

Research Article

Research on Dynamic Cost Prediction of Power Grid Technical Renovation Projects Based on Improved Transformer

Luo Yi*

Yueyang Power Supply Branch, State Grid Hunan Electric Power Co., Ltd., Yueyang, China

Abstract

Cost prediction for power grid technical renovation projects serves as a critical component of investment decision-making and optimal resource allocation for power enterprises, and its results directly affect investment control, budget formulation and risk assessment. To address the limitations of conventional methods, which struggle to capture the dynamic evolution of costs throughout the project life cycle and overlook the differences in cost characteristics across various construction phases, this paper proposes a dynamic cost prediction model based on the Gated Multi-Stage Transformer (GMST). The model divides technical renovation projects into three phases: planning, implementation and settlement. Separate Transformer encoders are established to extract temporal features of each phase, and a learnable gating network is introduced to adaptively aggregate multi-stage prediction information and realize dynamic weighting of features from different phases. Experiments on datasets of 1,876 real-world projects demonstrate that the GMST achieves an MAE of 123,700 CNY, an RMSE of 185,400 CNY and a MAPE of 5.82%, outperforming vanilla Transformer, LSTM and XGBoost by a notable margin. Ablation experiments verify the synergistic effect of multi-stage encoding and gated fusion. Phase-wise accuracy analysis reveals that the prediction error gradually converges from 9.21% to 5.82% as the project proceeds, indicating that the proposed model is suitable for the progressive cost estimation scenario of technical renovation projects..

Keywords

Transformer, Power Grid Technical Renovation, Dynamic Cost Prediction, Gating Mechanism, Multi-stage Learning

基于改进 Transformer 的电网技改项目动态成本预测研究

罗依*

国网湖南省电力有限公司岳阳供电分公司，岳阳，中国

*Correspondence: Luo Yi (2241830900@qq.com)

Received: 28 June 2026; **Accepted:** 24 July 2026; **Published:** 13 August 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

摘要

电网技改项目成本预测是电力企业投资决策与资源优化配置的关键环节，预测结果直接影响投资管控、预算编制与风险研判。针对传统方法难以捕捉项目全生命周期成本动态演化规律，忽略不同建设阶段成本特征差异性的问题，本文提出一种基于门控分阶段 Transformer (Gated Multi-Stage Transformer, GMST) 的动态成本预测模型。模型将技改项目划分为规划期、实施期和结算期三个阶段，分别构建独立 Transformer 编码器提取各阶段时序特征，引入可学习门控网络自适应聚合多阶段预测信息，实现不同阶段特征权重动态分配。在 1876 个真实项目数据上的实验表明，GMST 在 MAE、RMSE 和 MAPE 上分别达到 12.37 万元、18.54 万元和 5.82%，较标准 Transformer、LSTM 和 XGBoost 均有显著提升；消融实验验证了分阶段编码与门控融合的协同增效作用；分阶段精度分析表明预测误差随项目推进从 9.21% 递进收敛至 5.82%，能够适配技改项目渐进式成本预估场景。

关键词

Transformer, 电网技改项目, 动态成本预测, 门控机制, 多阶段学习

1. 引言

电网技改项目是电力企业维持设备可靠性和保障供电安全的重要手段，国家电网年均技改投资超千亿元，涵盖输电线路改造、变电站设备更换和配网自动化升级等多种类型，其成本管理直接影响企业经济效益。然而技改项目成本受设备价格波动、施工变更等多因素驱动，呈现显著非线性和动态演化特征，传统定额估算和类比法的预测偏差可达20%~30%[1]，灰色预测等参数法[2]难以建模复杂非线性关系。周磊等人[3]将多头自注意力机制引入电力时序预测，Transformer凭借全局依赖捕捉和并行化计算优势突破了RNN的长期依赖瓶颈。但现有变体[4-6]均面向连续传感器数据设计，电网技改项目属于“事件驱动型”数据——每个项目仅3~8个不等间隔阶段节点，各阶段特征空间与信息充分度存在本质差异，亟需分阶段建模方案。本文提出门控分阶段Transformer (GMST) 模型，将项目划分为规划、实施和结算三阶段独立编码，设计可学习门控网络自适应融合多阶段预测，在1876个真实项目上验证了有效性。

