

基于LSTM模型的轨道换乘站短期客流预测

陈思宇

重庆交通大学 重庆 400000

摘要: 为了提高城市轨道交通的运行效率,充分发挥运力且保证其运营的安全,来缓解道路交通拥堵的情况。本文通过对重庆市2018年1月客流数据进行分析处理,以重庆市轨道交通红旗河沟站和重庆北站南广场站两个换乘站为实验案例,提出了基于单变量及多变量的长短期记忆(LSTM)网络结构的轨道换乘站短时客流预测方法。最后,建立模型评价指标,将预测结果与BP(反向传播)神经网络模型进行误差对比分析,结果表明基于多变量LSTM网络模型在轨道换乘站短时客流量预测中效果更优,同时在将客流分组建立四种LSTM网络后,多变量较单变量LSTM模型预测精度得到了进一步提升。

关键词: LSTM模型; 多变量; 5分钟时间间隔客流量

一、引言

本文首先对重庆市轨道AFC数据进行预处理,通过作图分析,对客流特征进行提取来构造模型,然后利用LSTM网络模型进行轨道换乘站客流预测,与之其他模型进行比较分析后,得出基于LSTM模型预测效果更好。本次实验预测所采用的是2018年1、2、4月(三月数据损坏)重庆轨道三号线的运营数据来做模型的构建。由于此次题目所研究的是轨道换乘站客流,因此只是对换乘站客流进行预测研究,充分考虑工作日乘客出方式以及外出影响因素,来分析客流量时刻变化的趋势。通过利用相关数据分析软件对原始数据进行分析处理后,得到了三个月每天各个车站各时段的可客流量来做分析和预测。

本次研究的模型为选取前两个月的数据与处理过后,设置成训练集,后一个月的数据同样与处理过后,设置成测试集来构建出LSTM模型。然后将构建好的模型,选取1月前30天的客流数据做为训练集,后一天数据集做为测试集,采用单变量及多变量来预测后一天统计时间间隔为5分钟的客流,得出预测结果后,建立误差评价指标,计算误差。最后比较误差,得出更优的模型。

二、LSTM模型构建

综合考虑已经处理的数据样本集,在进行模型的分析时,基于多变量LSTM模型中,将数据集整理成数值矩阵,最终会形成一个90行204列矩阵输入进去训练好模型后可以来预测后一天时间间隔为5分钟的时间段的地铁客流量,因此LSTM网络模型预测可以使用的样本数据集的矩阵为90维。在接下来的具体短期客流预测,根据客流特征分析,可以选择合理的数据集输入到建立

好的模型构架中进行短期客流预测。

同时基于BP神经网络的进出站客流量预测,将整理好的数据集使用MATLAB软件,输入数据样本集到工作区后利用神经网络工具箱创建BP神经网络模型。本次使用重庆轨道3号线客流从2018年1月到2018年4月(三月客流数据损坏)这三个月中剔除异常数值,在整理好56天的工作日客流数据中选取10个星期一来做客流预测,每日6:15至23:15时间段时间间隔统计5分钟的进站客流,一共10天客流数据。先选取前8个星期一客流,则一共8天客流数据,其中80%数据作训练样本,10%数据作验证样本,10%数据作预测样本也组合成神经网络模型的数值矩阵,隐藏层选择10个,使用Levenberg-Marquardt算法进行BP神经网络训练好模型,最后根据实例分析选择相应的数据输入道模型框架中来进行短期客流预测。

三、实例仿真

1. 数据分析与预处理

将所得到的三个月客流数据首先按照国家规定的日历剔除节假日以及双休日客流。根据作图对比分析,由于一些突发事件或者发生意外,剔除异常时段客流以及特殊日期客流,然后利用软件进行分类汇总得到重庆轨道3号线每天运营6:15-11:15时段的客流数据,最后对这些初步处理的数据分时段进行分类统计,得到以5分钟的时间间隔统计客流。将这些各类数据整理成数据集,为后文客流分析以及短期客流预测做准备。

2. 评价指标

为评价BP神经网络模型和LSTM模型在轨道站点短时客流量预测问题中的算法预测结果精度,本文采用以

下二种评价指标:

(1) 均方根误差

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_i - x'_i)^2} \quad (5.1)$$

(2) 平均绝对误差百分比

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{x_i - x'_i}{x_i} \right| \times 100\% \quad (5.2)$$

其中, M 代表测试集样本数, x_i 代表真实值, x'_i 代表预测值。再加上最大绝对误差和平均绝对误差共五种方法是评价预测结果误差常用的方法, 这里利用其来评价 BP 神经网络模型和 LSTM 模型预测客流量与真实值客流量的误差, 可以有效地并且很直观的表现出算法预测结果精度。各项指标数值越小代表算法性能优越。

