

# 基于支持向量机的高速公路实时事故风险研判

游锦明, 王俊骅, 唐 棠, 方守恩

(同济大学 道路与交通工程教育部重点实验室, 上海 201804)

**摘要:** 采用 G60 高速公路(上海段)上布设的单组线圈检测器检测的车道级交通流数据对该路段上发生追尾事故可能性进行研究. 通过配对案例对照的方法, 分别对事故前 5~10 min, 10~15 min 和 15~20 min 的交通流数据建立了追尾事故实时预测支持向量机模型. 结论表明基于事故前 5~10 min 的交通流数据构建的支持向量机分类器能够有效的对事故进行实时预测, 总体事故预测精度为 84.85%, 误报率为 0.33%, 该支持向量机分类器具有较高的实用价值, 同时也表明了基于单流量检测器的交通流数据对事故进行实时预测的可靠性.

**关键词:** 高速公路; 追尾事故; 实时研判; 事故风险; 支持向量机

**中图分类号:** U491

**文献标志码:** A

## Support Vector Machines Approach for Predicting Real-time Rear-end Crash Risk on Freeways

YOU Jinming, WANG Junhua, TANG Tang, FANG Shouen

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

**Abstract:** The paper aims to study the relationship between rear-end crash potential and lane-level traffic data collected by single pair of loop detectors located on the G60 Freeway in Shanghai, China. The matched case-control method with support vector machines was applied to modelling the traffic data of different time slices before the crashes respectively, 5~10 minutes, 10~15 minutes and 15~20 minutes before the crash. Results indicate that support vector machines classifiers based on the traffic data of 5~10 minutes before the crashes have the highest crash prediction accuracy of 84.85% and a false alarming rate 0.33%. The model proves to be valid to predict the real-time crash risk, which is helpful in freeway traffic management.

**Key words:** freeway; rear-end crash; real-time prediction;

crash risk; support vector machines

高速公路交通安全由于其发生事故的随机性和严重性, 一直是交通安全领域研究的一个热门方向. 在高速公路事故中, 追尾事故发生的频率最高, 而且往往导致大面积的交通拥堵和严重的人员财产损失, 因此一直以来追尾事故都是高速公路安全研究的一个热点话题. 追尾事故因素主要由驾驶人、车辆特性、动态交通流、道路环境等因素构成.

传统的高速公路追尾事故安全研究集中在道路安全评价及微观行车风险研究. 基于道路安全评价的研究主要从道路安全审计的角度采用历史事故资料及高速公路运营数据研究速度与事故率之间的关系, 继而提出针对道路线形和安全设施的改善建议和交通管控方法. 该方法只能进行事后改善, 不能对高速公路交通运行风险实时做出评价, 也不能从根本上解决我国高速公路现阶段管理中存在的滞后性. 微观行车风险研究主要通过仿真软件从车辆运动状态入手分析高速公路跟驰状态下追尾事故发生的机理. 在实际管理工作中, 由于受周边车辆信息采集设备发展所限, 该部分研究成果并未能成功应用于高速公路安全实时管控.

为实现高速公路追尾事故风险实时研判, 本文使用我国现阶段高速公路信息化管理下的单流量检测器所采集的车道级交通流数据, 对事故前兆进行研究, 从而为后期国内在高速公路方面进行实时交通安全风险分析和主动安全管理提供借鉴.

## 1 研究综述

由于高速公路信息化较早, 国外的专家学者围绕事故可能性实时预测开展了大量的研究. Madanat

和 Liu<sup>[1]</sup>提出实时事故可能性预测的观点并利用高速公路流量、天气和事故数据进行建模. Lee 等人<sup>[2]</sup>提出事故先兆概念,发现速度离散性和交通流密度对事故发生有着较大影响. Abdel-Aty 等人<sup>[3]</sup>采用配对案例对照逻辑回归方法结合事故上游连续 5 个检测器和下游 1 个检测器的数据进行建模. 发现事故前 5~10 min 上游检测器的平均占有率和下游的速度协方差对事故影响最大,如果采用对数比率 1.0 模型有着 69% 的事故预测精度. 这些研究都说明了基于事故先兆的方法来进行事故可能性预测具有很高的可信度.

