

基于注意力-长短期记忆模型的偶发性 交通流异常检测

郑大庆¹, 林陈威², 王昂杰¹

(1. 上海财经大学 信息管理与工程学院, 上海 200433; 2. 复旦大学 经济学院, 上海 200433)

摘要: 提出了基于注意力(attention)机制的 LSTM(long short-term memory)交通异常检测模型, 并利用整体交通网格点数据对交通流量进行异常检测。模拟数据集的验证结果表明, Attention-LSTM 预测模型具有较好的检测效果; SKAB(skoltech anomaly benchmark)公开数据集进一步检验了模型的检测能力。以上海市出租车 GPS 实时数据代表实时交通流状况, 对交通流异常进行检测, 并对检测结果进行了分析, 验证了 Attention-LSTM 模型的有效性。

关键词: 交通流量异常检测; 偶发异常; 注意力机制; 长短期记忆

中图分类号: U491

文献标志码: A

Traffic Flow Occasional Anomaly Detection Based on Attention-LSTM Model

ZHENG Daqing¹, LIN Chenwei², WANG Bingjie¹

(1. School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China; 2. School of Economics, Fudan University, Shanghai 200433, China)

Abstract: This paper proposes an long short-term memory (LSTM) traffic anomaly detection model based on the attention mechanism, which uses the overall traffic grid point data to predict the future traffic flow at various points in order to achieve the purpose of anomaly detection. By evaluating the simulation data, the Attention-LSTM model has a better detection effect. Furthermore, this paper uses the SKAB dataset to test the anomaly detection ability of the model, and obtains good results. Finally, this paper conducts experiments on the actual GPS signal data of Shanghai taxis, empirically analyzes some of the actual traffic anomalies

detected, and proves the effectiveness of the model in detecting actual traffic anomalies.

Key words: traffic flow anomaly detection; occasional anomaly, attention mechanism; long short-term memory (LSTM)

随着我国经济和社会的不断发展, 大型和超大城市不断涌现, 与之相伴相生的“大城市病”也越来越严重。“交通拥挤”作为大城市病的重要“症状”之一, 受到广泛关注^[1], 尤其是一些偶发性的交通异常, 由于其不确定性, 给城市交通安全与运行效率带来严重影响, 甚至造成财产和生命损失, 如 2014 年的上海外滩踩踏事件。在资源有限和空间约束的条件下, 发展智慧城市, 利用信息技术解决城市管理中的难题成为一个现实可行的思路。在智慧交通领域, 交通流异常检测非常重要, 它是进行智慧交通管理的前提, 而随着手机、GPS 等可移动设备的广泛应用, 基于大数据开展的交通流异常检测成为新的趋势^[2-4]。

交通异常检测是交通运输工程领域的一个重要分支, 在大数据时代, 城市管理者希望将交通监控设备、车载 GPS 设备等采集的交通数据与现有算法相结合, 对道路状况进行分析预测, 从而达到智慧交通管理的目的。目前交通异常检测的难点主要有: ①模型运行效率与检测结果的准确性之间的权衡。随着深度学习的发展, 为了追求更高的检测准确性, 许多复杂的深层模型被提出, 希望通过增加模型复杂性来提高准确性, 结果导致模型的运行时间增加, 不利于实际投入使用, 但若为追求高效运行, 使用简单的模型又会影响检测的准确性。因此需要在模型运

收稿日期: 2022-02-22

基金项目: 国家社会科学基金(20BGL287); 上海市智能信息处理实验室开放基金(IIP201903)

第一作者: 郑大庆(1978—), 男, 副教授, 管理学博士, 主要研究方向为信息管理与信息系统。

E-mail: zhengdaqing@sina.com

通信作者: 王昂杰(1998—), 男, 硕士生, 主要研究方向为算法治理。E-mail: wangbj1207@163.sufe.edu.cn



论文
拓展
介绍

行效率与检测结果的准确性之间进行合理的权衡。

②异常检测的泛化能力。道路交通流量会受到许多不同因素的影响,且不同的区域受影响的权重存在差异^[5],因此需要异常检测算法有较强的泛化能力,能对不同交通情况进行准确的异常检测。

交通异常检测与交通流量预测有非常紧密的联系,这个领域的研究主要经历了统计方法、机器学习方法和深度学习方法三个阶段,其中,相较于统计方法和机器学习方法,深度学习方法通过深层次的神经网络层学习复杂的非线性关系,从而发掘出更深层次的特征关系,因而越来越多的研究采用深度学习方法。在交通流异常检测方面,目前较为常见的异常检测基础模型是由 Hochreiter 等^[6]提出的 LSTM(long short-term memory)模型,因为 LSTM 能够学习到较长时间序列数据的复杂关系。Zheng 等^[7]提出的基于 LSTM 模型的交通流量预测方法,通过在 LSTM 网络上增加由多个存储单元组成的二维网络学习交通数据的时空相关性,从而实现单位时间内交通流量的预测。该模型较为简易且运行效率较高,但模型的准确性有待提升。

