARX 神经网络模型结构图

### 4. 实例分析

笔者从江苏省工程造价信息网典型工程造价指标库中,按时间先后顺序,以每个季度为序列,选取某市从 2010-2016 年共 28 个典型市政工程工程造价指标。指标划分以市政道路工程经济指标采集为基础,根据项目特征的显性、直观、易于统计等方面考虑,最后选取道路面积、工程类别、道路土石方量、道路基础厚度、道路面层厚度,基础处理深度、侧缘石长度、侧缘石材质(1 为预制砼,2 为花岗岩)、桥梁经济占比、雨水管道长度、雨水管道平均管径、污水管道长度、污水管道平均管径等 13 个影响工程造价主要因素。

#### 4.1 数据描述

根据所收集的 28 个典型工程造价指标,具体数据如表 1 所示。

#### 4.2 主成分分析

本研究利用 SPSS19 软件的因子分析功能,对表 1 中 28 个典型工程样本与 13 个影响因子,构建原始数据样本  $X_{28 \times 13}$  进行因子分析。

表 1 市政工程造价指标数据样本

时间 (年季)	平米造价 (元/m <sup>2</sup> )	道路路面工程 积(m <sup>2</sup> )	道路土道 类别	路基路面 石方量(m <sup>3</sup> )	厚度 (mm)	厚度 (mm)	理深度 (cm)	长度 (m)	石质	材济占比	道长度 (m)	平均管径 (mm)	管道长度 (m)	平均管径 (mm)
201001	472.93	15100	3	16454	56	14	30	4567	1	0	1145	1200	0	0
201002	712.41	36377	2	151357	56	15	120	12353	2	0	2309	1350	1362	500
201003	568.43	12477	3	13256	52	11	100	1850	2	8.2	0	0	0	0
201004	699.95	64046	1	338672	58	13	120	13455	1	0	4284	1500	1673	500
201101	559.26	12355	3	0	96	10	0	3738	1	12.12	0	0	0	0
201102	752.82	64585	1	338687	58	13	120	13455	1	0	4284	1500	1673	500
201103	538.51	39176	2	0	56	12	60	16566	1	6.69	0	0	0	0
201104	648.34	14256	3	51312	55	11	40	1876	1	2.64	1356	1350	1174	600
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
201601	572.64	14453	3	19650	56	11	20	6305	2	15.48	1542	800	1418	400
201602	752.46	26880	3	52606	54	13	80	10188	2	0	1985	700	1326	400
201603	489.73	42947	2	67489	50	10	0	3021	2	13.4	1524	1200	1456	600
201604	632.54	17410	3	104285	52	11	0	5706	2	6.49	0	0	0	0

1、因子分析后相关检验结果如图 3，数据相关性检验中 KMO 值为 0.708，说明指标间相关性比较强；Bartlett 的球形度检验 Sig 值为 0.000，拒绝零假设，样本数满足要求。因此该样本数据进行主成分分析在技术上是可行的，13 个影响因子之间相关。

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。	0.708
Bartlett 的球形度检验 近似卡方	289.8062
df	78
Sig.	0.000

图 3 SPSS 影响因素数据相关性检验结果输出图

2、前 4 个因子的特征值大于 1，其累积贡献率达到 82.212%，超过 80%（见表 2），因此可确定提取的主因子数目为 4 个。

3、根据 SPSS19 软件分析，得出单位造价与主因子得分（归一化值）如表 3 所示。

### 4.3 NARX 神经网络预测

市政工程造价 NARX 神经网络预测构建基本思路：以主成分分析后所得到的主因子得分作为 NARX 神经网络输入值，单位工程造价作为输出值。

NARX 神经网络延迟阶数选择，阶数越大对复杂系统的建模法越准确，而神经元个

表 2 SPSS 因子分析解释总方差结果输出图（节选前 5）

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	5.356	41.200	41.200	5.356	41.200	41.200	4.358	33.526	33.526
2	2.384	18.340	59.539	2.384	18.340	59.539	2.797	21.516	55.042
3	1.774	13.649	73.188	1.774	13.649	73.188	2.275	17.501	72.543
4	1.173	9.023	82.212	1.173	9.023	82.212	1.257	9.669	82.212
5	0.763	5.866	88.077						

提取方法：主成份分析。

表 3 SPSS 因子分析解释单位造价与主因子得分表

时间 (年季)	主因子 1	主因子 2	主因子 3	主因子 4
201001	-0.23011	-1.02356	0.69834	-1.08087
201002	0.28099	0.58051	1.07643	0.53448
201003	-0.64567	-0.92024	-0.206	0.66053
201004	1.13366	0.71923	1.23908	-1.10932
.....	.....	.....	.....	.....
201601	-0.26082	0.58206	-1.5735	0.87228
201602	-0.28934	0.31909	0.36844	0.59377
201603	0.08281	1.10786	-1.7653	0.83854
201604	-0.34674	-0.9286	-0.8063	0.60764

