

# 基于 AM-LSTM 模型的共享单车时空需求预测

许淼<sup>1</sup>,刘宏飞<sup>1</sup>,初凯<sup>2†</sup>

(1. 吉林大学 交通学院,吉林 长春 130022;2.吉林警察学院 交通管理系 吉林 长春 130025)

**摘要:**为准确预测城市不同区域的共享单车需求量,解决区域间供需不平衡问题,在对上海市共享单车数据进行时空特征可视化分析的基础上,研究共享单车的出行分布规律.针对时间出行分布的非严格周期性,提出了一种引入注意力机制的长短时记忆网络预测模型 AM-LSTM.利用 Spearman 相关性分析法分析特征影响因素,提取模型特征值.分别构建不同输入序列的预测模型,与传统时序预测模型进行对比分析.结果表明,采用 30 min 时间间隔的输入序列具有较高的预测精度,AM-LSTM 模型能够较好地预测共享单车的出行需求量,预测精度优于单一的 LSTM 模型.最后对预测曲线进行相关度分析,验证了 AM-LSTM 模型的预测性能,可以为城市共享单车的调度及分配提供有效信息.

**关键词:**城市交通;需求预测;AM-LSTM;共享单车;时空特征;数据可视化

**中图分类号:**U491.1

**文献标志码:**A

## Spatiotemporal Demand Prediction of Bike-sharing Based on AM-LSTM Model

XU Miao<sup>1</sup>, LIU Hongfei<sup>1</sup>, CHU Kai<sup>2†</sup>

(1. College of Transportation, Jilin University, Changchun 130022, China;

2. Department of Traffic management, Jilin Police College, Changchun 130025, China)

**Abstract:** In order to accurately predict the demand for bike-sharing in different regions of a city and solve the problem of imbalance between supply and demand, the travel distribution law of bike-sharing in Shanghai were studied based on the visualization analysis of spatio-temporal characteristics. In view of the non-strict periodicity of time travel distribution, Attention Mechanism was introduced into the Long-short Term Memory (LSTM) network to build a demand forecasting model named AM-LSTM. Spearman correlation analysis method was used to analyze characteristic influencing factors and extract model characteristic values. The prediction models of different input sequences were constructed and compared with the traditional time series prediction models. The results showed that the input sequence with a time interval of 30 min had a higher prediction accuracy. AM-LSTM model can better predict the travel demand of bike-sharing, and the prediction accuracy was higher than that of the single LSTM model. Finally, the correlation analysis of the prediction curve was conducted to verify the prediction performance of AM-LSTM model, which can provide effective information for the scheduling and distribution of urban bike-sharing.

**Key words:** urban traffic; demand forecasting; AM-LSTM; bike sharing; spatio-temporal characteristics; data visualization

\* 收稿日期:2019-12-17

基金项目:国家重点研发计划项目(2018YFB1601600), National Key R&D Program of China(2018YFB1601600)

作者简介:许淼(1995—),女,山东菏泽人,吉林大学博士研究生

† 通讯联系人, E-mail:407676667@qq.com

共享单车作为一种短距离通勤工具,为人们提供了一种可持续、零排放的出行方式,有效缓解了城市交通拥堵,在世界各地的许多城市得到了广泛应用<sup>[1-2]</sup>。截至2018年,中国无桩式共享单车用户规模达到了2.35亿人。然而,共享单车在迅速发展的同时,其时空需求的波动,导致了城市区域间的单车分布不平衡,一方面,部分区域存在大量的“无效需求”;另一方面,却造成部分地区“供不应求”,增加了调度及分配的运营成本。

国内外学者为提高区域共享单车需求量的预测精度开展了大量研究工作。一方面,在出行模式的影响因素研究中,Gebhart等<sup>[3]</sup>分析了天气因素对有桩共享单车使用的影响,结果表明,寒冷、降雨和高湿度天气会降低需求量和骑行时间;Bachand-Marleau等<sup>[4]</sup>针对社会经济和空间因素开展调研,分析了其对有桩共享单车使用频率的影响。上述因素的研究虽然多针对有桩共享单车,但为分析无桩式共享单车的外部影响因素提供了很有价值的参考。另一方面,在预测模型的研究中,以大数据为支撑,深度学习在时间序列预测上得到了广泛应用<sup>[5-7]</sup>。史越<sup>[8]</sup>利用小波神经网络预测共享单车的短时借还需求;宋鹏等<sup>[9]</sup>通过Matlab构建了基于不同核函数支持向量机的共享单车需求预测模型,验证了基于径向基核函数的模型具有最好的预测效果;MA等<sup>[10]</sup>基于微波传感数据,利用LSTM(Long-short Term Memory network)模型预测交通速度,通过与其他常用的参数化和非参数化算法比较,表明了LSTM神经网络在精度和稳定性方面都能达到最佳的预测效果;XU等<sup>[11]</sup>将大数据分析和LSTM模型运用到共享单车问题中,实现了单车需求量的预测。上述研究均在一定程度上实现了交通状态的预测,但是单一的时序模型不能满足精度要求,尤其是对具有明显波动的非严格周期数据需要改进模型来实现更加精确的预测。

近年来,注意力机制(Attention Mechanism)在深度学习各个领域得到广泛使用。Bahdanau等<sup>[12]</sup>首次利用双向循环神经网络(Bidirectional Recurrent Neural Network, Bi-RNN)获取语义特征过程中引入注意力机制,改进了对长句子的翻译效率和准确度;XU等<sup>[13]</sup>分别建立“hard”和“soft”注意力机制,并定量解释了模型权重的生成过程。现有研究表明,AM对序列学习任务具有巨大的提升作用,在编解码器框架,即Encoder-Decoder框架内,通过在编码端或解码端引入AM,对源数据或目标数据序列进行数据加权变换,可以有效提高系统模型的性能。