随着研究的逐步深入,许多研究开始考虑将 LSTM 模型作为基础,与传统统计模型或其他深度学习模型相结合,从而实现更高准确度的异常检测。刘世泽等^[8]将 LSTM 网络、卷积残差网络和注意力机制中的挤压激励模块相结合,对交通异常事件进行高效检测。彭桐歆等^[9]将卷积神经网络、残差单元和门控循环单元相结合,对交通流量数据进行捕获,从而对异常事件进行检测。Kong 等^[10]利用 LSTM 模型预测不同区域异常得分,然后通过 OC-SVM(one-class support vector machine)模型对相应的异常区域进一步作异常检测。宋瑞蓉^[11]等分别使用卷积长短期记忆网络、双向长短期记忆网络和全连接层提取交通数据的时空、周期以及外部特征,最后利用卡尔曼滤波器预测交通的线性特征。廖挥若等^[12]使用卷积长短期记忆网络和注意力机制结合神经残差网络进行学习,最后再结合如天气等其他因素取得更高的精确度。王海起等^[13]通过网格划分构建交通流量时空矩阵,使用卷积神经网络模型获得交通数据的空间特征,使用基于注意力机制的 LSTM 模型获得交通数据的时间信息。以上几种方法将不同模型结合从而达到较好的效果,但它们都

面临着模型运行效率低下的问题,而且相比较而言,交通流偶发性异常检测具有更大的实际应用需求。

综合来看,目前的交通异常检测和流量预测主要是基于 LSTM 网络构建,且存在着模型不断复杂化的发展趋势。高度复杂的模型固然会带来较高的检测准确性,但不应忽视交通领域对偶发性交通异常所需的高效、实时的需求,需要兼顾模型运行效率与检测准确性。由此,本文提出的 Attention-LSTM 交通异常检测模型,在控制模型复杂度的基础上充分利用交通流量数据的时空特性;同时采用不同数据集对该模型进行比较分析,展现了其在不同数据集上的适应能力以及对异常事件高效准确的检测性能。

1 模型构建

长短期记忆(LSTM)神经网络由 Hochreiter 等^[6]于 1997 年首次提出,用于解决循环神经网络(recurrent neural network, RNN)中较常见的梯度消失问题。2000 年, Gers 等^[14]引入遗忘门机制,使得 LSTM 网络能够重置内部状态。在基础的 LSTM 神经网络中,每个神经元由输入门、遗忘门和输出门组成,使其拥有长期记忆的能力,可以处理长时间序列数据。由于在整个交通网络中,某一节点的交通流量会受到其他节点的交通流量影响,且交通流量的变化存在时序性^[15],因此考虑使用 LSTM 模型作为检测交通异常的基础组件。虽然 LSTM 模型通过门结构的设定,部分避免了在时间序列模型中的梯度消失的问题。但是其仍然存在着信息传递“瓶颈”的问题,尤其是在处理高维度的数据时(如本文实验中的 2 000 维以上的实际数据),仅通过一个输出的隐状态来预测未来,极易导致信息的丢失。此外,基础的 LSTM 还面临着计算费时^[16]和无法并行化处理^[17]等问题。

Attention 机制最早由 Bahdanau 等^[18]于 2015 年提出,用于机器翻译,同年,由 Xu 等^[19]将其引入图像标注(image caption)领域。Attention 机制的本质是查询(query)与一系列键值对(key-value)的相似度计算,两者之间的相似度越高会分配更多的注意力资源,相似度越小则会分配更少的注意力资源。在实际的交通网络中,也存在类似的情况:某些交通

“堵点”是由路网中多个毗邻或者较远距离的节点共同造成^[16,20],通过 Attention 机制关注影响交通流量的“关键点”显得非常必要。

本文提出模型结构如图1所示,由3部分组成,分别为输入降维、预测计算以及维度恢复。图3中, a 为输入的交通流量数据的时间序列长度, b 为交通

流量异常预测的时间序列长度, m 为交通流量数据的特征数量, r 为降维后交通流量数据的特征数量, h 为LSTM单元隐状态的维度,ReLU(rectified linear unit)为整流线性单元。相较于其他研究,该模型结构简洁,避免大量深层网络的堆砌,在确保准确率的同时,提高了模型的学习效率。

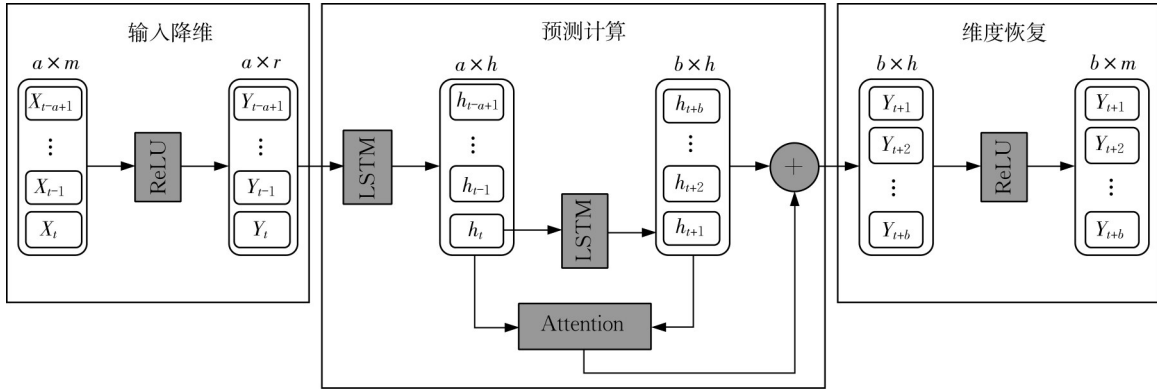


图1 Attention-LSTM交通异常检测模型

Fig.1 Attention-LSTM traffic anomaly detection model

1.1 输入降维与维度恢复

输入降维部分参考了Gugulothu等^[21]的研究,其研究发现,在对高维度的稀疏时间序列矩阵进行异常检测时,通过对输入矩阵进行维度降低可以提升模型异常检测的准确性。在本模型中,输入矩阵维度为 $(a \times m)$,即选取时间序列长度为 a 的 m 维数据进行计算,将输入矩阵的维度由 $(a \times m)$ 下降至 $(a \times r)$,其中满足条件 $a > r$ 。其数学表达式为