数关系到函数关系能否被网络准确学习，因此需要反复调试。通过不断调试，确定延迟阶数为 4，神经元个数为 12。激活函数为贝叶斯正规化算法 (Bayesian regularization) [22]，性能函数为：

$$msereg = \alpha \cdot msw + \beta \cdot mse \quad (3)$$

$$mse = \frac{1}{n} \sum_{i=1}^n (t_i - p_i) \quad (4)$$

$$mse = \sum_{i=1}^m (\omega_j^2) \quad (5)$$

式中，mse 为均方误差，n 为样本总数， $t_i$  为期望输出， $p_i$  为网络实际输出，msw 为网络权值均方，m 为网络权值总数， $\omega_j$  为网络权值， $\alpha$  和  $\beta$  为正则化系数，其大小对网络训练效果影响显著，本数据训练运用 MATLABR2016a 软件进行训练。

1、NARX 神经网络的训练效果判断：

图 4 中除零时延以外大部分相关函数值落在 95% 的置信区间，图 5 中的拟合优度都接近 1，得出模型具有良好的泛化能力和预测性能，判断模型的网络特性较好，技术可行。

2、本文研究选择时延阶数为 4，因此起始时是用 2010 年 1-4 季度的数据预测 2011 年 1 季度的造价，依此类推到 2016 年 4 季度。

由图 6 可以看出，训练样本的输入与输出值误差比较小，根据表 4 的数据对比，可以看出，相对误差在 10% 以内，远远优于工程投资估算的误差 30% 的要求。

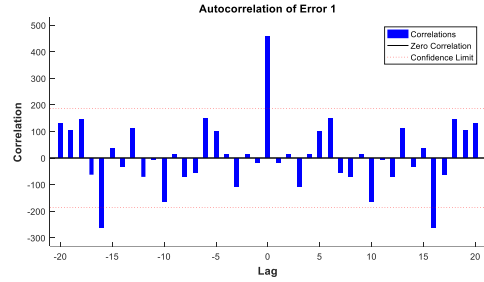


图 4 NARX 神经网络误差自相关图

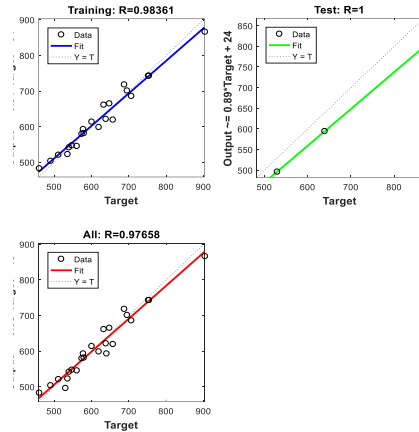


图 5 NARX 神经网络回归效果图

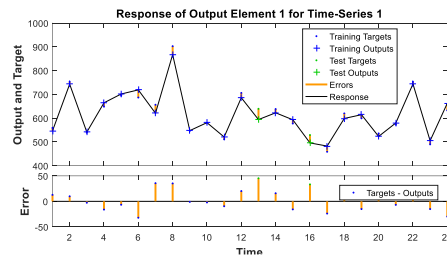


图 6 NARX 神经网络预测效果误差图

表 4 NARX 神经网络预测误差表

时间	实际值 (元/m <sup>2</sup> )	预测值 (元/m <sup>2</sup> )	绝对误差 (元/m <sup>2</sup> )	相关误差	相对误差 (%)
201101	559.26	546.48	-12.78	-0.0229	-2.29
201102	752.82	742.91	-9.91	-0.0132	-1.32
201103	538.51	541.24	2.73	0.0051	0.51
201104	648.34	664.55	16.21	0.0250	2.50
.....	.....	.....	.....	.....	.....
201601	572.64	579.55	6.91	0.0121	1.21
201602	752.46	743.84	-8.62	-0.0115	-1.15
201603	489.73	504.90	15.17	0.0310	3.10
201604	632.54	662.52	29.98	0.0474	4.74

3、同时本研究也在没有进行主成分分析时，用 13 个主要影响因素作为神经网络输入，即使不断调整神经元个数，虽然能够部分减少数据误差，但是都遇到收敛到局部极小点的情况（图 7），说明利用主成分分析可以有效解决数据冗余问题，进行主成分分析后运用 NARX 神经网络能够更好运算训练，更加有效。

