

doi: 10.7690/bgzdh.2024.09.008

# 基于马尔科夫链的建筑工程造价预测模型

金季岚

(厦门海迈科技股份有限公司, 福建 厦门 361008)

**摘要:** 针对建筑工程施工环节较多, 获取建筑工程造价预测指标难度加大, 导致预测准确性下降的问题, 设计基于马尔科夫链的建筑工程造价预测模型。通过问卷回收率、指标集中度、指标选取离散度等指标, 使用主成分分析方法处理指标数据, 获取数据特征向量, 构建建筑工程造价预测指标体系; 通过序列规律函数得到状态区间内的转移概率矩阵, 获取不同残差序列中的上下边界值, 得到建筑工程造价预测结果。实验结果表明: 基于马尔科夫链预测模型的平均误差为-0.000 25, 对比其他 2 种预测模型平均误差分别为-0.045 和-0.015, 该模型的相对误差小, 预测准确性更高。

**关键词:** 马尔科夫链; 建筑工程; 工程造价预测; 预测模型; 预测指标; 相对误差

**中图分类号:** TN06 **文献标志码:** A

## Construction Cost Prediction Model Based on Markov Chain

Jin Jilan

(Xiamen Hymake Technology Co., Ltd., Xiamen 361008, China)

**Abstract:** In order to solve the problem that there are many construction links in construction engineering, and it is more difficult to obtain the prediction index of construction engineering cost, which leads to the decline of prediction accuracy, a prediction model of construction engineering cost based on Markov chain is designed. Through the questionnaire recovery rate, index concentration, index selection dispersion and other indicators, the principal component analysis method is used to process the index data, obtain the data feature vector, and construct the construction cost prediction index system; through the sequence law function, the transition probability matrix in the state interval is obtained, the upper and lower boundary values in different residual sequences are obtained, and the construction cost prediction results are obtained. The experimental results show that the average error of the model based on Markov chain is -0.000 25, and the average errors of the other two models are -0.045 and -0.015, respectively. The relative error of the model based on Markov chain is smaller, and the prediction accuracy is higher.

**Keywords:** Markov chain; construction engineering; engineering cost prediction; prediction model; prediction index; relative error

## 0 引言

我国的国民经济中, 建筑行业占据重要地位, 不但可以减缓人口就业压力, 而且能够优化国家的资源配置, 在国家发展与经济建设中有着不可替代的作用。在建筑项目中, 工程造价的合理估测是优化一个工程项目成本的重要手段。如果工程预算超出建筑工程的实际造价, 则会造成大量的浪费。若工程预算小于建筑工程的实际造价, 则难以完全实现预期目标, 造成项目烂尾等。文献[1]以规范建筑企业工程造价成本为目标, 通过灰色模型的方法, 对造价成本进行预测。在 GM(1,1)模型的预测中, 结合预测结果通过标注人工费的投标标价, 实现了企业利润的最大化。该模型的预测结果只是一条平滑的曲线, 难以获取不同建筑工程造价的波动性, 存在较大的误差。文献[2]以输变电工程为实例, 对

其造价预测方法进行设计, 以 Mann-Kendall 趋势进行检验与验证, 通过平滑指数法计算数据指标的走势; 同时, 提出了一种基于 MK-TESM 方法的输变电工程造价预测模型。这一模型准确率较好, 可将平均绝对误差率控制在 10%以内。将该模型应用于建筑工程时, 受到人文因素的影响较大; 因此, 在一些特殊的时间段存在较大的波动性, 很难保证工程造价预测结果的准确性。为进一步提高建筑工程造价预测模型精度, 并保证工程造价预测的稳定性, 笔者在以上文献研究基础上, 设计了基于马尔科夫链的建筑工程造价预测模型。

## 1 构建工程造价预测指标体系

### 1.1 选取工程预测指标

在获取工程造价预测结果之前, 需在全面性、选择性、独立性、可操作性等原则基础上, 选择预

收稿日期: 2024-05-22; 修回日期: 2024-06-20

第一作者: 金季岚(1977—), 女, 江苏人, 硕士。

测指标<sup>[3-4]</sup>。通过问卷回收率、指标集中度、指标选取的离散度3种指标进行量化指标的选取。问卷回收率的计算公式为：

$$\delta_{JX} = X_j / X_p \quad (1)$$

式中： $\delta_{JX}$ 为专家对该问题的关注程度，当 $\delta_{JX} > 0$ 时，该指标选取的问卷调查结果成立； $X_j$ 为有限时间内提交问卷的数量； $X_p$ 为提交问卷的总人数<sup>[5]</sup>。