$$Y = \text{ReLU}(W_R X + B_R) \quad (1)$$

式中: X 为原始数据矩阵; W_R 和 B_R 为线性变换矩阵; Y 为降维后的数据矩阵。

维度恢复部分与输入降维部分相对应,需要将原本降低的输出维度恢复。为此本文采用与输入降维部分相同的方式,将输出维度由 $(b \times h)$ 恢复至 $(b \times m)$ 。其数学表达为

$$\bar{X} = \text{ReLU}(W_V \bar{Y} + B_V) \quad (2)$$

式中: \bar{Y} 为预测数据矩阵; W_V 和 B_V 为线性变换矩阵; \bar{X} 为维度恢复后的预测数据矩阵。

1.2 预测计算

预测计算部分具体过程如下:首先,将经过降维处理的输入矩阵依序代入第一层LSTM中计算,并将各时间点的输出值进行组合,得到维度为 $(a \times h)$ 的输出矩阵,其中 h 为LSTM单元隐状态的维度。

$$\{h_{t-a}, \dots, h_{t-1}, h_t\} = \text{LSTM}(Y_{t-a}, Y_t, \dots, Y_t) = \text{LSTM}(Y) \quad (3)$$

其次,将最后一个时刻的输出值 h_t 传入第二层LSTM层作为输入,由此得到此后 b 个时间段的输出矩阵,维度为 $(b \times h)$ 。

$$\{\bar{h}_{t+1}, \bar{h}_{t+2}, \dots, \bar{h}_{t+b}\} = \text{LSTM}(h_t) \quad (4)$$

最后,对两层LSTM层的输出进行Attention计算,其中, $\{\bar{h}_{t+1}, \bar{h}_{t+2}, \dots, \bar{h}_{t+b}\}$ 作为查询矩阵 Q ,键值对矩阵 K 和 V 均为 $\{h_{t-a}, \dots, h_{t-1}, h_t\}$ 。另外,受Vaswani等^[22]研究的启发,本文在Attention层之后使用残差连接,得到维度为 $(b \times h)$ 的预测值矩阵 \bar{Y} ,数学表达式如下:

$$K = V = \{h_{t-a}, \dots, h_{t-1}, h_t\} \quad (5)$$

$$Q = \{\bar{h}_{t+1}, \bar{h}_{t+2}, \dots, \bar{h}_{t+b}\} \quad (6)$$

$$\bar{Y} = \{\bar{Y}_{t+1}, \bar{Y}_{t+2}, \dots, \bar{Y}_{t+b}\} = \text{Attention}(Q, K, V) + V \quad (7)$$

1.3 损失函数

本文选用平均绝对误差 L_1 作为损失函数,计算某一时刻 t 预测值与实际值各维度的绝对差值的平均值,该值即为 t 时刻预测的损失值。将各个时间点的损失值进行求和取均值,可得到模型整体的损失值。

$$C_t = \frac{1}{m} \sum_{g=1}^m |\bar{X}_t^g - X_t^g|, 1 \leq t \leq n \quad (8)$$

$$\vartheta = \frac{1}{n} \sum_{t=1}^n C_t \quad (9)$$

为减小过拟合的可能性,避免模型对部分参数的过度依赖,本文对最后的目标函数引入参数正则化^[23]。其中 $W = \{W_R, W_L, W_V\}$, $B = \{B_R, B_L, B_V\}$, W_L, B_L 为 LSTM 层中的计算参数, $\|\cdot\|_1$ 则代表 L1 正则化,对模型中的所有参数进行绝对值求和,得到最优参数 W^* 和 B^* 。

$$W^*, B^* = \operatorname{argmin} \vartheta + \frac{\lambda}{n} (\|W\|_1 + \|B\|_1) \quad (10)$$

2 交通异常检测算法

2.1 交通流量数据的矩阵表达

本文将交通流量数据处理为在一定时间间隔内位于各网格中的出租车的数量矩阵。设置时间集合为 $T = \{t_1, t_2, t_3, \dots, t_n\}$, 共 n 个时间段, 设置网格点集合为 $G = \{g_1, g_2, g_3, \dots, g_m\}$, 共 m 个网格点。设 v_t^g 为在时间段 t 内于网格点 g 中的出租车数量, $V_t = \{v_t^1, v_t^2, \dots, v_t^m\}$ 为时间段 t 内所有网格点的全局出租车流量, 则 $V = \{v_t^g | 1 \leq g \leq m, 1 \leq t \leq n\}$ 为所有时间段内的全局交通流量, 其数学表达如下:

$$V = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} v_1^1 & v_1^2 & \dots & v_1^m \\ v_2^1 & v_2^2 & \dots & v_2^m \\ \vdots & \vdots & \ddots & \vdots \\ v_n^1 & v_n^2 & \dots & v_n^m \end{bmatrix} \quad (11)$$