国内外学者在该领域基于不同的事故类型和参数构建进行了深入的探索<sup>[4-9]</sup>,不同的模型有着不同的事故预测精度,其中 Pande<sup>[6]</sup>等人构建的模型能够识别 75% 的追尾事故. Mohamed Ahmed 等人<sup>[10-11]</sup>尝试使用最新的自动车辆识别系统(AVI)的数据进行建模,结果表明模型有着 75.93% 的追尾事故预测精度,如结合远程交通微波传感器(RTMS)所采集的数据构建数据融合模型,发现能够达到更高的精度,模型有着更高的鲁棒性和可靠性. 东南大学徐铨铨<sup>[12]</sup>提取美国加州 I-880 N 高速公路上采集的间隔

为 30 s 的实时交通流数据和事故数据,通过 Fisher 判别分析方法建立了判别危险交通流状态的交通流参数线性组合指标及 Logistic 模型,模型有着 65.7% 的事故预测精度,在后续研究<sup>[13]</sup>中发现增加天气参数以后能够提高事故风险模型的实时预测精度至 71.7%. 东南大学李志斌<sup>[14]</sup>根据高速公路常发瓶颈上游路段运动波传播过程中车辆行驶轨迹特征进行集计分析后构建了高速公路常发拥堵路段的追尾事故风险实时预测模型,并采用美国加州 I-880 洲际高速公路上的检测器数据进行了验证. 上述研究成果均表明了采用连续密集的检测器数据对事故风险进行预测的有效性,而且通过引入多元数据及高分辨率检测数据能够大幅度提高实时事故预测精度.

我国高速公路信息化发展尚处于初步阶段,众多高速公路上流量检测器等信息化设备布设密度较低,与多数研究中所采用的检测器布设密度(0.5 mile)相差甚远,而且这些设备主要部署在易发拥堵区等核心路段,因此基于国内高速公路运营数据的追尾事故研究较少. 有鉴于此,有必要在我国现阶段信息化条件下,对现有信息化数据挖掘,初步实现高速公路交通安全实时研判. 建模技术路线如图 1 所示.

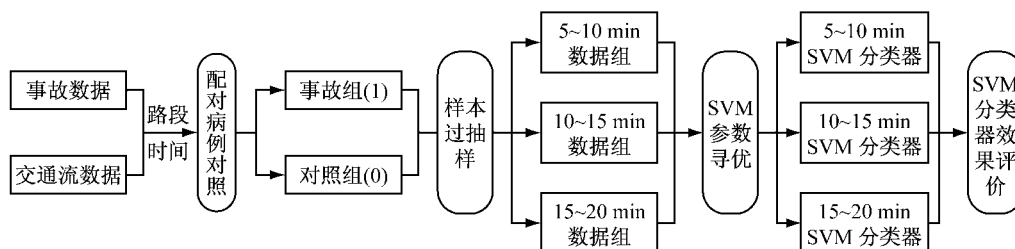


图 1 建模技术路线

Fig.1 Modeling process

## 2 数据准备

### 2.1 交通流数据

采用的交通流数据来源于 2014-01—2015-09 期间 G60 沪昆高速上海段. G60 上海段全长 48.7 km, 单向 3—5 车道, 设计车速  $120 \text{ km} \cdot \text{h}^{-1}$ . 该段高速上共布置了 9 组分车道的线圈流量检测器. 其中 4 组检测器布设在高速公路出入口匝道, 5 组检测器布设在高速公路主线上. 线圈每 20 s 对检测到的过车数据进行统计存储, 主要记录的数据字段有线圈编号、采集时间、数据有效性、车流量、平均速度、平均占有率、分车型的流量和速度、线圈状态等信息, 本文着重研究高速公路主线段的交通运行风险和检测

器数据之间的潜在关系,故仅选用 5 组主线线圈检测器采集的数据进行研究,对应的位置和所处道路环境如图 2 及表 1 所示.

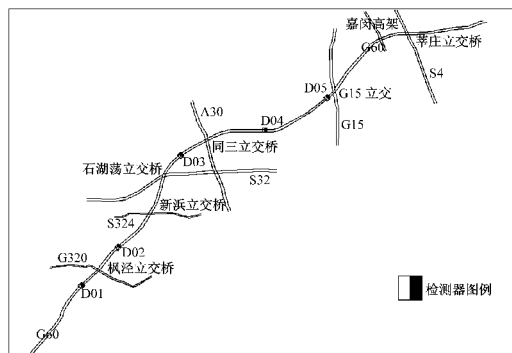


图 2 G60 高速上海主线段线圈检测器布设位置信息

Fig.2 Location of the loop detectors on G60 Freeway in Shanghai

表1 G60 高速上海主线段线圈检测器布置位置道路环境信息

Tab.1 Road environment of location of the loop detectors on G60 Freeway in Shanghai

