# 方法

## 模型概况

图2展示了本文提出的ST-Attn-LLM模型的整体结构，该模型采用了空间与时间特征提取-特征增强-预测框架。模型由三个部分组成：时空特征提取模块、特征增强和预测模块。各模块的主要功能如下所示：

（1）时空特征提取模块利用时空注意力机制提取各车站之间的空间特征以及进出站流的时间特征。

（2）特征增强模块整合时空特征提取模块所得到的时间特征和空间特征，利用大模型增强时空特征，并将其转换为稠密向量空间。

（3）预测模块以稠密向量空间为输入，通过时序预测头对未来的进出站流进行预测。

在接下来的章节中，我们将详细介绍各个模块的组成及其相互关系。



## 模型介绍

### 基于时空注意力机制的时空特征提取网络

为准确预估未来 k 个时间步的高铁站进出站流量，该网络旨在从进站和出站流数据中提取时空特征，以便更好地捕捉时序性和空间依赖关系。在交通流量预测等任务中，时空特征的提取对于准确建模客流变化至关重要。为了有效地从复杂的时空数据中学习并提取有用信息，本文提出了基于时空注意力机制的时空特征提取网络。既有研究[18-19]证明了时空注意力机制在捕捉时序数据的时空特征方面具有较大优势，其架构如图 所示。

具体而言，设定进站流为 ，出站流为 ，首先通过公式 将进出站流转换为时间特征数据Pt和空间特征数据Ps，通过公式 分别获取进出站流得查询嵌入（Query）、键嵌入（Key）以及值嵌入（Value）

具体可以表示为以下公式，

其中， 表示模型输入， 表示单元输出， 表示隐藏状态， 分别表示输入门、遗忘门以及输出门， 表示卷积权重， 是偏置项， 是sigmoid函数， 表示双曲正切函数，“\*”表示卷积算子，“ ”表示哈达玛乘积。



### 基于LLM的时空特征增强网络

已有研究针对地铁站进出站客流的关系进行了深入探讨，发现进站客流与出站客流之间存在显著的关联性[21]。同样地，在高铁站场景下，进站与出站流之间也展现出一定的因果关系与内在联系。具体而言，一个高铁站当前时刻的进站车流量，可能会对后续时刻其他站点的出站车流量产生影响，本文将这一现象定义为进站流对出站流的因果效应。以早高峰时段为例，当北京南站的进站车流骤然增加，意味着大量乘客正搭乘列车前往其他城市。经过一段时间后，这部分客流将在天津西站、济南西站等下游站点陆续出站，从而引发这些站点出站流量的增长，体现出上游进站流对下游出站流的直接影响。这种跨站点、跨时段的时空传导特性在高铁网络中尤为突出，尤其在节假日或重大活动期间表现得更为明显。此外，同一站点的进站流和出站流之间同样存在紧密联系。例如，南京南站在某一时段出现大量旅客进站乘车，往往与之前一段时间内出站旅客回流到站附近有关，或者在列车抵达后短时间内出现出站高峰，随后部分旅客经过换乘、候车等环节再次形成进站流。这种站内进出站流的动态交替反映了站点客流循环的特性，并受到站内设施条件、接驳交通能力及旅客出行行为等多方面因素的影响。综上所述，高铁站进出站流之间蕴含着丰富的相互信息，若能够充分挖掘并加以利用，将有助于提升进出站流协同预测的准确性与实用性。

为此，在利用时空注意力网络分别获得进出站交通流的隐藏状态后，本文进一步研究二者间的相互关系。具体来说，本研究嵌入了大模型挖掘进站流和出站流的关联，实现进站流隐藏状态和出站流隐藏状态间相互信息的传播，最终有效建模收费站进站流和出站流之间的内在交互过程。大模型[10]最早提出用于自然语言任务（NLP），其可以在不受节点距离影响的情况下，准确计算两个节点间的相关性，具有较强的相关性特征挖掘能力。本文采用低秩适应（LoRA）方法对大模型进行微调，通过冻结原有模型参数，仅引入可训练的低秩矩阵对关键层的权重矩阵进行调整，从而实现进站流与出站流隐藏状态之间的有效信息传播。具体而言，LoRA通过将权重更新项表示为两个低秩矩阵的乘积，在保持原有模型能力的同时，以较少的参数量有效捕捉进出站流之间的相关性特征，为进出站流的相互交互建模提供了一种高效的解决方式。