2.2 交通流量数据降维

考虑到交通流量数据中部分网格点处于车流量较小的区域, 如水域、工厂、农田等, 这些区域通常不容易发生交通异常事件, 即使发生异常事件对市区整体的交通情况影响也极小, 因此本文考虑通过设定交通流量的阈值来排除这些网格点对模型预测准确性的影响。设 $\bar{\mu}_g, \bar{\sigma}_g$ 分别为网格点 g 的交通流量值的均值与方差, 阈值为 Threshold, 当 $\bar{\mu}_g \times \bar{\sigma}_g < \text{Threshold}$ 时, 该点交通流量值为 0, 其数学表达式如下:

$$v_t^g = 0, \bar{\mu}_g \times \bar{\sigma}_g < \text{Threshold} \quad (12)$$

2.3 交通流量数据预测

根据本文建立的 Attention-LSTM 异常检测模型, 对于任意的 $\{a \leq t \leq n\}$, 使用已知的实际交通流量 $X = \{V_{t-a+1}, V_{t-a+2}, \dots, V_t\}$ 来预测未来的交通流量 $\bar{Y} = \{\bar{V}_{t+1}, \bar{V}_{t+2}, \dots, \bar{V}_{t+b}\}$, 其中参数 a 为用于预测的交通流量的时间序列长度, b 为模型的预测时间序列长度。在本文中, 设置 $a = 5, b = 1$ 。

通过对训练集的学习, 得到完成训练的模型参数, 并以此为基础, 代入测试集进行计算, 由此得到

对于任意时间点 t 的交通流量预测值, 以函数 F 代表模型的计算过程, 得到计算过程的数学表达如下:

$$\bar{Y}_{t+1} = \bar{V}_{t+1} = F(X) = F(V_{t-4}, V_{t-3}, V_{t-2}, V_{t-1}, V_t), t \in [5, n] \quad (13)$$

对于任意的网格点 i , 可以计算位于该点的交通流量的预测值与实际值的差的均值与标准差, 即对于数据集 $\{\bar{v}_t^i - v_t^i | 5 \leq t \leq n\}$, μ_i 与 σ_i 分别为数据集的均值与标准差。基于统计学中大多随机误差都服从正态分布, 因此本文假设模型的预测误差服从正态分布, 并由此建立模型预测误差的分布函数为 $N(\mu_i, \sigma_i)$ 。

$$\{\bar{v}_t^i - v_t^i | 5 \leq t \leq n\} \sim N(\mu_i, \sigma_i) \quad (14)$$

2.4 交通异常事件的检测

在得到了各时间点 t 的交通流量预测值以及各网格点的预测误差正态分布函数后, 进行一步计算异常值得分以对异常事件进行检测。

对于任意网格点 i , 模型预测值与实际值之间的误差应满足正态分布 $N(\mu_i, \sigma_i)$, 而当异常事件发生时, 预测值与实际值会出现较大的偏差。因此, 本文选择将预测值与实际值的差代入正态分布的密度函数中计算, 其数学式如下:

$$P_i^t = P\{\bar{v}_t^i - v_t^i \in N(\mu_i, \sigma_i)\} = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{(\bar{v}_t^i - v_t^i - \mu_i)^2}{2\sigma_i^2}} \quad (15)$$

由于正态分布的密度函数满足钟型对称结构, 当偏差值越偏离正态分布, 则 P_i^t 值将越小。为满足异常偏差越大, 异常值越大的需求, 本文以式(15)为基础, 将网格点 i 于时间点 t 的异常值得分定义为

$$S_t^i = \ln \frac{1}{P_i^t} = -\ln P\{\bar{v}_t^i - v_t^i \in N(\mu_i, \sigma_i)\} \quad (16)$$

当异常事件发生时, 与之相关的网格点的异常值得分将远大于其他的网格点。本文将时间点 t 的各网格点的异常值得分进行汇总, 得到该时间点异常值总得分为

$$S_t = \sum_{i=1}^m S_t^i \quad (17)$$

由此, 通过对各时间点的异常值总得分进行比较, 可以及时发现异常事件的发生时间, 进一步, 可以比较该时间点各网格点的具体异常值大小以确定异常事件的发生地点。对于是否存在异常的阈值的确定, 可以根据不同数据进行调整, 本文将比时间窗口内异常值总得分均值高 50% 的情况认为出现

异常。

3 数据实验

本文的数据实验由3部分组成,实验一是使用Net Logo软件,建立车辆流动规则,模拟在一定区域内交通流量的变化情况,并人为添加持续时间不同的交通异常事件。将本文提出的Attention-LSTM模型与自回归模型(AR)以及LSTM模型的异常事件检测效果进行比较。考虑到模拟数据具有虚构性,无法体现现实交通数据的复杂性与多变性,实验二使用SKAB这一在现实中收集的多维传感数据集,并将异常检测结果与排行榜上其他模型进行比较。实验结果验证了Attention-LSTM模型交通异常检测的优越性与对实际数据的优秀适应能力。在此基础上,实验三使用2014年12月28日至2015年1月10日14 d内的出租车GPS数据,对模型检测到的异常事件进行实证分析,证明模型对检测实际交通异常事件的可行性与准确性。

