

# 基于TCN-ITransformer-KAN模型的 短时交通流量预测方法研究\*

曾桐<sup>1</sup>,曹瑾鑫<sup>2,3</sup>,许振山<sup>2,3</sup>

(1. 内蒙古大学数学科学学院,呼和浩特 010021; 2. 内蒙古科学技术研究院,呼和浩特 010010;  
3. 内蒙古大学交通学院,呼和浩特 010070)

**摘要:**随着城市化进程加速与机动车保有量持续攀升,交通拥堵、安全隐患及环境污染等问题日益凸显,短时交通流量的精准预测已成为优化交通资源配置和提升路网运行效率的关键需求。本文以法国克雷泰伊(Europarc Creteil)环岛环形交叉口公开数据集为研究对象,提出一种融合时序卷积网络(TCN)、改进型Transformer(ITransformer)与知识增强网络(KAN)的混合预测模型(TCN-ITransformer-KAN)。为提升多模态时间序列预测的准确性与泛化能力,该模型通过TCN完成时间序列高效处理,利用ITransformer提供全局依赖建模能力,并基于KAN嵌入领域知识增强表示,实现多模态特征融合与领域知识驱动的协同优化。实验结果表明,相较于CNN、LSTM及其变体组合模型,TCN-ITransformer-KAN在预测精度上显著提升,其决定系数较CNN-Transformer、LSTM和TCN-Transformer模型分别提高3.99%、7.47%和53.54%。模型预测曲线与真实交通流量呈现高度吻合,验证了其在实际场景中的有效性与泛化能力。本研究为城市交通流量短时预测提供了一种基于时空特征解耦与知识注入的新范式,可为智能交通系统的实时决策提供理论支撑。

**关键词:**智能交通系统;短时交通流量预测;机器学习;TCN-ITransformer-KAN算法

**中图分类号:**U491.1+4 **文献标志码:**A

20世纪以后,随着科学技术快速发展以及经济水平不断提高,私家车和出租车的保有量出现了爆发式的增长,虽然多样化的交通方式给民众出行带来了便利,但是由此引发的交通拥堵、环境污染以及道路安全隐患等问题也日益明显,在这样的背景之下,建立智能交通系统(Intelligent transportation system, ITS)对于实现交通运行状态的精准感知、趋势预测与优化调控有着关键的现实意义<sup>[1]</sup>。作为ITS的核心功能模块,短时交通流量预测技术在道路安全防控、拥堵缓解以及应急疏散等应用场景中发挥着关键作用。因此,构建高精度的短时交通流量预测模型成为当前智能交通领域亟须解决的关键科学问题。

交通流量预测模型的发展历程可划分为3个主要阶段。早期研究主要基于统计学方法,包括差分自回归移动平均模型(ARIMA)<sup>[2]</sup>、卡尔曼滤波<sup>[3]</sup>和历史均值法等。韩超等<sup>[4]</sup>采用ARIMA( $p, d, 0$ )模型实现了短时交通流量的自适应预测,其中, $p$ 为自回归阶数(AR部分)、 $d$ 为差分阶数(Integrated部分)、 $q=0$ 为移动平均阶数(MA部分)。Xu等<sup>[5]</sup>通过整合ARIMA与卡尔曼滤波模型,建立了包含

\* 收稿日期:2025-05-12; 修回日期:2025-08-12

基金项目:国家自然科学基金项目(72461028);内蒙古自治区关键技术攻关计划项目(2019GG287)

作者简介:曾桐(2001—),男,湖北荆州人,2023级硕士研究生。E-mail:1475641945@qq.com

通信作者:曹瑾鑫(1982—),男,内蒙古呼和浩特人,教授,博士。主要从事智能优化和交通运输研究。

E-mail:imucjx@163.com

Kalman 滤波测量方程和更新方程的混合预测框架,一定程度上提升了模型预测精度。这些方法在早期具有一定实用性,尤其适用于交通状态相对稳定、变化趋势可控的场景。然而,由于其模型结构通常基于线性假设,难以充分刻画交通流量数据中普遍存在的非线性波动、时变规律与复杂的周期性特征,因此,在面对实际复杂多变的交通系统时,其泛化能力与预测精度受到较大限制。随着计算能力的突破,机器学习算法逐渐被应用于该领域,其中,支持向量机(SVM)<sup>[6]</sup>、决策树和随机森林等算法表现突出。肖思思等<sup>[7]</sup>采用 SVM 算法进行交通流量预测并验证了其分类效率。程政<sup>[8]</sup>针对 SVM 计算复杂度高的问题引入随机森林模型,实验证明该模型在预测准确率、泛化能力、易用性和可扩展性方面均优于 SVM。尽管机器学习方法在一定程度上克服了线性模型的局限性,但其建模过程往往依赖大量人工设计的特征工程,难以自动提取交通流量数据中的深层时空关联特征,同时在面对大规模、复杂结构数据时存在性能瓶颈,影响其在实际场景中的推广与应用。进入 21 世纪大数据时代后,深度学习算法在处理海量交通数据方面展现出独特优势,主要包括循环神经网络(RNN)<sup>[9]</sup>、卷积神经网络(CNN)<sup>[10]</sup>、图神经网络(GCN)<sup>[11]</sup>和 Transformer<sup>[12]</sup>等架构及其变体。魏荣<sup>[13]</sup>针对交通数据的非线性和动态特性,提出了融合注意力机制的 CNN-Attention-BiGRU 混合模型。张鸣剑等<sup>[14]</sup>开发了时间信息增强的 Transformer 模型,通过结合长短期记忆网络和 Transformer 架构,不仅有效提取了交通流量的时空特征,还显著降低了模型复杂度。刘晶鑫等<sup>[15]</sup>在考虑城市交通的运行时空特性上,评估实时交通态势,将门控循环单元(GRU)与传统的 LSTM 模型进行结果对比分析,进而提高了交通运行态势的预测精度。虽然深度学习方法显著突破了传统模型在特征提取方面的瓶颈,但仍存在诸如模型结构单一、泛化能力不足、空间依赖建模欠缺和训练成本较高等问题。

本研究提出了一种基于深度学习的组合预测框架,该框架创新性地融合了 Kolmogorov-Arnold 网络(KAN)<sup>[16]</sup>、时序卷积网络(Temporal convolutional network, TCN)和改进型 Transformer(ITransformer)3 种模型的优势。该框架采用多级特征提取与融合策略,TCN 模块借助因果膨胀卷积捕捉交通流量数据的局部时间依赖性和多尺度特征;ITransformer 编码器在传统 Transformer 架构的基础上进行了 3 项关键改进:一是引入因果注意力机制来提高时间序列建模能力,二是结合通道注意力模块实现特征维度的自适应加权,三是在前馈神经网络(FFN)中使用 GELU 激活函数替代 ReLU 激活函数,显著提升了模型的非线性表达能力;KAN 模块依靠其独特的自适应激活函数机制,实现对复杂非线性映射关系的精确逼近,并生成最终的预测结果。这种多模型协同的架构设计,充分发挥了各子模型的优势,还借助特征互补效应提高了短时交通流量预测的精度。