本研究在综合考虑共享单车用户出行需求的波动性及模型实效性的基础上,分析用户时空出行特征以及外部因素的影响。将注意力机制引入LSTM模型,构建改进的AM-LSTM预测模型,捕捉时间序列中的非严格周期性特征,实现不同时空共享单车资源需求量的精准预测,为城市共享单车的高效调度及分配提供方法支撑和研究途径。

## 1 时空数据处理及影响因素分析

通过时空数据可视化方法获取城市共享单车的出行特征<sup>[14]</sup>,对其影响因素进行相关性分析,提取特征因素,建立模型特征值,挖掘数据内在的隐含信息以及多源数据之间的实际关联关系。

### 1.1 时空数据可视化

#### 1.1.1 时间出行特征分析

首先对上海市MOBIKE 4个自然周的共享单车数据进行预处理。行程记录数据包括每辆共享单车的ID号、经度、纬度、时间戳和锁状态。其中,解锁状态记录为0,关锁状态为1。利用Python数据分析模块Pandas,清除冗余数据,删除无效信息,获取解锁状态的单车行程数据,并以此计算不同区域内各时间段共享单车的需求量。

以2个自然周内共享单车需求量为例进行分析,如图1所示,单车需求量按天、周两个模式呈周期性变化,并在工作日出现明显的早晚高峰。但由于周五、周六和周日受天气变化的影响,周对比曲线存在非一致性,需求量出现了局部时间变化现象,表明了上海市共享单车的出行特征具有短时和长时的周期特性,且存在局部时间周期性转移变化。

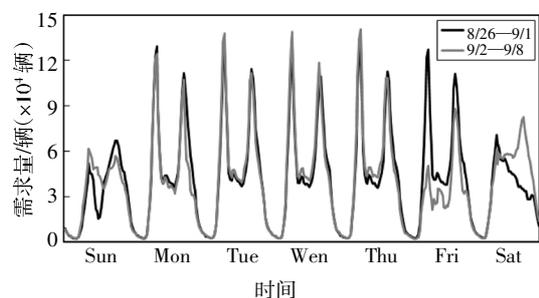


图1 共享单车需求量周对比曲线

Fig.1 Weekly demand curves for bike sharing

#### 1.1.2 空间出行特征分析

选取对日常通勤量影响较大的POI(Point of Interest),如地铁站、公交站、住宅等,研究上海市POI分布对共享单车空间出行的影响。以地铁站分布为

例,利用 GIS(Geographic Information System)点统计方法,对共享单车空间分布特征进行可视化.如图 2(a)所示,单车热点分布与 POI 密集度分布相似,需求量高值分布区域与地铁站位置密切相关,共享单车的使用围绕地铁站呈聚集现象,城市 POI 分布对共享单车出行特征具有显著影响.

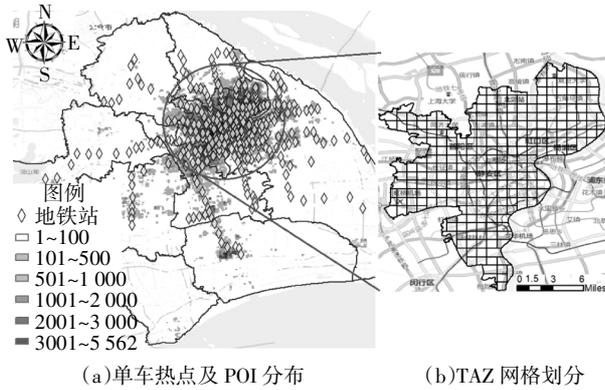


图 2 空间分布热力图及中心城区 TAZ  
Fig.2 Spatial distribution heatmap and central city's TAZ

由于上海市中心城区是 POI 高密度分布区域,共享单车的需求量预测研究将针对该区域进行.现有研究表明城市共享单车主要用于解决城市“最后一公里”出行问题<sup>[15]</sup>,传统的交通小区(Traffic Analysis Zone, TAZ)划分尺度会对共享单车的出行预测精度产生影响.因此,利用网格划分法,对中心城区划

分出符合共享单车出行特征的 TAZ.如图 2(b)所示,共构建 274 个平均大小为 1 平方千米的 TAZ,其中,边界区域进行了局部合并.

1.2 特征因素分析

上述分析表明,除时序数据中的时间相关因素外,外部因素包括 POI 分布、是否为工作日以及天气数据等也会对共享单车的使用产生影响.为明确影响因素与共享单车需求量的相互关系,对表 1 所示的 8 个影响因素及需求量进行 Spearman 相关系数分析.任意两个因素定义为大小为  $N$  的变量  $X$  和  $Y$ ,相关系数计算公式为:

$$\rho_{X,Y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (1)$$

将对应元素变量  $x_i$  和  $y_i$  转换为变量的秩次  $rg(x_i)$  和  $rg(y_i)$ ,且当所有秩次不相同,相关系数计算公式可以表示为:

$$\rho_{rg(X),rg(Y)} = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2-1)} = 1 - \frac{6 \sum_{i=1}^N (rg(x_i) - rg(y_i))^2}{N(N^2-1)} \quad (2)$$

表 2 所示的相关系数矩阵采用了双尾显著性检验的计算结果,可以看出,除压力因素  $h$  外,共享单

表 1 影响因素数据描述

Tab.1 Data description of influencing factors

离散型因素			连续型因素			
影响因素	取值范围	描述	影响因素	最小值	最大值	平均值
a 天气状况	1~4	{1:晴/多云;2:小雨/阴;3:中雨;4:大雨}	e 温度	22.00	36.00	28.93
b 星期	0~6	{0:周一;1:周二;...;6:周日}	f 湿度	36.00	100.00	74.77
c 小时	0~23	{0:0 时;1:1 时;...;23:23 时}	g 风速	0.00	29.00	14.46
d 是否工作日	0/1	{0:非工作日;1:工作日}	h 压力	998.60	1 014.60	1 005.93