2. Transformer 原理与应用

2.1. 自注意力机制

自注意力机制 (Self-Attention) 是Transformer架构的核心创新，其基本思想是让输入序列中的每个位置都能直接关注序列中的所有其他位置，从而捕获任意跨度的全局依赖关系。与RNN逐步递归处理序列不同，自注意力机制以整个序列为操作对象，通过计算位置之间两两的注意力权重实现信息聚合，具有完全并行化的计算特性。给定输入序列 $X = [x_1, x_2, \dots, x_n] \in R^{n \times d}$ ，首先通过三个可学习

投影矩阵 $W_Q \in R^{d \times d_k}$ 、 $W_K \in R^{d \times d_k}$ 、 $W_V \in R^{d \times d_k}$ ，将每个位置映射为查询向量 q_i 、键向量 k_i 和值向量 v_i ，其中 d_k 为注意力维度，通常设置 $d_k = d_v = d/h$ (h 为注意力头数)。在实际实现中，所有位置的投影通过矩阵乘法一次性完成： $Q = XW_Q$ 、 $K = XW_K$ 、 $V = XW_V$ 。每个位置的输出为所有位置值向量的加权和，权重由对应查询向量与键向量的缩放点积相似度经softmax归一化得到：

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

缩放因子 $\sqrt{d_k}$ 的作用是防止当 d_k 较大时点积值过大导致softmax函数进入饱和区，从而避免梯度消失问题。自注意力机制的时空复杂度为 $O(n^2d)$ ，当序列长度 n 较小时（如技改项目阶段内序列长度 $n \leq 8$ ），其计算开销完全可接受，且对全局依赖的捕捉能力远优于RNN类循环模型。通过注意力权重矩阵的可视化分析，可以直观地观察模型在进行成本预测时对不同时间步（如实施期不同月份）的注意力分配模式。

2.2. 多头注意力与编码器结构

单一注意力头仅能在一种投影子空间中捕获特征关联，为增强模型的表达能力，Transformer采用多头注意力 (Multi-Head Attention, MHA) 机制。其核心思想是将 Q 、 K 、 V 分别通过 h 组不同的线性投影映射到 h 个低维子空间，在每个子空间中并行执行缩放点积注意力，最后将各头输出拼接并经过线性变换融合：

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

多头机制使不同注意力头可以关注不同类型的特征关联——例如某头可能聚焦于相近时间步的局部趋势，另一头则捕获规划期预算与结算期实际成本之间的长程映射关系——多个子空间的互补信息经过拼接融合后形成更丰富的序列表征。Transformer编码器由L层结构相同的层堆叠而成，每层包含两个子层：多头自注意力子层和前馈网络（FFN）子层。每个子层均采用残差连接（Residual Connection）与层归一化（LayerNorm），以缓解深层网络的梯度退化问题并加速训练收敛[7]。FFN为两层全连接网络，激活函数为ReLU：

$$FFN(x) = ReLU(xW_1 + b_1)W_2 + b_2 \tag{4}$$

其中： $W_1 \in R^{d_{model} \times d_{ff}}$
 $W_2 \in R^{d_{ff} \times d_{model}}$

中间层维度 d_{ff} 通常设为 $4d_{model}$ 以提供足够的非线性变换容量。此外，因自注意力机制本身对序列位置不敏感，需在编码器输入层按下式叠加正弦位置编码使模型能感知输入序列的时序顺序[8-10]。

$$PE_{(pos,2i)} = \sin(\frac{pos}{10000^{\frac{2i}{d_{model}}}}) \tag{5}$$

$$PE_{(pos,2i+1)} = \cos(\frac{pos}{10000^{\frac{2i}{d_{model}}}}) \tag{6}$$

其中， pos 为序列中的当前位置编号。

3. GMST 模型设计

3.1. 问题形式化与阶段特征

电网技改项目全生命周期可划分为规划期（Stage 1）、实施期（Stage 2）和结算期（Stage 3）三个连续阶段[11]。设项目在阶段 $k(k=1,2,3)$ 可观测特征向量 $x_k \in R^{d_k}$ ， $d_1 = 9$ 、 $d_2 = 14$ 、 $d_3 = 7$ 。规划期仅含项目类型、电压等级、设备类别、计划工期、概算金额等静态属性；实施期追加施工进度、已发生成本、材料价格指数等动态特征；结算期进一步引入变更签证金额和审核调整率[12]。目标是利用截至阶段 k 的所有可用信息预测最终结算成本。各阶段特征如表1所示。