## 1 交通流量预测问题描述

交通流量预测是 ITS 中的核心任务之一,通常可划分为短时交通流量预测与长时交通流量预测两类。其中,短时交通流量预测主要面向分钟级到小时级时间尺度内的交通动态变化趋势,广泛应用于城市主干道路、交叉口、匝道及环形交叉口等局部路段的交通信号优化、拥堵预警与动态管控等场景。

本文关注城市交叉口层面的短时交通流量预测问题,目标是基于历史观测到的交通流量序列,预测未来若干时间段内的流量变化趋势,以辅助路口信号配时优化。该任务的数学表示为

$$\hat{Y}_{t+1:t+p} = F(X_{t-k+1:t}),$$

其中, $X_{t-k+1:t}$ 表示过去  $k$  个连续时间段的交通流量观测序列, $\hat{Y}_{t+1:t+p}$ 表示预测得到的未来  $P$  个时间段的交通流量序列; $F(\cdot)$ 表示建立映射关系的预测模型或函数。通过构建合适的映射函数  $F$ ,能够充分挖掘历史数据中的时序特性与潜在规律,实现对未来交通流量的有效预测。

## 2 模型建立

### 2.1 TCN-ITransformer-KAN 时间序列预测模型架构

基于现有时间序列预测模型存在长依赖建模能力不足、非线性表达能力弱和通道结构利用不充

分等情况,本文设计并实现了一个复合型预测架构 TCN-ITransformer-KAN,它集合了多尺度时序卷积、改进注意力建模与函数级非线性建模能力,由三大模块构成,各模块承担不同建模任务,依靠结构融合实现协同学习。本文的技术创新体现在3个方面:①在传统卷积结构中,时序预测任务常因填充出现“信息泄漏”现象,即未来信息影响当前输出。为了解决这一难题,我们在每层卷积后加入自定义的 Chomp1d 层,借助剔除多余 padding 实现信息屏蔽,保证仅用当前和历史时间步数据,再堆叠 3 层 TemporalBlock,采用膨胀因子使得感受野从局部向全局递增,在每个卷积层后引入 BatchNorm1d,用 ReLU+Dropout 提升非线性表达与抗过拟合能力,凭借残差连接提高深层梯度传递。设定卷积核大小为 4,实验显示该宽度在捕捉短期、中期模式间有较优平衡。通道设置与输出通道保持一致,防止通道爆炸,Dropout 设为 0.2,兼顾训练稳定性与模型容量。②尽管标准 Transformer 模型中的多头注意力机制在捕捉长距离依赖方面表现出显著优势,但其计算复杂度较高,且采用的原始位置编码方法无法充分反映时间序列的连贯性与顺序特征,此外,注意力结构对不同通道特征的响应差异性未被有效利用,限制了模型的表达能力。为此,我们改进基于因果与通道感知机制的 ITransformer 来建立时序敏感注意力建模机制,设置 ITransformer 层数为 3,Dropout 为 0.1,为了防止模型过拟合,设置头数为 4,发现该模块不仅对时间顺序敏感,还可以动态学习通道特征贡献度。③传统全连接层或多层感知机(MLP)在输出层的非线性拟合能力有限,尤其对周期性扰动和突发拐点等问题难以捕捉。我们通过激活高斯核函数构建可导可控的函数映射器,通过 KAN 网络增强输出的非线性表达能力。模型流程图如图 1 所示。

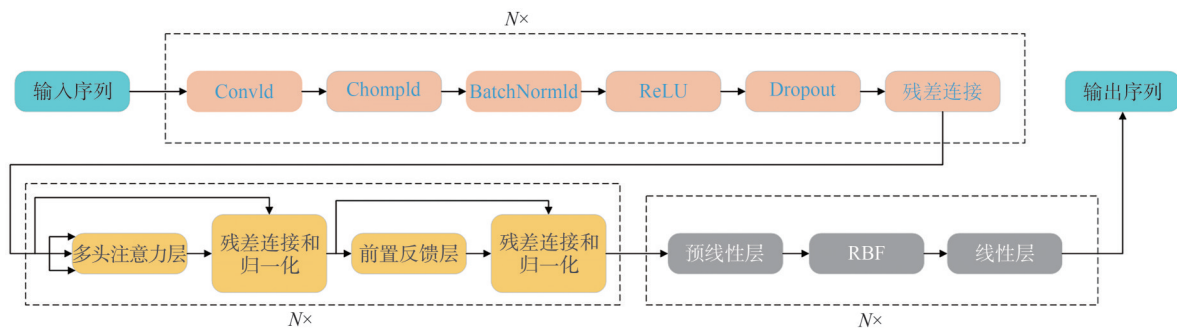


图1 TCN-ITransformer-KAN 模型结构

Fig. 1 TCN-ITransformer-KAN model structure

## 2.2 数据预处理和数据输入

选取法国克雷泰伊(Europarc Creteil)环岛环形交叉路口的车流量作为原始数据(<https://vehicular-mobility-trace.github.io>),并对其进行预处理,即进行归一化处理,公式为

$$x' = (x - x_{\min}) / (x_{\max} - x_{\min}),$$

其中, $x$ 表示原始数据, $x_{\max}$ 和 $x_{\min}$ 表示车流量的最大值和最小值。

为消除异常样本数据导致的不良情况,提高模型精度以及消除数值问题,将输入数据划分为 $(N, L, C)$ ,其中, $N$ 代表批量大小, $L$ 代表时间序列长度, $C$ 代表特征维度。

## 2.3 时序卷积网络

时序卷积网络(TCN)包含多个 TemporalBlock 模块,每个 TemporalBlock 模块中含有 1D 卷积、Chomp1d、BatchNorm1d、ReLU 激活函数、Dropout 和残差连接 6 个步骤。1D 卷积为了扩展感受野,提取局部时序特征;Chomp1d 用来裁剪填充,使得因果关系保持正确;BatchNorm1d 归一化用来提高模型训练稳定性;ReLU 激活函数增加非线性表达能力;Dropout 防止过拟合;残差连接促进模型梯度传播,防止梯度消失<sup>[17-18]</sup>。在时间序列建模中,1D 卷积公式为

$$y(t) = \sum_{i=0}^{k-1} W_i \times x(t-i),$$

其中,  $y(t)$  是输出特征序列,  $x(t)$  是输入时间序列,  $W_i$  是长度为  $k$  的卷积核权重。为了使模型能够捕捉长期时间依赖, 使用膨胀卷积, 其公式为

$$y(t) = \sum_{i=0}^{k-1} W_i \cdot x(t - d \times i),$$

其中,  $d$  是膨胀因子, 决定了当前时刻与间隔多久的历史时间数据有关。因果膨胀卷积进一步确保  $x(t)$  仅依赖于过去的的数据, 而与未来的时间序列数据无关。为了避免数据梯度消失, TCN 采用了残差连接, 公式为

$$X' = \text{ReLU} \left\{ \text{BatchNorm} \left[ \text{Conv1D}(X) \right] \right\} + X,$$