第2类指标为指标集中度，其计算公式为：

$$\gamma_{JZ} = \frac{\sum_{i=1}^5 P_j k_{ij}}{N_m} \quad (2)$$

式中： $\gamma_{JZ}$ 为问卷调查结果中对某指标的专注度，当 $\gamma_{JZ} \leq 3$ 时，认为该指标可用于工程造价的预测； $P_j$ 为特征值的种类，在本文中最大取值为5； $k_{ij}$ 为第 $j$ 个类型第 $i$ 项指标； $N_m$ 为有效问卷的总数量<sup>[6-7]</sup>。

指标选取离散度 $\lambda_{LS}$ 的计算公式为：

$$\lambda_{LS} = \sqrt{\frac{1}{N_m - 1} \sum_{j=1}^5 k_{ij} (P_j - \gamma_{JZ})^2} \quad (3)$$

式中 $\lambda_{LS}$ 为在问卷调查中评价的分散系数，当 $\lambda_{LS} \leq 1$ 时该指标可作为输入量<sup>[8]</sup>。问卷回收率、指标集中度、指标可选取离散度3个指标为后续的预测模型构架奠定坚实的基础。

## 1.2 数据主成分分析

选取一定量的样本，在各个指标特征的处理下，构成如下矩阵：

$$\mathbf{H}^T = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2m} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nm} \end{bmatrix} \quad (4)$$

式中： $n$ 为待分析样本的数量； $m$ 为指标数量。标准化处理矩阵中的样本数据，计算公式为：

$$\bar{h}_p = \frac{\sum_{i=1}^n h_{ij}}{n} \quad (5)$$

式中： $\bar{h}_p$ 为初始样本差异的平均值； $h_{ij}$ 为指标量纲的属性。

$$K_{hp} = \frac{\sum_{i=1}^n (h_{ij} - \bar{h}_p)^2}{n - 1} \quad (6)$$

式中 $K_{hp}$ 为标准化特征值。

通过以上2个公式可以生成标准矩阵 $\mathbf{H}_{n \times m}^*$ ：

$$\mathbf{H}_{n \times m}^* = (h_{ij})_{n \times m} \quad (7)$$

在标准矩阵的基础上计算主成分的数量：

$$N_k = \mathbf{H}_{n \times m}^* \mathbf{K}_{m \times f} \quad (8)$$

式中： $N_k$ 为主成分的数量； $\mathbf{K}_{m \times f}$ 为主成分对应的特征向量<sup>[9-10]</sup>。得到的数据通过聚类处理，可直接使用，从而实现构建工程造价预测指标体系。

预测指标大致可分为建筑特征、结构基础、装修、材料和工程特征5大类。其中建筑特征指标主要分为建筑面积、建筑层数、单位建筑层高、檐高、建筑进深、实际建筑面积、建筑功能等；基础结构则分为建筑结构种类、房间格式、地基类别、建筑抗震强度、户型、楼体结构等；装修指标包括内外立面装修、墙体材料、门窗家具、楼板屋顶、卫生标准等；建筑材料包括混凝土、钢筋、防水保温材料的价格与均价变化率；工程特征指标包括项目管理、工期、土方处理难度等。

## 2 基于马尔科夫链的工程造价预测模型

依据马尔科夫链的模型，可以在原始数据序列下获取同位的残差序列，序列的规律性用下述公式表示：

$$f(u) = t^{(0)}(u) - \hat{t}^{(0)}(u) \quad (9)$$

式中： $f(u)$ 为序列规律函数值； $t^{(0)}(u)$ 为原始序列的同位序列修正值； $\hat{t}^{(0)}(u)$ 为残差序列在不同状态下的序列修正值<sup>[11-12]</sup>。将其划分为不同的状态，可得到状态区间内的转移概率矩阵：

$$\mathbf{D}^{(0,1)} = \begin{bmatrix} d_{11}^{(1)} & d_{12}^{(1)} & \cdots & d_{1m}^{(1)} \\ d_{21}^{(1)} & d_{22}^{(1)} & \cdots & d_{2m}^{(1)} \\ \vdots & \vdots & & \vdots \\ d_{n1}^{(1)} & d_{n2}^{(1)} & \cdots & d_{nm}^{(1)} \end{bmatrix} \quad (10)$$