3.1 Net Logo模拟数据实验

本文使用Net Logo模拟特定区域中车辆的轨迹随时间的变化情况,借鉴Zhou等^[24]提出的交通流量变化规则,模拟数据遵循以下规则:

(1)该区域内的车辆总数固定,由参数CarNumber决定。

(2)车辆分为两种状态,受雇与失业,受雇车辆的比例由EmployRate决定。

(3)随机生成住宅与公司的位置,各住宅和各公司的车辆数分别由PeoplePerCommunity和PeoplePerCompany决定。

(4)将1 d划分为120个时间段,即5个时间段为1 h。每日的第110至次日第19时间段为休息时间,此时受雇车辆于家中休息,失业车辆于原地休息。

(5)每日的第20至第39时间段为上班时间,受雇车辆由住宅向公司移动;每日第40至第89时间段为工作时间,受雇车辆停留在公司;每日第90至第109时间段为下班时间,受雇车辆由公司向住宅移动。

(5)除休息时间外的所有时间段,失业车辆于区域内随机游荡。

(6)随机生成一定数量的异常事件,事件发生地点随机生成,事件数量与异常持续时间分别由参数EventNumber与EventDuration决定。

(7)当异常事件发生时,一定数量的受雇车辆和

失业车辆会受到影响,临近异常点的部分车辆会受到阻碍,由参数AttractRate决定。各参数的名称及含义见表1。

表1 Net Logo模拟数据参数
Tab.1 Parameters of Net Logo simulation data

参数名称	参数意义
CarNumber	区域内的车辆总数
EmployRate	区域内处于受雇状态的车辆比例
PeoplePerCommunity	各住宅的车辆数量
PeoplePerCompany	各公司的车辆数量
EventNumber	异常事件数量
EventDuration	异常事件持续时间
AttractRate	受异常事件影响的车辆比例

模拟数据分为训练数据和测试数据两部分,训练数据根据规则(1)~(5)生成无异常事件的车辆轨迹数据,代入不同模型中进行训练。测试数据则根据规则(1)~(7)生成含有一定数量的异常事件的车辆轨迹数据,代入已训练的模型中进行异常检测,由此比较不同模型的异常事件检测能力。

模拟数据设置时间长度为7 d,即共840个时间段。将区域划分为 11×11 共121个网格点。为比较模型对不同持续时长的异常事件的检测能力,本文设置了3种异常持续时间(EventDuration=5, 10, 20)。以下以EventDuration=5的情况为例,其中异常事件数量为9(EventNumber=9),模型的预测结果如图2所示

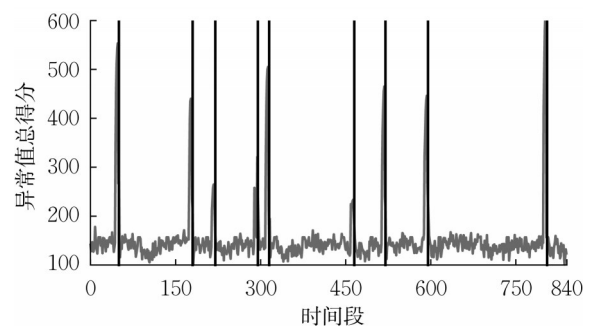


图2 Event Duration=5时的异常检测效果
Fig.2 Effect of anomaly detection at event duration of 5

竖直黑线所标记的是异常事件发生的时间点,可以发现在这些时间点的异常值总得分与其他时间的异常值总得分有较大的偏差,可以认为模型对异常事件的检测是显著的。

本文使用时间序列的传统统计模型自回归(AR)模型与深度学习中的LSTM预测模型作为比较基准。使用接受者操作特性曲线下方的面积大小

(AUC)作为比较标准,各方法在3种不同的异常持续时间(5/10/20)下的AUC值如表2所示

表2 各模型异常检测AUC比较

Tab.2 AUC of different anomaly detection models

模型	异常持续时间		
	5	10	20
AR($p=1$)	0.962 4	0.960 7	0.944 5
AR($p=2$)	0.961 3	0.954 4	0.948 6
LSTM	0.996 6	0.990 1	0.989 4
Attention-LSTM	0.999 7	0.992 1	0.997 4

在异常事件持续时间为5、10和20个时间段时, Attention-LSTM模型都有最好的效果。对于 Attention-LSTM模型与AR模型及LSTM预测模型间存在的差异,本文从两个角度进行分析。一是对深度学习模型与统计模型的差异分析:AR模型本质上是时间序列数据的线性回归,而深度学习中的LSTM模型能提供非线性函数,增加模型的复杂度,提升学习能力。由此发现无论是LSTM模型还是Attention-LSTM模型,表现都优于AR模型。二是对LSTM预测模型与Attention-LSTM模型的对比分析:交通流量数据存在周期性且受到交通网络中其他网格的影响,而Attention机制能找到与当前交通流量相似程度最高的历史时间点并赋予高权重,从而充分利用交通流量的历史信息,提升模型的准确性。

3.2 SKAB公开数据集实验

SKAB v0.9公开数据集为工业传感器收集到的实际多元时间序列数据。该数据按照事件顺序收集,具有时间序列的特征。传感器之前存在相关性关系,不同的传感器之前的相关性权重不同。由于多元传感器异常数据与交通流量数据存在类似性质,可以看做同一类数据进行分析处理。该数据集数据结构如表3所示。

