

增强随机集成的混合核 K 近邻算法的基站网络流量模型

孙宁^{1,2}, 李卓轩³, 时欣利^{1,3*}, 孙霏翀⁴, 许明杰^{1,5}, 曹进德³

(1. 东南大学网络空间安全学院, 江苏南京 211189; 2. 中国联合网络通信有限公司广州市分公司, 广东广州 510630;
3. 东南大学数学学院, 江苏南京 211189; 4. 华南理工大学信息网络工程研究中心,
广东广州 510641; 5. 国电南瑞科技股份有限公司, 江苏南京 211106)

摘要:面向5G/6G超密集组网的基站网络流量预测需求,提出一种增强随机集成混合核 K 近邻算法(enhanced random ensemble-based mixed kernel K -nearest neighbor algorithm, ER-MKKNN)。通过融合径向基函数与白噪声核构建混合核函数,突破了单一核函数在非线性和噪声抑制间的平衡瓶颈。创新性地引入样本-特征双重随机子采样与超参数区间随机化策略,显著提升了高维稀疏场景的泛化稳定性。基于袋外误差反演的动态权重分配机制,提升了算法对流量突变的鲁棒响应能力。配套设计的多层并行化架构,为超密集组网提供了可扩展的预测解决方案。实验表明,ER-MKKNN在均方根误差、平均绝对百分比误差和平均绝对误差三项指标上均优于所对比深度学习模型,为智能网络运维提供了新的技术路径。

关键词:基站网络流量预测;混合核 K 近邻算法;增强随机集成;多层并行架构

中图分类号:TP18 **文献标志码:**A **文章编号:**1001-2486(2025)06-024-12



论
文
拓
展

A model for base station network traffic prediction using an enhanced random ensemble-based mixed kernel K -nearest neighbor algorithm

SUN Ning^{1,2}, LI Zhuoxuan³, SHI Xinli^{1,3*}, SUN Peichong⁴, XU Mingjie^{1,5}, CAO Jinde³

(1. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China;

2. China United Network Communications Corporation Guangzhou Branch, Guangzhou 510630, China;

3. School of Mathematics, Southeast University, Nanjing 211189, China; 4. Information and Network Engineering Research Center, South China University of Technology, Guangzhou 510641, China; 5. NARI Technology Co., Ltd., Nanjing 211106, China)

Abstract: An ER-MKKNN (enhanced random mixed kernel K -nearest neighbors algorithm) was developed to meet the requirements of base station network traffic prediction in ultra-dense 5G/6G environments. A hybrid kernel function was formed by combining a radial basis function kernel with a white-noise kernel, thereby overcoming the trade-off between nonlinear relationship modeling and noise suppression that plagues single-kernel methods. Dual random subsampling of both samples and features, together with a randomized hyperparameter-interval strategy, was employed to bolster generalization stability in high-dimensional, sparse settings. A dynamic weight-allocation mechanism based on inversion of out-of-bag errors was introduced to improve robustness against abrupt traffic fluctuations. Finally, a multi-level parallel architecture was implemented to deliver a scalable prediction framework for ultra-dense network topologies. Experimental evaluations show that ER-MKKNN outperformed deep-learning models in root mean square error, mean absolute percentage error and mean absolute error, respectively, establishing a new technical pathway for intelligent network operations and maintenance.

Keywords: base station network traffic prediction; mixed kernel K -nearest neighbor algorithm; enhanced random integration; multi-layer parallel architecture

作为现代通信基础设施,蜂窝网络采用分布式架构,构建起覆盖广泛的无线通信体系。其核心构成为密集分布的蜂窝小区,作为独立服务单

元,每个小区依托基站实现区域内移动终端设备的通信管理与控制。在该系统中,“网络负载”特指基站承载的数据传输量及信道资源占用率,是

收稿日期:2025-06-05

基金项目:广州市重点领域研发计划2022年度“新一代信息技术”重大科技专项资助项目(202206070006)

第一作者:孙宁(1971—),男,广东梅州人,高级工程师,博士研究生,硕士生导师,E-mail:gdgzsun@139.com

*通信作者:时欣利(1988—),男,河南商丘人,副教授,博士,博士生导师,E-mail:xinli_shi@seu.edu.cn

引用格式:孙宁,李卓轩,时欣利,等.增强随机集成的混合核 K 近邻算法的基站网络流量模型[J].国防科技大学学报,2025,47(6):24-35.

Citation:SUN N, LI Z X, SHI X L, et al. A model for base station network traffic prediction using an enhanced random ensemble-based mixed kernel K -nearest neighbor algorithm[J]. Journal of National University of Defense Technology, 2025, 47(6): 24-35.

直接反映网络运行状态的关键指标。

当前移动通信技术正经历代际跃迁,5G 商用深化与 6G 研发推进正不断催生革命性应用场景。新一代通信技术将重构三大应用领域^[1]:一是智能交通领域,二是沉浸式交互领域,三是工业互联网领域。这些高带宽、低时延的创新应用将引发流量结构的根本性变革。更值得关注的是视频流量的爆发式增长,2025 年移动视频流量占比将攀升至 74%,年均复合增长率达 35%,这将持续加剧网络负载的动态不均衡特性^[2]。对于基站流量的预测,将成为解决网络负载问题的重点研究内容。

传统的基站网络流量预测方法通常采用统计时间序列方法,例如自回归积分移动平均模型^[3](autoregressive integrated moving average model, ARIMA)。随着人工智能方法的广泛应用^[4-8],越来越多的基站网络流量预测模型引入了机器学习方法。Rohini 等^[9]基于深度学习的长短期记忆(long short-term memory, LSTM)网络进行网络流量预测,因其长期记忆能力,被用于实现更精准的蜂窝网络流量预测,以支持动态资源分配和基站节能优化。Qiu 等^[10]提出了一种基于 LSTM 的多任务学习框架,该框架通过共享层提升泛化能力,通过融合注意力机制与卷积模块构建全局空间关联、全局时间关联及局部时空依赖的三维分析框架,在真实蜂窝流量数据集上实现了对复杂时空特征的全面挖掘与预测性能的显著提升。Huang 等^[11]提出一种基于条件概率建模的移动网络流量预测神经网络方法,通过对原始数据进行窗口化预处理、相邻窗口条件概率参数化及集成预测,构建了一种多尺度时序依赖建模框架。Zhang 等^[12]设计了一个 Inverted Transformer 模型来准确预测蜂窝流量。这些方法计算资源消耗大,动态适应性差,难以捕捉基站网络的动态拓扑变化。

除此之外,集成学习算法也受到研究人员的关注^[13-14]。Yamada 等^[15]在实时移动流量预测任务中使用随机森林(random forest, RF)预测网络流量,根据对预测准确性的贡献对流量日志数据进行优先级排序。Xia 等^[16]采用 RF 算法进行特征筛选,进一步将轻量级梯度提升机(light gradient boosting machine, LGB)应用于移动网络流量预测。Gijón 等^[17]基于全国范围内 30 个月的小区数据,对比了监督学习与时间序列分析方法在小区级月度高峰数据流量预测中的性能,结果表明随机森林、支持向量回归等监督学习模型具有更高的预测精度。在降低存储需求的同时超

越季节性自回归等传统方法,且非深度学习方法在计算效率上可与深度学习竞争,为长期流量预测提供了高效解决方案。

与上述工作不同的是,本文提出了一种增强随机集成混合核 K 近邻算法(enhanced random ensemble-based mixed kernel K -nearest neighbor algorithm, ER-MKKNN),在基站网络流量预测中有着更好的精度和鲁棒性。

1 增强随机集成混合核 K 近邻算法

1.1 混合核 K 近邻算法

1.1.1 算法原理

本文提出了一种混合核 K 近邻算法(mixed kernel K -nearest neighbor algorithm, MKKNN),突破传统 K 近邻算法(K -nearest neighbor algorithm, KNN)单一核函数的局限性,有效平衡非线性关系建模与噪声鲁棒性。通过将径向基函数(radial basis function, RBF)和白噪声核整合到统一框架中,混合核方法能够在异构特征空间中自适应地度量样本相似性。最终预测值通过对 k 个邻居的核值归一化加权平均获得。MKKNN 结构如图 1 所示。