表 2 影响因素相关系数矩阵

Tab.2 Correlation coefficient matrix of influencing factors

需求量	a	b	c	d	e	f	g	h	
需求量	1.000	-.676**	-.381*	.465**	.571**	.432**	-.424**	.356**	.056
a	—	1.000	.003	-.039	-.019	-.318**	.324**	-.126**	.060*
b	—	—	1.000	0.000	-.791**	-.004	.022	-.026	.013
c	—	—	—	1.000	0.000	.015	-.024	.046	.052
d	—	—	—	—	1.000	.099*	-.099*	.012	-.008
e	—	—	—	—	—	1.000	-.625**	.140*	-.270**
f	—	—	—	—	—	—	1.000	-.017	-.057*
g	—	—	—	—	—	—	—	1.000	.054*
h	—	—	—	—	—	—	—	—	1.000

注:\*\*.相关性在 0.01 级别显著(双尾);\*.相关性在 0.05 级别显著(双尾).

车需求量与其他各因素均显著性相关,且与天气状况和湿度因素呈负相关.但由于下划线标注的两处数据反映出  $b$  和  $d$  以及  $e$  和  $f$  均存在显著相关性,因素内部的强相关关系造成了影响因素的冗余,因此,不再将  $b$  和  $f$  作为模型特征因素进行处理.

## 2 AM-LSTM 模型原理及构建

### 2.1 注意力机制原理

注意力机制就是通过计算输入对输出的影响程度来确定当前输入的权重,研究采用 soft attention 模型<sup>[14]</sup>,结构如图 3 所示.

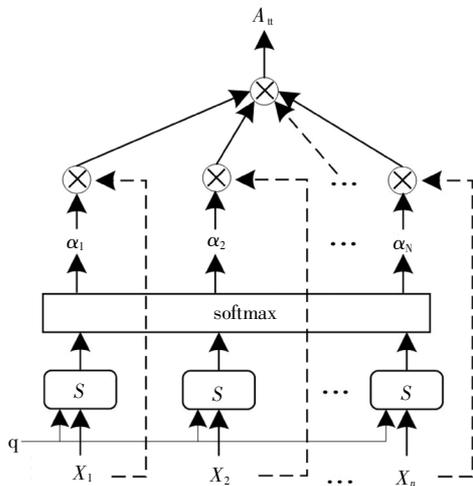


图 3 注意力机制原理图

Fig.3 Schematic diagram of attention mechanism

为便于描述注意力机制(AM)的原理,令 Encoder 模型为  $S$ ,输入  $X$  由  $n$  个大小为  $m$  的向量组成,在处理输入数据的过程中,计算不同输入  $x_i$  对隐藏层信息  $q$  的影响程度,获取权重  $\alpha_i$ ,总权重计算过程:

$$A_n = \text{attention}(x_{1:n}, q) = \sum_{i=1}^n \alpha_i x_i = \sum_{i=1}^n \text{softmax}(S(X_i, q)) x_i \quad (3)$$

$$S(X_i, q) = x_i^T W q = [x_1^T, x_2^T, \dots, x_n^T] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} q \quad (4)$$

式中:  $q$  为隐含层输入,  $\alpha_i$  为输入向量对应的权重,  $x_i$  为输入向量,  $W$  为模型学习权重.

利用注意力机制建立模型输入序列与隐含层输

入之间的关系,让模型自动学习输入序列数据的影响程度,对预测目标更关键的数据序列可以获取更高的权重.模型根据数据序列对应的权重对输入数据进行加权平均,通过分配关注度提高预测效率和精度.

### 2.2 AM-LSTM 模型构建

在对长时间序列数据进行预测时,LSTM 通过“门控制”可以解决训练过程中的梯度消失或爆炸问题<sup>[6]</sup>,但由于图 1 反映出的上海市共享单车时间出行特征上的非严格周期特性,需要引入注意力机制来捕捉周期性的影响,通过权重分配解决输入时间序列数据中的局部周期性转移问题.

针对文中的研究问题,图 3 中的 Encoder 模型  $S$  由 LSTM 替代,构建引入注意力机制的 LSTM 模型,简称 AM-LSTM.图 4 为系统预测模型,由输入模块 A、长时预测模块 B 和短时预测模块 C 三部分组成,子模块 b 是模块 B 中 AM-LSTM 的详细结构表示.

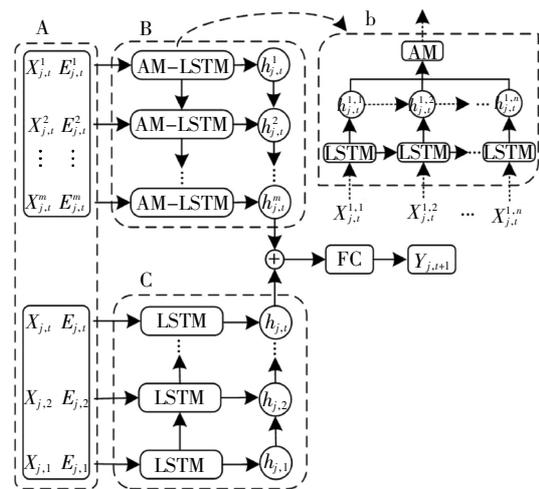


图 4 系统模型框架

Fig.4 Architecture of system model

模型预测目标为第  $T$  天的第  $t+1$  个时间段的单车需求量,选取当前预测日期的前  $m$  天,提取每天第  $t$  个时间段的前  $n$  个时间段进行模型计算分析,并使用  $\{1, 2, 3, \dots\}$  对共享单车 TAZ 进行标注.