3. 模型预测结果

经过不断的调试, 本文将 LSTM 预测模型设置成四层。它们分别为第一层输入层, 第二层 LSTM 层和第三层的全连接层及最后一层的输出层。其中, 将维度为一维时间序列客流数据输入到第一层, 即用过去 204 个时间步来预测未来一天的客流, 第二层的 LSTM 网络里包含了两层, 前一层为 LSTM 层, 这层需要加入门控单元, 主要是用来有效地缓解了梯度消失以及防止长距离依赖的问题, 另一层则为全连接层, 这层的主要作用是为了将数据从高维的映射降低维, 为了更好的表示及计算, 同时保留有效的信息。由于 Adam 优化器它可以使得模型训练的时候, 收敛速度加快以及计算更为简便高效等特点, 因此, 本次单变量的 LSTM 网络模型的优化算法可以初步选择 Adam 优化器。

将各个轨道换乘站点前 30d 的数据作为训练集, 最后 1d 的数据作为测试集。本次电脑使用的环境为基于 windows10 系统的 64 为处理器, 在基于所学习的 python3.7 编程语言, 编写出程序代码来实现预测模型, 所用的集成开发环境是 anaconda 里面的 Spyder, 引入了 pandas、numpy 和 keras 等第三方 python 库来协助开发运行, 通过 keras 深度学习库构建了整个模型以及测试, 同时将整理出来的 1 个月全部数据集作为训练集放入基于单变量的 LSTM 网络模型中进行训练, 此次训练的次數调整为 50 次。

通过单变量的 LSTM 模型对红旗河沟站、重庆北站南广场站两个轨道换乘站进站客流进行客流预测, 各个站点的预测结果有所差异。

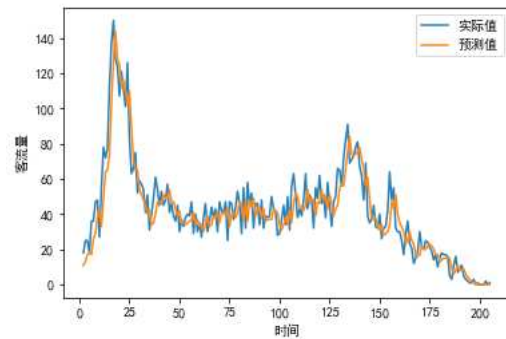


图1 红旗河沟站 (单变量)

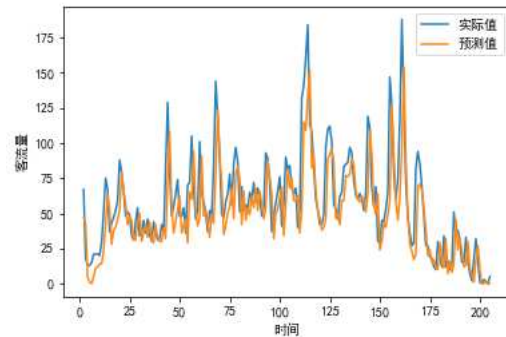


图2 重庆北站南广场站 (单变量)

如图1、图2所示, 基于 LSTM 模型单变量单步预测的结果来看, 整体上看模型的预测效果较好, 但是从单个时间段可以看出单变量单步预测的 LSTM 模型对换乘站一天客流的早高峰预测效果比其他平峰以及晚高峰好, 平峰效果相对来说较差。同时从之前所作出的特征图可以直观看出晚高峰客流变化的趋势是存在一致性的特征。晚高峰客流中的高峰时段针对不同的时间间隔统计也有所差异。以 5 分钟时间间隔统计集中在 17:35-17:55 和 18:00-18:30 两个时间段, 第一个峰形的峰值要比第二个略高一点, 以 10 分钟时间间隔统计集中在 17:20-17:50 和 17:50-18:20 两个时间段, 第一个峰形的峰值还是要比第二个高一点, 以 15 分钟时间间隔统计大致集中在 17:45-18:30, 其余时段相对来说也是都比较平稳。因此需要在此基础上考虑多变量的 LSTM 网络模型预测, 且按照不同时间段分成多组进行设置不同的模型参数。

在 LSTM 多变量单步预测模型中, 假设需要预测此刻时间段的轨道客流量为 $A_{i,j}$, 用 $(A_{i,j-1}, A_{i,j-2}, \dots, A_{i,j-m})$ 行向量表示在同一天此刻时间段前 m 个时段的地铁客流量, 体现地铁客流量数据对它是否存在随机性特征的影响; 用 $(A_{i-1,j}, A_{i-2,j}, \dots, A_{i-n,j})$ 列向量表示在不同的天内此刻时间段的前 n 天同一时间段的地铁客流, 这体现了地铁客流量数据对它是否存在周期性特征的影响。目前将 31 天数据按照统计时间为 5 分钟间隔来做预测, 此时分别

对应31维204个变量输入神经网络模型里面, 则可以使用LSTM网络预测模型针对输入样本数据集矩阵进行短期客流预测。矩阵表示如图3。

$$\begin{bmatrix} A_{i-n,j-m} & \cdots & A_{i-n,j-2} & A_{i-n,j-1} & A_{i-n,j} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ A_{i-2,j-m} & \cdots & A_{i-2,j-2} & A_{i-2,j-1} & A_{i-2,j} \\ A_{i-1,j-m} & \cdots & A_{i-1,j-2} & A_{i-1,j-1} & A_{i-1,j} \\ A_{i,j-m} & \cdots & A_{i,j-2} & A_{i,j-1} & A_{i,j} \end{bmatrix}$$