式中： $\mathbf{D}^{(0,1)}$ 为马尔科夫链模型在观测值状态下的转移统计矩阵； $d_{nm}^{(1)}$ 为状态转移后的数据。

据此可获取不同残差序列中的上下边界值：

$$\left. \begin{aligned} P_{\min} &= \min f(u) + (p_j - 1) / \sqrt{R_t} \\ P_{\max} &= \left( \max f(u) + p_j / \sqrt{R_t} \right) / t^{(0)}(u) \end{aligned} \right\} \quad (11)$$

式中： $P_{\min}$ 和 $P_{\max}$ 分别为工程造价预测值的下界和上界，即最小值与最大值； $p_j$ 为预测值的区间中点； $R_t$ 为预测转移系数<sup>[13-14]</sup>。

通过上文中的方法，可以基本得到建筑工程造价的预测结果，结合马尔科夫链设计建筑工程造价预测模型，具体流程如图1所示。

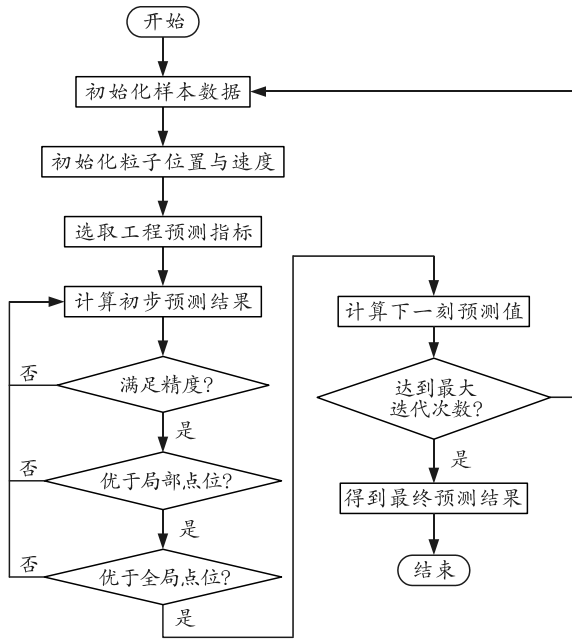


图 1 基于马尔科夫链得到建筑工程造价预测模型设计流程

如上图所示，最开始需要初始化预测指标样本数据，初始化粒子的位置与速度。此时的粒子为不同阶段、不同指标下建筑工程的造价预测指标数据，通过马尔科夫链计算初始的预测结果。验证该结果是否满足精度需求，是否优于局部点位，以及是否在全局角度优于点位。若未满足条件，需要重新计算预测结果，当满足条件时，可以计算下一刻的预测结果<sup>[15-16]</sup>。判断算法是否达到最大迭代次数，当达到最大迭代次数时，可以得到最终的建筑工程造价预测结果。

### 3 实验研究

#### 3.1 数据的收集与预处理

笔者设计了一种基于马尔科夫链的建筑工程造价预测模型，为验证该模型的实际应用效果，选择某已经完成修建的居住建筑作为本次实验的研究对象，收集该建筑建设过程中的各项造价信息。在该工程中，共需修建 62 组建筑，因此收集 62 组工程造价的数据。数据中包含冗余以及相互重叠的部分，剔除部分无用数据，得到 53 组有效数据，将这 53 组数据作为本次实验的样本数据集。在模型训练时，以 15 个工程预测的指标作为样本的输入值，并为其设置样本编号  $D_1-D_{15}$ 。将单一编号建筑的工程造价作为数据的输出值，设置编号为  $P$ 。对这 53 组数据以及 15 项工程指标进行聚类分析，通过 SPSS 软件，使用 Ward 系统聚类方法，可得到样本的聚类结果如表 1 所示。