表 1 项目各阶段特征。

阶段	符号	原始维度	特征内容与维度分解
规划期(Stage 1)	x1	d1=9	项目类型 one-hot(3)、电压等级(1)、设备类别(1)、计划工期(1)、概算金额(1)、建设规模(1)、地域标识(1)
实施期(Stage 2)	x2	d2=14	x1 全 9 维 + 施工进度%(1)、已发生累计成本(1)、材料价格指数(1)、实际开工偏移天数(1)、累计设计变更次数(1)
结算期(Stage 3)	x3	d3=16→7	x2 全 14 维 + 变更签证金额(1→log 变换)、结算审核调整率(1→分箱编码)

3.2. 分阶段 Transformer 编码器

GMST 模型按项目三个阶段分别构建独立的Transformer编码器 Enc_1 、 Enc_2 、 Enc_3 ，用于保留各阶段的专属特征模式，避免跨阶段信息的混淆学习[13]。具体流程如下所示：

（1）通过阶段特定的可学习投影矩阵 $W_p^k \in R^{d_k \times d_{model}}$ ，将原始阶段特征映射至统一的模型维度，以解决各阶段特征维度不同的问题。

（2）叠加标准正弦位置编码PE以保留阶段内的时序顺序信息。经过投影和位置编码的特征矩阵输入L层Transformer编码器，每层依次按式执行多头自注意力子层和前馈网络子层的残差连接与层归一化[14]：

$$H'_l = LayerNorm(H_{[l-1]} + MHA(H_{[l-1]})) \tag{7}$$

$$H_l = LayerNorm(H'_l + FFN(H'_l)) \tag{8}$$

（3）对编码器输出的完整时序特征序列，截取末尾时间步的隐向量作为整条序列的全局表征。该隐向量融合了序列内所有时间步的有效信息，可作为整条时序片段的聚合摘要。

3.3. 门控融合机制

分阶段编码器提取的三组阶段表征 h_1 、 h_2 、 h_3 蕴含不同层次、不同信息充分度的成本预测信号，若采用固定权重方案则缺乏针对不同项目和不同阶段的适应性调整能力。基于此类问题，GMST通过两级结构实现自适应融合。

第一级为阶段预测头：各阶段表征经独立的线性映射层将 d_{model} 维向量压缩为标量成本预测值，如下式所示：

$$\hat{y}_k = W_k^{pred} \cdot h_k + b_k^{pred}, \quad k = 1, 2, 3 \tag{9}$$

第二级为门控权重网络：以当前所处的项目阶段 $o_k \in R^3$ 和项目静态特征 $c \in R^5$ （由项目类型3维、电压等级和地域共5维组成，这些特征在项目全生命周期内保持不变）作为条件输入，经过一个轻量级的两层MLP输出三阶段融合权重，如下式所示：

$$(\alpha_1, \alpha_2, \alpha_3) = \text{softmax}(MLP_{gate}([o_k; c])) \tag{10}$$

门控MLP的网络结构为Linear(3+5→32)→ReLU→Linear(32→3)，包含约300个可学习参数，计算开销极低。最终的综合预测值为三阶段预测的加权融合，如下式所示：

$$\hat{y} = \alpha_1 \hat{y}_1 + \alpha_2 \hat{y}_2 + \alpha_3 \hat{y}_3 \tag{11}$$

门控网络的核心创新是以当前阶段标识 ok 为条件输入，使模型结合各阶段信息完备程度自适应调整多源信息融合权重。实验结果与设计预期一致：

（1）规划阶段（Stage 1）：门控输出 $\alpha_1 \approx 0.45$ ，优先采信历史先验预测结果，并为未观测的后期信息保留先验权重；

（2）实施阶段（Stage 2）：随施工进度、动态成本等实测数据接入， $\alpha_2 \approx 0.35$ ，模型由经验主导逐步过渡为实测数据主导；

（3）结算阶段（Stage 3）：掌握签证金额、审核调差等决定最终造价的完备信息， $\alpha_3 \approx 0.63$ ，该支路信息占据融合主导权重。