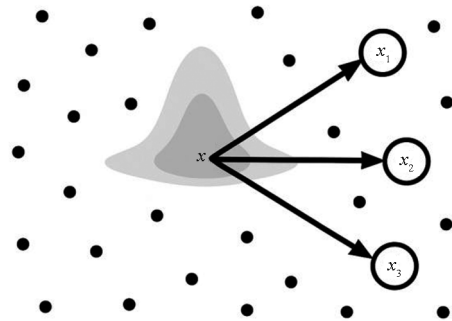


图 1 混合核 K 近邻

Fig. 1 Mixed kernel K -nearest neighbor

图 1 示意了 MKKNN 预测的基本原理。图中黑点表示特征空间中的已知训练样本,位于中央的 x 为当前待预测的目标样本(查询点)。灰色阴影区域表示以 x 为中心的核函数权重分布,颜色越深代表距离目标样本更近、权重值更高。 x_1 、 x_2 与 x_3 为根据度量从训练样本中选取的三个最近邻样本。

KNN 是一种经典的非参数回归与分类方法^[18]。传统 KNN 基于欧氏距离度量,相似性度量在存在特征尺度差异或非线性关系时可能表现不佳。核化 KNN 通过 Mercer 核替换距离度量^[19],但通常只使用单一核。本文将推广为仅由径向基函数和白噪声核构成的核混合形式:

$$K_{\text{mix}}(x, y) = w_{\text{rbf}} \exp(-\gamma \|x - y\|^2) + w_{\text{noise}} \sigma_n^2 \delta_{xy} \quad (1)$$

其中: $w_{\text{rbf}}, w_{\text{noise}} \geq 0$ 为对应权重; γ 为 RBF 核带宽参数; σ_n^2 为噪声方差。令 $X_{\text{train}} = \{x_i\}_{i=1}^n$, 那么 $\delta_{xy} = \sum_{i=1}^n \mathbb{I}(\|x - x_i\| < \epsilon) \mathbb{I}(\|y - x_i\| < \epsilon)$, 其中 \mathbb{I} 为定义的指示函数, ϵ 为阈值。RBF 核在小尺度内对相似样本赋予高权重, 能够精准刻画特征空间的平滑非线性关系。白噪声核仅在自相似位置贡献常量相似度, 等价于对离群样本施加最小影响, 显著提高整体鲁棒性, 该混合核实现了多尺度特征捕获。

设训练集为 $\{(x_i, y_i)\}_{i=1}^n$, 其中 $x_i \in \mathbb{R}^d, y_i \in \mathbb{R}$ 。对任意测试点 $x \in \mathbb{R}^d$, 先计算混合核相似向量:

$$\mathbf{k}(x) = [K_{\text{mix}}(x, x_1), \dots, K_{\text{mix}}(x, x_n)]^T \in \mathbb{R}^n \quad (2)$$

令 $N_k(x)$ 表示相似度排名前 k 的索引集合, 则预测值为

$$\hat{y} = \sum_{j \in N_k(x)} \frac{K(x, x_j)}{\sum_{\ell \in N_k(x)} K(x, x_\ell)} y_j \quad (3)$$

具体步骤如算法 1 所示。

算法 1 混合核 K 近邻算法

Alg. 1 Mixed kernel K -nearest neighbor algorithm

输入: 训练集 $\{(x_i, y_i)\}_{i=1}^n$, 邻居数 k

输出: 预测函数 $\hat{y}(x)$

将训练样本 $\{x_i, y_i\}$ 存于内存

for 每个测试样本 x

 计算混合核值:

$$K(x, x_i) \leftarrow w_{\text{rbf}} \exp(-\gamma \|x - x_i\|^2) + w_{\text{noise}} \sigma_n^2 \delta_{xx_i}$$

 找到相似度最高的 k 个索引集合 $N_k(x)$

 对每个 $i \in N_k(x)$, 计算归一化权重:

$$\omega_i = \frac{K(x, x_i)}{\sum_{j \in N_k(x)} K(x, x_j)}$$

 预测值 $\hat{y}(x) \leftarrow \sum_{i \in N_k(x)} \omega_i y_i$

end for

1.1.2 混合核正定性分析和距离分析

定义 1 (正定矩阵) 给定样本集 $\{x_1, \dots, x_n\} \subset \mathcal{X}$, 存在

$$\mathbf{K} = \begin{bmatrix} K(x_1, x_1) & K(x_1, x_2) & \dots & K(x_1, x_n) \\ K(x_2, x_1) & K(x_2, x_2) & \dots & K(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ K(x_n, x_1) & K(x_n, x_2) & \dots & K(x_n, x_n) \end{bmatrix}, \quad \mathbf{K}: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \quad (4)$$

则 \mathbf{K} 为正定矩阵。

定义 2 (半正定 (positive semi-definite, PSD) 核) 核函数若对任意样本集构造的正定矩阵 \mathbf{K} 满足

$$\begin{cases} \mathbf{K} = \mathbf{K}^T \\ \forall \mathbf{c} \in \mathbb{R}^n, \mathbf{c}^T \mathbf{K} \mathbf{c} \geq 0 \end{cases} \quad (5)$$

则称 \mathbf{K} 为 PSD 核。

定理 1 设 $\mathbf{K}_1, \mathbf{K}_2$ 均为 PSD 核, 且权重 $w_1, w_2 \geq 0$ 满足 $w_1 + w_2 = 1$, 则

$$K_{\text{mix}}(x, y) = w_1 K_1(x, y) + w_2 K_2(x, y) \quad (6)$$

亦为 PSD 核。

证明: 对任意 $\mathbf{c} \in \mathbb{R}^n$, 有

$$\mathbf{c}^T \mathbf{K}_{\text{mix}} \mathbf{c} = w_1 \mathbf{c}^T \mathbf{K}_1 \mathbf{c} + w_2 \mathbf{c}^T \mathbf{K}_2 \mathbf{c} \geq 0 + 0 = 0,$$

则 \mathbf{K}_{mix} 为 PSD 核。 □

定理 2 (Mercer 定理) 若 \mathbf{K} 为连续 PSD 核, 则存在一个希尔伯特空间 \mathcal{H} 及映射 $\phi: \mathcal{X} \rightarrow \mathcal{H}$, 使得

$$K(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}} \quad (7)$$

证明: 对连续的 PSD 核 \mathbf{K}_{mix} , 可在 $\mathcal{L}^2(\mathcal{X})$ 上做谱展开:

$$K_{\text{mix}}(x, y) = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i(y), \quad \lambda_i \geq 0$$

其中, $\{\psi_i\}$ 为正交归一的特征函数, 对应特征值 $\{\lambda_i\}$ 。定义映射

$$\phi: x \mapsto (\sqrt{\lambda_1} \psi_1(x), \sqrt{\lambda_2} \psi_2(x), \dots) \in \ell^2$$

则在 ℓ^2 中有

$$\langle \phi(x), \phi(y) \rangle_{\ell^2} = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i(y) = K(x, y)$$

取 $\mathcal{H} = \overline{\text{span}\{\phi(x)\}}$, 即得所需再生核希尔伯特空间。 □

定理 3 (Representer 定理) 在核正则化的经验风险最小化问题

$$\min_{f \in \mathcal{H}} \sum_{i=1}^n \ell(y_i, f(x_i)) + \lambda \|f\|_{\mathcal{H}}^2 \quad (8)$$

中, 其解 f^* 可表示为

$$f^*(\cdot) = \sum_{i=1}^n \alpha_i K_{\text{mix}}(x_i, \cdot) \quad (9)$$

证明: 令 $\{K_{\text{mix}}(x_i, \cdot)\}_{i=1}^n$ 在 \mathcal{H} 中生成子空间 \mathcal{H}_0 , 任取 $f \in \mathcal{H}$, 可分解为 $f = f_0 + f_{\perp}$, $f_0 \in \mathcal{H}_0$, $f_{\perp} \in \mathcal{H}_0^{\perp}$ 。由于

$$f_{\perp}(x_i) = \langle f_{\perp}, K_{\text{mix}}(x_i, \cdot) \rangle_{\mathcal{H}} = 0$$