输入模块 A 包含数据预处理后获得的共享单车需求量  $X_{j,t}$  (第  $j$  个 TAZ 中第  $t$  个时间段内的需求量)和特征因素数据  $E_{j,t}$ .其中,利用 Python 数据处理库 Pandas 对离散变量  $a, c, d$  进行 one-hot 编码,对连续变量  $e, g$  进行归一化处理,根据 1.2 中分析确定的特征因素建立模型特征值.

作为模块 B 的输入,  $X_{j,t}^d = [X_{j,t}^{d,1}, X_{j,t}^{d,2}, \dots, X_{j,t}^{d,m}]^T$  和  $E_{j,t}^d = [E_{j,t}^{d,1}, E_{j,t}^{d,2}, \dots, E_{j,t}^{d,m}]^T$  均是具有 3 个时间变化和 1 个空间变化的四阶张量, 任意一天  $d$  的输入由两个

$$X_{j,t}^d = \begin{bmatrix} x_{1,1}^d & x_{1,2}^d & \dots & x_{1,t}^d \\ x_{2,1}^d & x_{2,2}^d & \dots & x_{2,t}^d \\ \vdots & \vdots & & \vdots \\ x_{j,1}^d & x_{j,2}^d & \dots & x_{j,t}^d \end{bmatrix} = \begin{bmatrix} [x_{1,1}^{d,1}, x_{1,1}^{d,2}, \dots, x_{1,1}^{d,n}] & [x_{1,2}^{d,1}, x_{1,2}^{d,2}, \dots, x_{1,2}^{d,n}] & \dots & [x_{1,t}^{d,1}, x_{1,t}^{d,2}, \dots, x_{1,t}^{d,n}] \\ [x_{2,1}^{d,1}, x_{2,1}^{d,2}, \dots, x_{2,1}^{d,n}] & [x_{2,2}^{d,1}, x_{2,2}^{d,2}, \dots, x_{2,2}^{d,n}] & \dots & [x_{2,t}^{d,1}, x_{2,t}^{d,2}, \dots, x_{2,t}^{d,n}] \\ \vdots & \vdots & & \vdots \\ [x_{j,1}^{d,1}, x_{j,1}^{d,2}, \dots, x_{j,1}^{d,n}] & [x_{j,2}^{d,1}, x_{j,2}^{d,2}, \dots, x_{j,2}^{d,n}] & \dots & [x_{j,t}^{d,1}, x_{j,t}^{d,2}, \dots, x_{j,t}^{d,n}] \end{bmatrix}$$

$$E_{j,t}^d = \begin{bmatrix} e_{1,1}^d & e_{1,2}^d & \dots & e_{1,t}^d \\ e_{2,1}^d & e_{2,2}^d & \dots & e_{2,t}^d \\ \vdots & \vdots & & \vdots \\ e_{j,1}^d & e_{j,2}^d & \dots & e_{j,t}^d \end{bmatrix} = \begin{bmatrix} [e_{1,1}^{d,1}, e_{1,1}^{d,2}, \dots, e_{1,1}^{d,n}] & [e_{1,2}^{d,1}, e_{1,2}^{d,2}, \dots, e_{1,2}^{d,n}] & \dots & [e_{1,t}^{d,1}, e_{1,t}^{d,2}, \dots, e_{1,t}^{d,n}] \\ [e_{2,1}^{d,1}, e_{2,1}^{d,2}, \dots, e_{2,1}^{d,n}] & [e_{2,2}^{d,1}, e_{2,2}^{d,2}, \dots, e_{2,2}^{d,n}] & \dots & [e_{2,t}^{d,1}, e_{2,t}^{d,2}, \dots, e_{2,t}^{d,n}] \\ \vdots & \vdots & & \vdots \\ [e_{j,1}^{d,1}, e_{j,1}^{d,2}, \dots, e_{j,1}^{d,n}] & [e_{j,2}^{d,1}, e_{j,2}^{d,2}, \dots, e_{j,2}^{d,n}] & \dots & [e_{j,t}^{d,1}, e_{j,t}^{d,2}, \dots, e_{j,t}^{d,n}] \end{bmatrix}$$

$n$  个时间段, 不同时间段的权值表示对第  $t$  个时间段预测影响的大小, 通过权值捕捉时间序列中的局部周期性转移特征.

模块 B 通过  $m$  个子模块  $b$  对前  $m$  天进行长时预测, 将当前时间段  $t$  的输出状态与前  $m$  天中的  $n$  个时间段按照式(6)逐个进行相似度计算, 通过式(7)正则化表达计算注意力权重, 得到前  $m$  天中时间段  $t$  的预测结果, 计算过程为:

$$h_{j,t}^{m,n} = \text{LSTM}([x_{j,t}^{m,n}; e_{j,t}^{m,n}], h_{j,t}^{m,n-1}) \quad (5)$$

$$\text{score}(h_{j,t}^{m,n}, h_{j,t}^{m,n}) = \mathbf{v}^T \tanh(W_1 h_{j,t}^{m,n} + W_2 h_{j,t}^{m,n}) \quad (6)$$

$$\alpha_{j,t}^{m,n} = \frac{\exp(\text{score}(h_{j,t}^{m,n}, h_{j,t}^{m,n}))}{\sum_{n \in N} \exp(\text{score}(h_{j,t}^{m,n}, h_{j,t}^{m,n}))} \quad (7)$$

$$h_{j,t}^m = \sum_{n \in N} \alpha_{j,t}^{m,n} h_{j,t}^{m,n} \quad (8)$$

式中:  $\text{score}$  为相似度计算函数,  $h_{j,t}^{m,n}$  为第  $j$  个 TAZ 的预测时间段  $t$  在第  $m$  天中时间段  $n$  的输出,  $\tanh$  为激活函数,  $W_1, W_2, \mathbf{v}$  为模型参数,  $\mathbf{v}^T$  是转置矩阵.