设定进站流的隐藏状态为 ​ ，出站流的隐藏状态为 ，首先通过式 将进站流的隐藏状态和出站流的隐藏状态合并，通过预训练参数矩阵和低秩矩阵的共同作用后的到查询嵌入（Query）、键嵌入（Key）以及值嵌入（Value）









其中，Concat表示向量操作，W表示Q、K、V转换权重，在该模型中，它们应该使用预训练参数，并不参与更新，R表示低秩矩阵，其拥有比W更少的参数，在该模型中，仅对其进行参数更新与进出站流的隐藏状态相同，查询嵌入、键嵌入以及值嵌入均具有相同的维度。接着，根据注意力机制，我们可以计算它们之间的注意力分数：



其中，Softmax表示激活函数，T表示矩阵的转置。在求得它们之间的注意力分数后，多头注意力机制的计算如下：



在获得进站流与出站流的注意力结果后，紧接着通过Layer norm、Feed Forward、Layer norm层，整个过程嵌套12次，最终获得进站流和出站流的增强隐藏状态。

通过进出站流间的信息交互，不仅可以增强进站交通流与出站交通流的隐藏状态，而且有利于捕捉进站流与出站流间的因果性和关联性，进而辅助模型准确预测未来的进出站交通流。



# 实验与结果

这部分从多个角度对所提出的模型进行了全面的对比与分析，包括整体预测性能、在不同高铁站点的表现，以及通过消融实验进行的鲁棒性分析。对比内容涵盖了模型的预测准确性和鲁棒性。

## 数据描述

本论文采用2023年京沪高铁各站点进出站流数据集。该数据集一共包含24个站点，从二月份开始连续八周的数据，即2023年2月6日-2023年4月2日。每条数据中均记录了高铁站点名称、站点ID、日期、小时、进站流以及出站流信息。由于存在不可能避免的数据收集误差，在提取进出站流时间序列数据之前，需要对原始数据进行数据清洗，例如一些缺失站点信息的进出站流数据或者较长时间内进出站流为零的数据。在完成数据清洗后，依据章节1.1问题描述部分对于时间序列数据的定义，以60分钟为时间粒度，分别提取高铁站进站交通流和出站交通流的时间序列数据。

## 模型配置

本研究所构建的模型采用 Pytorch 框架实现，针对进站与出站交通流数据，按照 5:3:2 的比例划分为训练集、验证集和测试集。经过多次实验调优，最终确定批次大小为 32，学习率设定为 0.0001，优化器选择 Adam。在模型训练过程中，采用均方误差（MSE）作为主要的损失函数。由于模型采用了多任务学习策略，其总损失函数由进站交通流和出站交通流的均方误差加权组合而成，两部分的权重保持一致，具体形式如公式所示。

|  |  |
| --- | --- |
|  |  |

其中，分别用和表示进站交通流与出站交通流的样本数量，和 代表进站与出站交通流的真实值，对应的预测值则记为和 。此外，为了全面评估模型的预测性能，本文采用均方根误差（RMSE）、平均绝对误差（MAE）以及加权平均绝对百分比误差（WMAPE）作为评价指标，其具体计算方法见公式 (11)-(13)。为了简化符号表示，后续统一用表示进出站交通流的真实值，表示对应的预测值，样本总数统一记作。

|  |  |  |
| --- | --- | --- |
|  |  | (11) |
|  |  | (12) |
|  |  | (13) |

## 基准实验设置

(1) **SVR (Support Vector Regression)**：支持向量回归作为经典的机器学习方法之一，广泛应用于交通流预测任务 [22]。其核心思想是利用支持向量机对时间序列数据进行回归建模。本研究中，SVR模型选用径向基函数（rbf）作为核函数，误差容忍参数设为 epsilon=0.005，正则化系数 C 设置为 3。