经验损失 $\sum_i \ell(y_i, f(x_i)) = \sum_i \ell(y_i, f_0(x_i))$ 与 f_{\perp} 无关, 但正则项 $\|f\|_{\mathcal{H}}^2 = \|f_0\|_{\mathcal{H}}^2 + \|f_{\perp}\|_{\mathcal{H}}^2$ 随 $\|f_{\perp}\|$ 增大而增大, 因此最优解必有 $f_{\perp} = 0$, 即

$$f^* \in \mathcal{H}_0$$

令 $f^*(\cdot) = \sum_{i=1}^n \alpha_i K(x_i, \cdot)$, 得证。 \square

推论 1 由 PSD 核可定义核距离

$$d_{K_{\text{mix}}}(x, y) = \sqrt{K_{\text{mix}}(x, x) + K_{\text{mix}}(y, y) - 2K_{\text{mix}}(x, y)} \quad (10)$$

满足度量性质,有助于 KNN 中最近邻查询的理论保证。

证明:对任意 $x, y, z \in \mathcal{X}$, 令

$$d_{K_{\text{mix}}}(x, y) = \|\phi(x) - \phi(y)\|_{\mathcal{H}}$$

则有非负性:

$$\|\phi(x) - \phi(y)\| \geq 0 \text{ 且 } d_{K_{\text{mix}}}(x, y) = 0 \Leftrightarrow \phi(x) = \phi(y)$$

对称性:

$$\|\phi(x) - \phi(y)\| = \|\phi(y) - \phi(x)\|$$

三角不等式:

$$\|\phi(x) - \phi(z)\| \leq \|\phi(x) - \phi(y)\| + \|\phi(y) - \phi(z)\|$$

由此 $d_{K_{\text{mix}}}$ 满足距离的三要素。 \square

综上所述, MKKNN 所用的混合核 $K_{\text{mix}}(x, y)$ 因其 PSD 的性质,不仅保证了对应的正定矩阵在再生核希尔伯特空间中的合法内积表示,也使得基于该核函数的距离度量 $d_{K_{\text{mix}}}(x, y) = \sqrt{K_{\text{mix}}(x, x) + K_{\text{mix}}(y, y) - 2K_{\text{mix}}(x, y)}$ 满足伪度量甚至度量的条件。进一步地, Representer 定理说明了在正则化经验风险最小化框架下,任意解都可在有限样本核基上展开,从而为 MKKNN 的可解释性和算法实现提供了理论支持。基于以上理论保证, MKKNN 在多核自适应融合、异常点抑制与核回归平滑估计等方面具有坚实的数学基础,为后续的一致性与收敛性分析奠定了良好基础。

1.2 增强随机集成

为了提高算法的预测精度,本文提出一种基于随机子集策略的增强集成回归框架,将 MKKNN 作为基学习器,该算法被称为 ER-MKKNN。

设训练集为 $D = \{(x_i, y_i)\}_{i=1}^N, x_i \in \mathbb{R}^d$ 。对第 m 个基学习器,从 $\{1, \dots, N\}$ 中随机不放回抽取 αN 个索引,记为实例子集 I_m ;从 d 维特征中随机选取 βd 维,记为特征子集 F_m ;设置随机采样超参数 $k_m \sim U(k_{\min}, k_{\max}), \gamma_m \sim U(\gamma_{\min}, \gamma_{\max})$;在子集 $\{(x_i^{(F_m)}, y_i) \mid i \in I_m\}$ 上训练 MKKNN,得到回归函数 f_m 。双重随机化打破了特征间的伪相关性,特别适用于基站流量数据中的时空耦合特征。

对每个基学习器,计算其袋外(out-of-bag,

OOB)样本误差样本集 $I_m^{\text{ob}} = \{1, \dots, N\} \setminus I_m$ 上的均方误差:

$$E_{\text{MSE}}^m = \frac{1}{|I_m^{\text{ob}}|} \sum_{i \in I_m^{\text{ob}}} (y_i - f_m(x_i^{(F_m)}))^2 \quad (11)$$

定义原始权重

$$r_m = \frac{1}{E_{\text{MSE}}^m + \varepsilon} \quad (12)$$

$$\omega_m = \frac{\exp(\ln r_m - \max_j \ln r_j)}{\sum_{j=1}^M \exp(\ln r_j - \max_k \ln r_k)} \quad (13)$$

其中, ε 为防零常数。该方法有效利用未参与训练的 OOB 样本评估模型性能,指数加权强化高性能模型贡献。

最终集成预测为

$$\hat{y}(x) = \sum_{m=1}^M \omega_m f_m(x^{(F_m)}) \quad (14)$$

具体步骤如算法 2 所示。

算法 2 增强随机集成混合核 K 近邻算法

Alg. 2 Enhanced random mixed kernel

K-nearest neighbor algorithm

输入:训练集 $\{(x_i, y_i)\}_{i=1}^N$, 基学习器数 M , 采样比例 α 、 β , 邻居数范围 $[k_{\min}, k_{\max}]$, 核宽度范围 $[\gamma_{\min}, \gamma_{\max}]$
输出:加权集成回归函数 $\hat{y}(x)$

for $m = 1$ to M

从 $\{1, \dots, N\}$ 中不放回抽取 αN 个实例索引 I_m

从特征维度中随机选取 βd 个索引 F_m

随机采样超参数 k_m, γ_m

在子集 $\{(x_i^{(F_m)}, y_i) \mid i \in I_m\}$ 上训练 MKKNN 得到 f_m

计算 OOB 索引 I_m^{ob} 上误差 E_{MSE}^m

end for

对每个 $i \in N_k(x)$, 计算归一化权重:

$$\omega_i = \frac{K(x, x_i)}{\sum_{j \in N_k(x)} K(x, x_j)}$$

预测值 $\hat{y}(x) \leftarrow \sum_{i \in N_k(x)} \omega_i y_i$

通过在实例子集与特征子集上随机采样并随机化超参数,生成多样化基学习器;利用无放回样本的 OOB 误差反演为权重,再经 Stable Softmax 归一化,实现加权融合,能够有效提升模型的多样性和在高维稀疏空间中的泛化能力。

1.3 多级并行化策略

为进一步提升训练与预测效率,本文在增强随机子集集成框架中引入多级并行化策略,包括基学习器级并行、OOB 误差并行计算、预测阶段

的并行融合和流水线与异步调度。多级并行化策略结构如图 2 所示。由于各基学习器互相独立,训练过程可按实例子集和特征子集并行执行,对 M 个基学习器同时调度:

$$\{f_m\}_{m=1}^M = P_{\text{parallel}}(n_{\text{jobs}}, T_{\text{MKKNN}}(I_m, F_m, k_m, \gamma_m))$$

其中, P_{parallel} 是并行执行算子, 参数 n_{jobs} 根据中央处理器 (central processing unit, CPU) 核数动态设定。

在所有基学习器训练完毕后, 针对各自的 OOB 子集独立计算均方误差 E_{MSE}^m 。同样可利用并行框架批量执行:

$$\{E_{\text{MSE}}^m\} = P_{\text{parallel}}(n_{\text{jobs}}, \ell_{\text{MSE}}(f_m, f_m^{\text{ob}})) \quad (15)$$

其中, $\ell_{\text{MSE}}(f_m, f_m^{\text{ob}})$ 为第 m 个基学习器在 OOB 样本上的均方误差损失, f_m 为模型预测值, f_m^{ob} 为对应的 OOB 实测值。并行完成后再统一进行 Stable Softmax 权重归一化。对新样本集 $\{x^{(F_m)}\}$, 各基学习器的预测 $\{f_m(x^{(F_m)})\}$ 同样可并行计算。最终加权求和仅涉及向量内积, 无须全局同步。在大规模数据场景下, 可将训练和 OOB 验证两步流水线化: 异步启动子任务, 先训练 f_m , 训练完成即提交其 OOB 误差计算, 再根据结果动态调整剩余基学习器的资源分配, 进一步提高集群利用率。