模型的训练损失函数为标准均方误差：

$$\mathcal{L} = \left(\frac{1}{N}\right) \sum_{i=1}^N (\hat{y}_i - y_i)^2 \tag{12}$$

4. 实验分析

4.1. 数据集与实验设置

实验数据来源于某省级电网公司2019-2024年已完成结算的1876个技改项目，涵盖输电线路改造（603个，32.1%）、变电站设备更换（768个，40.9%）和配网自动化升级（505个，26.9%）。按项目开工时间8:1:1划分为训练/验证/测试集，连续特征Z-score标准化，目标为结算成本（万元）。

4.2. 综合评价

为验证GMST模型的优势性，选用XGBoost、LSTM、GRU、标准Transformer等方法进行同步预测[15, 16]，并采用MAE、RMSE、MAPE等指标进行综合评价，各方法的成本预测结果综合评价如表2所示。

表 2 各预测方法结果综合评价。

模型	MAE(万元)↓	RMSE(万元)↓	MAPE(%)↓
XGBoost	22.13±0.45	31.26±0.72	10.21±0.18%
LSTM	18.45±0.38	26.93±0.65	8.67±0.15%
GRU	17.82±0.41	25.61±0.58	8.35±0.14%
标准 Transformer	15.82±0.32	22.67±0.48	7.18±0.12%
GMST(Ours)	12.37±0.28	18.54±0.43	5.82±0.10%

由上表可知，首先，LSTM、GRU标准Transformer预测结果在三个评价指标上均优于XGBoost，MAPE的绝对降幅在1.5%~3.0%之间，充分验证了将项目成本建模为动态演化过程并使用时序模型进行预测的必要性。其次，标准Transformer结果优于LSTM和GRU，说明自注意力机制的全局依赖建模能力能够同时更好的捕捉规划期静态属性与实施期动态变量之间的长期关联。最后，GMST的MAPE为5.82%，为所有方法中的最优，相比标准Transformer降低19.0%（7.18%→5.82%），相比LSTM降低

32.9%（8.67%→5.82%），相比XGBoost降低43.0%（10.21%→5.82%），充分证明了分阶段独立编码与门控自适应融合机制的有效性。

4.3. 消融实验

为量化GMST各组成模块的单独作用，采用完整的消融实验进行验证，消融结果如表3所示。

表 3 GMST 消融结果。

消融模型	MAE	RMSE	MAPE	相对完整模型 ΔMAPE
GMST（完整模型）	12.37	18.54	5.82%	—
等权平均	15.15	22.41	7.34%	↑26.1%（恶化）
单共享编码器	14.48	21.76	6.91%	↑18.7%（恶化）
标准 Transformer	16.82	24.35	8.03%	↑37.9%（恶化）

由上表可知，完整GMST模型MAE、RMSE、MAP 均为最优。移除门控融合改用等权平均后MAPE上升26.1%，去掉分阶段编码换单共享编码器上升18.7%，说明门控融合对性能增益更大；无两大核心模块的标准Transformer MAPE恶化37.9%，该损失低于两者单独恶化幅度之和，证明分阶段编码与门控融合存在正向协同，二者搭配可显著降低预测误差。

5. 结论

本文面向电网技改项目全生命周期动态成本预测难题，构建了门控分阶段 Transformer（GMST）预测模型，针对技改项目分阶段特征差异大、信息完备度随施工推进逐步提升的业务特性，设计独立分阶段编码器与自适应门控融合模块，实现多阶段特征差异化提取与动态加权聚合。基于 1876 项真实电网技改数据开展多组对比实验，结果表明 GMST 在 MAE、RMSE、MAPE 三项指标上全面优于 XGBoost、LSTM、标准 Transformer 等主流基线模型，预测误差得到大幅削减。配套消融实验进一步厘清两大核心组件的独立作用与耦合关系，证实门控融合对模型精度的提升贡献高于分阶段编码，且二者存在显著正向协同效应，分阶段编码输出高质量阶段表征，门控机制动态调配各阶段预测权重，二者相互赋能共同优化预测效果。

本次研究可为电网企业技改项目投资估算、造价管控提供精准量化支撑。后续可进一步引入项目外部宏观经济、材料价格波动等外部变量拓展特征维度，同时轻量化模型结构，使其适配工程现场轻量化部署场景，持续提升成本预测的实用性与泛化能力。

参考文献

[1] 谢旭峰, 陈上上, 潘攀, 等. 电力工程项目造价管理研究 [J]. 中国招标, 2024, (6): 99-101.