其中,  $X$  是 TCN 的输出序列; BatchNorm 为归一化数据, 为了让训练变得更加稳定; ReLU 是激活函数; 残差连接保留了输入的信息, 避免了梯度消失。

#### 2.4 改进的 Transformer 网络

传统 Transformer 编码器架构中, 每个编码器包含多个 EncoderLayer, 将 EncoderLayer 模块中的自注意力层替换为因果注意力层来增强时间序列建模能力<sup>[19-20]</sup>, 公式为

$$Q = X'W_Q, K = X'W_K, V = X'W_V,$$

其中,  $X'$  是映射到 ITransformer 空间后的输入。然后计算因果掩码注意力, 公式为

$$\text{Attention}(Q, K, V) = \text{Softmax} \left( \frac{QK^T}{\sqrt{d_k}} + M_{i,j} \right) \times V,$$

其中,  $\sqrt{d_k}$  是缩放因子;  $M_{i,j}$  是因果掩码, 用来屏蔽未来时间步的信息, 采用此种方式的原因在于: 在时间序列建模中, 时间序列任务往往是单向因果的, 也就是说当前时刻的交通流量预测不能使用未来的信息。因果掩码  $M_{i,j} \in \mathbf{R}^{L \times L}$  是一个上三角形矩阵, 对应未来时刻的值为  $-\infty$ , 当前以及以前时刻的值为 0, 即

$$M_{i,j} = \begin{cases} 0, & j \leq i \\ -\infty, & j > i \end{cases}$$

通过 Softmax 后, 当前时刻只能关注到自身和以前时刻, 符合实际场景, 可以防止未来时间信息泄露, 从而提升模型的效果和准确率。最后进行多头注意力、残差连接和归一化, 具体公式为

$$\begin{aligned} \text{MHA}(X') &= \text{Concat}(h_1, \dots, h_h) \times W^0, \\ X'_{\text{MHA}} &= \text{LayerNorm}(X' + \text{MHA}(X')), \end{aligned}$$

其中,  $\text{MHA}(X')$  是  $X'$  经过多头注意力的输出;  $h_i$  是第  $i$  个注意力头的输出 ( $i = 1, 2, \dots, h$ ); Concat 将所有注意力头的输出拼接在一起; LayerNorm 表示层归一化;  $X'_{\text{MHA}}$  是  $X'$  经过残差连接和层归一化处理的编码结果 (文中变量说明详见附录 A)。同时, 结合通道注意力, 使模型更加关注关键特征维度, 计算过程包含平均池化和最大池化, 计算公式为

$$\begin{aligned} X'_{\text{avg}} &= \frac{1}{L} \sum_{i=1}^L X'_{\text{MHA}, i}, \\ X'_{\text{max}} &= \max_i X'_{\text{MHA}, i}, \end{aligned}$$

其中,  $X'_{\text{avg}}$  表示  $X'_{\text{MHA}}$  在时间维度上的平均值;  $X'_{\text{max}}$  表示  $X'_{\text{MHA}}$  在时间维度上的最大值。然后进行权重计算, 则

$$S = \sigma \left\{ H_2 \cdot \text{ReLU} \left[ H_1 \cdot (X'_{\text{avg}} + X'_{\text{max}}) \right] \right\},$$

再重新加权输入:

$$X'_{\text{CA}} = X'_{\text{MHA}} \cdot S,$$

其中,  $H_1$  和  $H_2$  是第一个全连接层和第二个全连接层的权重。  $\sigma$ : Sigmoid 激活函数将输出限制在

(0, 1)之间,作为每个通道的注意力权重因子。采用此方法的原因在于:在时间序列中,可能只有少部分特征维度对时间序列具有重要影响,所以采用通道注意力机制可以帮助模型学会忽略不重要的因素和维度,这样通道注意力就可以和前面提到的因果注意力互补,用因果注意力来负责时间依赖关系,强调时序关系,用通道注意力解决维度选择,选取最重要的维度,这样可以提升模型的泛化能力,让模型更加聚焦,减少无用数据和特征的影响,提高模型的精度,提升模型对复杂动态数据的理解能力。而且此方法采用两层MLP,使得中间层压缩通道数,降低计算复杂度,提高时间效率。最后在前馈网络中采用GELU激活函数代替ReLU激活函数,提高模型的非线性表达能力。计算过程为先线性变换:

$$Z = H_1 X'_{CA} + b_1,$$

然后GELU激活:

$$U(Z) = 0.5 \times Z \times \tanh\left(\sqrt{\frac{2}{\pi}}(Z + 0.044715Z^3)\right),$$

再进行Dropout和线性变换:

$$Y = H_2 \cdot \text{Dropout}(U(Z)) + b_2,$$

其中, $U(Z)$ 是对 $Z$ 进行GELU处理, $Y$ 是前馈神经网络的输出。最后进行残差连接和归一化:

$$X'_{FFN} = \text{LayerNorm}(X'_{CA} + Y).$$

因为GELU函数倒数连续,且平滑处处可导。采用此激活函数使得模型表达能力更强,而且在零点附近能平滑过渡,保留了信息,使得模型梯度下降更稳定,更能提高训练效果。

## 2.5 KAN网络

当前主流的深度学习模型,包括CNN、RNN、Transformer及近年来流行的大规模预训练模型,均可视为在MLP基础结构上的演化与拓展。MLP的基本原理是对输入数据进行线性变换后施加非线性激活函数,并通过多层结构的堆叠与参数优化,实现复杂的非线性映射能力,函数表示为

$$\text{MLP}(X) = f(x_1, x_2, \dots, x_n) = (M_n, \sigma_{n-1}, \dots, \sigma_1 \circ M_1)(X),$$

其中, $M_i$ 是第 $i$ 个线性组合函数,具有可学习权重参数; $\sigma_i$ 是第 $i$ 个确定的激活函数。KAN网络模型结构是将MLP边缘上的权重参数替换为单变量函数,并采用样条函数的形式化<sup>[21]</sup>。KAN网络模型的函数表示为

$$\text{KAN}(X) = g(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left( \sum_{p=1}^n \phi_{q,p}(x_p) \right),$$

其中, $\phi_{q,p}$ 表示输入变量的单变量映射; $\phi_{q,p}(x_p)$ 表示可训练的激活函数; $\Phi_q: \mathbf{R} \rightarrow \mathbf{R}$ 。MLP和KAN网络模型结构如图2所示。

从模型整体来看,传统MLP大多采用激活函数的一种固定形式,这使得它受限于非线性关系的建模能力。在实际应用中,为了提升结构的深度和表达能力,MLP模型往往会采用加深网络层数的方法。但是,采用ReLU等非平滑的激活函数,容易出现梯度不连续和数值不稳定的问题,使得模型易出现梯度弥散和梯度爆炸,难以达到收敛,最终会降低模型的预测能力。综上所述,传统MLP结构对于复杂时间序列的非线性表达具有一定的限制。