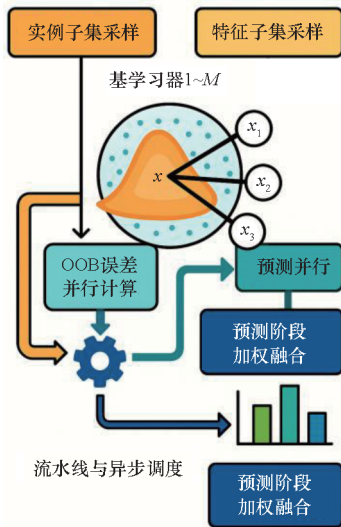


图 2 多级并行化策略改进的增强随机集成混合核 K 近邻算法

Fig. 2 Enhanced random mixed kernel K -nearest neighbor algorithm with improved multi-level parallelization strategy

2 实验与分析

本实验收集了深圳市 18 个小区 2018 年 3 月 1 日至 4 月 19 日的基站网络小时级流量数据, 选取前 4 周数据作为训练集, 第 5、6 周数据作为测试集。图 3 展示了样例基站一周的流量数据情况。具体统计描述如表 1 所示。

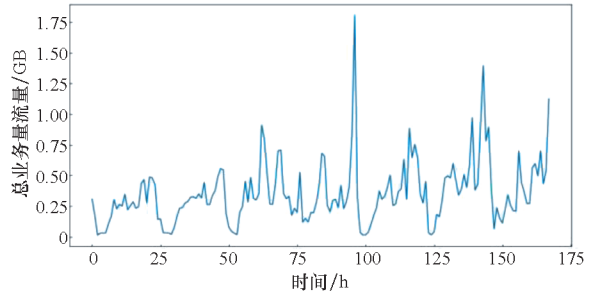


图 3 样例基站一周的流量数据

Fig. 3 A traffic data of the sample base station for one week

表 1 基站网络小时级流量数据统计描述

Tab. 1 Base station network hourly traffic data statistics description

单位: GB					
基站编号	均值	方差	最大值	最小值	中位数
1	0.535 5	0.583 8	4.904 5	0.000 028	0.105 8
2	0.185 0	0.036 1	1.299 0	0.000 686	0.128 3
3	0.064 0	0.012 7	1.129 0	0.000 190	0.029 8
4	0.259 9	0.119 1	2.632 7	0.000 531	0.152 2
5	0.524 5	0.185 5	3.166 6	0.003 189	0.464 2
6	0.173 5	0.211 7	10.471 9	0.000 001	0.029 3
7	0.178 5	0.031 5	1.157 4	0.001 170	0.122 2
8	0.515 1	0.167 8	3.510 6	0.002 791	0.458 5
9	0.253 1	0.173 9	9.611 8	0.000 335	0.155 9
10	1.156 0	0.871 7	10.823 8	0.029 684	0.902 3
11	0.089 6	0.008 8	0.898 0	0.000 290	0.067 9
12	0.441 0	0.130 9	3.375 2	0.005 253	0.364 8
13	0.606 7	0.154 7	2.475 4	0.002 017	0.626 7
14	1.323 4	0.662 0	4.748 4	0.014 504	1.312 9
15	0.702 7	0.712 8	5.504 2	0.000 361	0.480 5
16	1.462 9	1.620 2	6.069 2	0.001 714	1.225 0
17	0.049 3	0.004 5	0.629 7	0.000 012	0.028 1
18	0.074 7	0.006 8	1.145 7	0.000 325	0.057 1

本文在实时预测时, 利用过去 7 个时段 (共 7 h) 的流量数据来预测下一时段 (1 h) 的流量, 同时采用均方根误差 (root mean square error, RMSE)、平均绝对误差 (mean absolute error, MAE) 和平均绝对百分比误差 (mean absolute percentage error, MAPE) 作为基站网络流量的预测精度评价指标。

2.1 基站流量数据分析

首先, 对不同小区的基站流量数据进行了相关性分析, 采用皮尔逊相关系数对不同基站流量的相关性进行计算, 并绘制热度图, 如图 4 所示。

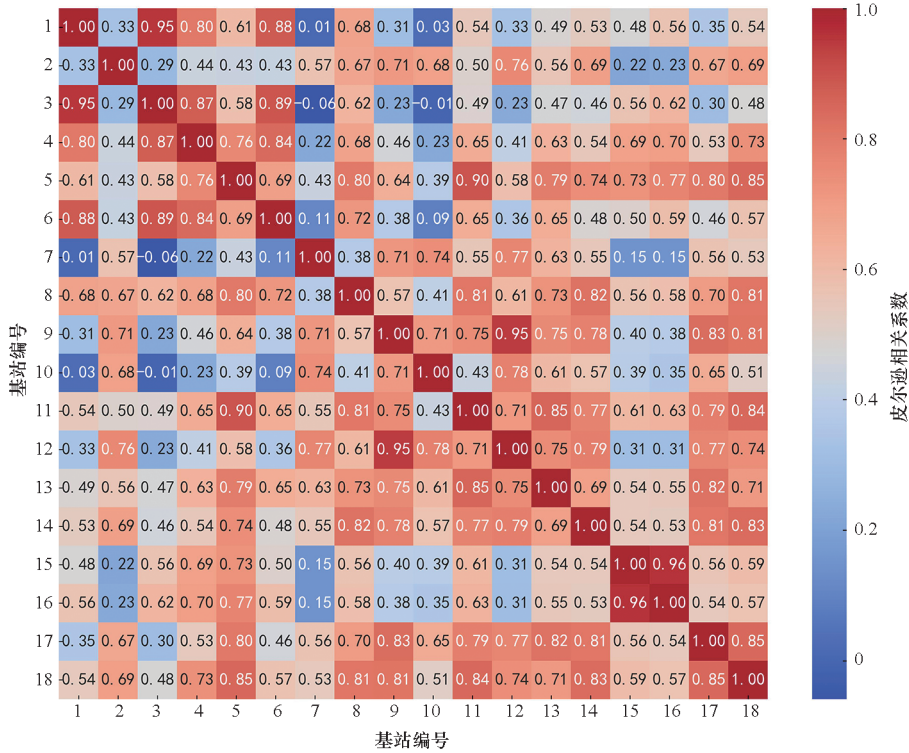


图 4 不同基站流量的相关性热度图

Fig. 4 Correlation heat map of traffic at different base stations

从图 4 中可以发现,大多数基站间的皮尔逊相关系数都在 0.4~0.7,说明大部分基站的小时流量模式具有一定相似性。这可能源自相邻基站覆盖区域内用户行为的共性(如上下班高峰、午间休息、夜间流量低谷等)相互重叠。少数基站间的相关系数接近 0(比如 1 与 7),提示它们的时段行为几乎不相关。

接下来,对不同小区的流量分布情况进行分析,测试正态、对数正态、韦伯和伽马分布,采用赤池信息准则选择小区基站流量的最优分布,所有基站的流量分布情况如表 2 所示,其中: k_{Γ} 为伽马分布的形状参数,决定峰度与偏态; θ_{Γ} 为伽马分布的尺度参数; σ_{LN} 为对数正态分布的形状参数; μ_{LN} 为对数正态分布的尺度参数; k_w 为韦伯分布的形状参数; λ_w 为韦伯分布的尺度参数; δ 表示分布的平移量。

2.2 MKKNN 的表现

为了验证混合核方法的有效性,本文分别与核 K 近邻算法(kernel K-nearest neighbor algorithm, KKN)、采用欧氏距离的 KNN(KNN - 欧氏距离)、采用马氏距离的 KNN(KNN - 马氏距离)、采用余弦距离的 KNN(KNN - 余弦距离)和采用动态时间规整(dynamic time warping, DTW)距离的 KNN(KNN - DTW 距离)进行比较,所有

