



基于路网结构的交通事故短期风险预测方法

张延孔, 卢家品, 张帅超, 姬晓鹏, 罗月童, 陈为

引用本文:

张延孔, 卢家品, 张帅超, 等. 基于路网结构的交通事故短期风险预测方法[J]. 智能系统学报, 2020, 15(4): 663–671.

ZHANG Yankong, LU Jiapin, ZHANG Shuaichao, et al. A short-term risk prediction method for urban traffic accidents based on road network[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(4): 663–671.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201910002>

您可能感兴趣的其他文章

基于生成式对抗网络的道路交通模糊图像增强

Enhancement of blurred road-traffic images based on generative adversarial network

智能系统学报. 2020, 15(3): 491–498 <https://dx.doi.org/10.11992/tis.201903041>

深度学习的双人交互行为识别与预测算法研究

Human interaction recognition and prediction algorithm based on deep learning

智能系统学报. 2020, 15(3): 484–490 <https://dx.doi.org/10.11992/tis.201812029>

融合迁移学习和神经网络的皮肤病诊断方法

A skin diseases diagnosis method combining transfer learning and neural networks

智能系统学报. 2020, 15(3): 452–459 <https://dx.doi.org/10.11992/tis.201811015>

受成本约束的电力信息物理相依系统鲁棒性优化设计

Robust optimization design of an electrical cyber-physical system based on constrained cost

智能系统学报. 2020, 15(3): 623–632 <https://dx.doi.org/10.11992/tis.201812034>

海底声呐图像智能底质分类技术研究综述

Survey of the intelligent seabed sediment classification technology based on sonar images

智能系统学报. 2020, 15(3): 587–600 <https://dx.doi.org/10.11992/tis.202004026>

在线学习的大规模网络流量分类研究

Large-scale network traffic classification based on online learning

智能系统学报. 2016, 11(3): 318–327 <https://dx.doi.org/10.3969/j.issn.1673-4785.201603033>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201910002

基于路网结构的城市交通事故短期风险预测方法

张延孔¹, 卢家品¹, 张帅超², 姬晓鹏², 罗月童¹, 陈为²

(1. 合肥工业大学 计算机与信息学院, 安徽 合肥 230000; 2. 浙江大学 计算机科学与技术学院, 浙江 杭州 310018)

摘 要: 城市交通事故一般都发生在公共道路上, 然而现有的交通事故风险预测算法都通过对预测区域进行规则网格化来确定预测空间单位, 导致预测精度不高且实用价值较低。本文将道路路段作为预测单位, 采用图卷积和长短期记忆网络, 构建了一种基于路网结构的城市交通事故短期风险预测方法 (traffic accidents risk prediction based on road network, TARPBRN)。该方法能对指定路段短期内的交通事故风险进行预测, 从而可以有针对性地进行治理, 减少交通事故的发生。本文使用杭州市西湖区的交通事故数据对模型进行了训练, 并与 4 种常用的计量经济学模型和 3 种已有的深度学习预测算法进行了对比。实验结果证明本文算法在准确度、正确率和漏报率等方面都优于已有算法。

关键词: 图卷积; 交通事故; 事故模式; 多源数据; 风险预测; 路网结构; 长短期记忆网络; 智慧城市

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2020)04-0663-09

中文引用格式: 张延孔, 卢家品, 张帅超, 等. 基于路网结构的城市交通事故短期风险预测方法 [J]. 智能系统学报, 2020, 15(4): 663-671.

英文引用格式: ZHANG Yankong, LU Jiapin, ZHANG Shuaichao, et al. A short-term risk prediction method for urban traffic accidents based on road network[J]. CAAI transactions on intelligent systems, 2020, 15(4): 663-671.

A short-term risk prediction method for urban traffic accidents based on road network

ZHANG Yankong¹, LU Jiapin¹, ZHANG Shuaichao², JI Xiaopeng², LUO Yuetong¹, CHEN Wei²

(1. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230000, China; 2. College of Computer Science and Technology, Zhejiang University, Hangzhou 310018, China)

Abstract: Urban traffic accidents usually occur on public roads. However, the existing traffic accident risk prediction algorithms determine the prediction space unit by regularizing grid of the prediction area, which leads to low prediction accuracy and low practicability. Taking road sections as the prediction unit, this paper constructs a short-term traffic accident risk prediction method based on road network structure (TARPBRN) by using graph convolution and long short-term memory network. This method can predict the traffic accident risk in a short period of the designated section, so as to carry out targeted governance and reduce the occurrence of traffic accidents. In this paper, traffic accident data from Xihu District, Hangzhou city are used to train the model, and four econometric models and three existing deep learning prediction algorithms are compared. The experimental results show that the proposed algorithm is superior to the existing ones in accuracy, precision and false negative rate (FNR).

Keywords: GCNN; traffic accident; accident mode; multi-source data; risk forecasting; road network structure; LSTM; smart city

收稿日期: 2019-10-08.

基金项目: 国家自然科学基金项目 (61602146); 浙江大学 CAD&CG 国家重点实验室开放课题 (A1814); 中央高校基本科研业务费专项 (75104-036002).

通信作者: 张延孔. E-mail: zhangyankong@hfut.edu.cn.