编号	线圈数目	断面车道数	断面线形	交通安全设施
D01	8	8	直线	完好
D02	8	8	直线	完好
D03	8	8	直线	完好
D04	8	8	直线	完好
D05	10	10	直线	完好

## 2.2 事故数据

原始事故数据为 2014-01—2015-09 期间 G60 沪昆高速上海段接报事故数据. 原始事故信息存储在 excel 工作簿中, 记录的字段有日期、事件发生时间、位置、涉及车辆数目、事故形态、伤亡情况和处置结束时间. 在统计期间共计发生事故 913 起. 根据接报的事故形态对事故进行初步统计, 如图 3 所示.

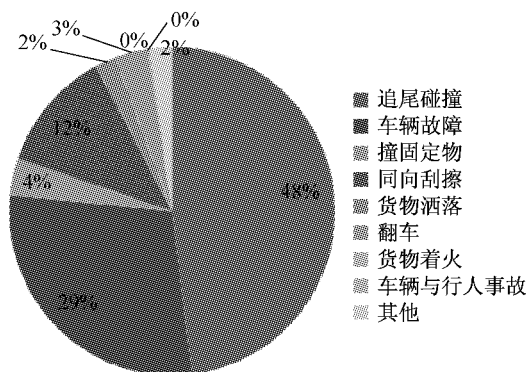


图3 G60 上海段事故形态统计

Fig.3 Detailed crash types of the crashes on G60 Freeway in Shanghai

其中追尾碰撞和同向刮擦累计占比 60%，除去车辆故障、撞固定物、货物洒落、翻车等因为驾驶人操作不当和车辆突发故障等不可预期因素引发的事故以外, 可以认为与交通流因素有关的事故均为追尾碰撞事故. 因此, 着重对该段高速公路所产生的追尾碰撞事故进行建模研究.

## 2.3 交通流数据与事故数据配对

从图 2 可以看出, 长达 48.7 km 的高速公路主线路段上仅布置了 5 组线圈检测器, 两检测器间最近的距离为 2.63 km, 最远的为 18.6 km, 平均间距 8.7 km, 与国外布设密度 0.5 mile(约 800 m)相比, 相差甚远. 因此, 尝试从单检测器采集的交通流数据入手, 以高速公路出入口和立交枢纽作为节点, 以节点中间部分作为基础路段, 通过单检测器所在的桩号位置判定检测器隶属路段, 并按接报事故中的桩号信息判定事故所处的路段, 应用这种空间配对方

式对事故数据和交通流数据进行配对. 例如, 检测器 D02 安装在枫泾立交和新浜立交中间, 若记录的事故根据记录的桩号和方向判别发生在枫泾立交和新浜立交中间, 则通过数据库建立链接, 可以查询到事故对应路段对应时间的交通流数据, 以完成配对.

## 2.4 时间片段选取

由于 20 s 的交通流数据过于离散, 在研究中, 往往需要把这些交通流数据集计之后再行研究. Pande<sup>[15]</sup> 等人在研究中表明, 采用 5 min 集计的交通流数据来对事故进行预测比 3 min 集计的交通流数据结果更加显著. 将 20s 的数据集计至 1 min 的数据, 然后按照 5 min 一段时间(共 5 个连续 1 min 集计数据)对数据再次进行集计, 并构建研判指标体系. 由于预测需具备一定的前瞻性, 而在现状交通调度管理过程中, 从事故发生至接到报警救援电话中间大致有 3 min 的延迟, 故选取记录时间前的 5~20 min 这 15 min, 共计 3 个 5 min 的时间片段进行研究, 并选取最优的预测时间片段.

# 3 基于单检测器数据的追尾事故风险研判支持向量机模型

## 3.1 支持向量机模型

支持向量机(Support Vector Machines, SVM), 作为机器学习领域最新研究热点, 能够较好的解决传统统计学习理论遇到的高维数据、小样本、非线性和局部极小点等实际问题<sup>[16]</sup>. 近年来在各大应用中使用的文字识别、网页文本分类、语音识别、人脸识别、数据压缩等功能中都有使用 SVM 算法.