图3 数值矩阵

经过不停地实验, 根据前面将LSTM模型分别设置成双层的LSTM网络, 第一层改变成使用ReLU的激活函数, 设置0.01的学习率, 另一层则改变使用成tanh的激活函数, 也设置0.01的学习率。随后为dropout=0.2的随机失活层, 以免模型训练过程当中出现了过拟合的情况, 其次为全连接层, 有6个神经元, 输入层的维度为204维(即一天按5分钟间隔统计数据), 时间步长为1, 输出层为1个神经元, 训练次数也设定为50次。其余参数及优化器基于单变量模型来建立LSTM多变量模型。在此基础上, 将一天的204个数据分成4组, 每组51个分别进行预测。4组数据分别放入四种不同参数的LSTM模型中去预测, 最终预测效果如下;

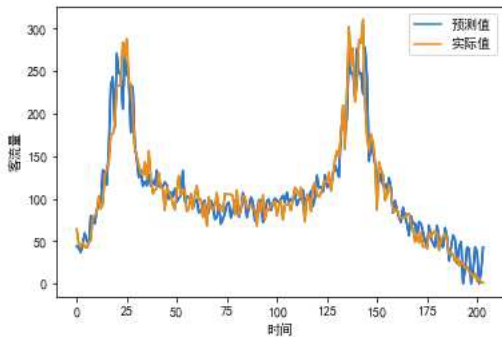


图4 红旗河沟站 (多变量)

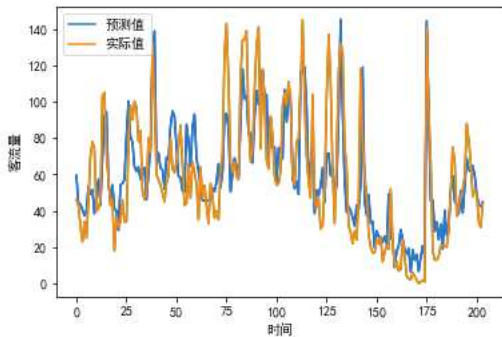


图5 重庆北站南广场站 (多变量)

以图4、图5可以只管看出不同轨道换乘站点的预测结果有所差异。具体分析如下表1所示。从表中可以清楚地看出, 所建立LSTM模型曲线效果更好, 即模型预

测精度优于BP神经网络, 预测出的曲线更接近实际值。

在设置多变量过后以及提升高峰期的预测精度, 将客流数据分为了5组得出基于多变量LSTM模型整体预测效果比单变量较好, 预测精度进一步提升了约5%左右。

表1 预测结果评价指标

模型\站点		红旗河沟站	重庆北站南广场站
BP神经网络	RMSE	43.02	48.62
	MAPE	27.99	35.06
LSTM (单)	RMSE	20.08	23.67
	MAPE	16.28	19.78
LSTM (多)	RMSE	16.48	19.01
	MAPE	15.34	16.57

四、结论

本文基于轨道换乘站客流提出了基于单变量及多变量的长短期记忆网络(LSTM)模型, 利用此模型预测出的地铁进站客流量相对与其他模型更准确。首先, 对重庆轨道刷卡数据进行了预处理, 基于对站点客流量的分析, 对提取出的客流特征进行构建模型。通过对两个轨道换乘站客流量特征提取分析后, 设计了基于单变量、多变量LSTM网络模型以及BP(反向传播)神经网络模型, 将得出的结果建立效果评价指标, 做出误差分析表。结果表明, LSTM模型预测性能明显优于BP(反向传播)神经网络。然后对于LSTM模型中, 将客流数据按照5分钟时间间隔统计, 同时为了在高峰期表现出更好的效果, 将数据平均分成四组, 最后得出基于多变量LSTM模型预测性能高于基于单变量LSTM模型, 这也充分证明了基于多变量的LSTM模型不仅适用于轨道换乘站短期客流预测, 同时模型的预测精度更优。

参考文献:

- [1]罗向龙, 李丹阳, 杨彧等. 基于KNN-LSTM的短时交通流预测[J]. 背景工业大学学报, 2018, 44(12):55-61.
- [2]王兴川, 姚恩建, 刘莎莎. 基于AFC数据的大型活动期间城市轨道交通客流预测[J]. 北京交通大学学报, 2018, 42(1):87-93.
- [3]李梅, 李静, 魏子健, 等. 基于深度学习长短期记忆网络结构的地铁站短时客流量预测[J]. 城市轨道交通研究, 2018(11):49.
- [4]毛静. 城市轨道交通客流短期预测方法及实证研究[D]. 北京: 北京交通大学, 2012.
- [5]林瑞华. 基于时间序列分析的轨道交通客流预测. 现代经济信息, 2013(22):370.