世界卫生组织在 2015 年发表的《全球道路安全状况报告》^[1]中指出每年约有 125 万人死于交通事故。通过交通大数据, 识别交通事故中的规律并加以治理和预防已成为交通领域重要的研

究课题。

建立有效的交通事故风险预测系统是交通事故预防的关键。如果某一区域的交通事故发生风险是可以预测的,就可以将这一信息传播给附近的司机,提醒他们选择一条不那么危险的道路;交警部门也可以在该区域有针对性地安排巡逻人员,达到减少事故发生的目的。然而交通事故发生风险的准确预测是非常困难的,交通事故的诱发因素多种多样,大致可以分为人、车、路和环境四大类^[2-3],其中人的因素是首要因素。人的行为主观性较强,不可预测,导致虽然许多研究人员已经把注意力集中在交通事故相关的关键因素识别上,交通事故风险的有效预测仍然是一个具有挑战性的问题。

在过去,针对交通事故预测的研究主要聚焦在粗粒度、宏观方面上的研究。研究空间粒度精度聚焦在省、市层次,时间粒度聚焦在年、月份层次^[4-8]。因此研究成果的针对性不高,实用性较差,仅能提供参考性意见。近年来,随着大数据和深度学习技术的发展^[9-14],交通事故的短期实时预测成为了可能。2016年,Chen等^[15]通过Stack去噪自动编码器来提取人的移动特征构建深度学习网络来预测日本境内交通事故风险;2018年,Ren^[16]综合考虑交通事故的周期性和空间分布模式并结合其他重要因素如天气、空气质量等,采用LSTM模型,构建了短期北京市的交通事故风险预测系统。以上方法都对预测区域进行规则化网格划分,针对每一个网格统计其数据特征,预测网格区域短期事故发生风险。2018年,Bao等^[17]在划分的网格的基础上,利用卷积神经网络(convolutional neural networks, CNN)来计算网格的局部区域特征,从而得到了更好的预测效果。

但这些基于网格划分的交通事故短期预测方法,将规则化网格作为预测的空间单位,这些网格内包含大量的非道路区域,如居民区、商场区等。非道路区域通常不会有交通事故的发生,这在一定情况下影响了预测的准确度。交警在进行巡逻、救援等执勤任务时都以路段为单位进行划分,划分网格的预测结果只能在任务分配时提供一定程度的参考;以网格作为交通事故风险预测单位也不能满足司机根据事故发生风险来选择行进路线的需求。因此基于划分网格的短期交通事故预测方法准确度普遍不高且实用价值较低。

因此,本文提出基于路网结构的城市交通事

故短期风险预测算法。该算法以真实路段而不是规则网格作为空间预测单位,使用深度学习技术预测真实路段在短期内的事故风险,完成了逐天和逐小时的交通事故风险预测任务。

路网是一个复杂的不规则网络, CNN等处理规则化数据的深度学习方法无法提取复杂网络的局部特征,因此本文引入图卷积神经网络(graph convolutional neural networks, GCNN)模型来计算路网结构的局部特征。GCNN可以处理不规则的网络数据,在深度学习领域得到了广泛的应用^[18-20]。在预测模型方面,2018年, Lin等^[20]将GCNN和LSTM结合用于纽约市共享单车站点每小时单车需求量的预测,取得良好的效果。

与迭代神经网络(recurrent neural network, RNN)相比,长短期记忆网络(long short-term memory network, LSTM)因其可以进行长时间预测的特点而在时序数据预测中广泛使用。但LSTM在处理超长序列时会导致训练时间大大增长,同时在超长的序列中进行反向传播也可能导致梯度消失,从而削弱模型的可靠性。本文结合交通事故的时序特征,设计了双链LSTM(double link LSTM, DLSTM),处理交通事故风险预测中超长输入的问题,取得了较好的效果。

综上所述,本文的主要贡献有以下几点:

1) 提出基于路网结构、以真实路段作为空间预测单元,逐日、逐小时对城市交通事故短期风险进行预测,提高了预测结果的准确性和实际应用价值。

2) 采用深度学习技术构建了基于路网的短期交通事故风险预测模型。在模型中使用GCNN提取交通事故预测属性的空间特征,使用优化后的DLSTM提取预测属性的时间特征,并将两者结合提取其时空特征。

3) 利用杭州市的多源交通数据对模型进行了训练,并与已有的计量经济学模型和深度学习模型进行了对比,实验证明本文算法在准确度、正确率和漏报率上均优于已有算法。

1 多源交通数据及预处理

1.1 多源交通数据

本文采用浙江省杭州市西湖区总共5类与交通有关的数据进行交通事故风险的短期预测。

交通事故数据 交通事故数据由杭州市交警支队提供,数据时间跨度为2015年7月11日—2015年12月21日。交通事故数据主要包括交通事故发生的时间和位置信息,格式如表1所示。

表1 交通事故数据格式
Table 1 Traffic accident data format

字段名	内容介绍	格式示例
ID	事故编号	2015071100060612
Time	发生时间	2015-07-11 00:06:06
ADDR	发生地点	环城北路杭州大厦
Coord_x	发生经度	120.161984
Coord_y	发生纬度	30.272192
Content	事故描述	两车刮擦,公交车逃逸, 现双方在武林广场处

路网结构数据 本文所使用路网结构数据从OpenStreetMap网站下载,并经过简化处理获得,其数据格式如表2所示。路网结构数据中记录各个道路路段的关键节点的经纬度坐标,及该路段邻接路段的编号。

表2 路网结构数据格式
Table 2 Road network structure data format

字段名	内容介绍	格式示例
ID	路段ID	13
Xlng	路段起点经度	120.161984
Xlat	路段起点纬度	30.272192
Ylng	路段终点经度	120.161984
Ylat	路段终点纬度	30.272192
Way	路段所属道路名	秋石高架
neiborID	与该路段相邻的所有路段id,用“,”隔开	13,16,45
Distance	路段长度	130.134
Type	所属区域类型	商业区
Level	道路等级	一级公路
flow_num	车道数	三车道