(2) **XGBoost (eXtreme Gradient Boosting)** [23]：梯度提升树算法近年来在交通预测领域表现优异，其通过集成多个弱学习器形成一个强学习器，从而显著提升整体模型效果。本实验采用‘gbtree’作为基础提升器，树的最大深度设定为6，学习率设为0.1。

(3) **BPNN (Back Propagation Neural Network)**：反向传播神经网络作为最早提出的神经网络结构之一，已被证明能够有效建模交通流中的非线性特性 [24]。本文所使用的BPNN由两层全连接层构成，分别包含1024个与512个神经元，采用Adam优化器，学习率设置为0.0005。

(4) **CNN (Convolutional Neural Network)**：卷积神经网络擅长挖掘数据中的空间相关性 [25]，能够很好地捕获时间序列数据内部的空间依赖。本研究设计的CNN模型首先通过一个卷积核大小为3×3的卷积层提取空间特征，随后接入全连接层进行预测，并使用ReLU作为激活函数。

(5) **LSTM (Long Short-Term Memory)**：长短时记忆网络因其擅长处理长时间依赖问题，在交通客流预测领域有着较好表现 [26]。本文搭建的LSTM模型包含一层LSTM单元，设置4个隐藏层，后接一个全连接层用于输出预测结果，激活函数为ReLU。

(6) **GRU (Gated Recurrent Unit)**：门控循环单元作为循环神经网络的变种，能够有效缓解传统RNN在长时间依赖问题中出现的梯度消失现象 [27]。与LSTM相比，GRU结构更为简洁，计算效率更高。本研究所构建的GRU模型包括3个隐藏层，每层含有128个神经元，并在末尾连接一个全连接层。

(7) **ST-ResNet (Spatial Temporal Residual Network)**：该模型通过引入残差卷积单元来建模交通流的空间与时间特性，同时采用三个分支分别学习近期、周期性及趋势性时间特征。本实验设计的ST-ResNet包含一层卷积核为3×3的残差卷积层，之后接全连接层完成预测任务。

(8) **Transformer**：最初由 [10] 于2017年提出，Transformer模型因其能够高效捕捉序列中任意位置的依赖关系，近年来在时间序列预测领域得到了广泛应用。本文搭建的Transformer框架包含2层编码器与2层解码器，多头注意力机制的头数设置为4，特征维度 dmodeld\_{model}dmodel​ 设为256，输出部分连接全连接层用于最终预测。

本节将从模型的总体预测效果、车站级预测[28]效果以及消融实验等几个方面对本文提出的模型进行全面分析。

## 总体预测分析

基于前文所述京沪高铁进站流和出站流数据集，所有模型均采用固定输入步长（历史12个时间步）的流量数据作为输入，预测未来一小时（1个时间步）、两小时（2个时间步）、三小时（3个时间步）的流量数据。基准模型均采用独立且相同的模块分别预测高铁站站未来进站流量和出站流量。预测效果如表1和表2所示。

首先，基于机器学习的SVR和XGBoost模型在不同预测时间步数下的预测性能均低于其他模型。由于这些模型进行多步预测时，每个预测时间步均对应一个独立的模型，无法直接处理时间序列数据的时间依赖性，表明不适合高速公路节假日多步进出站流量预测。

其次，虽然BPNN能够捕捉节假日流量的非线性关系，但由于缺乏对时空依赖性的建模，导致其表现仅优于基于机器学习的模型。LSTM、GRU和Transformer通过捕捉时间依赖性，实现了比BPNN更好的预测性能。Transformer利用自注意力机制捕捉长时间依赖，表现优于LSTM和GRU。CNN通过卷积操作捕捉节假日流量的局部站点间空间关系和时间依赖，但随着时间步数的增加，其预测性能下降，因为CNN难以捕捉全局流量信息。ST-ResNet通过全局建模时间和空间依赖性，预测精度优于其他模型，但未能考虑进出站流量之间的交互关系。

最后，为了捕捉节假日期间复杂的高速公路流量模式，本文提出的模型采用了Bi-ConvLSTM同时捕捉空间依赖性和全局时间依赖性。同时，进站流量和出站流量具有显著的相关性，本文通过Cross-attention机制捕捉节假日高速公路进站流量与出站流量之间的关联，实现了进出站流量特征的共享。结果表明，与基准模型相比，模型在不同预测时间步上表现出最佳的预测性能，证明了本文所提出模型的预测准确性和鲁棒性。