表 2 不同小区基站流量的分布情况

Tab. 2 Distribution of traffic at different residential base stations traffic

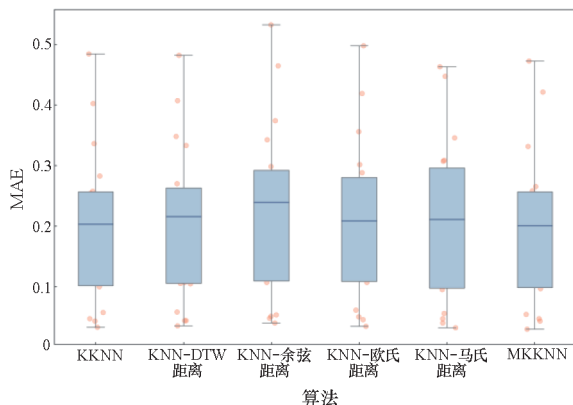
基站编号	最优分布	参数	分布特征描述
1	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.401\ 2, 2.760\ 9 \times 10^{-5}, 0.820\ 9)$	短时高频、强右长尾
2	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.876\ 5, 0.000\ 685\ 7, 0.207\ 2)$	单调下降、无峰值
3	对数正态分布	$(\sigma_{LN}, \delta, \mu_{LN}) = (1.970\ 1, 0.000\ 168\ 4, 0.018\ 22)$	偶有极端大值
4	韦伯分布	$(k_w, \delta, \lambda_w) = (0.753\ 9, 0.000\ 530\ 6, 0.214\ 5)$	早期密集、后期稀疏
5	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.937\ 3, 0.003\ 189\ 2, 0.556\ 0)$	最优但残差大
6	对数正态分布	$(\sigma_{LN}, \delta, \mu_{LN}) = (3.070\ 5, 3.424\ 6 \times 10^{-7}, 0.016\ 17)$	强尾性、偶发极大值
7	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (1.076\ 9, 0.001\ 149\ 95, 0.164\ 7)$	中度偏态、次峰可预测
8	韦伯分布	$(k_w, \delta, \lambda_w) = (1.111\ 6, 0.002\ 317\ 8, 0.530\ 5)$	后期事件增多呈“磨损”特征

(续表)

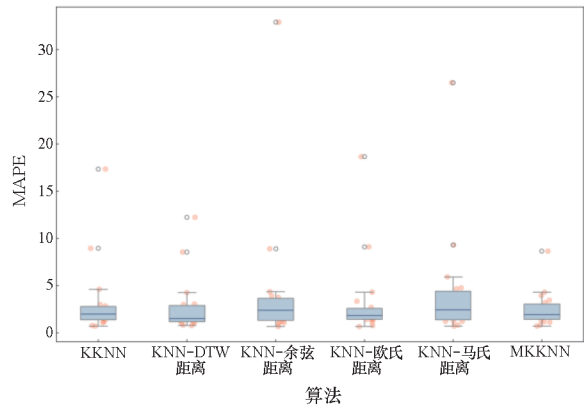
基站编号	最优分布	参数	分布特征描述
9	对数正态分布	$(\sigma_{LN}, \delta, \mu_{LN}) = (1.003\ 2, -0.014\ 04, 0.162\ 2)$	正偏、偶有小概率负偏
10	对数正态分布	$(\sigma_{LN}, \delta, \mu_{LN}) = (0.651\ 6, -0.155\ 35, 1.063\ 5)$	中度波动伴测量误差
11	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.951\ 1, 0.000\ 29, 0.093\ 81)$	平滑单调衰减
12	韦伯分布	$(k_w, \delta, \lambda_w) = (1.223\ 3, 0.004\ 778\ 5, 0.465\ 4)$	中后期事件略增
13	韦伯分布	$(k_w, \delta, \lambda_w) = (1.670\ 5, -0.059\ 04, 0.741\ 7)$	强“磨损”效应
14	韦伯分布	$(k_w, \delta, \lambda_w) = (1.806\ 0, -0.126\ 55, 1.625\ 9)$	延迟平移、后期拖尾明显
15	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.563\ 1, 0.000\ 360\ 7, 1.298\ 8)$	初期高密度、缓慢衰减
16	伽马分布	$(k_{\Gamma}, \delta, \theta_{\Gamma}) = (0.759\ 4, 0.001\ 713\ 6, 1.881\ 3)$	短时集中、尾部较厚
17	韦伯分布	$(k_w, \delta, \lambda_w) = (0.762\ 7, 1.155\ 3 \times 10^{-5}, 0.051\ 16)$	超短时初期集中爆发
18	韦伯分布	$(k_w, \delta, \lambda_w) = (0.882\ 7, 0.000\ 325\ 4, 0.071\ 28)$	与 17 类似但尾部稍厚

算法的超参数 $K = 5$ 。不同 KNN 变体在 RMSE、MAPE 和 MAE 上的性能比较如图 5 所示。不同 KNN 变体在 18 个基站网络流量预测中的平均表现如表 3 所示。

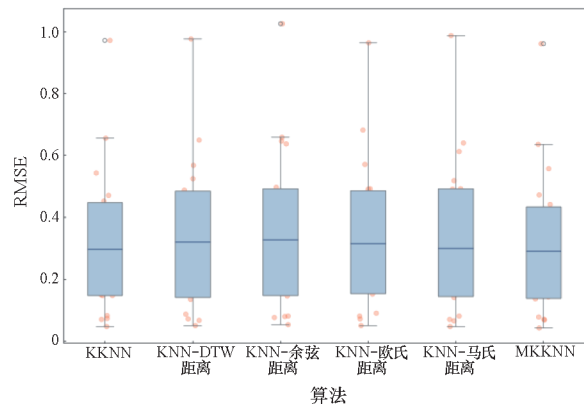
从表 3 可见,在相同超参数 $K = 5$ 条件下, MKKNN 在 18 个基站的平均流量预测中, RMSE 为 0.317 4、MAPE 为 2.518 7、MAE 为 0.195 4, 三项指标均优于单一距离度量的 KNN 方法。与其他模型相比, MKKNN 在 RMSE 上降低 2.1% ~



(a) MAE



(b) MAPE



(c) RMSE

图 5 不同 KNN 变体的性能对比箱线图

Fig. 5 Performance comparison boxplots of different KNN variants

表 3 不同 KNN 变体的性能对比

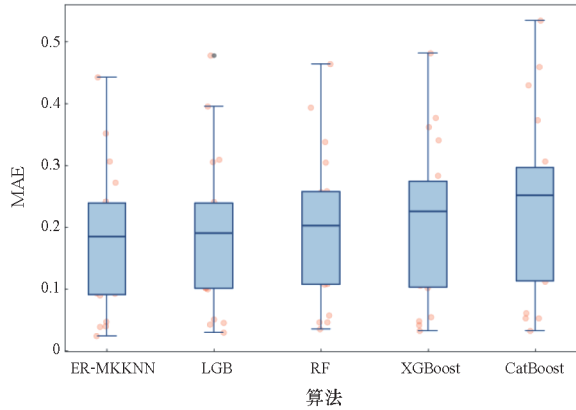
Tab. 3 Performance comparison of different KNN variants

模型	平均 RMSE	平均 MAPE	平均 MAE
KKNN	0.324 3	3.223 8	0.198 0
KNN - DTW 距离	0.339 7	2.758 2	0.203 4
KNN - 余弦距离	0.369 5	4.357 3	0.227 6
KNN - 欧氏距离	0.339 4	3.259 4	0.207 8
KNN - 马氏距离	0.339 5	4.257 6	0.206 1
MKKNN	0.317 4	2.518 7	0.195 4

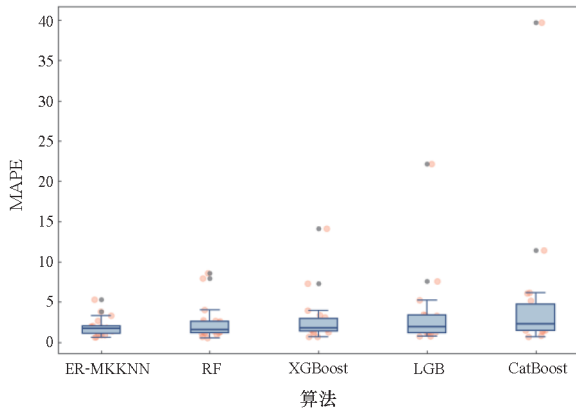
14.1%, MAPE 降低 8.7% ~ 42.2%, MAE 降低 1.3% ~ 14.1%。余弦距离因忽略幅度信息,在 MAPE 上波动最剧烈;DTW 虽强调时序对齐,但对突变与周期性波动敏感性不足, MKKNN 相比 KNN - DTW 距离仅降低 MAPE 8.7%。MKKNN 通过综合核策略兼顾周期性与非线性特征,显著抑制噪声,使 MAPE 降低高达 42.2%。这充分验证了混合核方法在多尺度特征捕捉与预测稳定性上的优势,而 KKNN 固定单一核的特性限制了其在多样流量模式下的适应性,相比 KNN, RMSE 仅降低约 2.1%。