[2] 孔祥靖, 韩松, 郝成亮, 等. 基于改进灰色模型的电力数据预测方法研究 [J]. 电子设计工程, 2021, 29(5): 139-143. <https://doi.org/10.14022/j.issn1674-6236.2021.05.029>

[3] 周磊, 竺筱晶. 基于 MA-CNN-LSTM 和自注意力机制的单变量短期电力负荷预测 [J]. 科学技术与工程, 2024, 24(22): 9408-9416. <https://doi.org/10.12404/j.issn.1671-1815.2305659>

[4] 刘朋远, 苗光尧, 张洁. 基于 LSTM-XGBoost 的短期电力负荷预测研究 [J]. 自动化仪表, 2023, 44(8): 79-83+90. <https://doi.org/10.16086/j.cnki.issn1000-0380.2021040043>

[5] 李柏杰, 朱贤伟, 王伟, 等. 结合 Transformer 与 TimeGAN 的智慧电厂电力负荷预测方法 [J]. 国外电子测量技术, 2025, 44(4): 211-216. <https://doi.org/10.19652/j.cnki.femt.2504033>

[6] 李佳兴. 基于并行 CNN-LSTM 时空特征提取及强化的短期电力负荷预测研究 [D]. 南昌大学, 2023. <https://doi.org/10.27232/d.cnki.gnchu.2023.003134>

[7] 毛远宏, 孙琛琛, 徐鲁豫, 等. 基于深度学习的时间序列预测方法综述 [J]. 微电子学与计算机, 2023, 40(4): 8-17. <https://doi.org/10.19304/J.ISSN1000-7180.2022.0725>

[8] 吴海翔. 基于跨时空关联的Transformer多元时序分析方法研究 [D]. 江苏大学, 2025. <https://doi.org/10.27170/d.cnki.gjsuu.2025.001909>

[9] 陈柯. 基于序列分解的长时序预测模型研究 [D]. 海南大学, 2025. <https://doi.org/10.27073/d.cnki.ghadu.2025.000398>

[10] 柯望. 基于注意力机制的多变量时间序列预测方法研究 [D]. 上海财经大学, 2024. <https://doi.org/10.27296/d.cnki.gshcu.2024.001060>

[11] 李嘉. 电网企业大修技改项目全生命周期风险管理体系研究 [J]. 工程技术研究, 2021, 6(17): 207-208. <https://doi.org/10.19537/j.cnki.2096-2789.2021.17.090>

[12] 刘婉婷. 精益化管理导向下电网工程预算编制的效率提升策略 [J]. 中国市场, 2026, (18): 60-63. <https://doi.org/10.13939/j.cnki.zgsc.2026.18.015>

[13] 彭延俊. 基于预训练技术的风功率预测研究 [D]. 内蒙古大学, 2025. <https://doi.org/10.27224/d.cnki.gnmdu.2025.001612>

- [14] 冀泽宇. 分布式轻量化非侵入式电力负荷监测方法研究 [D]. 吉林大学, 2023.
<https://doi.org/10.27162/d.cnki.gjlin.2023.001028>
- [15] 张巧艳. 基于深度学习的电气工程中电力设备状态监测与预警系统研究 [J]. 电工技术, 2024, (S2): 102-104+108.
<https://doi.org/10.19768/j.cnki.dgjs.2024.26.036>
- [16] 吴英男, 赵思宇. 基于深度学习的工程造价预测模型研究 [J]. 现代工程科技, 2025, 4(13): 1-4.
<https://doi.org/10.26929/j.cnki.issn.2097-1672.2025.13.001>