表 1进站流预测结果(Inflow Dataset)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **进站流** | | | | | | | | | |
| 模型 | 1天 (24个时间步) | | | 2天 (48个时间步) | | | 3天 (72个时间步) | | |
| RMSE | MAE | WMAPE | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE |
| SVR | 205.53 | 129.47 | 41.93% | 204.09 | 128.69 | 41.31% | 201.46 | 127.33 | 40.71% |
| XGBoost | 184.70 | 100.13 | 39.28% | 184.64 | 101.93 | 38.56% | 184.44 | 102.43 | 42.74% |
| BPNN | 161.26 | 85.82 | 33.06% | 191.70 | 103.15 | 39.35% | 187.66 | 96.23 | 34.56% |
| CNN | 158.84 | 84.72 | 30.77% | 166.27 | 87.72 | 31.51% | 184.91 | 99.41 | 35.73% |
| LSTM | 163.29 | 84.88 | 30.66% | 159.66 | 83.48 | 29.82% | 161.41 | 81.69 | 29.16% |
| GRU | 180.55 | 94.12 | 34.06% | 175.07 | 90.44 | 32.36% | 159.82 | 84.30 | 30.14% |
| ST-ResNet | 151.07 | 79.28 | 28.78% | 161.18 | 85.18 | 30.58% | 180.75 | 99.57 | 35.76% |
| Transformer | 145.81 | 87.24 | 31.68% | 148.63 | 86.34 | 30.98% | 153.32 | 80.37 | 28.86% |
| ***Our Model*** | ***136.89*** | ***66.91*** | ***28.55%*** | ***142.32*** | ***70.27*** | ***28.47%*** | ***148.34*** | ***72.08*** | ***27.47%*** |

表2 出站流预测结果(Outflow Dataset)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **出站流** | | | | | | | | | |
| 模型 | 1天 (24个时间步) | | | 2天 (48个时间步) | | | 3天 (72个时间步) | | |
| RMSE | MAE | WMAPE | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE |
| SVR | 177.89 | 114.13 | 37.42% | 232.94 | 150.77 | 44.32% | 175.65 | 111.58 | 36.16% |
| XGBoost | 161.32 | 89.46 | 37.51% | 194.20 | 110.49 | 40.12% | 164.31 | 91.28 | 37.76% |
| BPNN | 142.34 | 75.65 | 28.41% | 158.47 | 86.28 | 33.53% | 152.34 | 81.43 | 29.87% |
| CNN | 126.68 | 69.58 | 25.69% | 145.30 | 81.07 | 29.74% | 161.56 | 94.59 | 34.69% |
| LSTM | 133.85 | 69.04 | 25.49% | 132.42 | 69.41 | 25.45% | 139.51 | 72.53 | 26.61% |
| GRU | 140.69 | 73.59 | 27.18% | 136.94 | 72.15 | 26.47% | 130.81 | 73.20 | 26.85% |
| ST-ResNet | 132.01 | 71.03 | 26.23% | 146.56 | 82.07 | 30.10% | 162.27 | 92.54 | 33.94% |
| Transformer | 117.49 | 75.56 | 27.89% | 127.39 | 70.71 | 25.94.% | 132.44 | 76.64 | 27.38% |
| ***Our Model*** | ***116.63*** | ***54.70*** | ***24.28%*** | ***122.86*** | ***59.54*** | ***25.37%*** | ***127.39*** | ***61.99*** | ***24.76%*** |

4.2 车站级预测效果分析

为验证模型在单个高铁站层面的具体预测效果，本文选取三个典型高铁站，该三个高铁站在不同预测时间步数量下预测流量与真实流量的对比如图5-图7所示。

第一个收费站为临沂市蒙阴收费站，节假日期间该站的进站流量和出站流量相当，为均衡性收费站。如图5所示，该收费站的进出站流量在工作日具有明显的早晚高峰特征，节假日进出站流量显著高于工作日流量。在不同预测时间步下，预测值均与真实值拟合效果较好，表明模型具有良好预测性能。