道路微波检测数据 道路微波检测数据由杭州市交警支队提供,数据时间跨度同交通事故数据一致,为2015年7月11日—12月21日。微波检测数据由两部分构成:微波检测器分布数据和微波检测器输出数据。微波检测器分布数据如表3所示。

表3 微波检测器分布数据格式
Table 3 Microwave detector distributed data format

字段名	内容介绍	格式示例
ID	微波检测器编号	23
Name	位置描述	庆春路秋涛路西口
Lng	微波检测器位置经度	120.194988
Lat	微波检测器位置纬度	30.252766

微波检测器每5s记录一次道路路况,每条记录包含该路段在5s内通过的车辆数目,以及通过车辆的平均速度,微波检测器检测数据格式如表4所示。

表4 微波检测器检测数据格式
Table 4 Microwave detector detection data format

字段名	内容介绍	格式示例
ID	检测记录编号	283706900
WaveID	微波检测器编号	890
Totalflow	5秒内车辆数量	6
Speed	5秒内车辆平均速度	30
Collectdate	记录时间	2015-07-12 00:00:00

出租车轨迹数据 本文使用的出租车轨迹数据来自于杭州市出租车公司。时间跨度为2015年6月22日—6月30日,为期9天的数据。因为数据时间跨度很小,故本文未将其用作预测交通事故风险的属性信息,只将其作为构建路网图结构的支撑数据。出租车轨迹数据的格式如表5所示。

表5 出租车轨迹数据格式
Table 5 Taxi trajectory data format

字段名	内容介绍	格式示例
ID	出租车记录ID	15
Date	发生时间	2015-06-22 23:59:00
City	所属城市	HZ
carID	出租车牌号	75636
Lng	出租车所在位置经度	120.171984
Lat	出租车所在道路纬度	30.272192
Speed	出租车速度	42
Angel	出租车前轮角度	60
Isfull	出租车是否载客	1

天气数据 本文所使用的数据由爬虫从天气后报网(<http://www.tianqihoubao.com/>)爬取获得,天气数据的时间跨度为2015年7月11日—12月21日,同交通事故数据一致,天气数据的主要格式如表6所示。

表6 天气数据格式
Table 6 Weather data format

字段名	格式介绍	示例
Province	省	浙江
City	城市	杭州
Date	日期	2015-06-22
Tainqi	白天天气、夜晚天气	晴/多云
Temperature	最高气温、最低气温	30℃/25℃
Feng	白天夜晚风力、风向	北风 微风 / 无持续 风向 微风

1.2 数据预处理

本文提出的基于路网结构的城市交通事故短期风险预测算法以真实路段作为空间预测单元,因此使用的所有属性特征应该按照路段进行划分。由2.1节可知,所有数据的位置信息以经纬度来描述,因此首先要将经纬度所代表信息绑定到相应路段。假设预测路段集合为 S ,待绑定的位置为 $P_{(lng,lat)}$,位置与路段绑定的算法如下。

1) 针对于路段集合 S 中第 i 条路段,其长度为 l_i ,位置 $P_{(lng,lat)}$ 与其起点和终点的距离分别为 $d_{i-start}$ 、 d_{i-end} ,距离采用欧几里德距离公式计算,则可得路段差值 $d_i = d_{i-start} + d_{i-end} - l_i$ 。

2) 寻找路段 m 使得 d_m 为所有路段差值的最小值。

3) 定义阈值 ε ,若 d_m 小于阈值 ε ,则该位置与路段 m 绑定;若 d_m 不小于阈值 ε ,则该位置与预测集合中的任何路段都不绑定。

根据以上位置点与路段的绑定算法,可以将交通事故、微波检测器及出租车位置信息绑定到相应路段。从而可得路段在预测交通事故时的所有属性特征。本文将路段属性数据分为3类:空间相关数据、时间相关数据、时空相关数据。

空间相关数据 仅与路段相关的特征,对于任意时间段数据值都一样。该类数据有路段长度、道路等级、车道数、道路所属区域等。

时间相关数据 仅与时间段相关联的数据,对于所有的路段数据值都一样。该类特征有天气、风力、风向、气温等。

时空相关数据 是与时间段、路段都有关联的数据。同一路段在不同时间、不同路段在相同时间都具有不同的数据值。该类数据有单位时间道路通过车辆数、单位时间道路车辆通过平均速度、单位时间内事故发生数量等。

观察数据可知,以上数据数值范围都不统一,因此并不能直接作为神经网络的输入。因此针对于所有的枚举类特征,例如天气、风向、道路所属区域等,本文采用 one-hot 编码。对于所有的数值型特征,本文统计最大值、最小值,然后根据最大值最小值将其映射到区间 $[0,1]$ 。于是所有的特征值都处在 $[0,1]$ 区间内,使得深度学习框架能获得良好的效果。

在此说明,本文仅对输入特征数据进行量纲统一处理,对于预测结果,本文使用单位时间内事故发生的次数并没有映射到区间 $[0,1]$,这也与已有算法的交通事故风险描述一致。

2 交通事故预测深度学习模型

本节首先详细地描述了基于路网的交通事故风险预测模型(traffic accidents risk prediction based on road network, TARPBRN)的总体架构,然后分别对每个主要模块进行了具体介绍。