模块 C 仅通过 LSTM 模型对第  $T$  天第  $t$  个时间段进行短时预测, 计算公式为:

$$h_{j,t} = \text{LSTM}([x_{j,t}; e_{j,t}], h_{j,t-1}) \quad (9)$$

式中:  $h_{j,t}$  表示第  $j$  个 TAZ 在  $t$  时间段的输出.

其中, 模块 C 的输入  $X_{j,t}$  和  $E_{j,t}$  比模块 B 中输入减少了前  $m$  天的长时间变化, 是具有 1 个时间变化和 1 个空间变化的二阶张量, 以矩阵形式表示:

三阶张量表示:

结合图 3 注意力机制原理图, AM-LSTM 具体的模型结构如图 4 中子模块 b 所示. AM 作用于每天各

$$X_{j,t} = [X_{j,1}, X_{j,2}, \dots, X_{j,t}]^T = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,t} \\ x_{2,1} & x_{2,2} & \dots & x_{2,t} \\ \vdots & \vdots & & \vdots \\ x_{j,1} & x_{j,2} & \dots & x_{j,t} \end{bmatrix}$$

$$E_{j,t} = [E_{j,1}, E_{j,2}, \dots, E_{j,t}]^T = \begin{bmatrix} e_{1,1} & e_{1,2} & \dots & e_{1,t} \\ e_{2,1} & e_{2,2} & \dots & e_{2,t} \\ \vdots & \vdots & & \vdots \\ e_{j,1} & e_{j,2} & \dots & e_{j,t} \end{bmatrix}$$

最后, 将长时预测模块 B 中前  $m$  天内时间段  $t$  的输出和短时预测模块 C 中第  $T$  天时间段  $t$  的输出结合, 通过一层全连接层生成模型的最终预测结果:

$$Y_{j,t+1} = \tanh(W_{fc}[h_{j,t}^M, h_{j,t}] + b_{fc}) \quad (10)$$

式中:  $W_{fc}, b_{fc}$  为学习参数.

### 3 模型预测及结果分析

#### 3.1 实验数据

##### 3.1.1 数据描述

研究所使用的多源数据包括上海市的共享单车 GPS 数据、地理数据、天气数据和其他影响数据. 实验数据的具体信息如表 3 所示, 研究的共享单车共 354 233 辆, 数据收集时间为 2018 年 8 月 20 日到 9 月 16 日, 研究的时间范围为全天. 取上海市中心城区作为模型的预测地点, 将上文 1.2 中筛选出的天气状况、温度、风速、小时和是否工作日作为模型特征因素数据.

表3 模型数据集具体信息

Tab.3 Specific information of datasets

数据集	具体信息
空间数据	上海中心城区(274个TAZ)
时间数据	08/20/2018 - 09/16/2018 (时间间隔:30 min,60 min)
特征因素	取值范围及描述
天气状况	1~4 {1:晴/多云;2:小雨/阴;3:中雨;4:大雨}
小时	0~23 {0:0时;1:1时;...;23:23时}
是否工作日	0/1 {0:非工作日;1:工作日}
温度	[22.00,36.00] 单位:℃
风速	[0.00,29.00] 单位:m·s <sup>-1</sup>

### 3.1.2 数据处理

为确保模型对比的公平性,便于不同单位或量级的输入数据进行比较和加权,利用最小-最大标准化方法将不同类型的输入数据映射在[0,1]范围内,计算公式如下:

$$x^* = \frac{x - \min}{\max - \min} \quad (11)$$

式中: $x$ 代表输入样本值, $x^*$ 为标准化的值, $\max$ 为样本数据最大值, $\min$ 为样本数据最小值。

将标准化后的实验数据集划分为两种不同时间间隔的数据集进行对比预测.最后,将各模型的预测结果通过反标准化方法还原为原始预测值,进而进行预测性能的评价。

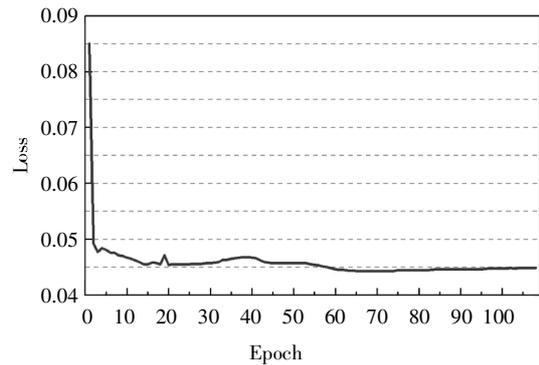
### 3.2 实验设置

实验环境为 Windows 10 操作系统,处理器型号 Intel(R) Core(TM) i9-9900KF CPU @ 3.60GHz,GPU 采用 Quadro P4000,基于开源 Python 环境管理平台 Anocinda1.9.7,利用 TensorFlow 中的 Keras 2.2.4 模块进行模型的搭建及预测<sup>[17-18]</sup>。