第二个收费站为济南市埠村收费站，节假日期间该站的进站流量大于出站流量，为流入型收费站。如图6所示，该收费站的进站和出站流量在工作日具有显著的规律性，表现出明显的早晚高峰特征。节假日期间，进站流量略高于工作日，而出站流量与工作日相当，节假日期间的流量增长并不显著。在所有预测时间步长下，预测曲线与真实值高度吻合，预测效果良好。

第三个泰安市彭集收费站，节假日期间该站出站流量大于进站流量，为流出型收费站。如图7所示，该站的工作日进出站流量明显低于前两个收费站，但在节假日期间，进出站流量增长幅度显著。尽管如此，本文所提出的模型可以准确捕捉交通流变化趋势，具有较强的鲁棒性。

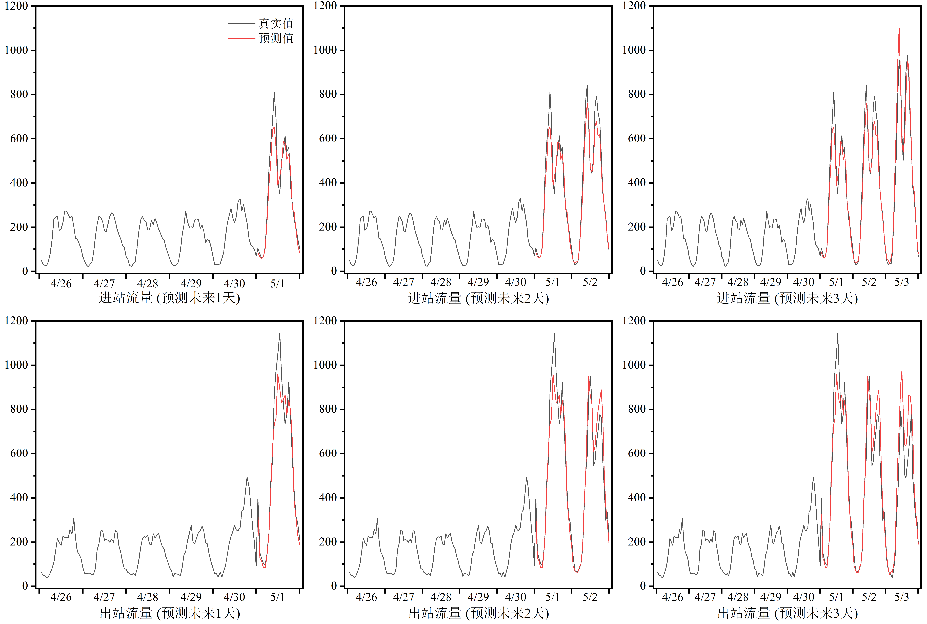


图5 蒙阴收费站流量预测值-真实值对比

4.3 消融实验

为了验证本文所提出的时空注意力交互模型中各个模块的有效性与优越性，本节进行了一系列消融实验，具体变体如下所示。

**变体1: ST-Attn-LLM-LSTM：**使用卷积长短时记忆网络模块替换时空注意力交互模型中的双向卷积长短时记忆网络模块，其余配置基本不变，探究双向建模方式对于特征挖掘的影响。

**变体2: ST-Attn-LLM-Decoder：**使用双向长短时记忆网络（Bi-LSTM）模块替换时空注意力交互模型中的双向卷积长短时记忆网络模块，其余配置基本不变，探究卷积运算在捕捉时空特性方面的能力。

**变体3: ST-Attn-LLM-No Attn：**在所提出模型中，删除注意力交互模块，探究交互注意力在提升特征表示方面的能力。需要注意的是，在去除该模块后，进站交通流和出站交通流的隐藏特征不再进行交互，此时模型可以视为两个单任务学习模型的简单合，由于模型的损失函数中仍然包含进站交通流部分和出站交通流部分，因此与单任务学习还是存在区别。

**变体4: ST-Attn-LLM-No LLM：**为学习节假日特征，本研究提出了进出站交通流增广时间序列矩阵，利用历史假期交通流辅助预测未来假期的交通流。此处使用原始的进出站交通流时间序列矩阵对未来交通流进行预测，旨在探究历史节假日交通流数据对与当前节假日交通流预测效果的提升。