表 1 数据样本聚类结果

类别	样本编号	样本数量
1	1、2、3、5、6、8、9、11、12、13、14、	34
	18、19、20、23、24、26、27、28、32、	
	33、34、38、39、40、41、42、46、47、	
	48、50、51、52、53	
2	4、7、10、15、16、17、21、22、25、29、	19
	30、31、35、36、37、43、44、45、49	

由于本实验中样本总量较小，可通过样本聚类的方式，减少样本误差，将 53 个建筑依据工程指标分为 2 类：第 1 类 34 组数据，第 2 类 19 组数据。本实验选择第 1 类数据进行模型预测。

#### 3.2 工程指标特征降维

使用 SPSS 软件作为指标降维处理的工具，以上 34 组数据作为标准化的初始数据，分别对其 15 项工程指标进行因子分析，获得特征指标的相关矩阵  $X^T$ ：

$$X^T = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,14} & X_{1,15} \\ X_{2,1} & X_{2,2} & \dots & X_{2,14} & X_{2,15} \\ \vdots & \vdots & & \vdots & \vdots \\ X_{14,1} & X_{14,2} & \dots & X_{14,14} & X_{14,15} \\ X_{15,1} & X_{15,2} & \dots & X_{15,14} & X_{15,15} \end{bmatrix} \quad (12)$$

式中  $X_{1,1}$ 、 $X_{2,2}$ 、 $X_{15,15}$  等指标均为 1，提取其主成分，绘制不同成分分析中的特征值碎石图，如图 2 所示。

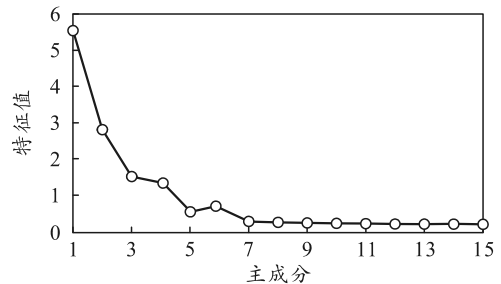


图 2 特征值碎石(马尔科夫链)

由上图可知，在 15 个主成分中，所有特征值均大于零，其工程造价的预测结果满足整体主成分分析的需求；且在主成分 1—6 之间，原始信息的累计贡献率占据整体信息的 90% 以上，基本可以在可靠性指标预测的范围之内；因此，可将主成分 1—6 作为本实验中预测模型的输入指标。

#### 3.3 模型训练与测试

通过 6 个主成分，分别得到 34 组数据的输入数据集，并计算其造价预测的输出集。依据样本数据，选择 34 组数据集的前 26 组作为样本数据的训练集，后 8 组数据作为样本数据的测试集。设置最大迭代

次数为 500，动量因子为 0.42，将 SIM 函数作为训练拟合函数，得到不同预测模型下造价预测结果与实际造价的对比结果如图 3 所示。

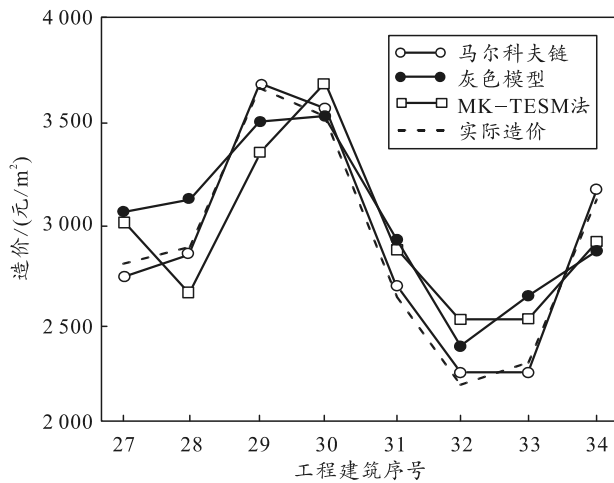


图 3 造价预测与实际造价对比

上图中，8 个项目实际的工程造价分别为 2 803、2 854、3 711、3 529、2 702、2 231、2 277、3 208 元/m<sup>2</sup>，4 种不同工程造价预测模型的预测结果如表 2 所示。