支持向量机算法是基于结构风险最小化原则下, 通过对线性可分的分类问题建立一个最优分类界面, 令该分类界面两侧和最近的两类样本之间的距离相等且距离最大化, 从而使支持向量机具有良好的泛化能力. 支持向量机通过引入核函数, 将高维数据映射到低维空间, 很好地解决了“维数灾难”的问题. 由于对于事故的预测只存在会发生事故(1)和不会发生事故(0)两种结果, 实际上便为一个二元分类问题, 同时由于事故样本量较小, 因此采用支持向量机构建分类器能够取得较优结果. 采用 C-SVC(C-Support Vector Classification)模型作为最优分类界面求解模型. 那么该模型便为解决下列描述的优化问题.

$$\begin{aligned} \min_{\omega, b, \xi} & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i \\ \text{s. t. } & y_i (\omega^T \phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, l \end{aligned}$$

式中:  $\phi(x_i)$  为核函数, 它将  $x_i$  投射到更高的维度;  $C$  是惩罚参数, 是一个大于零的常量, 调节该值可改变模型的偏向;  $\omega, b, \xi$  三个参数为该优化问题的变量, 可在模型的优化求解过程中求解。

C-SVC 模型经数学方法求解最终可得到分类的决策函数, 其形式如下:

$$f(x) = \text{sgn}(\omega^T \phi(x) + b) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b\right)$$

式中:  $x_i, y_i \alpha_i \forall i, b$  为模型优化求解得到的参数;  $K(x_i, x)$  为高维映射选用的核函数. 根据核函数的不同, C-SVC 模型决策函数的最终形式也不相同。

C-SVC 模型支持常见的线性、多项式、RBF(径向基)、Sigmoid 四种核函数, 研究经验表明, 采用 RBF 核函数作为 C-SVC 模型高维映射变换函数的分类预测模型的分类效果普遍优于其他三类. 因此选用 RBF(径向基)核函数作为样本高维映射函数. RBF 核函数的数学形式为

$$K(x_i, x) = e^{-\gamma \|x_i - x\|^2}$$

将核函数代入 C-SVC 模型决策函数中得到采用 RBF 核函数的 C-SVC 模型的决策函数, 即

$$f(x) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i e^{-\gamma \|x_i - x\|^2} + b\right)$$

### 3.2 支持向量机分类精度评价指标

SVM 的分类精度一般采用三个指标来进行评价, 分别为总体预测精度、事故预测精度和误报率。

(1) 总体预测精度指的是 SVM 对所有样本数据(含事故组样本和对照组非事故样本)的正确预测样本数占总数的百分比, 即

$$A = \frac{T_P + F_N}{P + N} \times 100\%$$

式中:  $A$  为总体预测精度;  $T_P$  为事故样本预测为事故的样本数;  $F_N$  为非事故样本预测为非事故的样本数;  $P$  为事故样本数;  $N$  为非事故样本数。

(2) 事故预测精度  $T_{PR}$  指的是 SVM 正确预测的事故样本数占总事故样本数的百分比, 即

$$T_{PR} = \frac{T_P}{P} \times 100\%$$

(3) 误报率  $F_{PR}$  指的是 SVM 将非事故样本数预测为事故的样本数  $F_P$  占非事故样本总数的百分比, 即

$$F_{PR} = \frac{F_P}{N} \times 100\%$$

下文将综合考虑以上三个指标值, 对模型分类精度进行评价。

### 3.3 基于单检测器数据的追尾事故风险建模

#### 3.3.1 参数描述

G60 高速上海段, 单向 3—5 车道, 建模过程将考虑到模型的普适性, 并充分利用单断面线圈检测器所采集的车道级交通流信息(如图 4 所示)。

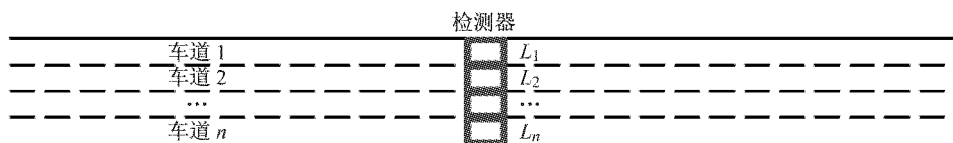


图 4 车道线圈检测器布设

Fig.4 Detectors of each lane

每个车道上的检测器采集的数据有流量( $f$ )、占有率(Occ)和速度( $V$ ), 将这些参数进行集计, 集计至 1 min 的数据, 然后按照 5 min 一段时间(共 5 个连续 1 min 集计数据)对数据再次进行集计, 车道  $n$  第  $t$  个的 1 min 集计数据分别以  $f_{nt}$ ,  $O_{cnt}$  和  $V_{nt}$  表示, 构建如表 2 的研判指标体系。

