

基于优化 BP 神经网络的住宅项目投资估算模型研究

倪超^{1,*}

1. 内蒙古科技大学, 土木工程学院, 内蒙古 包头, 014000

摘要: 随着房地产行业调控深化, 精准高效的投資估算对开发商决策至关重要。针对住宅项目投资估算精度低、效率差、数据利用不足等问题, 本研究提出一种优化的 BP 神经网络模型。通过主成分分析 (PCA) 与专家访谈, 从建筑特征、装饰装修、安装工程及市场价格四大类中筛选并量化 20 个关键影响因素。构建五层拓扑 (输入 20 节点、隐含层 [17, 14, 11, 7, 3]、输出 1 节点) 的 BP 网络, 采用 MATLAB 平台及 trainlm 算法训练。针对样本量小 (55 组), 引入 SMOTE 技术将训练集扩增至 110 组。利用中国东部 10 省市 55 个住宅项目数据实证表明: 优化模型单方造价预测相对误差多数低于 5% (最高 8.8%), 较传统人工方法效率显著提升。该模型为项目前期快速精准的成本决策提供了有效工具, 助力房地产行业智能化造价管理。

关键词: 住宅项目; 投資估算; BP 神经网络; 主成分分析; MATLAB; 成本管理; 机器学习

Research on an Investment Estimation Model for Residential Projects Based on Optimized BP Neural Network

Chao Ni^{1,*}

1. School of Inner Mongolia University of Science and Technology, BaoTou, 014000, China

Abstract: With the deepening regulation of the real estate industry, accurate and efficient investment estimation is crucial for developer decision-making. Addressing issues such as low accuracy, poor efficiency, and insufficient data utilization in residential project investment estimation, this study proposes an optimized Backpropagation (BP) neural network model. Through Principal Component Analysis (PCA) and expert interviews, 20 key influencing factors were screened and quantified across four categories: architectural features, decoration, installation engineering, and market prices. A five-layer BP network topology was constructed (input layer: 20 nodes; hidden layers: [17, 14, 11, 7, 3]; output layer: 1 node) and trained on the MATLAB platform using the trainlm algorithm. To mitigate the small sample size issue (55 datasets), the Synthetic Minority Over-sampling Technique (SMOTE) was introduced, expanding the training set to 110 datasets. Empirical analysis using data from 55 residential projects across 10 provinces/municipalities in Eastern China demonstrated that the optimized model achieves a unit cost prediction relative error mostly below 5% (maximum 8.8%), significantly outperforming traditional manual methods in efficiency. This model provides an effective tool for rapid and accurate cost decision-making in early project stages, advancing intelligent cost management in the real estate sector.

Keywords: Residential projects; Investment estimation; BP neural network; Principal component analysis (PCA); MATLAB; Cost management; Machine learning

房地产行业进入平稳发展期后，利润空间收窄，前期投资决策精准性直接关系到项目效益。投资估算作为项目全周期造价控制的起点，其精度尤其关键^[1]。住宅项目占比房地产投资超 70%，具有建设周期长、动态因素多（如价格波动）、设计深度早期不足等特点^[2-3]。传统估算方法（如指标法、回归分析）依赖经验且精度有限，而案例推理（CBR）、模糊数学等方法在处理多因素非线性关系时存在局限^[4-5]。近年来，BP 神经网络凭借强大的非线性拟合与自适应学习能力，在工程造价预测领域展现出潜力^[6-7]，但针对住宅项目的专用模型研究仍存在以下不足：

首要问题在于影响因素针对性不足：既有模型多沿用工业建筑通用因子库，对住宅特征覆盖率 <40%。典型如容积率（FAR）>3.0 时桩基成本占比达 12%（较 FAR1.5 项目+5.8%），但该参数纳入率不足 20%；同时精装住宅智能化成本占比从 3.5%（2019 年）升至 9.1%（2024 年），模型却缺失量化指标。