基于此,引入KAN模型。KAN模型能利用核激活机制建立时间依赖图,能描述时间的非线性变化,对复杂函数的拟合能力也更强,和传统方法相比,KAN模型在建模时间依赖关系和描述时间非线性变化等方面性能更好,但模型整体计算量较大,训练难度较高。为提高模型性能和稳定性,对KAN模型中的核激活函数进行改进。传统KAN模型的核函数一般选择B样条函数,B样条函数在局部范围内具有良好的近似函数,能较好地逼近局部变化,但对时间序列中一些全局依赖的建模或

序列, B 样条函数并不能很好地进行描述。本文将 KAN 模型中的核函数更改为径向基函数(RBF 函数), 该函数又称为高斯核函数<sup>[22]</sup>, 其性质在整个定义域内是局部连续且无穷可微的, 并且无论在哪个点, RBF 都有良好的映射连续性, 从而能全局创造一个稳定且平滑的函数。这个特点有助于提高模型对复杂的输入-输出关系的拟合能力, 改善梯度传播的连续性和稳定性, 提高收敛速度, 可以增强训练过程的鲁棒性和泛化能力。其核激活函数形式为

$$\phi(x) = \exp(-\gamma x^2),$$

其中,  $\gamma$  是可调节的超参数 ( $\gamma > 0$ ), 用来控制核的宽度;  $x$  是输入特征变量。该激活函数的运用旨在提升模型对非线性特征的表达能力和对复杂结构的建模效果。

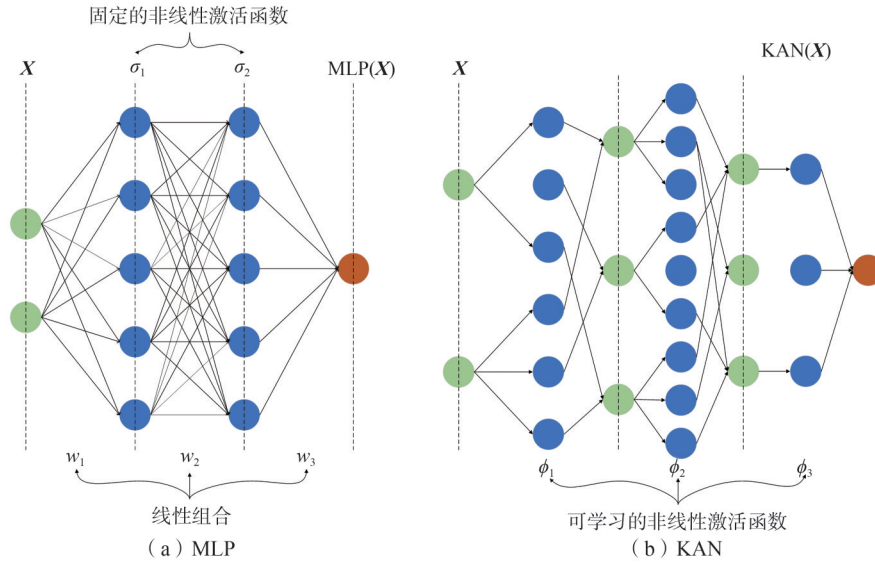


图 2 MLP 和 KAN 的网络结构  
Fig. 2 Network architectures of MLP and KAN

### 3 实验结果分析

#### 3.1 实验配置和参数设计

本实验采用的电脑配置 CPU 为 13th Gen Intel(R) Core(TM) i5-13500H 2.60 GHz, GPU 为 NVIDIA GeForce RTX 4050 Laptop GPU, 内存为 16 GB, python 版本为 3.11.4, CUDA 版本为 12.0。本实验选择深度学习组合模型的参数如表 1 所示。

表 1 模型相关参数  
Table 1 Model parameters

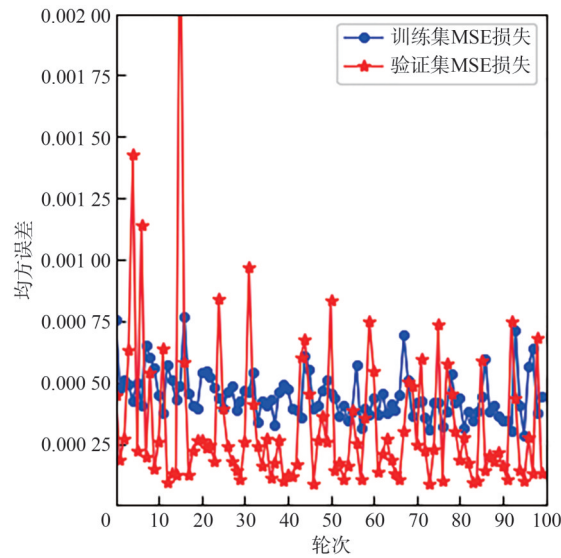
名称	参数
TCN 每一层通道数	128
卷积核的大小	4
ITransformer 隐藏层维度	128
因果多头注意力的头数	4
ITransformer 的层数	3
Adam 优化器学习率	0.0005
Epoch	100

### 3.2 损失函数设计

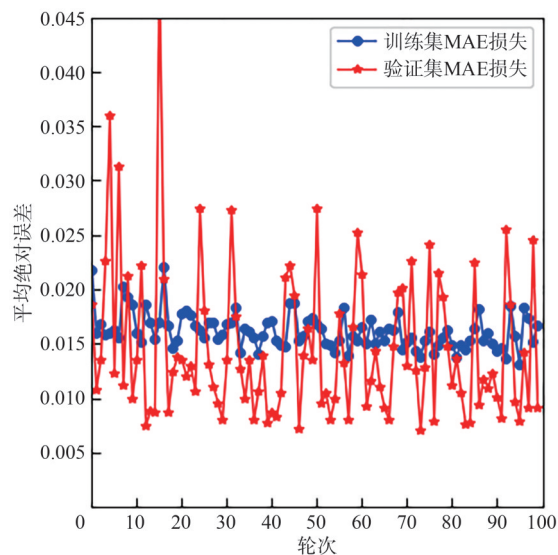
采用平均绝对误差(MAE,  $E_{MA}$ )和均方误差(MSE,  $E_{MS}$ )作为模型训练和验证过程中的损失函数,公式为

$$E_{MA} = \frac{1}{n} |y_i - \hat{y}_i|, E_{MS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2。$$

MAE损失函数对于异常值不太敏感,但是能较好地反映整体预测误差所呈现的分布特征,MSE损失函数对异常值比较敏感,但却难以全面地呈现误差的分布状况。在实际建模时,大多会将这两种损失函数结合起来,以此来兼顾模型的稳健性以及对于异常情况的响应能力。实验损失函数结果如图3所示。



(a) MSE损失曲线



(b) MAE损失曲线

图3 MSE和MAE损失曲线

Fig. 3 MSE and MAE loss curves

借助图3可发现,训练集的MSE、验证集的MSE,以及训练集的MAE与验证集的MAE,它们的整体波动趋势较为一致,刚开始误差损失较大,当训练轮数不断增加时,MSE损失在相对较低的水