2.1 TARPBRN 模型架构

由1.2节可知,本文将交通事故的诱发因素从数据特征出发主要分为三大类:时间相关特征、空间相关特征和时空相关特征,针对每一类特征,分别采用不同的网络结构进行特征提取。TARPBRN模型的总体架构如图1所示。首先,对3类属性数据进行特征提取操作。对于空间相关数据,本文使用多层GCNN来提取空间特征。对于时间相关数据采用DLSTM提取时间特征。对于时空相关数据,本文将GCNN和DLSTM相结合来提取时空特征。然后,将3类属性数据通过相应特征提取层获得的特征传入特征融合层,将3类属性特征融合为一个整体。最后,将融合后的特征传入全连接网络,通过全连接网络输出最终的预测结果。

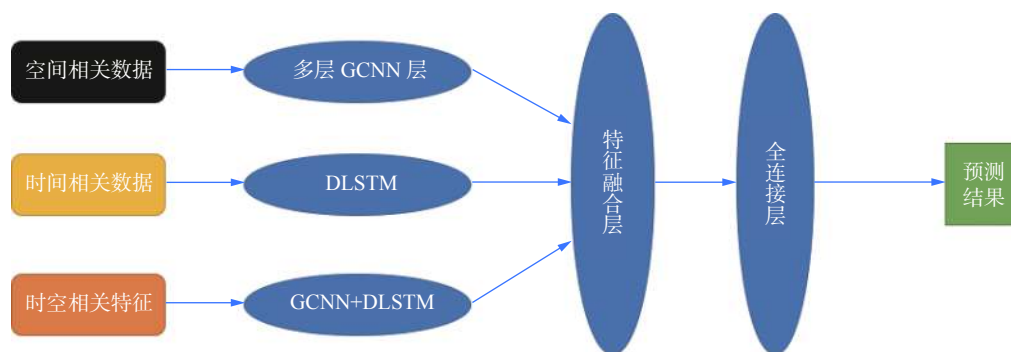


图1 交通事故风险预测框架

Fig. 1 Traffic accident risk prediction framework

2.2 多层GCNN层提取空间属性特征

GCNN是一种能对图数据进行深度学习的方

法,在对图进行卷积神经网络学习时,可以同时考虑到节点的特征信息以及结构信息。这使得

GCNN 广泛用于非规则数据的深度学习。图卷积网络的主要计算方式如下:

给定图 $G=(V, \mathbf{x}, \varepsilon, \mathbf{A})$, V 是 N 个顶点的集合, $\mathbf{x} \in R^N$ 是每一个顶点特征向量的集合。 ε 是边的集合, \mathbf{A} 是邻接矩阵, 其中 A_{ij} 表示两个顶点间的关联度。假设 GCNN 模型从输入层到输出层共有 m 层。图网络中的每一个顶点在 l 层中都有一个长度为 C^l 的特征向量。对于每一层 $l(l=1, 2, \dots, m-1)$, 图卷积模型 GCNN 以如下的规则从输入层到输出层进行传播:

$$\mathbf{H}^l = \sigma(\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{H}^{l-1} \mathbf{W}^l)$$

其中 \mathbf{D} 是一个对角矩阵, $D_{ii} = \sum A_{ij}$, $\mathbf{W}^{l-1} \in R^{C^{l-1} \times C^l}$ 是每一层可训练的权重矩阵, $\sigma(\cdot)$ 是一个激活函数, $\mathbf{H}^l \in R^{N \times C^l}$ 是第 l 层的激活输出矩阵。 $\mathbf{H}^{l-1} \in R^{N \times C^{l-1}}$ 则是第 $l-1$ 层的激活输出矩阵。对于第 1 层输入层:

$$\mathbf{H}^1 = \sigma(\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{H}^0 \mathbf{W}^1)$$

\mathbf{H}^0 即是图网络结构中各顶点的输入特征 \mathbf{x} 。

综上所述, 图卷积进行特征提取是需要输入两部分数据, 图的邻接矩阵 \mathbf{A} 及特征属性数据 \mathbf{x} 。特征属性数据 \mathbf{x} 由第 2 节经过数据预处理已经获得, 图的邻接矩阵可由路网结构转化得到。

本文以路段作为空间预测单元, 路段在路网结构中通常以边的形式存在, 而 GCNN 模型主要计算图结构的顶点特征。因此本文首先要将路网中的路段转化为图结构中的顶点, 将路网中的交叉点转化为边。转化方式如图 2 所示。

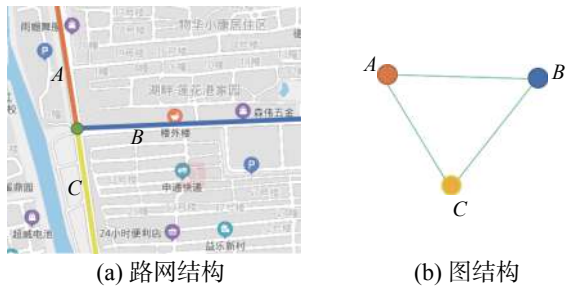


图 2 路网转化为图结构示意

Fig. 2 Road network transformation into graph structure sketch

图 2(a) 中有 3 条路段 A、B、C 交汇于一点, 当该路网结构转化为图结构时, A、B、C 路段转化为 3 个顶点, 如图 2(b) 所示。路段 A 可通过交汇点到达 B, 因此图结构中, 顶点 A 与顶点 B 之间有边相连, 同理, 顶点 A、B, 顶点 B、C 之间都有边相连。

将路网结构转化为图结构后就可以得出路网结构的邻接矩阵, 但此时邻接矩阵的只有 0、1 两个数值。0 表示不连接, 1 表示连接。为了更明确地表示相邻路段之间的关联度, 本文根据出租车轨迹数据, 计算出出租车经过相邻路段的次数, 用