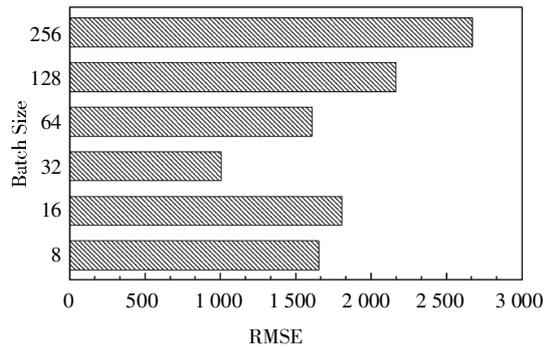
由于模型输入是时间序列数据,按机器学习中通用比例 6:2:2<sup>[19]</sup>,对数据集进行训练集、验证集和测试集划分.通过多次训练利用验证集调整模型超参数来提高模型的泛化能力,最后利用测试集评估模型的精确度。

### 3.3 模型超参数优化

为了优化所提出的模型结构,提高预测性能,使用学习率  $lr=0.01$  的 Adam 优化器来训练模型<sup>[16]</sup>,Adam 算法中超参数设定值为  $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10e^{-8}$ ,进而对模型中的超参数进行分析,结果如图 5 所示。



(a) 训练周期数



(b) 样本批量大小

图5 模型超参数敏感性分析

Fig.5 Sensitivity analysis of model hyperparameters

从图 5(a)中可以看出,训练周期达到 65 左右时出现最低损耗,且在 70 之后获得恒定的损失值,在模型训练过程中,取 Epoch=60,足以使得 Adam 优化算法收敛.为发挥 GPU 最佳性能,实验中 Batch Size 取图 5(b)所示的大小,但数值过大或过小都会导致模型的预测误差增加,故取 Batch Size=32 以获取最佳的模型预测效果。

### 3.4 模型评价指标

以 RMSE、平均百分比误差(MAPE)和决定系数  $R^2$  为各预测模型的评价指标,具体计算过程如下:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_{\text{test}}^i - y_{\text{predict}}^i)^2} \quad (12)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \frac{|y_{\text{test}}^i - y_{\text{predict}}^i|}{y_{\text{predict}}^i} \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_{\text{predict}}^i - \bar{y}_{\text{predict}}^i)^2}{\sum_{i=1}^m (y_{\text{predict}}^i - \bar{y}_{\text{predict}}^i)^2} \quad (14)$$

式中: $m$ 为样本个数, $y_{\text{test}}^i$ 为样本标签实际值, $y_{\text{predict}}^i$ 为样本预测值, $\bar{y}_{\text{predict}}^i$ 为平均样本预测值。

### 3.5 预测结果对比分析

在预测实验中,与传统时间序列预测模型 RNN

和 LSTM<sup>[13]</sup>进行对比以更好地验证所提模型的预测性能. 由于数据时间间隔分为 30 min 和 60 min, 对应模型输入序列有 48 和 24 两种, 则每个模型对应两种变体形式. 同时, 为进一步验证特征因素对预测结果的影响, 建立了去除特征因素  $E_{j,i}$  的变体模型 AM-LSTM1.

因此, 实验中共构建 8 个模型 AM-LSTM\_24、AM-LSTM\_48、AM-LSTM1\_24、AM-LSTM1\_48、LSTM\_24、LSTM\_48、RNN\_24 和 RNN\_48 进行对比分析, 其中, 相同输入序列模型的输入数据保持一致.

表 4 所示的预测结果是各预测模型对中心城区 TAZ 预测结果的平均值, 通过图 6 和图 7 可以更直观地分析不同模型的整体预测效能.

表 4 模型预测结果

Tab.4 Prediction results of models

模型	MAPE/%	RMSE	R <sup>2</sup>
RNN_24	22.90	71.07	0.69
RNN_48	20.96	67.39	0.68
LSTM_24	15.61	52.73	0.73
LSTM_48	14.97	49.65	0.73
AM-LSTM_24	11.38	39.81	0.81
AM-LSTM_48	9.28	35.87	0.82
AM-LSTM1_24	32.39	91.41	0.62
AM-LSTM1_48	31.07	86.20	0.64

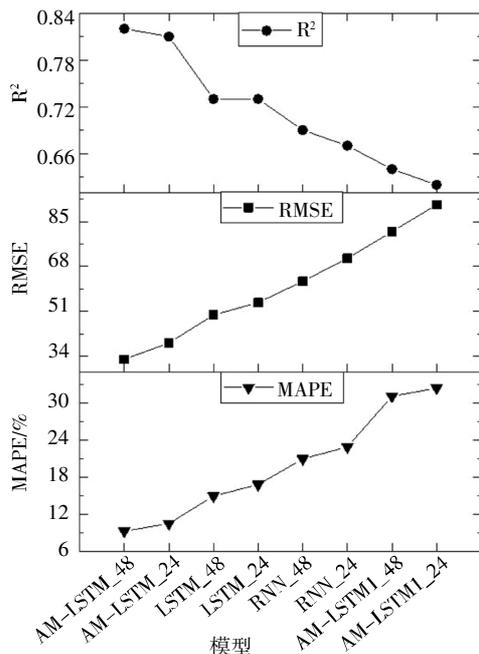


图 6 不同模型预测性能对比

Fig.6 Comparison of prediction performance of different models

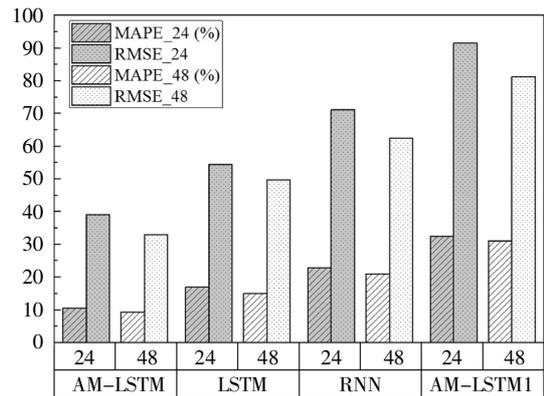


图 7 不同输入序列预测结果对比

Fig.7 Comparison of prediction results of different input sequences

### 3.5.1 注意力机制对模型的影响

从图 6 中各模型预测结果的对比曲线可以看出, 在不同输入序列的模型中, 引入注意力机制的 AM-LSTM 模型的预测精度均显著高于传统时间序列预测模型.