表3 SKAB数据集参数及意义

Tab.3 Parameters and meanings of SKAB dataset

参数名	意义
Datetime	写入数据库的日期和时间
Accelerometer1RMS	振动加速度
Accelerometer2RMS	振动加速度
Current	电动机的安培数
Pressure	水泵后的回路压力
Temperature	发动机机体的温度
Thermocouple	循环回路中流体的温度
Voltage	电动机上的电压
RateRMS	回路内流体的循环流量
Anomaly	该点是否异常(0或1)

该数据集提供了用于评估模型检测能力的指标算法,其采用的指标为F1分数、漏报率(MAR)与误报率(FAR)。将SKAB数据集应用于本文模型,并与排行榜上的其他模型比较检测效果,其结果如表4所示。

表4 SKAB leaderboard模型比较

Tab.4 Model comparison of SKAB leaderboard

模型	F1	误报率/%	漏报率/%
Perfect detector	1.00	0	0
T-squared+Q (PCA)	0.67	13.95	36.32
Attention-LSTM	0.67	15.73	35.73
MSCRED	0.64	13.56	41.16
T-squared	0.56	12.14	52.56
Autoencoder	0.45	7.56	66.57
Isolation forest	0.40	6.86	72.09
Null detector	0	0	100.00

由表4可知,本文算法以0.67的F1分数与排行榜中的T-squared+Q(PAC)方法并列榜首。

综合来看,实验一在模拟的交通数据上证明了本文提出的Attention-LSTM模型较于传统模型中的AR模型与深度学习中的LSTM模型有更好的异常检测能力。为排除模拟数据的影响,实验二使用了多维传感器现实数据,值得注意的是,SKAB数据集具有与交通流量数据集相似的时空特性。在该数据集上,Attention-LSTM模型表现出很好的检测效果,其在不同的变量维度,不同的时间序列长度甚至不同的应用场景上,有着很好的适应能力。由此可以得出结论,本文提出的基于Attention的交通异常检测模型较传统方法中的自回归模型与深度学习方法中的LSTM预测模型存在优势,本文进一步将模型用于实际的上海市出租车GPS轨迹数据中以检验其对实际交通异常事件的检测能力。

3.3 上海市出租车GPS数据实验

本文使用了上海市出租车的实际GPS轨迹数据进行实验。数据的时间跨度为2014年12月28日至2015年1月10日,共14 d的数据。本文将后7天即2015年1月4日至2015年1月10日的数据作为训练集,将前7天即2014年12月28日至2015年1月3日的数据作为测试集。

为充分减少交通流量值较低的地区对模型效果的影响,本实验只选取了上海外环线以内的地区,具体覆盖范围如图3所示,面积约为900km²(30 km×30 km)。在网格点的划分上,由于原始数据是每辆出租车在某一时刻所在地点的经纬度信息,因此本文并没有按照具体尺寸将网格点分为长宽相同的方

格,而是使用了相同的经纬度度数进行划分,每个格子的经纬度跨度都是 0.005° ,大致相当于东西宽480 m、南北长550 m的网格,由此整个地图区域被划分为3 348(54×62)个同等大小的网格点。在时间间隔上,本文选择每隔10 min统计处于每个网格内的出租车数量,将训练集与测试集都分为1 008($7 \times 24 \times 60/10$)个时间段。最终处理好的数据集为 $1\,008 \times 3\,348$ 的矩阵形式。

本文选取的区域内有河流等出租车无法到达的

地方,同时为了消除部分交通流量较小的网格点的影响,本文设最低流量阈值 $\text{ThreLow}=50$ 。图3为选取地区的交通流量热力图,黑色区域就是被消除的交通流量较小的网格点,颜色越亮表示交通流量越大。对比实际地图来看,与预想相符,消除了水域与偏僻地区的交通流量影响。在消除这些网格的影响后,参与计算的网格数量减少为1 903个,约为总网格数的58%。数据集转化为 $1\,008 \times 1\,903$ 的矩阵形式。