该数据来描述相邻路段的关联度, 在相邻路段间经过的次数越多, 相邻路段之间的关联度越高, 转换公式为

$$f(x) = \lg x / \lg \max$$

式中: x 为车辆在相邻路段经过的次数; \max 为车辆在相邻路段经过次数的最大值; $f(x)$ 即为转化后的关联度。

2.3 DLSTM 层提取时间特征

长短期记忆 (long short-term memory, LSTM) 是一种特殊的 RNN, 相比普通的 RNN, LSTM 能够在长序列数据中有更好的表现, 它在算法中加入了一个判断信息有用与否的结构单元。如图 3 所示, 结构单元中被放置了 3 扇门, 即输入门、遗忘门和输出门。当一个信息进入 LSTM 的网络当中, 可以根据规则来判断是否有用, 实现信息的选择性通过。LSTM 更细节的信息可以参考文献 [23]。

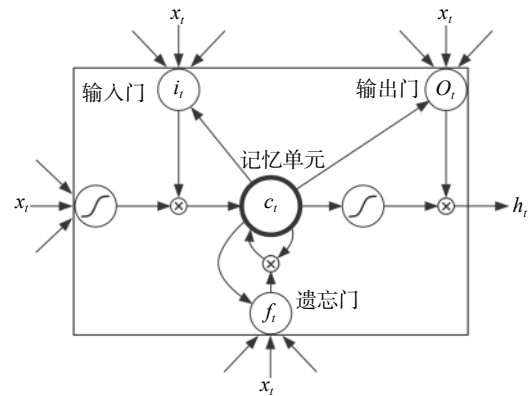


图 3 LSTM 结构单元

Fig. 3 LSTM structural unit

虽然通过 LSTM 可以获取长期记忆的特征, 但 LSTM 在处理超长序列时会导致训练时间大大增长, 同时在超长的序列中进行反向传播也可能导致梯度消失, 从而削弱模型的可靠性。部分情况下, 还可能因为冗余信息太多, 无法取得良好效果。

因此, 基于交通事故在时序上呈现的周期性和渐近性的特征, 本文提出了双链 LSTM(DLSTM) 网络。如图 4 所示。假设预测 t 时刻的时序特征 (假设序列的周期性为 5), 则提取 $t-1$ 、 $t-2$ 、 $t-3$ 、 $t-4$ 时刻的特征 (图中蓝色区域特征) 作为渐近特征。提取 $t-5$ 、 $t-2 \times 5$ 、 $t-3 \times 5$ 、 $t-4 \times 5$ 时刻的特征 (图中绿色区域的特征作为周期特征)。然后将渐近特征和周期特征分别传入 LSTM 网络, 最后将两个 LSTM 链中的特征进行融合, 作为 DLSTM 网络的输出特征。注意这里因为两种特征同属于多层 LSTM 层输出的时序特征, 相互之间影响较大, 故以相乘的方式进行特征融合。

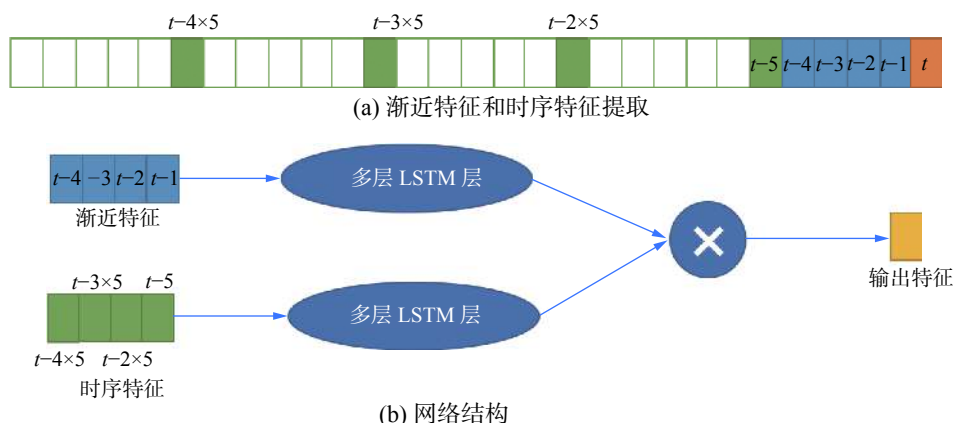


图 4 DLSTM 结构示意图

Fig. 4 Structural schematic diagram of DLSTM

2.4 GCNN 与 DLSTM 相结合提取时空特征

在城市交通事故风险短期预测模型中, 交通事故风险、道路车流量、道路行车平均速度不仅与时间相关, 还与所属路段及空间有关。因此, 本文将 DLSTM 与 GCNN 模型进行整合, 构建了新的时空特征提取层。在整合过程中, 本文提出两种方案:

1) GCNN+DLSTM 模型, 先对时空特征在每个时间段上进行图卷积求出局部特征, 然后针对每个时间段上的图卷积特征, 构建时间序列, 传入到 DLSTM 求取其时间特征。

2) DLSTM+GCNN 模型。首先根据交通事故时空特征所属路段的不同进行分类, 然后将每一路段的时空序列特征传入 DLSTM 模型, 计算其时序特征, 最后将所有路段的时序特征传入 GCNN 模型, 求取其空间特征。

本文对两种方案都进行了比较。为了介绍方便, 本文在除模型对比外, 将这两种方案作为一种, 在下文中都采用 GconvLSTM 模型表示。

2.5 特征融合与全连接层

在特征融合层, 将 GCNN 提取的空间特征、DLSTM 神经网络提取的时间特征、GconvLSTM 神经网络提取的时空特征串联成一个密集的向量, 构建输入全连接网络的高维特征。然后通过全连接层将矢量转换为模型输出。因此, 模型预测输出可转化为公式:

$\hat{y}_t = W_{\text{gcn}} X_t^{\text{gcn}} + W_{\text{DLstm}} X_t^{\text{lstn}} + W_{\text{GconvLstm}} X_t^{\text{GconvLstm}} + b_t$ (1)
其中 X_t^{gcn} 、 X_t^{lstn} 、 $X_t^{\text{GconvLstm}}$ 分别表示在 t 时刻的空间特征、时间特征和时空特征; W_{gcn} 、 W_{DLstm} 、 $W_{\text{GconvLstm}}$ 、 b_t 表示各个属性相关的权重和偏差。

2.6 损失函数

在交通事故风险预测中, 绝大部分时间、绝大部分路段的交通事故发生风险值都为 0, 事故风险为 0 与事故风险非 0 的样本数量差距相当

大, 非常不平衡。为了缓解样本间的不平衡问题, 本文在目标函数均方差误差的基础上进行了改进, 加强事故风险不是 0 的样本的损失函数值。为防止在训练过程中出现过拟合现象, 本文引用了 L2 规则化范数。因此最终的损失函数为

$$\min_{w,b} (\|y_t - \hat{y}_t\|_2^2) \times (y_t + 1) + \alpha \|W\|_2^2$$

式中: W 代表式 (1) 中所有的权重参数, α 为正则化参数。

2.7 评价函数

现有的基于深度学习的城市短期交通事故预测模型均采用划分网格的方式来进行预测, 在进行方法对比时均采用 MAE、MRE、RMSE 这 3 种度量方式来对预测模型进行度量, 本文同样如此。

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ \text{MRE} &= \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \end{aligned}$$

式中: y_i 为路段 i 的交通事故真是发生风险, \hat{y}_i 为路段 i 交通事故发生风险的预测值; n 为所预测道路的总条数。

3 分析结果

3.1 交通事故数据的周期特征

交通事故是人类活动产生的结果, 因此必将在时间上呈现规律性变化。本文统计了交通事故 1 天和 1 周的分布情况, 分别如图 5、6 所示。

由图 5 可知, 交通事故多发生在日间, 在夜间事故发生次数较少。且事故多发生在早上 7~10 点及晚上 4~6 点。由图 6 可知, 交通事故在周末发生较多, 在工作日发生事故较少。综上可知, 交通事故分别以天和星期为周期呈现出一定的周期

性。因此,在逐小时风险预测模型中,可以设置 DLSTM 的步长为 24;在逐日风险预测模型中,可以设置 DLSTM 的步长为 7。

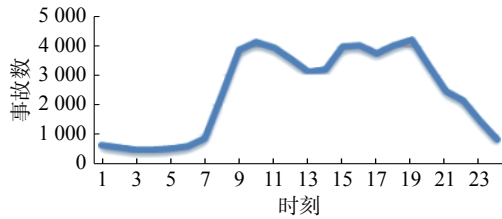


图5 交通事故24小时分布图

Fig. 5 Hourly distribution map of traffic accidents

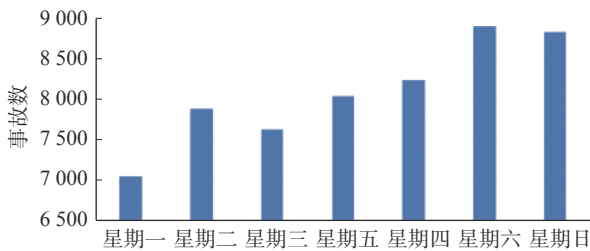


图6 交通事故1周分布图

Fig. 6 Sequence chart of traffic accident distribution with week

3.2 结果分析

本研究建立了两类短期碰撞风险预测模型:逐日风险预测模型和逐小时风险预测模型。在每种预测模型中又根据时空特征提取的顺序分为 GCNN+DLSTM 模型和 DLSTM+GCNN 模型。预测数据集按 7:3 的比例分为训练数据集和测试数据集。在模式测试过程中,本文假设所有空间特征为 S,所有时间特征为 T,时空特征为 ST,针对不同模型输入不同特征,则模型预测结果如表 7、8 所示。

分析表 7、8 可知,当 ST 特征不参与计算时, $TARPBRN_{GCNN+DLSTM}$ 与 $TARPBRN_{DLSTM+GCNN}$ 为同一种模型,因此结果一致。在逐小时事故风险预测结果 $TARPBRN_{GCNN+DLSTM}$ 表现较好,而在逐天事故风险预测结果 $TARPBRN_{DLSTM+GCNN}$ 有更好的表现,因此在进行事故预测时,可以根据预测时间的粒度,来选择合适的算法。在 3 类特征对比中可知,时空属性相关特征对预测结果影响最大,其次为时间相关特征,再次为空间相关特征。

表 7 逐小时事故风险预测结果

Table 7 Hourly accident risk prediction result

特征	$TARPBRN_{GCNN+DLSTM}$			$TARPBRN_{DLSTM+GCNN}$		
	MSE	MRE	RMSE	MSE	MRE	RMSE
S+T+ST	0.28	0.25	0.37	0.30	0.27	0.39
T+ST	0.34	0.27	0.41	0.32	0.29	0.43
S+ST	0.35	0.31	0.45	0.36	0.33	0.47
S+T	0.47	0.38	0.52	0.47	0.38	0.52