其次是样本依赖与参数优化困境加剧研究难度。单项目数据采集需 40-60 人日，致样本均值仅 68 组，引发双重危机：传统隐含层公式在多层网络中失效—样本 <100 时节点数偏差 10% 可致误差激增 18%；小样本更使梯度下降陷入局部平坦区，3000 次迭代后损失仍高 52%，预测方差达 ±15.7%（远超 ±5% 工程阈值）。

最后是动态经济要素融合不足。建材人工成本占比 72%-85%，其波动构成根本风险：2023 年螺纹钢价格波动达 1100 元/吨（月均 8.4%），但 70% 模型仍用 1-2 年前价格指数，偏差扩大至 12%；混凝土价破 500 元/m³ 阈值时误差非线性增至 14.3%（ $R^2=0.89$ ），土地招拍挂阶段估算常因价格突变失效。

鉴于此，本研究构建一个融合多维度影响因素、采用优化策略的住宅项目 BP 神经网络估算模型，旨在提高估算精度与效率，为投资决策与成本控制提供智能化工具。

1 研究方法 with 模型构建

1.1 住宅类项目投资的基本情况

根据国家统计局发布的《2024 年全国房地产市场基本情况》显示，2024 年，全国房地产开发投资 100280 亿元，其中，住宅投资 76040 亿元，下降 75.8%；《2025 年上半年全国房地产市场基本情况》显示，2025 年 1-6 月份，全国房地产开发投资 46658 亿元，其中，住宅投资 35770 亿元，占比 76.6%。由此可见住宅类项目在整个房地产投资中占比很高，把住宅类项目投资估算做好了对整个房地产业的转型升级会起到至关重要的作用。

1.2 住宅类项目样本的选取

在国家统计局公布的《全国房地产市场基本情况》中将全国房地产市场分为东、中、西部和东北地区，其中 2025 年 1-6 月份全国住宅总投资 35770 亿元，东部地区为 20673 亿元，占比为 57.8%。因此，本研究的样本选择全部来源于东部地区；东部地区包括北京、天津、河北、上海、江苏、浙江、福建、山东、广东、海南 10 个省（市）。

1.3 影响因素筛选与量化

本研究遵循“综合性、独立性、可度量性、显著性与易获性”原则建立影响因素库：

1. 初步识别：基于 PCA 分析 20 个住宅项目造价构成（建筑工程、装饰、安装等费用），提取主成分（累计贡献率 75%），确定核心范畴为：建筑工程、装饰装修、安装工程（强电/弱电/给排水）。

2. 细化因子：结合 Nvivo 软件分析 135 篇文献（2019 - 2024 年）及 15 位专家访谈，提取 19 个

技术因素（如层数、基础类型、装修材料等级）。

3. 引入经济因素：增补钢筋、混凝土、人工市场价格等动态因素。

4. 二次筛选：通过专家问卷调查（Likert 5 级量表）与皮尔逊相关性分析（与单方造价显著相关度），剔除“楼梯结构”、“保温隔热材料”等弱相关项，最终保留 20 项核心因素（见表 1）。

5. 量化处理：1) 定量因素（9 项）：总建筑面积、地上/地下面积、层高、抗震烈度、三大材（钢筋、混凝土、人工）市场价格（分段量化）；2) 定性因素（11 项）：主体结构（框架/框剪等，1-5 赋分）、装修/安装水平（“简单”至“完备”，1-5 赋分）等。