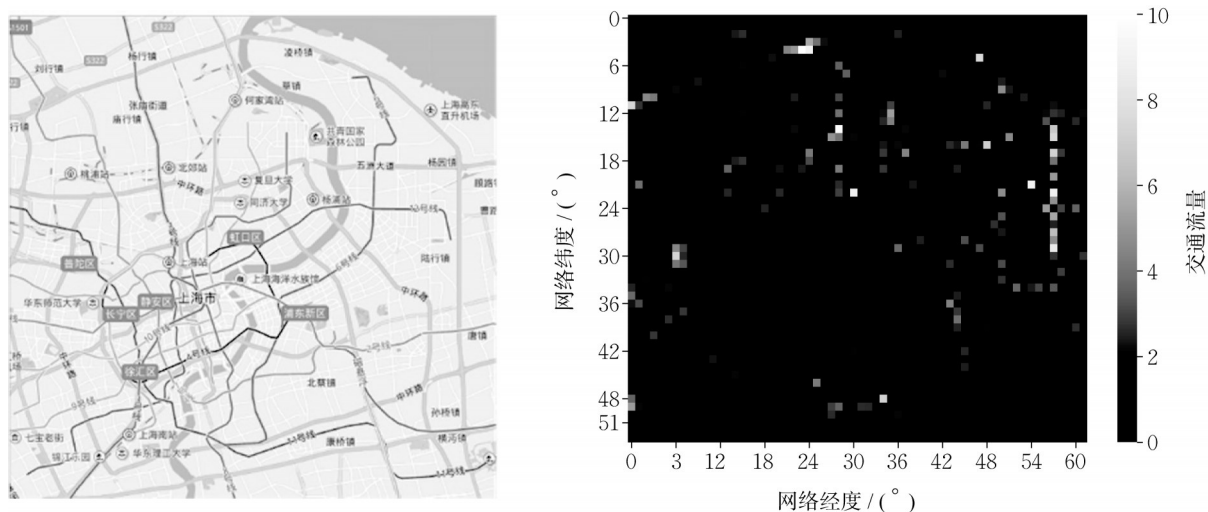


图3 选取的实际地图范围与模型检测输出

Fig.3 Real map and model detection output of selected area

由于本实验中选取的区域面积较大,覆盖较广,即使对网格进行降维处理后仍有1 903个网格点。而在现实情况中,当异常事件发生时,异常事件通常集中在少部分网格点,如果采用对模拟数据的检验方法,将各网格点的异常值大小进行累加,会导致噪音很多,异常事件检测效果很差。因此,本文将每个时间点的异常值最高的前5个网格的平均值作为当

前时间点整个地区的异常值得分。图4中,水平线为平均异常值得分。可以发现,这些高异常值得分主要分布在2个时间范围,一个是在2014年12月28日上午,而另一个则分布在2014年12月31日上午。

对两个时间段的检测结果进一步分析,本文发现2014年12月28日上午的异常事件是由于当天举办硕士研究生招生考试,大量考生乘坐出租车前往

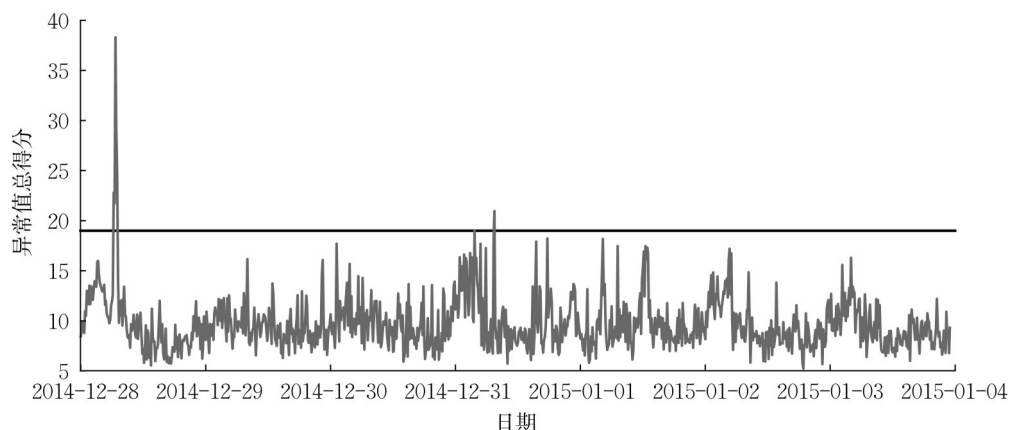


图4 交通异常检测结果

Fig.4 Result of traffic anomaly detection

考点,进而导致在部分考点周边路段出现大规模堵车;而2014年12月31日上午的异常事件是由于当天有许多跨年庆祝活动在上海市区举办,市民由郊区涌向市区,导致上海绕城高速与外环高速往市区方向路段的持续拥堵。

由此,本文证明了该模型能对实际GPS数据进行处理分析,并对现实中发生的交通异常事件进行有效检测,且能快速精确地定位异常事件的发生位置。

4 结语

本文针对交通异常事件的检测问题,将Attention机制引入交通异常事件检测领域中,将Attention-LSTM模型与其他模型进行对比分析,证实了Attention-LSTM模型的优越性。进一步将模型用于实际GPS数据中,对现实中的异常事件有很好的检测效果,证实了本文模型实际应用的有效性。通过Attention机制的注意力权重分配特性,使得算法对交通异常事件更为敏感,在模拟数据集、公开数据集以及上海出租车GPS数据集上,检测准确性均更高,可见,Attention-LSTM模型具有良好的适应能力。与此同时,Attention-LSTM模型尽量精简了结构,降低了模型训练时的成本,使得其在处理2 000维左右的高维度交通数据时,也能够较为快速地完成。本文也形成了一套完整的从模型训练、方法迁移到实际应用的设计方案,比较适合无标签的数据集分析。

作者贡献声明:

郑大庆:方法构思,实验设计和论文撰写。

林陈威:数据整理,代码设计和论文撰写。

王昃杰:数据整理,代码设计和论文撰写。

参考文献:

- [1] 沈洁, 张可云. 中国大城市病典型症状诱发因素的实证分析[J]. 地理科学进展, 2020, 39(1): 1.
SHEN Jie, ZHANG Keyun. An empirical analysis of factors leading to typical urban problems in China [J]. Progress in Geography, 2020, 39(1): 1.
- [2] EAGLE N. Behavioral inference across cultures: using telephones as a cultural lens [J]. IEEE Intelligent Systems, 2008, 23(4): 62.
- [3] GASSON M N, KOSTA E, ROYER D, *et al.* Normality mining: privacy implications of behavioral profiles drawn from GPS enabled mobile phones [J]. IEEE Transactions on Systems, Man, and Cybernetics: Part C (Applications and Reviews), 2011, 41(2): 251.
- [4] GAO Jun, ZHENG Daqing, YANG Su. Sensing the disturbed rhythm of city mobility with chaotic measures: anomaly awareness from traffic flows [J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 11(6): 1.
- [5] 曾阳艳, 苏雅, 张琪慧. 基于EMD-LSTM神经网络的交通流量预测模型[J]. 商学研究, 2021, 28(3): 110.
ZENG Yangyan, SU Ya, ZHANG Qihui. Traffic flow prediction model based on empirical mode decomposition and long short-term memory neural network [J]. Commercial Science Research, 2021, 28(3): 110.
- [6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735.
- [7] ZHENG Zhao, CHEN Weihai, WU Xingming, *et al.* LSTM network: a deep learning approach for short-term traffic forecast [J]. IET Intelligent Transport Systems, 2017, 11(2): 68.
- [8] 刘世泽, 秦艳君, 王晨星, 等. 基于深度残差长短记忆网络交通流量预测算法[J]. 计算机应用, 2021, 41(6): 1566.
LIU Shize, QIN Yanjun, WANG Chenxing, *et al.* Traffic flow prediction algorithm based on deep residual long short-term memory network [J]. Journal of Computer Applications, 2021, 41(6): 1566.
- [9] 彭桐歆, 韩勇, 王程, 等. 基于混合深度学习模型的短时地铁客流量预测[J/OL]. 计算机工程. [2021-08-10]. <https://doi.org/10.19678/j.issn.1000-3428.0061309>.
PENG Tongxin, HAN Yong, WANG Cheng, *et al.* Short-term metro passenger flow prediction based on hybrid deep-learning model [J/OL]. Computer Engineering. [2021-08-10]. <https://doi.org/10.19678/j.issn.1000-3428.0061309>.
- [10] KONG Xiangjie, GAO Haoran, ALFARRAJ O, *et al.* HUAD: hierarchical urban anomaly detection based on spatio-temporal data [J]. IEEE Access, 2020, 8: 26573.
- [11] 宋瑞蓉, 王斌君, 全鑫, 等. 融合多维时空特征的交通流量预测模型[J]. 科学技术与工程, 2021, 21(31): 13439.
SONG Ruirong, WANG Binjun, TONG Xin, *et al.* Traffic flow forecasting model based on multi-dimensional spatial and temporal characteristics [J]. Science Technology and Engineering, 2021, 21(31): 13439.
- [12] 廖挥若, 杨燕. 基于注意力的时空神经网络城市区域交通流量预测[J]. 计算机应用研究, 2021, 38(10): 2935.
LIAO Huiruo, YANG Yan. Predicting citywide traffic flow using attention-based spatial-temporal neural network [J]. Application Research of Computers, 2021, 38(10): 2935.
- [13] 王海起, 王志海, 李留珂, 等. 基于网格划分的城市短时交通流量时空预测模型[J/OL]. 计算机应用. [2022-02-07]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20210930.0905.002.html>.

- WANG Haiqi, WANG Zhihai, LI Liuke, *et al.* Spatial-temporal prediction model of urban short-term traffic flow based on grid division [J/OL]. Journal of Computer Applications . [2022-02-07]. <http://kns.cnki.net/kcms/detail/51.1307.TP.20210930.0905.002.html>.
- [14] GERS F, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM [J]. Neural Computation, 2000, 12(10): 2451.
- [15] YANG Su. On feature selection for traffic congestion prediction [J]. Transportation Research Part C, 2013, 26: 160.
- [16] SHEWALKAR A, NYAVANANDI D, LUDWIG S A. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU [J]. Journal of Artificial Intelligence and Soft Computing Research, 2019, 9 (4): 235.
- [17] MOHAMED B, MOHAMED M, RICHARD D, *et al.* Parallel long short-term memory for multi-stream classification [C]// 2016 IEEE Spoken Language Technology Workshop. San Diego: IEEE, 2016: 218-223
- [18] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [C]// 3rd International Conference on Learning Representations. San Diego: OpenReview.net, 2015: 1-15.
- [19] XU K, BA J, KIROS R, *et al.* Show, attend and tell: neural image caption generation with visual attention [C]// 32nd International Conference on International Conference on Machine Learning. Lille: JMLR.org, 2015: 2048 - 2057.
- [20] GAO Jun, ZHENG Daqing, YANG Su. Perceiving spatiotemporal traffic anomalies from sparse representation-modeled city dynamics [J/OL]. Personal and Ubiquitous Computing [2021-10-10]. <https://doi.org/10.1007/s00779-020-01474-4>.
- [21] GUGULOTHU N, MALHOTRA P, VIG L, *et al.* Sparse neural networks for anomaly detection in high-dimensional time series [C]// AI4IOT Workshop at IJCAI 2018. Stockholm: Morgan Kaufmann, 2018: 1-7.
- [22] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]// 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates, 2017: 6000 - 6010.
- [23] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society, Series B (Methodological), 1996, 58(1): 267.
- [24] ZHOU Wenbin, YANG Su. Outlier detection on large-scale collective behaviors [C]// 4th International Joint Conference on Computational Sciences and Optimization. Kunming: IEEE, 2011: 635-639.