表 8 逐天事故风险预测结果

Table 8 Daily accident risk prediction result

Feature	$TARPBRN_{GCNN+DLSTM}$			$TARPBRN_{DLSTM+GCNN}$		
	MSE	MRE	RMSE	MSE	MRE	RMSE
S+T+ST	0.58	0.45	0.67	0.57	0.44	0.65
T+ST	0.64	0.48	0.70	0.62	0.33	0.47
S+ST	0.65	0.49	0.72	0.64	0.29	0.43
S+T	0.70	0.53	0.79	0.70	0.53	0.79

3.3 模型对比

与已有的基于划分网格的城市短期交通事故风险预测系统不同,本文所提出的城市短期交通事故风险预测系统以真实的路段作为预测对象。因此本文使用的数据结构并不能直接用于已有算法,且因为预测对象不同,并不能使用 MAE、MRE、RMSE 3 种度量方式在不同模型间进行对比。为了在不同的预测模型间比较,经过与交警部门交流,本文提出使用准确率 (acc, accuracy),精准率 (p, precision) 和漏报率 (FNR) 3 项来评估模型的预测性能。3 种度量方式的计算方式如下:

$$acc = \frac{TN + TP}{TN + FP + FN + TP}$$

$$p = \frac{FP}{TN + FP}$$

$$FNR = \frac{FN}{FN + TN}$$

式中的变量由表 9 确定。

表 9 变量定义

Table 9 Variable definition table

按类型划分	预测高风险	预测低风险
实际发生事故	TP	FN
实际未发生事故	FP	TN

本文以几种已有模型为基准,与所提出的模型进行了比较。所选的基准测试方法可以分为两种类型:计量经济模型和深度学习模型。所选择的计量经济模型包括自回归综合移动平均模型 (autoregressive integrated moving average mode, AR-IMA)、随机参数 (random-parameter) 模型、随机效应 (random-effects) 模型和地理加权回归模型 (geographical weighted regression, GWR) 模型,所选择的机器学习模型包括堆叠降噪自动编码器 (stacked denoising auto encoder, SDAE) 模型、基于 LSTM 的交通事故风险预测模型 (traffic accident risk prediction method based on LSTM, TARPML) 和基于时空卷积的长短期记忆网络 (spatiotemporal convolutional long short-term memory network, STCL-Net)。

ARIMA 是一种应用广泛的时间序列分析方

法,它可以同时集成移动平均模型、自回归部分和移动平均部分^[14]。随机效应模型假设预测结果与独立的变量不相关,可以通过观察很好地解释观测之间的空间相关性、时间相关性或者时空相关性^[15]。近年来,随机参数模型和GWR模型是解决空间异质性问题的两种潜在的碰撞建模技术,它们允许碰撞计数与预测因子之间存在不同的关系^[16]。SDAE模型是一种基于降噪编解码器的深度学习模型,在日本城市交通预测中取得了较好的效果。TARPMML是一种结合多源数据,基于LSTM与全连接网络的深度学习模型,曾用于北京市的城市交通风险预测。STCL-Net在LSTM模型的基础上引入了CNN模型,充分考虑数据的空间特征对城市交通事故风险进行了预测。3种深度学习方法都基于划分网格的方法对交通事故风险进行预测。

为了进行公平的比较,所有的基准方法都采用同一批数据,并在同样的训练周期数下进行调优。但因为本文的算法为基于路段的算法,数据结构不能直接应用于基准测试方法,因此对于不同的方法数据都进行了相应的修改。例如:在划分网格的方法中,本文采用500 m×500 m网格大小对预测区域进行网格划分,并对所有网格进行预测。模型对比的预测效果如表10所示。总体来看,本文提出的算法在准确率、正确率及漏报率方面都优于其他基准模型。逐天预测模型的整体准确率要高于逐小时预测,正确率逐渐增大,但漏报率也呈变大的趋势。

表 10 模型结果对比表
Table 10 Model result contrast table %

模型	逐天预测结果			逐小时预测结果		
	准确率	正确率	漏报率	准确率	正确率	漏报率
TARPBNN _{GNN+DLSTM}	95.28	25.30	40.25	85.58	20.33	33.27
TARPBNN _{DLSTM+Genn}	95.36	25.34	40.35	85.46	20.25	32.75
STCL-Net	94.05	20.45	43.78	83.96	17.45	30.45
TARPMML	93.98	18.47	44.92	83.12	16.37	36.51
SDAE	93.56	17.56	39.21	82.95	16.78	35.79
ARIMA	92.09	15.03	49.82	82.11	13.59	39.81
Random-parameter	92.02	15.12	47.63	82.34	14.65	38.62
Random-effects	92.19	15.29	45.27	82.02	15.26	40.37
GWR	92.10	14.78	49.84	81.56	14.73	41.58

4 结束语

本文以路段作为空间预测单元,预测某路段

短期内的交通事故发生风险,从而给司机及交警部门在道路选择和事故治理方面提供决策依据。与已有的计量经济学模型和深度学习算法相比,本文算法在准确率、正确率和漏报率方面都表现出极大的优势。

虽然提出的基于路网的城市交通事故短期风险预测模型显示出了巨大的潜力,但在本研究中仍有几个局限性需要解决。交通事故成因复杂,涉及方方面面的因素,仅凭已有的数据能起到的预测效果有限,未来应该收集更多数据(例如车辆数据、违章数据、行人信息、司机信息等),来进行事故预测,以求取得更精确的预测效果。与传统的计量经济学模型相比,本文提出的TARPBNN模型工作起来像一个黑匣子,缺乏因子分析,这也是深度学习方法普遍存在的问题。交通事故属于小概率事件,这导致了交通事故预测中存在着严重的样本偏差问题,虽然本文通过重新构建损失函数来处理这一问题,但仍需投入大量的工作来解决这一问题。