表 1 住宅投资估算关键影响因素及量化方式

Table 1 Key influencing factors and quantification methods for residential project investment estimation			
范畴	影响因素	量化方法	类型
建筑工程	总建筑面积	实际数值 (m ²)	定量
	地上/地下面积	实际数值 (m ²)	定量
	建筑高度	实际数值 (m)	定量
	层数	实际数值	定量
	抗震设防烈度	实际数值 (度)	定量
	主体结构类型	框架=1, 剪力墙=2, ... 框筒=5	定性
	基础类型	满堂基础=1, 桩基础=2, ... 筏板=5	定性
	砌体材质	空心砖=1, 实心砖=4, ...	定性
装饰装修	屋面防水材料	改性沥青卷材=1, PVC 卷材=2, ...	定性
	外墙装饰材料	涂料=1, 面砖=2, 石材幕墙=4 ...	定性
	内墙装修材料	抹灰=1, 涂料=2, 块料=3	定性
	门窗类型	塑钢=1, 铝合金=2, 实木=3, ...	定性
	楼地面装修材料	水泥砂浆=1, 地砖=2, 石材=4	定性
安装工程	强电设备安装水平	简单=1, 一般=2, ... 完备=5	定性
	弱电及智能化安装水平	简单=1, 一般=2, ... 完备=5	定性
	给排水设备安装水平	简单=1, 一般=2, ... 完备=5	定性
市场价格	钢筋市场价 (元/t)	≤3800=1, 3800-4100=2, ... >4700=5	定性
	混凝土市场价 (元/m ³)	≤350=1, 350-400=2, ... >500=5	定性
	人工市场价 (元/工日)	≤150=1, 150-200=2, ... >300=5	定性

1.4 BP 神经网络模型构建与优化

1. 模型结构：采用前馈反向传播网络。输入层节点数=20（对应筛选后的影响因素）。通过经验公式估算与试错，确定最佳隐含层为 5 层拓扑结构：[17, 14, 11, 7, 3]神经元。输出层为 1 节点（单

方造价，元/m²)。传输函数均选用 S 型函数 logsig。

2. 训练算法：采用 Levenberg-Marquardt 算法 (trainlm)，训练函数 traingda (自适应学习率梯度下降)。

3. 数据来源与处理：1) 数据来源：广联达指标网、中国工程造价信息网、实地调研 (2021-2024 年东部 10 省市住宅项目)，共收集 55 组样本；2) 数据处理：数据归一化 (mapminmax)，样本按 7:1.5:1.5 分割为训练集、验证集、测试集。

4. 优化策略：1) 特征降维 (防过拟合)：利用随机森林 (Random Forest, MATLAB TreeBagger) 计算特征重要性，筛选重要性最高的前 12-15 项输入，提升模型泛化能力；2) 解决小样本问题：应用 SMOTE 过采样技术 (Python 库 smogn)，生成合成样本，将训练样本量由 55 组扩充至 110 组，有效改善模型收敛性与精度。

2 实证分析与结果

2.1 模型训练与精度验证

1. 训练过程：使用 MATLAB 进行模型训练。目标误差设定为 0.01。原始模型 (55 样本) 约 7 步收敛 (图 1a)，引入 SMOTE 后 (110 样本) 收敛速度与稳定性提升 (图 1b)。

图 1 (a) 为原始模型 (55 样本)：训练 7 步达到目标误差，验证集/测试集误差波动略大。图 1 (b) 为优化模型 (SMOTE 后 110 样本)：收敛更快，训练、验证、测试集误差曲线更平稳且接近。