将 LSTM 与 AM-LSTM 进行对比. 当输入序列为 24 时, AM-LSTM 比 LSTM 在单车需求量上的 MAPE 降低了 4.23%, RMSE 减少了 12.92; 当输入序列为 48 时, MAPE 降低了 4.69%, RMSE 减少了 13.78. 结果表明, 注意力机制在处理较长时间序列数据时可以表现出更大的优势, AM-LSTM 模型的预测性能得到更大的提升. 注意力机制的引入, 可以使改进的 LSTM 模型的预测精度最多提高 4.69%.

### 3.5.2 输入序列对模型的影响

从图 7 可以看出, 对于同一模型, 当模型输入序列为 48 时, 其预测结果要优于输入序列为 24 时的预测结果. 例如 LSTM 模型的预测性能优于 RNN, 尤其是在输入序列为 48 时, 长时间序列预测结果表明, LSTM 比 RNN 的 MAPE 降低了 5.99%, RMSE 降低了 17.74, 决定系数 R<sup>2</sup> 提高了 6.85%.

对 4 类模型不同输入序列的预测结果分析可知, 由于共享单车的使用多是为了解决市内短距离出行, 骑行时间相对较短, 较长的时间间隔输入必然会导致模型预测性能下降. 而针对 30 min 以下的时间间隔, 已有研究<sup>[11]</sup>表明, 过短的时间间隔数据会造成数据噪声和无用波动的增加, 同样会导致预测精度下降. 因此, 模型选取 30 min 时间间隔的输入序列具有最好的预测性能.

### 3.5.3 特征因素对模型的影响

将 AM-LSTM1 与其他模型进行对比. 当模型输

入不考虑融合特征因素时,AM-LSTM1的预测性能与AM-LSTM相比显著下降,平均百分比误差分别达到了31.07%和32.39%。

即使在引入注意力机制的基础上,当输入序列为48时,AM-LSTM1比传统时序预测模型LSTM和RNN的MAPE分别增加了16.10%和10.11%,RMSE增加了36.55和18.81。这表明注意力机制的引入也不能弥补特征因素对模型预测性能的影响,进一步说明了预测模型中构建影响因素特征值的重要性。

综上所述,当输入序列为48时,以传统LSTM模型为基础,引入注意力机制,同时融合外部影响因素数据,可以使模型的预测性能大大提高。

### 3.6 AM-LSTM 预测效果验证

为验证AM-LSTM模型的有效性,将序列大小为48的测试集数据输入训练好的模型中,对上海市中心城区共享单车的需求量进行整体预测。

如图8所示,测试集中的预测结果与实际值的变化趋势基本保持一致,在白天期间,尤其是早晚高峰时段附近,预测效果较好。但是晚间至凌晨期间,由于共享单车的使用率大大降低,预测结果出现波动现象,模型预测准确度有所降低。

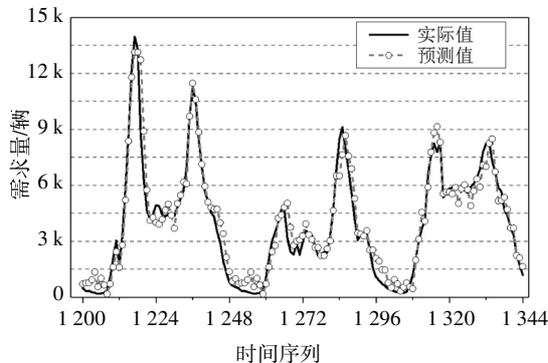


图8 AM-LSTM 模型预测结果

Fig.8 Prediction results of AM-LSTM

$$\rho = \frac{M \sum y_{\text{val}} y_{\text{pred}} - \sum y_{\text{val}} \sum y_{\text{pred}}}{\sqrt{M \sum y_{\text{pred}}^2 - (\sum y_{\text{pred}})^2} \sqrt{M \sum y_{\text{val}}^2 - (\sum y_{\text{val}})^2}} \quad (15)$$

式中: $M$ 为验证集样本数, $y_{\text{val}}$ 为实际值, $y_{\text{pred}}$ 为预测值。

为定量分析模型预测性能,通过Pearson相关分析法分析预测曲线与实际曲线的相关度,表5为相关度参数及计算结果,其中 $P$ 值远小于0.01,且相关系数达到了0.9709,两曲线显著相关,说明模型具有

较好的预测性能,可以满足城市共享单车需求量的预测。

表5 预测曲线与实际曲线相关度参数  
Tab.5 Parameters of correlation between predicted and actual curve

	平均值	最小值	最大值	相关系数 $\rho$	$P$ 值
预测值	4 140	174	13 151	0.970 9	9.701 1E91
真实值	3 958	172	13 967		

## 4 结论

1)针对上海市共享单车出行特征的非严格周期性,提出了一种引入注意力机制的AM-LSTM需求预测模型。AM的引入可以更好地捕捉模型输入数据的时序特征,改进的模型最高预测精度达到90.72%,比传统LSTM模型的预测精度提高了4.69%。