#### 3.3.2 配对案例对照方法构建样本

高速公路交通事故因素主要由驾驶人、车辆特性、动态交通流、道路环境等因素构成, 而驾驶人、车

辆特性、道路环境因素方面的数据无法获取, 因此仅对动态交通流和事故之间的关系进行研究, 在构建模型时, 需要对驾驶人、车辆特性及道路环境等因素进行控制变量研究. Ghosh 和 Chen<sup>[17]</sup> 指出, 当对照组和案例组比例超过 4:1 时, 统计结果并不会显著增加. 因此本文也采用对照组和案例组 4:1 的比例. 对照组根据事故发生的时间进行选定, 分别选取其前 14 d 相应时间、前 7 d 相应时间、后 7 d 相应时间和后 14 d 相应时间的交通流数据作为对照组, 由于线

表2 研究中所用的参数描述

Tab.2 Parameters and symbols used in the study

参数	参数描述(5 min 内)	计算公式	参数	参数描述(5 min 内)	计算公式
断面流量	本组线圈检测器检测到的车流量总和	$\sum_{t=1}^5 \sum_{n=1}^n f_{nt}$	车道流量方差累计	每个车道流量值的方差总和	$\sum_{n=1}^n D_n(f)$
平均占有率	各车道线圈检测到占有率均值	$E(\sum_{t=1}^5 \sum_{n=1}^n Oc_{nt})$	车道占有率方差累计	每个车道占有率值的方差总和	$\sum_{n=1}^n D_n(O_{cc})$
平均速度	各车道线圈检测到的过车平均速度	$\frac{\sum_{t=1}^5 \sum_{n=1}^n (f_{nt} \times V_{nt})}{\sum_{t=1}^5 \sum_{n=1}^n f_{nt}}$	车道速度方差累计	每个车道速度值的方差总和	$\sum_{n=1}^n D_n(V)$
车道间流量累计均方差	每 min 不同车道流量值方差之和	$\sum_{t=1}^5 D_t(f)$	最大车道流量方差	各车道中最大的流量方差	$\text{Max}(D_n(f))$
车道间占有率累计均方差	每 min 不同车道占有率值方差之和	$\sum_{t=1}^5 D_t(O_{cc})$	最大车道占有率方差	各车道中最大的占有率方差	$\text{Max}(D_n(O_{cc}))$
车道间速度累计均方差	每 min 不同车道速度值方差之和	$\sum_{t=1}^5 D_t(V)$	最大车道速度方差	各车道中最大的速度方差	$\text{Max}(D_n(V))$

圈检测器的偶发故障,所以部分交通流数据采集值为 NaN 空值或者线圈检测器上传的数据字段中含有“invalid”和“badloop”(无效和线圈故障)信息,构建样本时将该部分无效数据剔除,分别对 5~10 min,10~15 min,15~20 min 三组时间段的数据进行样本构建,样本详细信息如表 3 所示:

表3 样本详细信息

Tab.3 Detailed information of the samples

时间片段	5~10 min		10~15 min		15~20 min	
	事故组	对照组	事故组	对照组	事故组	对照组
样本数量	307	1 210	309	1 200	305	1 210

### 3.3.3 SMOTE 抽样技术

剔除无效样本后,各组样本的事故组和对照组的比例约为 1:4. 在机器学习领域,不均衡数据由于其数据比例的不均衡,导致各式分类器对样本进行分类时会有明显的偏向性,导致少数类的分类精度过低<sup>[18]</sup>. 为了避免不均衡数据对建模带来的影响,采用 SMOTE (Synthetic Minority Over-sampling Technique)来对样本进行改进.

SMOTE 过抽样技术是 Chawla<sup>[19]</sup>提出的少数样本合成抽样技术. SMOTE 方法的基本假设为同类样本在模式空间中必定互相靠近,因此在一个少数类样本最近邻随机选择一个样本,在两者之间的连线上随机选一点作为新合成的少数类样本. 若抽样率是  $m$ ,对于每个少数类样本  $x_i$ ,找出它  $k$  个少数类近邻点. 从中任选  $m$  个近邻点  $y_{ij} (j=1,2,\dots,m)$ ,具体插值方法如下:

$$P_j = x_i + \text{rand}(0,1) \times (y_{ij} - x_i)$$

通过 SMOTE 算法产生少数类样本,扩大了少数类的分类决策区域,达到了平衡数据集、提高少数类识别率的目的.