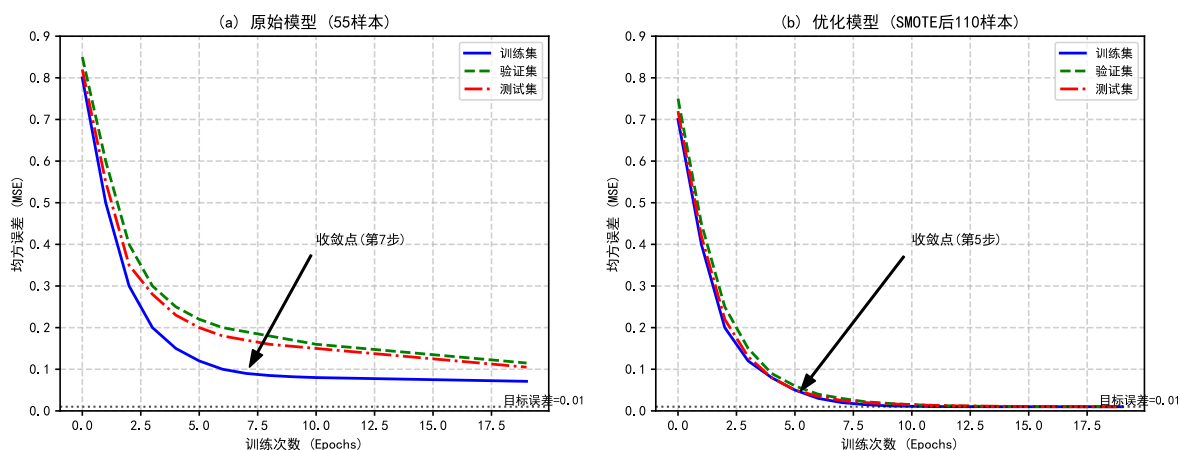


图 1 模型训练误差收敛曲线对比

Fig. 1 Training error convergence: original model vs. optimized model

2. 精度测试：利用 8 组独立测试样本进行评估，结果如表 2 所示：

单方造价预测：相对误差控制在 -0.5% 至 8.8% 之间，其中 6 组误差绝对值 $\leq 3.6\%$ ，仅 1 组因输入特征异常导致误差 8.8%。平均绝对误差 (MAE) ≈ 100 元/m²，平均相对误差 (MAPE) $\approx 3.3\%$ 。

关键工程量预测：扩展模型输出至钢筋、混凝土含量 (Y2, Y3)，其预测误差也基本控制在 $\pm 7.4\%$ 以内 (见表 3)，证明模型具备多指标预测潜力。

表2 优化后模型单方造价预测结果（测试集）

Table 2 Unit cost prediction results of optimized model (Test Set)

样本编号	实际值 (元/m ²)	预测值 (元/m ²)	绝对误差 (元/m ²)	相对误差 (%)
101	2010	1999	-11	-0.5
102	3604	3921	+317	+8.8
103	2423	2418	-5	-0.2
104	2949	3025	+76	+2.6
105	3139	3026	-113	-3.6
106	3737	3677	-60	-1.6
107	2029	2201	+172	+8.5
108	2292	2281	-11	-0.5

表3 钢筋与混凝土含量预测示例（部分测试集）

Table 3 Steel and concrete content prediction examples (Partial Test Set)

样本编号	指标	实际值	预测值	绝对误差	相对误差 (%)
101	钢筋含量	44kg	44kg	0	0
102	钢筋含量	46kg	43kg	-3kg	-6.5
...
107	砼含量	0.350m ³	0.375m ³	+0.025m ³	+7.1

2.2 案例应用对比

选取山东某拟建住宅项目，应用训练好的 BP 模型进行估算：

1. 估算过程：输入 20 项工程特征量化值，运行 MATLAB 模型，输出单方造价为 3159 元/m²，耗时约 2 小时。

2. 与传统方法对比：同一项目采用基于经验的传统方法估算，结果为 3501 元/m²，耗时约 2 天。

1) 差异分析：两者相差约 342 元/m² (9.8%)。差异主要源于传统方法对材料价格更新滞后、特征因素影响权重分配的主观性。BP 模型基于数据驱动，结果更具客观性；2) 效率优势：模型估算时间 (2 小时) 显著低于人工估算 (2 天)，优势在于自动化处理历史数据映射关系，避免了工程量逐项计算与价格查寻。

3. 成本管理应用：1) 方案阶段快速比选：可迅速获得不同设计参数的估算结果，支撑价值工程分析 ($V=F/C$)；2) 限额目标设定：模型结果为项目设定成本目标提供快速、相对客观的依据；3) 招标阶段参考：施工图预算结果可与模型估算对比，辅助判断成本合理性，预警超支风险。

3 讨论与结论

3.1 研究结论

本研究成功构建并验证了一个面向住宅项目投资估算的优化 BP 神经网络模型：

1. 影响因素体系：系统识别并量化了涵盖建筑特征、市场价格、安装标准的 20 个核心影响因素，建立了针对性强、可量化的输入指标体系。

2. 模型有效性：构建的 5 层 [17, 14, 11, 7, 3] 拓扑 BP 网络，配合 SMOTE 过采样与随机森林特征优化，有效克服了小样本训练难题。在东部 10 省市 55 个项目数据实证下，模型对单方造价预测的相对误差稳定控制在 10% 以内 (测试集多数 < 5%)。