2)不同的时间间隔划分,即模型的输入序列对预测性能会产生影响。针对共享单车的需求量预测模型AM-LSTM,采用30 min时间间隔的输入序列比60 min的预测精度提高了1.25%。

3)模型特征值对预测性能具有显著影响,输入数据中融合外部影响因素是必要的。不考虑特征影响因素的AM-LSTM1\_48比原模型的预测精度下降了22.79%。

AM-LSTM模型实现了共享单车需求量的精确预测,可以为城市共享单车系统的调度与分配提供有效信息,解决区域间分配不平衡问题。下一步工作会继续进行注意力机制的研究,进一步探究AM与时序预测模型的融合性能,并通过其他类型的交通时序数据验证模型的普适性。

## 参考文献

- [1] WANG M S, ZHOU X L. Bike-sharing systems and congestion: evidence from US cities[J]. *Transport Geography*, 2017, 65: 147-154.
- [2] ELLIOT F, 魏贺. 公共自行车近期文献综述[J]. *城市交通*, 2015, 13(6): 80-94.  
ELLIOT F, WEI H. Recent literature review on public bicycles[J]. *Urban Traffic*, 2015, 13(6): 80-94. (In Chinese)
- [3] GEBHART K, NOLAND R B. The impact of weather conditions on bikeshare trips in Washington, DC [J]. *Transportation*, 2014, 41(6): 1205-1225.
- [4] BACHAND-MARLEAU J, LEE B H Y, EL-GENEIDY A M. Better

- understanding of the factors influencing the likelihood of using shared bicycle systems and frequency of use[J]. *Transportation Research Record*, 2012, 2314(1): 66—71.
- [5] 刘智慧, 张泉灵. 大数据技术研究综述[J]. *浙江大学学报(工学版)*, 2014, 48(6): 957—972.
- LIU Z H, ZHANG Q L. Big data according to the technical research review[J]. *Journal of Zhejiang University (Engineering Science)*, 2014, 48(6): 957—972. (In Chinese)
- [6] 尹新, 周野, 何怡刚. 基于混合算法优化神经网络的混沌时间序列预测[J]. *湖南大学学报(自然科学版)*, 2010, 37(6): 41—45.
- YI X, ZHOU Y, HE Y G. Optimization of chaotic time series prediction of neural network based on hybrid algorithm[J]. *Journal of Hunan University (Natural Sciences)*, 2010, 37(6): 41—45. (In Chinese)
- [7] 李嘉, 刘春华, 胡赛阳, 等. 基于交通数据融合技术的行程时间预测模型[J]. *湖南大学学报(自然科学版)*, 2014, 41(1): 33—38.
- LI J, LIU C H, HU S Y, *et al.* A travel time prediction model based on traffic data fusion technology [J]. *Journal of Hunan University (Natural Sciences)*, 2014, 41(1): 33—38. (In Chinese)
- [8] 史越. 共享单车需求预测及调度方法研究[D]. 北京: 北京交通大学经济管理学院, 2019.
- SHI Y. Research on demand forecasting and rebalancing strategy of free-floating bike sharing system[D]. Beijing: School of Economics and Management, Beijing Jiaotong University, 2019. (In Chinese)
- [9] 宋鹏, 黄同愿, 刘渝桥. 基于 SVM 的共享单车需求预测[J]. *重庆理工大学学报(自然科学版)*, 2019, 33(7): 187—194.
- SONG P, HUANG T Y, LIU Y Q. Study on demand forecasting of shared bicycle based on SVM [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2019, 33 (07): 187—194. (In Chinese)
- [10] MA X L, TAO Z M, WANG Y H, *et al.* Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. *Transportation Research Part C: Emerging Technology*, 2015, 54: 187—197.
- [11] XU C C, JI J Y, LIU P. The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets[J]. *Transportation Research Part C: Emerging Technology*, 2012, 95: 47—60.
- [12] BAHDANAU D, CHO K H, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// *Proceedings of the Conference on ICLR*. San Diego, CA: Computational and Biological Learning Society, 2015.
- [13] XU K, BA J, KIROS R, *et al.* Show, attend and tell: Neural image caption generation with visual attention [J]. *Computer Science*, 2015, 2048—2057.
- [14] 付鑫, 孙茂棚, 孙皓. 基于 GPS 数据的出租车通勤识别及时空特征分析[J]. *中国公路学报*, 2017, 30(7): 134—144.
- FU X, SUN M P, SUN H. Taxi commuting identification and spatiotemporal characteristics analysis based on GPS data [J]. *China Journal of Highway and Transport*, 2017, 30 (7): 134—144. (In Chinese)
- [15] 周荣, 王元庆, 朱亮, 等. 基于时空数据的共享单车出行特征研究[J]. *武汉理工大学学报(交通科学与工程版)*, 2017, 43(1): 159—163.
- ZHOU R, WANG Y Q, ZHU L, *et al.* Research on characteristics of bike sharing trips based on spatiotemporal data [J]. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2017, 43(1): 159—163. (In Chinese)
- [16] SEPP H, JURGEN S. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735—1780.
- [17] ABADI M, AGARWAL A, BARHAM P, *et al.* Tensorflow: large-scale machine learning on heterogeneous distributed systems [EB/OL]. [2016-03-14]. <https://arxiv.org/pdf/1603.04467v1.pdf>.
- [18] CHPLLET F. Keras [EB/OL]. [2015]. <https://github.com/fchollet/keras>.
- [19] ZHANG A, 李沐, LIPTON Z C, 等. 动手学深度学习 [EB/OL]. [2019-08-02]. <https://zh.d21.ai>.
- ZHANG A, LI M, LIPTON Z C, *et al.* Hands-on learning deep learning [EB/OL]. [2019-08-02]. <https://zh.d21.ai>. (In Chinese)