表 3 消融实验进站交通流预测结果精度

Table 3 Inflow Prediction Result Evaluation of Variants

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **进站流** | | | | | | | | | |
| 变体 | 1天 (24个时间步) | | | 2天 (48个时间步) | | | 3天 (72个时间步) | | |
| RMSE | MAE | WMAPE | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE |
| 变体1(No Bi) | 137.55 | 68.56 | 29.25% | 143.21 | 72.04 | 29.19% | 155.94 | 87.61 | 33.42% |
| 变体2(LSTM) | 152.92 | 83.85 | 35.87% | 158.29 | 86.19 | 34.96% | 160.47 | 85.26 | 32.53% |
| 变体3(No Attn) | 139.84 | 74.35 | 31.71% | 145.76 | 77.67 | 31.47% | 169.76 | 83.35 | 31.80% |
| 变体4(No holidays) | 165.83 | 76.84 | 32.10% | 160.47 | 75.03 | 32.08% | 153.39 | 74.12 | 30.01% |
| ***ST-Cross-Attn*** | ***136.89*** | ***66.91*** | ***28.55%*** | ***142.32*** | ***70.27*** | ***28.47%*** | ***148.34*** | ***72.08*** | ***27.47%*** |

表 4 消融实验出站交通流预测结果精度

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **出站流** | | | | | | | | | |
| 变体 | 1天 (24个时间步) | | | 2天 (48个时间步) | | | 3天 (72个时间步) | | |
| RMSE | MAE | WMAPE | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE |
| 变体1(No Bi) | 117.23 | 55.88 | 24.79% | 124.22 | 61.72 | 26.87% | 132.51 | 72.22 | 28.82% |
| 变体2(LSTM) | 135.35 | 71.09 | 31.55% | 139.64 | 73.63 | 31.36% | 142.27 | 73.24 | 29.24% |
| 变体3(No Attn) | 119.10 | 60.01 | 26.62% | 125.71 | 64.83 | 27.61% | 162.37 | 79.87 | 31. 89% |
| 变体4(No holidays) | 167.52 | 65.37 | 29.02% | 154.16 | 72.71 | 30.83% | 138.08 | 71.69 | 28.63% |
| ***ST-Cross-Attn*** | ***116.63*** | ***54.70*** | ***24.28%*** | ***122.86*** | ***59.54*** | ***25.37%*** | ***127.39*** | ***61.99*** | ***24.76%*** |

表3和表4记录了所有消融实验结果，根据上述实验结果，得出以下的实验结论：

(1) **变体1将双向卷积长短时记忆网络替换成单向卷积长短时记忆网络，仅从单一方向挖掘交通流的时空依赖性。**由预测结果分析可知，模型的预测精度出现下降，这是因为单向建模无法充分挖掘时序数据的时间关联性和因果性，因此证明了双向建模方式的有效性。同时，变体1随着预测时间步的增加，预测误差逐渐变大，展示了双向建模在多步预测上的优越性，表明在面对较长预测时间步长的情况下，双向建模可以充分挖掘长时的时空依赖性，避免时间步数过长导致特征学习不足。

(2) **变体2将双向长短时记忆网络替换成双向长短时记忆网络，仅挖掘交通流时间序列数据的时间依赖性。**由预测结果分析可知，模型的预测精度明显下降，这表明使用LSTM进行特征挖掘时，模型只能捕捉有限的时间特征，无法充分建模收费站进出站交通流的空间特征。事实上，收费站进出站交通流间具有复杂的空间特征，例如市域通勤收费站点间存在显著的相关性，这些站点的进站交通流与出站交通流之间的潮汐特征明显。卷积长短时记忆网络可以有效利用卷积操作捕捉到目标站点的空间特征，从而有效提升预测精度。