表 2 不同模型下单位面积工程造价预测值

建筑工程	实际造价	马尔科夫链	灰色模型	MK-TESM 法
27	2 803	2 755	3 074	3 025
28	2 854	2 812	3 136	2 433
29	3 711	3 724	3 502	3 396
30	3 529	3 534	3 528	3 732
31	2 702	2 744	2 969	2 886
32	2 231	2 256	2 415	2 534
33	2 277	2 251	2 621	2 514
34	3 208	3 249	2 876	2 913

### 3.4 预测精度与稳定度分析

以上 27—34 号建筑工程的 8 组工程施工实际造价与 3 种模型下的造价预测结果，可依据公式，计算其相对误差：

$$f_n = (f_i - \hat{f}_i) / f_i \quad (13)$$

式中： $f_n$  为工程造价预测结果与实际结果的相对误差； $f_i$  为该工程项目中实际造价； $\hat{f}_i$  为该工程项目在某模型下的预测造价。经过计算，可得到如表 3 所示的计算结果。

表 3 相对误差计算结果

工程建筑	马尔科夫链	灰色模型	MK-TESM 法
27	0.017	-0.097	-0.079
28	0.015	-0.099	0.148
29	-0.004	0.056	0.085
30	-0.001	0.001	-0.058
31	-0.016	-0.099	-0.068
32	-0.011	-0.082	-0.136
33	0.011	-0.151	-0.104
34	-0.013	0.103	0.092

上表在马尔科夫链的预测模型中，预测结果的相对误差分布区间为[-0.016, 0.017]，均控制在±0.02 以内，且相对误差的平均值为-0.000 25。在灰色模型的预测算法中，预测结果的相对误差分布区间为[-0.151, 0.103]，其平均值为-0.046。在 MK-TESM 方法的预测模型中，相对误差分布区间为[-0.136, 0.148]，平均值为-0.015。通过上述数据可知，本文中方法的建筑工程造价预测模型预测精度最高。

## 4 结束语

笔者构建了基于马尔科夫链的建筑工程造价预测模型，通过对指标的处理，获取了更准确的初始数据。对数据指标进一步处理，使用马尔科夫链构建预测模型获取建筑工程造价预测结果。在实验过程中，通过对比实际造价与预测造价，并计算其相对误差，本文中的预测模型精度优于现有的几种预测模型，可得到更准确的工程造价预测结果。

## 参考文献：

- [1] 程晓丽. 基于灰色模型的主体结构价格预测研究——以邢台住宅项目为例[J]. 安徽建筑, 2021, 28(11): 183-184, 187.
- [2] 丁政中, 彭露苇. 基于 MK-TESM 法的输变电工程造价数据预测方法[J]. 沈阳工业大学学报, 2021, 43(2): 126-131.
- [3] 董娜, 卢泗化, 熊峰. 大数据背景下基于 ABC-SVM 的建筑工程造价预测[J]. 技术经济, 2021, 40(8): 25-32.
- [4] 乔慧婷, 文上勇, 黄球. 电网输电工程项目数据插补及造价预测融合模型[J]. 沈阳工业大学学报, 2021, 43(5): 481-486.
- [5] 刘沁. 基于深度学习和聚类算法的电力工程建设数据分析系统设计[J]. 电子设计工程, 2021, 29(3): 27-30, 35.
- [6] 陈飞宇, 董利飞, 王苗, 等. 基于灰色马尔科夫模型的地基承载力预测[J]. 地质科技通报, 2022, 41(3): 222-227.
- [7] 郭书翊. 基于灰色关联度及支持向量机理论的公路工程造价预测研究[J]. 北方交通, 2020(11): 84-86, 89.
- [8] 牛衍亮, 高立扬, 段晓晨, 等. 基于余弦相似度-神经网络模型的高铁土建工程造价估算[J]. 土木工程与管理学报, 2020, 37(5): 82-87, 94.
- [9] 王杰, 卢毅. 基于 ANN 贡献分析及 GEP 算法的地铁车站土建造价预测模型[J]. 铁道科学与工程学报, 2020, 17(8): 2152-2161.