平进行波动,整体波动的范围较小,基本都维持在 0.0005 的水平,MAE 损失则稳定在大约 0.015 的位置附近,模型在训练集以及验证集上的拟合效果较为良好,拥有较强的泛化能力。MSE 与 MAE 在中心部分轮数上出现了一定程度的波动,这或许反映出交通流量数据里存在噪声或者突变特征,不过从整体情况来看,模型表现较为稳定,具有不错的应用价值。

### 3.3 实验结果分析及评价

选取法国克雷泰伊(Europarc Creteil)环岛环形交叉路口的车流量作为原始数据,将原始数据的 80% 作为初始模型训练集,20% 作为测试集。再将初始模型训练集的 75% 作为训练集,25% 作为验证集导入 TCN-ITransformer-KAN 模型中进行训练、调参,以便达到模型的最优状态,最终通过测试集评估模型的泛化能力。选择 CNN、CNN-Transformer、LSTM、LSTM-Transformer、TCN-Transformer 与 TCN-ITransformer-KAN 模型,以平均绝对误差、均方误差、均方根误差(RMSE)以及决定系数( $R^2$ )为评测指标开展对比实验,实验结果如表 2 所示。

表 2 模型评测指标对比

Table 2 Comparison of model predictions

评价指标	模型					
	CNN	CNN-Transformer	LSTM	LSTM-Transformer	TCN-Transformer	TCN-ITransformer-KAN
MAE	0.1128	0.0125	0.0151	0.0128	0.0394	0.0092
MSE	0.0137	0.0003	0.0004	0.0003	0.0017	0.0001
RMSE	0.1169	0.0159	0.0189	0.0161	0.0417	0.0115
$R^2$	-3.5490	0.9161	0.8813	0.9138	0.4206	0.9560

从表 2 可以看出,TCN-ITransformer-KAN 模型在各项评测指标上整体表现优于 CNN、CNN-Transformer、LSTM、LSTM-Transformer 以及 TCN-Transformer 模型,其中 CNN-Transformer 模型的综合性能次之。

在 MAE 指标方面,虽然 CNN-Transformer 模型也呈现出了较优的性能,但 TCN-ITransformer-KAN 模型的误差更小,和 CNN-Transformer 以及 LSTM-Transformer 模型相比较而言,误差分别降低了 26.4% 和 28.1%。而 CNN 以及 TCN-Transformer 模型在该指标上表现欠佳,说明 TCN-ITransformer-KAN 模型在预测精度以及异常值处理方面有更高的稳定性和可靠性。

在 MSE 和 RMSE 指标方面,和 CNN 以及 TCN-Transformer 模型相比,TCN-ITransformer-KAN 模型同样呈现出较优的表现,其在误差控制能力方面更为稳定,预测结果与真实值之间的偏差水平明显更低。而 LSTM 模型在这两项指标上表现平平,展现出 TCN-ITransformer-KAN 模型对于本数据集拥有更好的适应性以及误差控制能力。

在  $R^2$  指标方面,TCN-ITransformer-KAN 模型与 LSTM-Transformer 模型都取得了比较不错的拟合效果。TCN-ITransformer-KAN 与 CNN-Transformer、LSTM 和 TCN-Transformer 模型相比, $R^2$  值分别提高了 4.35%、8.49% 和 127.3%,进一步验证了本文所提出的模型在拟合精度以及预测准确性方面有着明显优势。

### 3.4 实验预测图与模型拟合效果散点图

通过图 4 可知,测试集实际值曲线波动较大,具有明显的周期性和非线性特征,对应图中深蓝色曲线。CNN 模型预测曲线整体偏离真实值,存在高频振荡,预测精度很差,与实际数据有较大的区别,不能准确预测测试集的数据变化趋势以及交通流量。CNN-Transformer 比 CNN 模型预测精度有所提升,拟合曲线更加平滑,但是在局部或者波峰和波谷处偏差依旧很大。LSTM 模型能较好地捕捉时序依赖,整体趋势也跟随真实值波动,但是波动具有一定的误差,数据预测有滞后现

象,预测结果中规中矩。LSTM-Transformer在LSTM基础上结合了Transformer模型,使得全局依赖建模能力更强,效果也优于单一LSTM,大部分预测值和真实值比较接近。TCN-Transformer预测曲线走势符合真实趋势,在波谷位置预测值和真实值相近,说明TCN捕捉了局部时序特征,Transformer使得全局依赖建模能力更强,两者优势互补,但是预测值和真实值有一定的差距。TCN-ITransformer-KAN模型在测试集上预测效果较好,预测值和真实值几乎重合,误差极小。为了进一步验证本文提出的混合模型的预测值和真实值的一致性,绘制各模型拟合效果散点图,结果如图5所示。

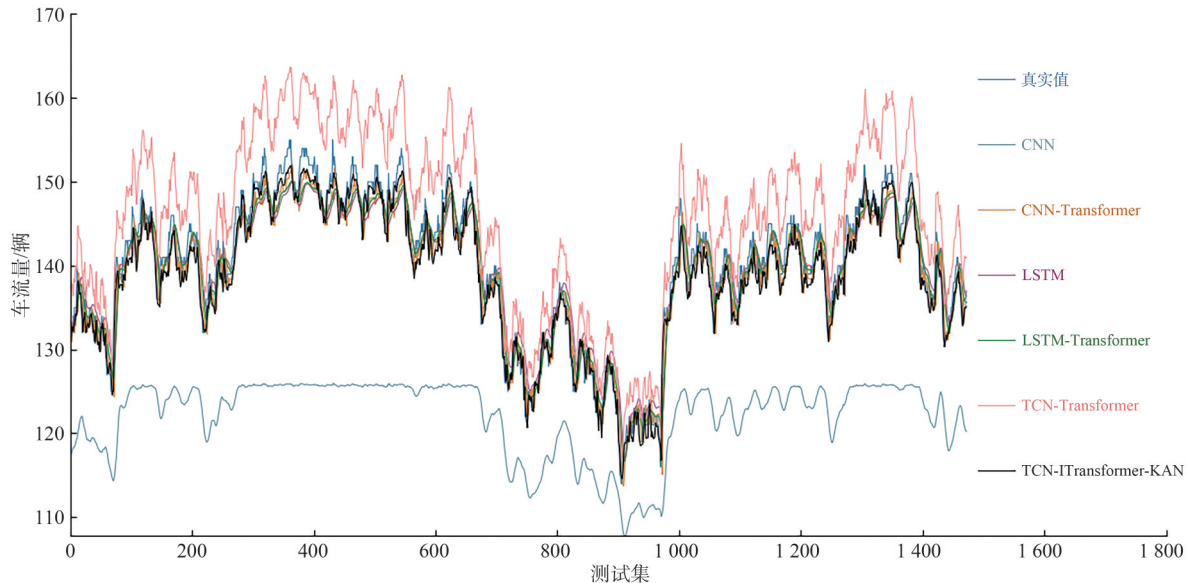


图4 实验预测对比图

Fig. 4 Comparisons for experimental prediction curves

实验结果表明,TCN-ITransformer-KAN模型在整体趋势拟合方面具有优异的表现,预测结果与实际观测值基本一致,预测曲线围绕真实数据小幅波动,说明该模型具备良好的时间序列建模能力,也具备较高的预测准确性。该方法能够从交通流量数据中有效提取潜在特征,并学习其变化规律,为解决交通流量预测问题提供了一种新方法。在对验证集和测试集的预测结果进行详细分析后,可以看出模型在不同数据集上表现出较强的稳定性。同时,它也展现出了良好的泛化能力。与传统的单一预测模型以及现有的一些混合模型相比,该模型在预测精度方面具有明显优势,且在稳定性方面也较为出色,而且对交通流量数据中的波动趋势响应更敏感。