(3) **变体3删除了原模型的注意力交互机制，不考虑进站流和出站流之间的关联性。**由实验结果分析可知，该变体在不同预测步长下的预测精度均出现下降，且随着预测步长的增加，预测精度下降程度增大。这是因为注意力机制在捕捉目标节点与其他节点的相关性时可以忽略节点间的距离长度，从而有效计算长距离外节点的相关性。在删除交互注意力机制后，模型计算目标节点与距离较远节点间相关性的能力减弱，随着预测步长增加，无法有效计算目标节点与更远的时间序列节点间的相关性，导致预测精度大幅下降。同时，交互注意力机制可以共享进站交通流与出站交通流的时空信息，增强特征表示，在删除该模块后，进站交通流与出站交通流的特征信息不再共享，无法实现时空特征交互，使预测精度下降。因此，交互注意力机制不仅有利于计算不同时间序列节点间的相关性，还可以共享进出站流特征信息，增强特征表示，从而提升预测精度。

(4) **变体4的输入由进出站交通流的增广时间序列矩阵替换为了进出站交通流时间序列矩阵，不考虑历史节假日的进出站交通流时间序列信息。**由实验结果分析可知，当模型不考虑历史节假日进出站交通流信息后，预测精度出现不同程度的下降，且当预测时间步为24时，预测效果最差，随着预测步长增加，预测效果提升。分析原因，主要是因为节假日特征通常出现在假期的第一天，当天交通流往往骤增，而后交通流相较日常的增幅变小，特征不如第一天明显。在不考虑历史节假日特征的情况下，预测时间步为24恰好为假期第一天，模型难以学习假期特征，所以第一天的预测效果最差；随着预测时间步增加，后两天的假期特征不如第一天明显，所以模型在后面两天的预测精度逐步提升。因此，出现了上述的实验结果，也进一步证明了本文构建的增广时间序列矩阵的有效性。

最后，与表格1对比可以看出，虽然消融实验中提出的变体模型的预测精度有所下降，但预测效果优于大部分基准模型，表明模型中的每个模块都是不可或缺的，可以有效提升模型的预测性能。