2.3 增强随机集成 MKKNN 的表现

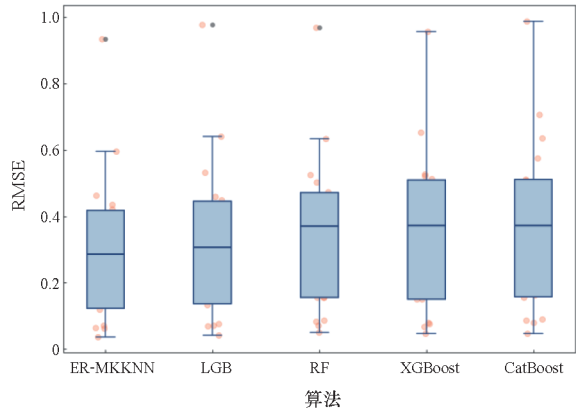
为了验证增强随机集成 MKKNN 的有效性,本文分别与先进的集成学习算法(包括 RF、XGBoost^[20]、LGB 和 CatBoost^[21])进行比较。对比较的集成学习算法参数采用的是默认推荐参数。集成学习算法在 RMSE、MAPE 和 MAE 上的性能比较如图 6 所示。集成学习算法在 18 个基站网络流量预测中的平均表现如表 4 所示。



(a) MAE



(b) MAPE



(c) RMSE

图 6 集成学习算法性能对比箱线图

Fig. 6 Performance comparison boxplots of ensemble learning algorithms

表 4 集成学习算法性能对比

Tab. 4 Performance comparison of ensemble learning algorithms

模型	平均 RMSE	平均 MAPE	平均 MAE
LGB	0.321 0	3.534 5	0.191 1
RF	0.342 1	2.506 5	0.199 6
XGBoost	0.347 7	2.917 2	0.208 4
CatBoost	0.365 9	5.203 7	0.231 5
ER-MKKNN	0.296 3	1.949 2	0.179 9

与 RF、XGBoost、LGB 和 CatBoost 四种先进集成学习方法进行对比,ER-MKKNN 在 18 个基站的网络流量预测中的 RMSE (0.296 3)、MAPE (1.949 2) 和 MAE(0.179 9) 均显著领先:与次优的 LGB 相比, RMSE 降低了约 7.7%、MAE 降低了约 5.9%;与 RF 相比, MAPE 降低了约 22.2%、MAE 降低了约 9.9%,表明通过随机子集成多核函数,不仅能更精准地捕捉时序模式,还能在峰谷波动中提供更稳定的预测性能。

基于 KNN 作为基础学习器,通过混合核函数实现多尺度特征映射,再结合增强随机集成机制以提高模型的泛化性能与稳健性。一方面利用核函数捕捉多尺度的局部相似性,另一方面通过随机化增强模型多样性,降低过拟合风险。相比树模型,核方法能更直观地以距离度量进行预测,因此在时序平滑、周期性和突发性特征并存时,表现更加稳定。ER-MKKNN 将 MKKNN 与双重随机化子集集成有机结合,在单模型层面实现多尺度特征捕捉,在集成层面最大化模型多样性并降低过拟合。两者协同作用下,不仅带来 RMSE、MAPE、MAE 的显著降低,也令模型在峰谷突变与噪声干扰中表现得尤为稳定与鲁棒,充分验证了其在网络流量预测中的创新价值。

2.4 多级并行化策略的表现

为验证多级并行化策略的有效性,构建了包含 20 000 个样本的合成数据集(特征维度 $d = 10$),其中 75% 的样本作为训练集,25% 的样本作为测试集,对其进行了对比实验,硬件环境为 32 核 CPU 集群,所有测试均在独占计算节点环境下进行。固定基学习器数量为 100,采用控制变量法分别测试 {1, 2, 4, 8, 16} 核的并行性能。每个配置重复运行 10 次以消除随机波动,最终结果取平均值,结果如图 7 所示,具体表现如表 5 所示。

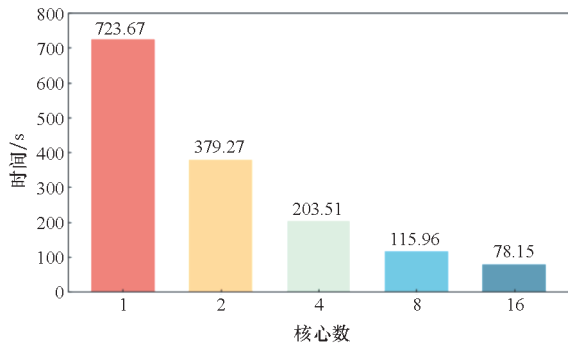


图7 多级并行化加速性能表现

Fig.7 Performance of multi-level parallelization acceleration

表5 多级并行化加速效果

Tab.5 Multi-level parallelization acceleration effect

核数	平均耗时 时间/s	理想加速 时间/s	实际 加速比
1	723.67	723.67	1.00
2	379.27	361.84	1.91
4	203.51	180.92	3.56
8	115.96	90.46	6.24
16	78.15	45.23	9.26

从表5中发现,多级并行化策略能够在保持高并行度(大于95%)的前提下,在中等规模核数(2~8核)上实现近线性加速(2核约1.9倍、4核约3.6倍、8核约6.2倍)。但当并行核数进一步

增至16核时,由于线程间同步、通信及内存带宽竞争等开销急剧增加,并行效率由8核时的77%下降至16核时的58%,出现明显的收益递减效应,但是整体的性能却进一步提升,从而验证了本文提出的多级并行化的有效性。

2.5 与不同的机器学习和深度学习模型比较

为系统验证ER-MKKNN的泛化能力,本研究设计了多维度对比实验。在传统机器学习模型方面,选取ARIMA、多层感知器^[22](multi-layer perceptron, MLP)、决策树^[23](decision tree, DT)和支持向量机^[24](support vector machine, SVM)作为基准模型,其实现基于Python的Scikit-learn^[25]库,均采用推荐默认参数配置。在与深度学习模型对比方面,进一步引入了堆叠自编码器^[26](stacked autoencoders, SAEs)、循环神经网络^[27](recurrent neural network, RNN)、门控循环单元^[28](gated recurrent unit, GRU)、LSTM、Transformer^[29]和时间卷积网络^[30](temporal convolutional network, TCN)前沿时序预测模型。实验设计中,所有深度学习模型均通过Adam优化算法进行模型训练,初始学习率设为0.001,以保持参数配置的一致性。特别地,通过构建涵盖传统机器学习与深度学习的多层次对照体系,充分确保了算法验证的全面性和结果的可信度。以小区1为例,多模型网络流量预测结果如图8所示。