### 3.5 超参数分析

为了更清晰地剖析验证集中MAE与MSE在中心部分训练轮数出现波动的缘由,本研究运用超参数敏感性分析方法展开探讨,在每次只调整一个超参数的情况下,记录该超参数对应的预测结果同真实观测值之间MAE和MSE的变化情形,依靠逐个分析各超参数对预测误差的影响,可更直观地辨别出它们对模型稳定性以及预测效果的作用。本研究从模型结构与训练机制两个方面挑选了6个关键超参数作为敏感性分析的主要对象,这些参数覆盖:Num\_channel,即TCN中每一层所运用的通道数;Kernel\_size,即TCN中卷积核的大小;Hidden\_dim,即隐藏层的维度大小;Transformer\_heads,即改进Transformer结构中多头注意力机制的头数;Transformer\_layers,即Transformer编码器的堆叠层数;Dropout,即训练过程中随机丢弃的比例。

在实验过程中,依次对这6个超参数进行单变量调整,并评估其对模型输出结果的具体影响,以便掌握各参数变化对预测性能的作用方向和程度,结果如图6所示。

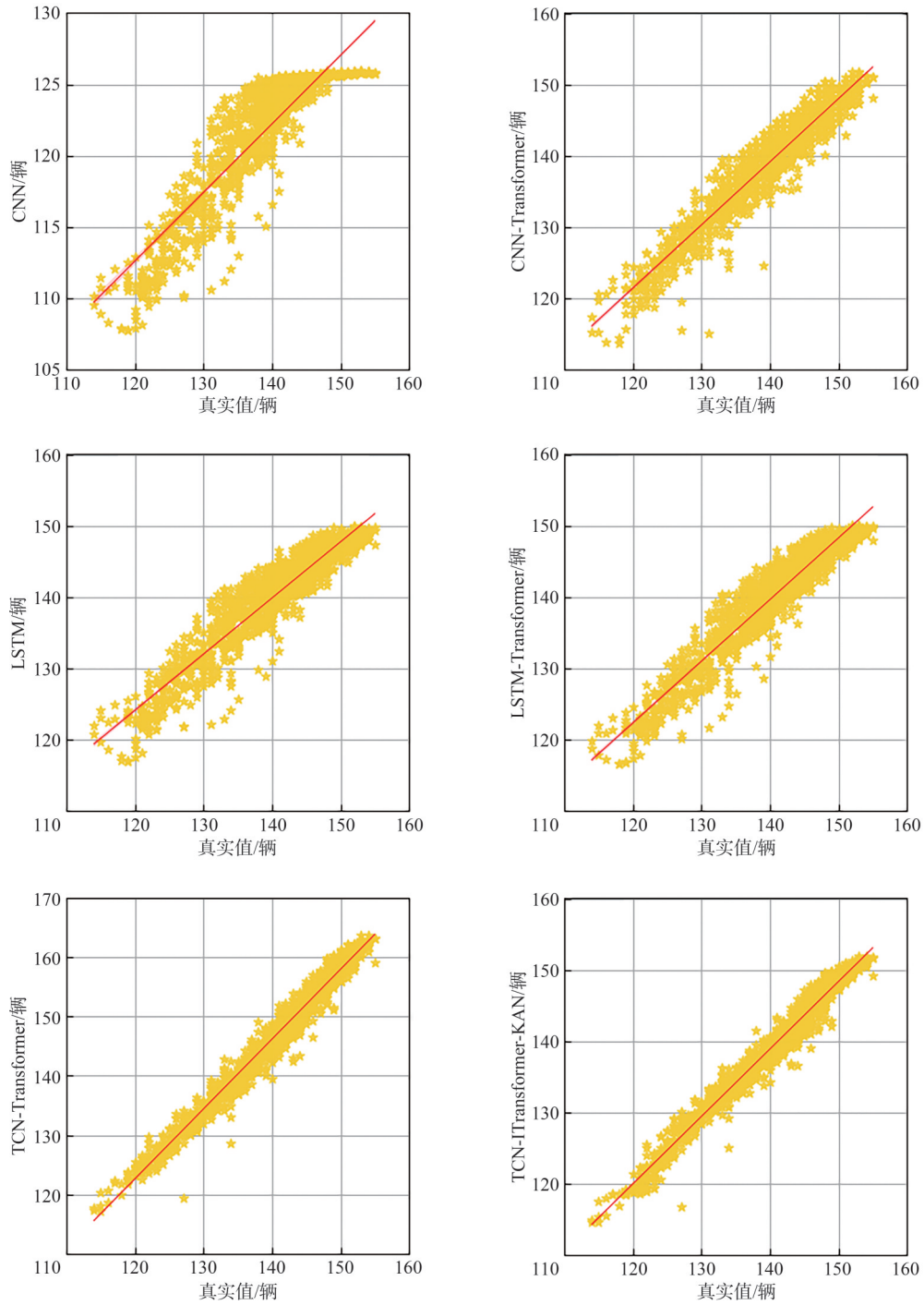


图 5 各模型拟合效果散点图

Fig. 5 Scatters plots of the fitting performance of each model

由图 6 可以观察到模型性能随不同超参数取值呈现出的变化趋势,该可视化结果为合理确定模型参数给予了明确依据,通过详细分析各超参数对模型性能的影响,可在模型训练进程中有针对性地实施参数调整,进而有效提高模型的预测精度,并且提高模型的鲁棒性。

该方法可减少因参数选择不当引起的过拟合或欠拟合问题,另外,该分析还可揭示模型针对不同关键超参数的响应特性,借助识别这些敏感性特征,可为后续的模型优化提供直接参考,也为模型在更为复杂或变化的数据环境中维持稳定性奠定一定基础。