参考文献:

- [1] World Health Organization. Global status report on road safety 2015[R]. World Health Organization, 2015.
- [2] ZHANG Guangnan, KELVIN KW Yau, CHEN Guanghan. Risk factors associated with traffic violations and accident severity in China[J]. Accident analysis & prevention, 2013, 59: 18–25.
- [3] 敖曼, 翟润平. 气象条件对道路交通的影响分析[J]. 公路与汽运, 2011(2): 58–62.
AO Man, ZHAI Runping. Analysis of the influence of meteorological conditions on road traffic[J]. Highways & automotive applications, 2011(2): 58–62.
- [4] NOLAND R B. Traffic Fatalities and Injuries: the effect of changes in infrastructure and other trends[J]. Accident analysis & prevention, 2003, 35: 599–611.
- [5] LI Z, WANG W, LIU P, et al. Using geographically weighted roisson regression for county-level crash modeling in california[J]. Safety science, 2013, 58: 89–97.
- [6] 秦利燕, 邵春福, 赵亮. 道路交通事故宏观预测模型[J]. 武汉理工大学学报: 交通科学与工程版, 2010, 34(1): 158–161.
QIN Liyan, SHAO Chunfu, ZHAO Liang. Macro prediction model of road traffic accident based on neural network[J]. Journal of Wuhan University of Technology (transportation science & engineering edition), 2010, 34(1): 158–161.
- [7] 张辉, 陈柯羽. 高速公路减速波事故的前兆特征与概率模型[J]. 交通运输研究, 2017, 3(5): 26–32.
ZHANG Hui, CHEN Keyu. Precursory characteristics and

- probability model of deceleration wave accidents on expressways[J]. *Transportation research*, 2017, 3(5): 26–32.
- [8] 王健, 卢锡凤. 基于聚类分析的高速公路事故黑点鉴别及成因分析研究[J]. *公路交通技术*, 2016, 32(5): 114–119.
- WANG Jian, LU Xifeng. Identification and cause analysis of expressway accident black spots based on cluster analysis[J]. *Highway traffic technology*, 2016, 32(5): 114–119.
- [9] BAO J, LIU P, YU H, et al. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas[J]. *Accident analysis & prevention*, 2017, 106: 358–369.
- [10] SINNOTT R O, YIN S. Accident black spot identification and verification through social media[C]// *IEEE International Conference on Data Science & Data Intensive Systems*. Sydney, Australia, 2015.
- [11] 李娟, 邵春福. 基于 BP 神经网络的交通事故预测模型[J]. *交通与计算机*, 2006, 24(2): 34–37.
- LI Juan, SHAO Chunfu. Traffic accident prediction model based on BP neural network[J]. *Traffic and computer*, 2006, 24(2): 34–37.
- [12] 张志豪, 杨文忠, 袁婷婷, 等. 基于 LSTM 神经网络模型的交通事故预测[J]. *计算机工程与应用*, 2019, 55(14): 249–253.
- ZHANG Zhihao, YANG Wenzhong, YUAN Tingting, et al. Traffic accident prediction based on LSTM neural network model[J]. *Computer engineering and applications*, 2019, 55(14): 249–253.
- [13] 胡立伟, 张婷, 郭凤香. 基于灰色 BP 神经网络的道路交通事故车型分担率预测及其预防策略研究[J]. *武汉理工大学学报(交通科学与工程版)*, 2018, 42(3): 388–392.
- HU Liwei, ZHANG Ti, GUO Fengxiang. Traffic accident split rate of vehicle types prediction and prevention Strategies study based on gray BP neural network[J]. *Journal of Wuhan university of technology (transportation science & engineering edition)*, 2018, 42(3): 388–392.
- [14] MA X, TAO Z, WANG Y, et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. *Transportation research part C: emerging technologies*, 2015, 54: 187–197.
- [15] CHEN Qunjun. Learning deep representation from big and heterogeneous data for traffic accident inference[C]// *Thirtieth AAAI Conference on Artificial Intelligence*. Phoenix, Arizona, United States, 2016: 338–344.
- [16] REN Honglei. A deep learning approach to the citywide traffic accident risk prediction[C]// *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, United States, 2018: 3346–3351.
- [17] BAO J, LIU P, UKKUSURI S V. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data[J]. *Accident analysis & prevention*, 2019, 122: 239–254.
- [18] DEFFERRARD Michaël, XAVIER Bresson, PIERRE rgheynst. Convolutional neural networks on graphs with fast localized spectral filtering[J]. *Advances in neural information processing systems*, 2016: 3844–3852.
- [19] KIPF T N, MAX W. Semi-supervised classification with graph convolutional network[J]. *arXiv preprint arXiv: 1609.02907*, 2016.
- [20] LIN Lei, HE Zhengbing, SRINIVAS Peeta. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach[J]. *Transportation research part C: emerging technologies*, 2018, 97: 258–276.
- [21] GEORGE E P, DAVID A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. *Journal of the american statistical association*, 1970, 65: 1509–1526.
- [22] NAZNIN F. Application of a random effects negative binomial model to examine tram-involved crash frequency on route sections in Melbourne, Australia[J]. *Accident analysis & prevention*, 2016, 92: 15–21.
- [23] BAO Jie. Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas[J]. *Accident analysis & prevention*, 2017, 106: 358–369.

作者简介:



张延孔, 讲师, 博士研究生, 主要研究方向为数据可视化、数据可解释分析。



卢家品, 硕士研究生, 主要研究方向为数据可视化。



张帅超, 博士研究生, 主要研究方向为交通利益预测与评估、大数据分析。