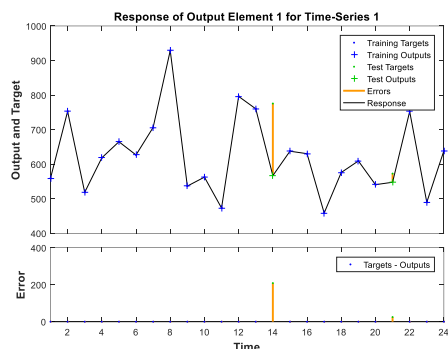


图 7 NARX 神经网络预测效果误差图

## 5. 结语

本研究对影响市政工程造价的 13 个主要影响因素进行主成分分析，得出 4 个主因子，以主因子得分作为神经网络输入，利用 NARX 神经网络建立了单位面积市政工程

造价预测模型，进行基于时间序列的工程造价预测，相对误差有效控制在 10% 以内，满足工程投资估算精度要求，也可以用于项目投资审核，可有效提高工作效率。

## 参 考 文 献

- [1]武艳红.工程造价管理信息化建设的探索[J].产业与科技论坛, 2013,12(21): 252-253.
- [2]黄伟.工程造价信息平台的构建研究[D].重庆: 重庆大学, 2009.
- [3]秦中伏等.基于 SVM 和 LS-SVM 的住宅工程造价预测研究[J].浙江大学学报(理学版), 2016(3): 357-363.
- [4]胡晓娟.多元线性回归模型参数的 STLS 估计法及其在工程造价预测中的应用[J].四川建筑科学研究, 2016(8):142-147.
- [5]胡伟勋.基于 G S T 的工程造价预测模型理论研究[J].中南林业科技大学学报, 2011(4): 146-150.
- [6]王佼,刘艳春.应用灰关联分析的 PSO-SVR 工程造价预测模型[J].华侨大学学报(自然科学版), 2016(6): 708-713.
- [7]胡志根.基于模糊预测的工程造价估算模型研究[J].系统工程理论与实践, 1997(2): 50-55.
- [8]于靖等.基于模糊数学的工程造价动态快速估算方法[J].广东工业大学学报, 2007(2): 107-110.
- [9]叶青等.基于 RN 和 GN 的两种 RBF 神经网络的工程造价预测模型[J]. 数学的实践与认识,2016(7):25-30.
- [10]任 宏, 周其明.神经网络在工程造价和主要工程量快速估算中的应用研究[J].土木工程学报, 2005(8): 135-138.
- [11]冯为民等.基于案例模糊推演的土木工程造价估算方法研究[J].土木工程学报, 2003(3): 51-56.
- [12]蒋均杰, 沈飞.基于时间序列分析的工程造价指数预测[J].项目管理技术, 2011(12): 85-89.
- [13]王学军等.关于工程造价 Bayes 估算方法的探讨[J].武汉水利电力大学学报, 2000(6): 115-117.
- [14]梁喜,刘雨.基于模糊神经网络的建筑工程造价预测模型[J].技术经济, 2017(3): 109-113.
- [15]王建平.基于模糊自适应 BP 算法的工程造价估算[J].西北建筑工程学院学报(自然科学版), 2002(3): 5-8.

- [16]杨永明等.基于灰关联-神经网络的电力工程造价估算[J].重庆大学学报, 2013(11): 15-20.
- [17]李永明等.数据挖掘和神经网络技术的电力工程造价应用[J].重庆大学学报, 2008(6): 663-667.
- [18]张晓东等.基于 MVC 模式的工程造价信息化平台建设架构[J].土木工程与管理学报, 2017(5): 1-6.
- [19]唐勇波等.PCA 和 KICA 特征提取的变压器故障诊断模型[J].高电压技术, 2014,40(2): 557-563.
- [20]卢辉斌等.PSO 优化 BP 神经网络的混沌时间序列预测[J].计算机工程及应用, 2015,51(2): 224-264.
- [21]王红军等.基于 NARX 神经网络的日光温室湿度预测模型研究[J].广东农业科学, 2014(20): 170-172、177.
- [22]赵良杰等.基于 NARX 神经网络的岩溶地下河日流量预测[J].水电能源科学, 2015(5): 19-21、25.

**作者简介:** 张晓东 (1972—), 男, 汉, 安徽芜湖人, 副教授、高级工程师, 造价工程师、一级建造师、咨询工程师、监理工程师, 硕士, 从事建筑经济与管理的研究与教学工作。联系电话: 15358159728; e-mail:466877977@qq.com

2016 年江苏省高校自然科学研究面上项目-----《基于 BIM 的房建工程工程量清单自动生成系统研究》(16KJD560001) 研究成果