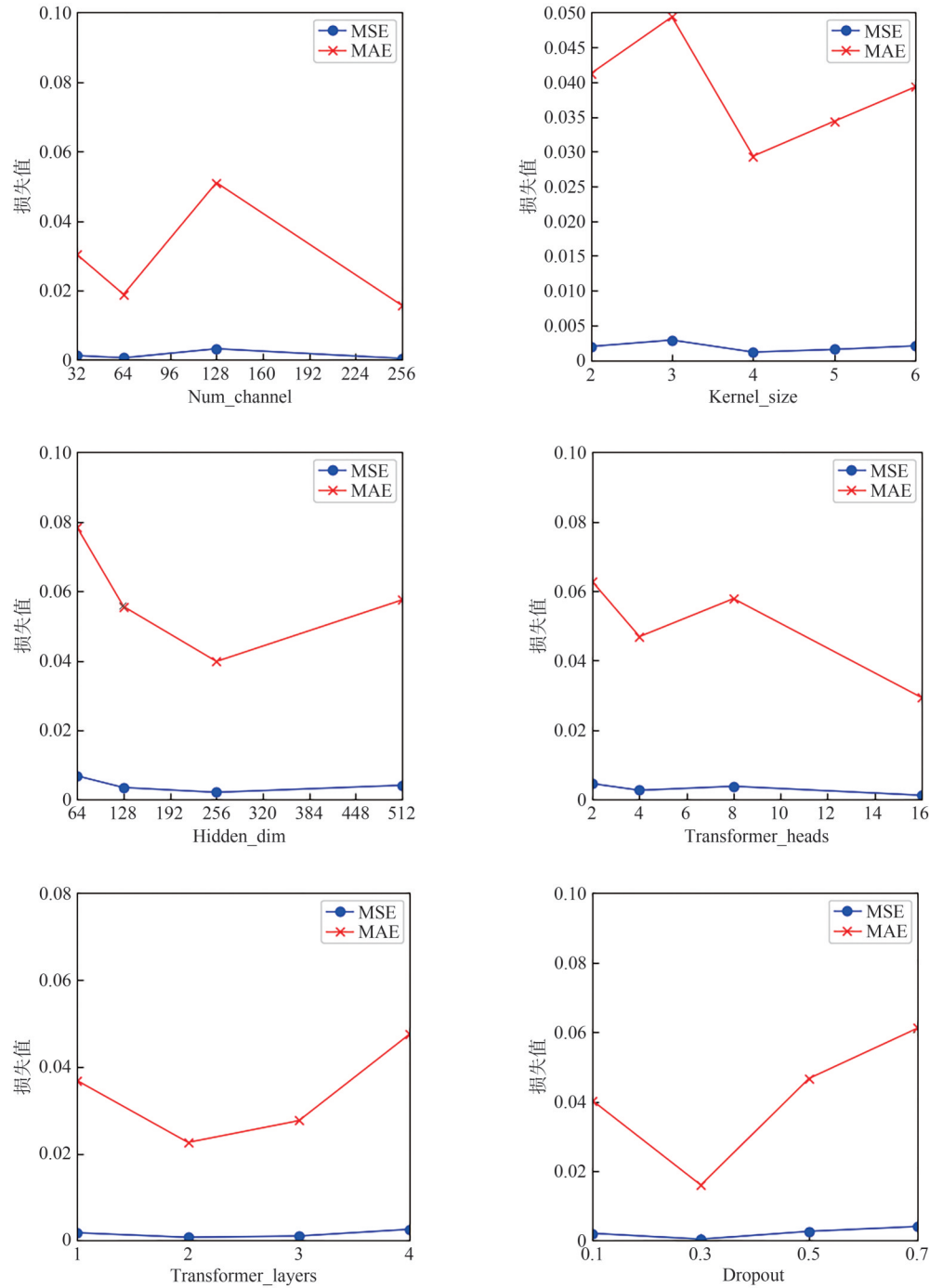


图6 超参数分析

Fig. 6 Hyperparameter sensitivity analysis

## 4 结论

本研究提出了一种集成式时间序列预测模型——TCN-ITransformer-KAN,将时序卷积网络(TCN)、改进型因果多头注意力机制(ITransformer)与径向基函数激活的神经网络(KAN)相结合,实现了对交通流量序列中时间依赖结构、关键信息通道及非线性动态关系的协同建模。实验结果表明,与当前主流预测方法相比,该模型在多个性能指标上均有显著优势,验证了其在短时交通流量预测任务中的应用价值。

值得强调的是,本文仅采用历史交通流量作为输入特征,一方面,基于其在短时间尺度上的显著时序依赖性,足以反映主要变化趋势;另一方面,避免了外部因素难以获取和编码不一致等带来的噪声干扰。该策略保证了建模可控性与实验可重复性,但也存在局限,如数据仅覆盖法国克雷泰伊(Europarc Creteil)环岛环形交叉口的两小时窗口,缺乏时空多样性与外部信息支持,因而模型泛化能力仍有待进一步验证。

针对上述不足,后续研究可从以下两个方面拓展优化。其一,扩大数据采集范围,引入如天气状况、节假日安排和交通事件等多源异构数据,以增强模型对复杂外部环境变化的鲁棒性与泛化能力<sup>[23]</sup>;其二,在现有模型框架基础上引入图神经网络结构,结合交通网络拓扑信息<sup>[24]</sup>,进一步提升空间依赖建模能力,实现对交通流量在时空维度上的全面预测。此外,深入剖析模型内部各模块间的交互机制,提升可解释性,将有助于构建更加透明、可靠且具有实用价值的智能交通预测系统。

## 参考文献:

- [1] WANG B M, ZHENG H T, QIAN K, et al. Edge computing and AI-driven intelligent traffic monitoring and optimization[J]. *Applied and Computational Engineering*, 2024, 77: 225-230.
- [2] MIN X, QIN Z F. A novel hybrid ARIMA and regression tree model for the interval-valued time series[J]. *Journal of Statistical Computation and Simulation*, 2021, 91(5): 1000-1015.
- [3] WANG T L, TAO X R, ZHANG J M, et al. Predicting traffic congestion time based on Kalman filter algorithm[J]. *Advances in Research and Reviews*, 2020, 1: 7.
- [4] 韩超, 宋苏, 王成红. 基于 ARIMA 模型的短时交通流实时自适应预测[J]. *系统仿真学报*, 2004, 16(7): 1530-1532, 1535.
- [5] XU D W, WANG Y D, JIA L M, et al. Real-time road traffic state prediction based on ARIMA and Kalman filter [J]. *Frontiers of Information Technology & Electronic Engineering*, 2017, 18(2): 287-302.
- [6] FENG X X, LING X Y, ZHENG H F, et al. Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction[J]. *IEEE Transaction on Intelligent Transportation Systems*, 2019, 20(6): 2001-2013.
- [7] 肖思思, 杨宏业. 基于 SVM 的短期交通状态预测识别研究[J]. *内蒙古工业大学学报(自然科学版)*, 2017, 36(2): 149-154.
- [8] 程政. 城市道路短时车流量预测模型研究[D]. 合肥: 中国科学技术大学, 2016.
- [9] FANG W, CHEN Y P, XUE Q Y. Survey on research of RNN-based spatio-temporal sequence prediction algorithms[J]. *Journal on Big Data*, 2021, 3(3): 97.
- [10] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [DB/OL]. *ArXiv*, 2018: 1803.01271 (2018-05-19) [2025-04-12]. <https://arxiv.org/abs/1803.01271>.
- [11] MATINFAR M M, BAG-MOHAMMDI M, KARAMI M. Unmasking the brain in cocaine use disorder: A deep learning approach with graph convolutional networks and principal component analysis[J]. *Next Research*, 2025, 2(2): 100304.
- [12] 鲁思源, 沈琴琴, 包银鑫, 等. 基于时空感知 Transformer 的交通流预测模型[J]. *电子测量技术*, 2024, 47(10): 85-92.
- [13] 魏荣. 基于注意力机制的短时交通流预测模型研究[J]. *交通科技与管理*, 2024, 5(20): 42-44.
- [14] 张鸣剑, 叶宝林, 董睿, 等. 基于时间信息增强 Transformer 的短时交通流预测模型[J]. *软件工程*, 2024, 27(8): 41-45, 50.
- [15] 刘晶鑫, 唐婕, 曹瑾鑫. 基于机器学习的城市交通态势评估与预测方法研究[J]. *内蒙古大学学报(自然科学版)*, 2021, 52(2): 198-205.