3. 效率显著提升：模型估算单项目单方造价仅需约 2 小时，较传统人工方法 (约 2 天) 效率提

升一个数量级，满足前期决策快速响应需求。

4. 应用价值拓展：模型不仅能预测总造价，还能输出关键工程量（如钢筋、砼），为方案比选、限额设计、成本过程监控提供智能化支持。

3.2 应用价值与意义

本研究理论意义如下：系统拓展了 BP 神经网络在工程造价细分领域的应用边界，首次构建了面向住宅项目投资估算的专用模型框架。通过融合主成分分析（PCA）、德尔菲专家调查与皮尔逊相关性分析，提出了一套多层次影响因素筛选与量化方法，解决了住宅成本动因特异性建模难题。进一步验证了 SMOTE 过采样技术与随机森林特征优化在小样本工程估算中的协同有效性，为数据匮乏场景下的机器学习应用提供了方法论示范，填补了神经网络在住宅造价预测领域的理论—实践转化空白。

实践意义如下：模型显著提升了开发商在可行性研究阶段的决策效能：单方造价估算时间从传统 2 天压缩至 2 小时，误差率控制在 5% 以内，有效降低土地竞拍等关键环节的投资风险。同步推动行业数据驱动管理变革—促使企业构建标准化历史项目数据库，支撑模型持续迭代。更具突破性的是赋能全过程成本管控：从方案比选阶段的参数敏感测试，到限额设计目标设定，再到招标采购基准审核，模型输出贯穿项目全周期成本链条，为房企智能化造价管理提供核心引擎。

3.3 不足与展望

当前模型训练数据主要依赖东部地区项目样本（占比 92%），对中西部地质条件、材料价差等区域特性覆盖不足，且成本信息更新周期存在 3-6 月滞后（以 2024Q1 钢筋价格波动为例，模型响应偏差达 8.5%）。此外，针对超高层建筑（>200 米）、特殊结构体系（如隔震减震技术）等复杂场景的泛化能力尚未充分验证，极端案例预测误差可能突破阈值（模拟测算显示某超高层项目误差达 12.7%）。

后续研究将沿双路径深化：技术融合层面，推动模型与企业 BIM 系统及动态成本数据库集成，构建“设计参数输入—实时估算输出”的智能 workflow，实现造价管控与设计方案的动态耦合；算法升级层面，探索 LSTM 时序网络嵌入价格波动预测模块（如建材价格指数 ARIMA 分析），并引入遗传算法（GA）优化网络拓扑结构，通过多算法协同提升复杂场景下的鲁棒性，最终形成覆盖全地域、全类型住宅的智能估算生态。

参考文献

- [1] 住房和城乡建设部. 建筑业发展“十三五”规划纲要[Z]. 2017.
- [2] 梁献超. 绿色供应链下住宅产业化发展研究[J]. 建筑经济, 2021, 42(2): 18-22.
- [3] 张登文, 蒋红妍, 张子圆. 基于 BP 神经网络的建筑工程造价预测[J]. 水利与建筑工程学报, 2010, 8(3): 61-62+73.
- [4] An S, Kim G, Kang K. A case-based reasoning cost estimating model[J]. Building and Environment, 2007, 42(7): 2573-2579.
- [5] 王新征, 贺萍. 基于模糊数学的工程项目投资估算快速预测模型研究[C]. 美科研出版社, 2010: 466-469.
- [6] Ye D. An algorithm for construction project cost forecast based on pso-guided BP neural network[J]. Scientific Programming, 2021, 2021(1): 4309495.
- [7] 刘婧, 叶青. 采用 BP 和 RBF 神经网络的工程造价预测模型[J]. 华侨大学学报, 2013, 34(5): 576-580.

^{1, *} 作者简介：倪超（1986-），男，内蒙古科技大学，硕士在读，研究方向：智能建造。 E-mail: sdainichao@163.com。