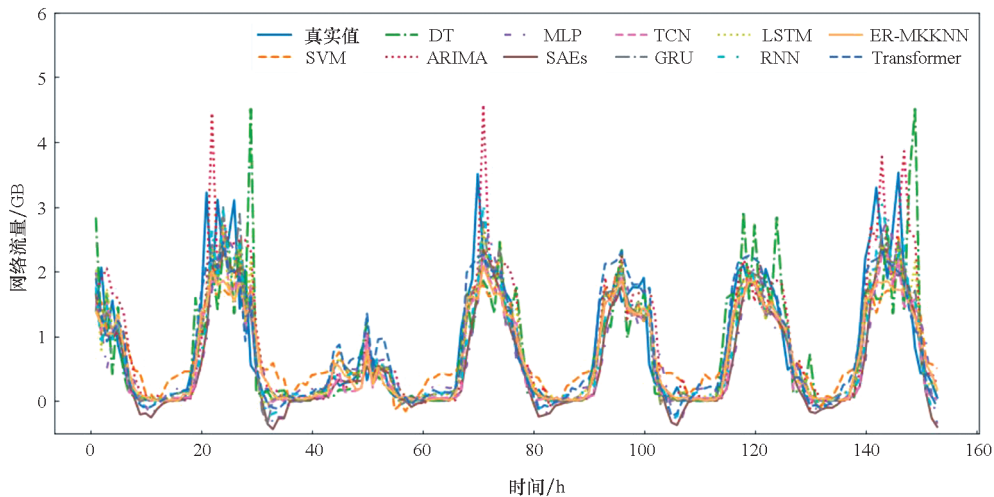


图8 多模型网络流量预测结果

Fig.8 Network traffic prediction results of multiple models

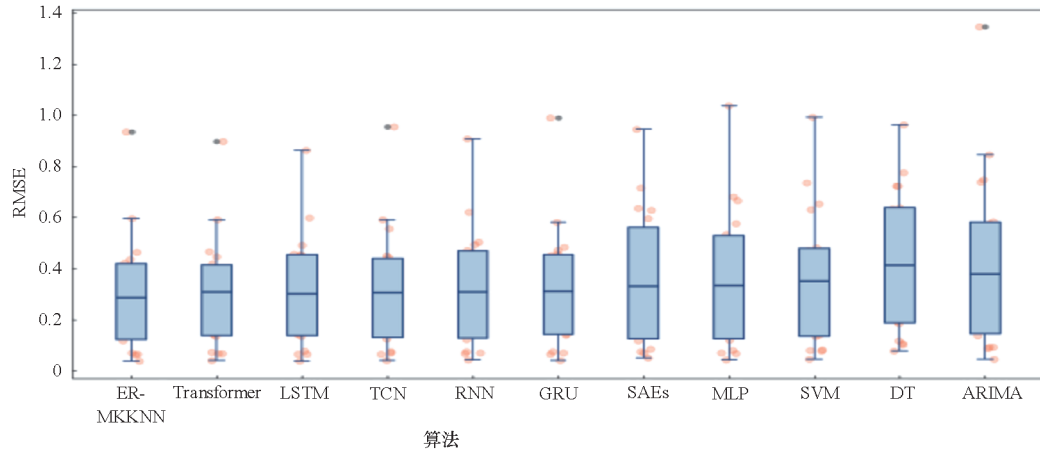
多模型在RMSE、MAPE和MAE上的性能比较如图9所示。多模型在18个基站网络流量预测中的平均表现如表6所示。

实验数据充分验证了ER-MKKNN的综合优

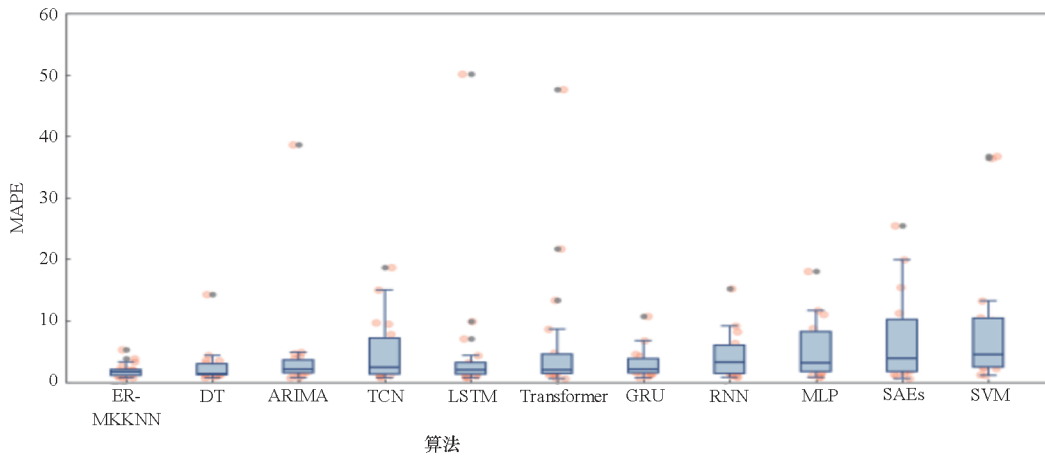
越性,其以0.2963的平均RMSE、1.9492的平均MAPE和0.1799的平均MAE全面领先所有对比模型。相较于传统机器学习领域表现最佳的MLP模型,ER-MKKNN在关键指标MAPE上实现

82.4% 的大幅降低,显著克服了 MLP 对数据异常值敏感的问题;面对深度学习领域表现最优的 LSTM 和 Transformer 模型,ER-MKKNN 在保持 RMSE 优势的基础上,其 MAPE 指标分别优于二者 63.5% 和 70.8%,凸显了混合架构在时序特征

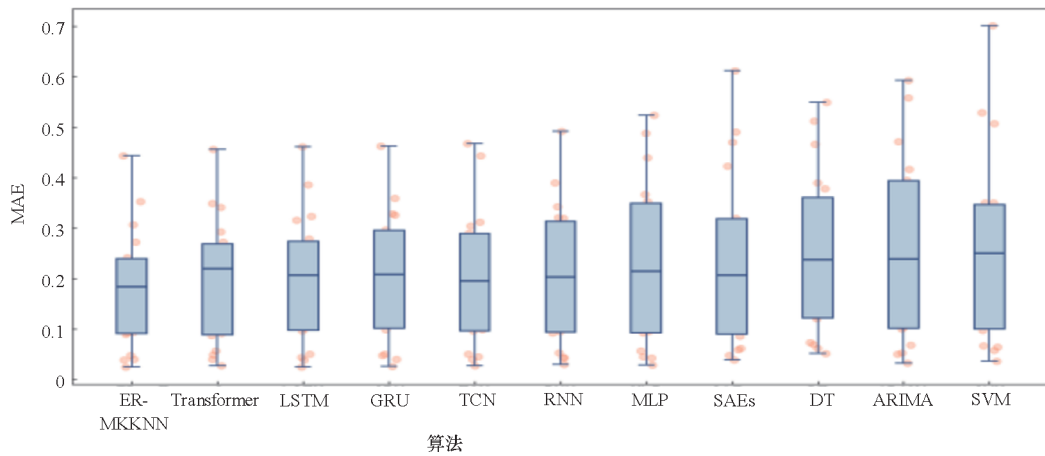
提取与误差控制方面的双重突破。特别是在处理基站流量数据的非线性特征时,有效规避了 DT 模型的高系统性偏差缺陷,同时解决了 SAEs 等生成式模型在判别特征学习上的局限性,最终达成预测精度与稳定性的协同优化。



(a) RMSE



(b) MAPE



(c) MAE

图 9 多模型综合性能对比箱线图

Fig. 9 Comprehensive performance comparison boxplots of multiple models

表 6 多模型综合性能对比

Tab. 6 Comprehensive performance comparison of multiple models

模型	平均 RMSE	平均 MAPE	平均 MAE
SVM	0.364 0	22.072 0	0.260 6
DT	0.422 8	2.586 1	0.247 9
MLP	0.358 0	11.082 0	0.232 6
ARIMA	0.428 1	4.460 9	0.257 6
ER-MKNN	0.296 3	1.949 2	0.179 9
Transformer	0.300 1	6.679 0	0.194 4
TCN	0.315 6	5.053 9	0.201 6
SAEs	0.357 2	11.744 4	0.239 5
GRU	0.322 6	8.255 6	0.200 6
LSTM	0.310 6	5.338 0	0.196 7
RNN	0.321 9	9.886 2	0.207 3

3 讨论

本文主要聚焦于基于时序模型的流量预测方法。为了更加全面地理解基站流量的波动特征,在后续的数据收集和模型构建方面,有必要进一步考察影响流量变化的多维因素,如发射功率和网络状况等。

发射功率越高,覆盖半径和小区容量越大,理论上可承载更多用户并分担热点区域的流量压力;但过高功率也会导致小区间干扰增大,反而影响用户体验。可通过采集不同功率档位下的基站业务量数据,构建功率-流量曲线;结合信号强度评估覆盖边缘用户接入成功率与重传率对流量的拉动或抑制作用。网络状况主要包括负载率、上行/下行时延、丢包率等,高负载率会使得调度资源紧张,用户等待时延增加,可能导致短时流量峰值被抑制或错峰;而丢包与重传会引发额外流量消耗,形成“假性高流量”。通过对多维因素的深入分析与建模,实现对基站负载的精准评估与动态资源调度。