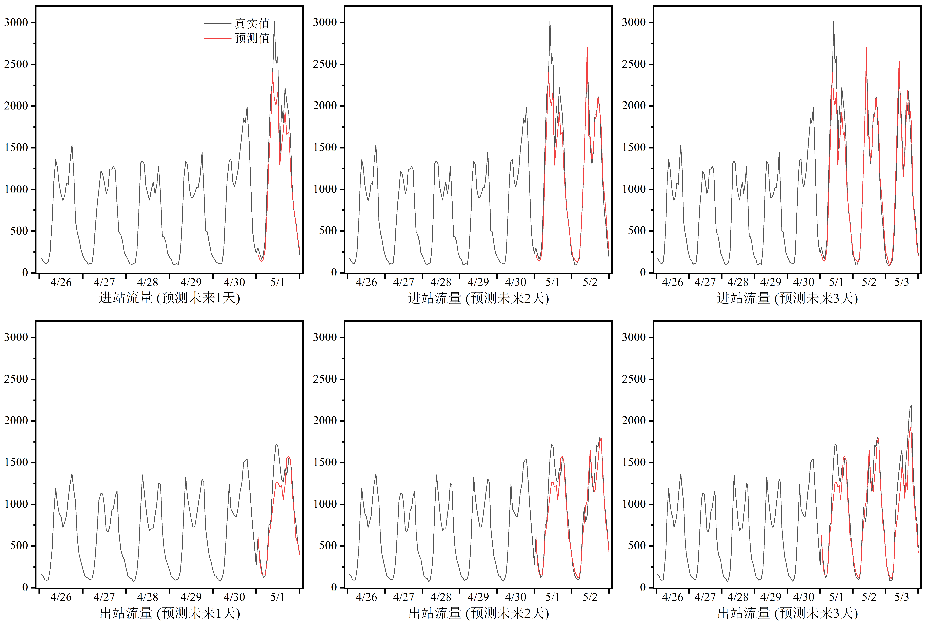


图6 埠村收费站客流预测值-真实值对比

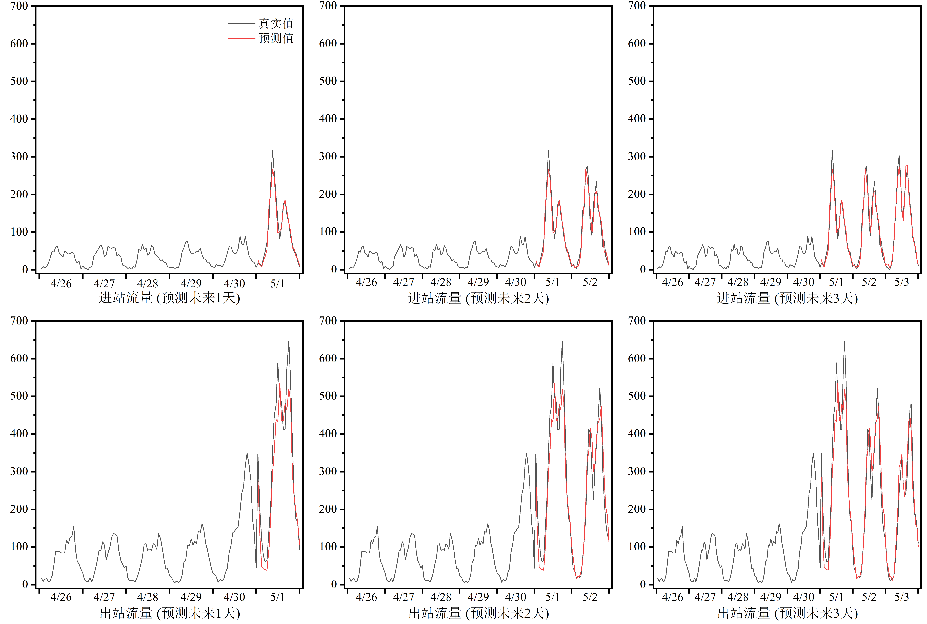


图7 彭集收费站流量预测值-真实值对比

4.4 多任务学习效果分析

本研究采用多任务学习框架对模型进行训练，通过共享进站与出站交通流的时空特征信息，有效提升两者之间的关联性，并充分挖掘进出站交通流的时空依赖特征，实现对进出站交通流的协同精准预测。为进一步验证所采用多任务学习策略的优势，本文将多任务学习方法与传统的单任务学习方式在预测精度上的表现进行了对比分析。具体而言，除了利用本文提出的多任务学习模型同步预测进出站交通流外，还分别采用单任务学习模式，对进站交通流和出站交通流各自独立进行预测，并对比两种策略下模型的预测效果。在单任务学习模式中，模型仅基于进站流或出站流数据，通过时空注意力机制提取其时空特征，随后借助大规模网络增强特征表达能力。最终，通过实验得到两种学习方式下的预测精度，并将结果如表所示呈现。

表5 多任务学习与单任务学习进出站交通流预测结果精度

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1天 (24个时间步) | | | 2 天(48个时间步) | | | 3天(72个时间步) | | |
|  | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE | RMSE | MAE | WMAPE |
| 多任务(进站流) | *136.89* | *66.91* | *28.55%* | *142.32* | *70.27* | *28.47%* | *148.34* | *72.08* | *27.47%* |
| 单任务(进站流) | 138.44 | 71.10 | 30.32% | 144.04 | 73.90 | 29.94% | 149.66 | 75.03 | 28.59% |
| 多任务 (出站流) | *116.63* | *54.70* | *24.28%* | *122.86* | *59.54* | *25.37%* | *127.39* | *61.99* | *24.76%* |
| 单任务(出站流) | 139.93 | 74.74 | 31.87% | 145.42 | 76.95 | 31.18% | 152.67 | 80.72 | 30.78% |

从上述预测结果可以看出，在各个预测步长条件下，采用多任务学习策略的进出站交通流协同预测模型，其预测表现均优于单任务学习模式下的预测效果。这充分说明，多任务学习通过进出站交通流的时空特征共享与交互，能够有效捕捉二者之间的内在关联性，增强特征表达能力，从而提升整体的协同预测性能。进一步分析结果发现，出站交通流在多任务与单任务学习模式下的预测效果存在明显差距，表明进站交通流所蕴含的时空特征信息对于出站交通流的预测具有重要的补充作用，有助于模型更准确地把握出站流的变化规律和时空分布特性。