- [16] 刘灿锋,孙浩,东辉. 结合 Transformer 与 Kolmogorov Arnold 网络的分子扩增时序预测研究[J]. 图学学报, 2024,45(6):1256-1265.
- [17] SHEN Z P,ZHANG Y M,LU J W, et al. A novel time series forecasting model with deep learning[J]. Neurocomputing, 2020,396:302-313.
- [18] GASPARIN A,LUKOVIC S,ALIPPOL C. Deep learning for time series forecasting: The electric load case[J]. CAAI Transactions on Intelligence Technology, 2022,7(1):1-25.
- [19] LIVIERIS I E,STAVROYIANNIS S,PINTELAS E, et al. A novel validation framework to enhance deep learning models in time-series forecasting[J]. Neural Computing and Applications, 2020,32(23):17149-17167.
- [20] VIDA A,TAHERI H G,MOJTABA S S, et al. Water level forecasting using deep learning time-series analysis: A case study of red river of the north[J]. Water, 2022,14(12):1971.
- [21] 袁立宁,冯文刚,刘钊. 基于 Kolmogorov-Arnold 网络的节点分类算法[J]. 计算机科学与探索, 2025,19(3): 645-656.
- [22] 左小伟. 基于 RBF 核的多分类 SVM 围岩变形易发性评价模型研究[J]. 铁道建筑技术, 2024(8):158-162.
- [23] 张宏,王智远,程昊,等. 寒旱区隧道出入口路桥过渡段路面病害成因机理[J]. 内蒙古大学学报(自然科学版), 2014,45(4):435-440.
- [24] QIAN C,TANG C K,LI C, et al. An innovational GNN-based urban traffic-flow measuring approach using dynamic vision sensor[J]. Measurement, 2025,252:117262.

(责任编辑 李成兵)

## Research on Short-Term Traffic Flow Prediction Method Based on TCN-ITransformer-KAN Model

ZENG Tong<sup>1</sup>, CAO Jinxin<sup>2,3</sup>, XU Zhenshan<sup>2,3</sup>

(1. School of Mathematical Sciences, Inner Mongolia University, Hohhot 010021, China;

2. Inner Mongolia Academy of Science and Technology, Hohhot 010010, China;

3. Institute of Transportation, Inner Mongolia University, Hohhot 010070, China)

**Abstract:** With the acceleration of urbanization and the continuous increase in vehicle ownership, issues such as traffic congestion, safety hazards, and environmental pollution have become increasingly prominent. Accurate short-term traffic flow prediction has emerged as a critical requirement for optimizing traffic resource allocation and improving road network efficiency. The study adopts the publicly available dataset of the Europarc Créteil roundabout in France as the research subject and proposes a hybrid prediction model (TCN-ITransformer-KAN) that integrates temporal convolutional network (TCN), improved transformer (ITransformer), and knowledge-augmented network (KAN). To improve the accuracy and generalization capability of multimodal time series prediction, the proposed model employs the TCN for efficient temporal sequence processing, leverages the ITransformer to model global dependencies, and integrates domain knowledge via the KAN to enhance feature representation. The approach enables effective multimodal feature fusion and facilitates knowledge-driven collaborative optimization. Experimental results demonstrate that compared to CNN, LSTM, and their variant hybrid models, TCN-ITransformer-KAN achieves significant improvements in prediction accuracy. Specifically, the coefficient of determination ( $R^2$ ) increases by 3.99%, 7.47%, and 53.54% over CNN-Transformer, LSTM, and TCN-Transformer models, respectively. The predicted curves exhibit high consistency with real traffic flow data, validating the model's effectiveness and generalization capability in practical scenarios. The research provides a novel paradigm for urban short-term traffic prediction based on spatio-

temporal feature decoupling and knowledge injection, offering theoretical support for real-time decision-making in intelligent transportation systems.

**Key words:** intelligent transportation system; short-term traffic flow prediction; machine learning; TCN-ITransformer-KAN algorithm

## 附录 A

符号	符号意义及格式标注
$N$	批量大小, 整数
$L$	时间序列长度, 整数
$C$	特征维度, 整数
$C'$	TCN 输出通道数, 整数
$d$	卷积的膨胀系数, 整数
$k$	卷积核大小, 整数
$x(t)$	时间步 $t$ 处的输入特征, 向量 ( $R^C$ )
$y(t)$	时间步 $t$ 处的 TCN 输出特征, 向量 ( $R^{C'}$ )
$W_i$	第 $i$ 个卷积权重, 矩阵 ( $R^{C' \times C}$ )
$X$	TCN 的输出序列, 三维张量 ( $R^{N \times L \times C'}$ )
$Q, K, V$	多头注意力机制中的查询、键、值向量, 三维张量 ( $R^{N \times L \times C'}$ )
$d_k$	每个注意力头的维度, 等于输入数据的总维度与注意力头的数量的比值, 整数
$M_{i,j}$	注意力 mask, 用于防止信息泄露, 矩阵
Attention( $Q, K, V$ )	多头注意力结果, 三维张量
$X'_{MHA}$	MHA 输出, 三维张量 ( $R^{N \times L \times d_{model}}$ )
$X'_{avg}, X'_{max}$	沿时间维度的平均池化与最大池化结果, 三维张量 ( $R^{N \times 1 \times d_{model}}$ )
$H_1, H_2$	Channel Attention 模块中的全连接层权重, 权重矩阵
$S$	通道注意力生成的缩放权重, 三维张量 ( $R^{N \times 1 \times d_{model}}$ )
$X'_{CA}$	通道注意力的输出, 三维张量
$Z$	前馈网络的输入, 三维张量
$U(Z)$	GELU 近似变换结果, 三维张量
$Y$	Feedforward 模块输出, 三维张量
$X'_{FFN}$	ITransformer 层的输出, 三维张量
$\gamma$	RBF 激活函数的带宽参数, 整数
$b_1$	偏置项, 大小与 $Z$ 相同, 向量
$b_2$	输出线性层的偏置项, 用于控制每个输出维度的偏移, 向量