4 结论

针对 5G/6G 时代基站网络流量预测面临的动态不平衡性与高维时空复杂性挑战,本文提出了一种融合混合核函数与增强随机集成的多级并行 K 近邻算法。

通过融合径向基函数与白噪声核构建混合核函数,模型既能捕捉相关性,也能针对流量突变进行局部“去噪”,实现了“信号”与“噪声”之间的有效平衡;结合样本-特征双重随机子采样与超

参数区间随机化策略,每次基模型只在部分基站和时间段数据上训练,使得各基模型之间的训练误差具有较低相关性,有效降低整体方差,对核函数的超参数在给定区间内随机采样,使得不同基模型聚焦于不同的时空尺度,从而进一步提升集成多样性;通过基于 OOB 误差反演的动态权重分配机制,增强了系统对流量突变的响应能力;同时设计的多级并行化架构,为 5G/6G 超密集组网的负载预测提供了高精度、强鲁棒、可扩展的一体化解决方案。

尽管所提多级并行 ER-MKNN 在预测精度与鲁棒性方面表现优异,但仍存在不足之处,ER-MKNN 架构天然关注局部相似性,难以捕捉更长范围的时序趋势与全局周期,该模型对多级时序依赖(如跨天、跨周)的显式建模能力仍较为有限。虽然在核函数中可融入空间维度信息,但 ER-MKNN 并未显式利用基站间的地理或网络拓扑关系。在后续的工作中,将通过时序分解融合和混合深度学习方法进行改进。

参考文献 (References)

- [1] MA B, GUO W S, ZHANG J. A survey of online data-driven proactive 5G network optimisation using machine learning[J]. IEEE Access, 2020, 8: 35606-35637.
- [2] MORLEY J, WIDDICKS K, HAZAS M. Digitalisation, energy and data demand: the impact of Internet traffic on overall and peak electricity consumption [J]. Energy Research & Social Science, 2018, 38: 128-137.
- [3] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control [M]. 5th ed. Hoboken, New Jersey: John Wiley & Sons, Inc., 2015.
- [4] 李敏, 李卓轩, 时欣利, 等. 基于可解释性集成学习的 RIOHTrack 车辙预测模型及驱动因素研究[J]. 应用数学和力学, 2025, 46(1): 92-104.
LI M, LI Z X, SHI X L, et al. Research on driving factors of the RIOHTrack rutting prediction model based on interpretable ensemble learning[J]. Applied Mathematics and Mechanics, 2025, 46(1): 92-104. (in Chinese)
- [5] 李卓轩, 林凯迪, 郭建华, 等. 基于车联网数据的运输车辆安全评价模型[J]. 南通大学学报(自然科学版), 2020, 19(1): 26-32, 47.
LI Z X, LIN K D, GUO J H, et al. Transportation vehicle safety evaluation model based on vehicle network data [J]. Journal of Nantong University (Natural Science Edition), 2020, 19(1): 26-32, 47. (in Chinese)
- [6] 张承焯, 李卓轩, 曹进德. 基于随机 k -近邻集成算法的网络流量入侵检测[J]. 南通大学学报(自然科学版), 2023, 22(3): 26-32.
ZHANG C Y, LI Z X, CAO J D. Network intrusion detection based on random k -nearest neighbor ensemble algorithm [J]. Journal of Nantong University (Natural Science Edition), 2023, 22(3): 26-32. (in Chinese)
- [7] 张玺君, 郝俊. EEMD + BiGRU 组合模型在短时交通流量预测中的应用[J]. 国防科技大学学报, 2023, 45(2):

- 73–80.
- ZHANG X J, HAO J. Application of EEMD + BiGRU combination model in short-term traffic flow prediction [J]. *Journal of National University of Defense Technology*, 2023, 45(2): 73–80. (in Chinese)
- [8] 李卓轩, 何桂仲, 卫一恒, 等. 基于分数阶优化的极限学习机交通流短时预测模型 [J]. *控制理论与应用*, 2025, 42(6): 1191–1199.
- LI Z X, HE G Z, WEI Y H, et al. Short-term traffic flow prediction model based on extreme learning machine with fractional-order optimization [J]. *Control Theory & Applications*, 2025, 42(6): 1191–1199. (in Chinese)
- [9] ROHINI G, GNANA K C, NAGENDRA S, et al. Autonomous forecasting of traffic in cellular networks based on long-short term memory recurrent neural network [J]. *Cybernetics and Systems*, 2025, 56(1): 33–45.
- [10] QIU C, ZHANG Y, FENG Z Y, et al. Spatio-temporal wireless traffic prediction with recurrent neural network [J]. *IEEE Wireless Communications Letters*, 2018, 7(4): 554–557.
- [11] HUANG J, XIAO M. Mobile network traffic prediction based on seasonal adjacent windows sampling and conditional probability estimation [J]. *IEEE Transactions on Big Data*, 2022, 8(5): 1155–1168.
- [12] ZHANG J W, TAN C C, CAI Z Y, et al. Cellular traffic forecasting based on inverted transformer for mobile perception dual-level base station sleep control [J]. *Ad Hoc Networks*, 2024, 161: 103505.
- [13] LI Z X, SHI X L, CAO J D, et al. CPSO-XGBoost segmented regression model for asphalt pavement deflection basin area prediction [J]. *Science China Technological Sciences*, 2022, 65(7): 1470–1481.
- [14] ZHOU Z, CAO J D, SHI X L, et al. Probabilistic rutting model using NGBost and SHAP: incorporating other performance indicators [J]. *Construction and Building Materials*, 2024, 438: 137052.
- [15] YAMADA Y, SHINKUMA R, SATO T, et al. Feature-selection based data prioritization in mobile traffic prediction using machine learning [C]//*Proceedings of the IEEE Global Communications Conference*, 2018.
- [16] XIA H W, WEI X, GAO Y, et al. Traffic prediction based on ensemble machine learning strategies with bagging and LightGBM [C]//*Proceedings of the IEEE International Conference on Communications Workshops*, 2019.
- [17] GIJÓN C, TORIL M, LUNA-RAMÍREZ S. Data-driven estimation of throughput performance in sliced radio access networks via supervised learning [J]. *IEEE Transactions on Network and Service Management*, 2023, 20(2): 1008–1023.
- [18] CHEN G H, SHAH D. Explaining the success of nearest neighbor methods in prediction [J]. *Foundations and Trends® in Machine Learning*, 2017, 10(5/6): 337–588.
- [19] YU K, JI L, ZHANG X G. Kernel nearest-neighbor algorithm [J]. *Neural Processing Letters*, 2002, 15(2): 147–156.
- [20] SEVGICAN S, TURAN M, GÖKARSLAN K, et al. Intelligent network data analytics function in 5G cellular networks using machine learning [J]. *Journal of Communications and Networks*, 2020, 22(3): 269–280.
- [21] HANCOCK J T, KHOSHGOFTAAR T M. CatBoost for big data: an interdisciplinary review [J]. *Journal of Big Data*, 2020, 7(1): 94.
- [22] ISABONA J, IMOIZE A L, OJO S, et al. Development of a multilayer perceptron neural network for optimal predictive modeling in urban microcellular radio environments [J]. *Applied Sciences*, 2022, 12(11): 5713.
- [23] HAIDER U, WAQAS M, HANIF M, et al. Network load prediction and anomaly detection using ensemble learning in 5G cellular networks [J]. *Computer Communications*, 2023, 197: 141–150.
- [24] PERIHANOGLU G M, KARAMAN H. Spatial prediction of received signal strength for cellular communication using support vector machine and k -nearest neighbours regression [J]. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2024, 48: 291–297.
- [25] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in Python [J]. *Journal of Machine Learning Research*, 2011, 12: 2825–2830.
- [26] LV Y S, DUANY J, KANG W W, et al. Traffic flow prediction with big data: a deep learning approach [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(2): 865–873.
- [27] ELMAN J L. Finding structure in time [J]. *Cognitive Science*, 1990, 14(2): 179–211.
- [28] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. (2014–09–03) [2025–05–05]. <https://arxiv.org/abs/1406.1078>.
- [29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.
- [30] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [EB/OL]. (2018–04–19) [2025–05–05]. <https://arxiv.org/abs/1803.01271>.