样本经过 SMOTE 后样本详细信息见表 4.

表4 SMOTE 抽样后样本详细信息

Tab.4 Detailed information of the samples after SMOTE

时间片段	5~10 min		10~15 min		15~20 min	
	事故组	对照组	事故组	对照组	事故组	对照组
样本数量	1 228	1 210	1 236	1 200	1 220	1 210

经过抽样以后事故组和对照组比例约为 1:1. 下面便开始基于此样本进行支持向量机建模.

### 3.3.4 模型分析

研究基于 Matlab 的 LIBSVM 工具箱和 Mysql 数据库对样本数据进行处理建模,对 5~10 min,10~15 min 和 15~20 min 三组样本分别构建支持向量机模型. 随机将各组样本按照训练集:测试集 7:3 的比例进行分配. 用训练集数据进行训练支持向量机,用测试集的数据进行模型测试.

在 C-SVC 模型的构建时,引入了常数  $C$  作为惩罚系数控制损失的大小. 模型求解中  $C$  可作为调节参数,影响训练模型的性能. 此外, RBF 核函数中  $\gamma$  参数也是模型训练前需输入的常数,该参数的数值也会明显影响模型的性能. 所以应用 SVM 方法解决分类问题还需解决 SVM 模型参数及核函数参数的寻优问题,得到分类效果最佳的一组  $C$ 、 $\gamma$  参数. 过程中采用网格分析法及交叉验证对支持向量机中的 RBF 核函数进行参数寻优,通过编程枚举的方式对不同参数下的模型预测效果进行对比,得出最优的参数对  $(C, \gamma)$  从而建立 SVM 分类器. 将训练集作为模型训练数据输入 LIBSVM 中对 C-SVC 模型最优分类问题进行求解,同时进行  $C$ 、 $\gamma$  参数的寻优. 对寻优过程中枚举的不同  $C$ 、 $\gamma$  组合下的模型交互检验的预测准确率进行列举并绘制得到如图 5 的参数寻优三维视图,各分类器的分类精度如表 5

所示。

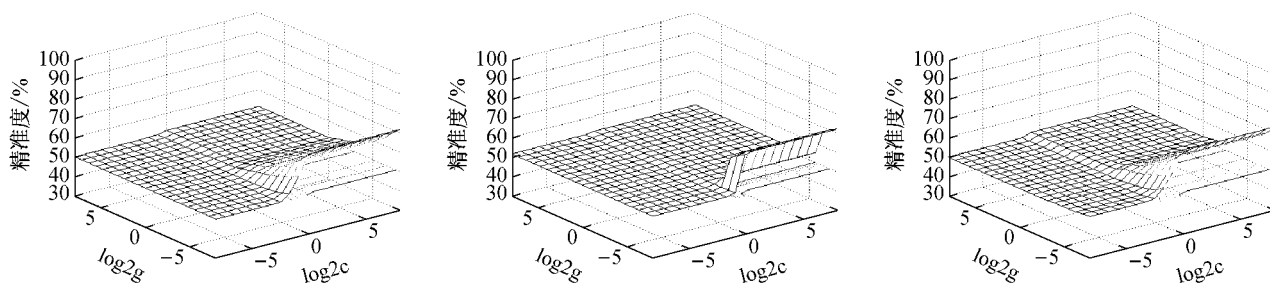


图 5 SVM 核函数参数寻优结果

Fig.5 Training results of the optimal kernel function parameters

表 5 支持向量机分类效果

Tab.5 Performance of the SVM classifiers

	5~10 min			10~15 min			15~20 min		
	$T_{PR}$	$F_{PR}$	$A$	$T_{PR}$	$F_{PR}$	$A$	$T_{PR}$	$F_{PR}$	$A$
训练集	100% 852/852	0 0/855	100% 1 707/1 707	100% 882/882	0 823/823	100% 1 705/1 705	100% 850/850	0 851/851	100% 1 701/1 701
测试集	50.53% 190/376	1.13% 4/355	74.01% 541/731	48.31% 171/354	2.92% 11/377	73.46% 537/731	45.41% 168/370	1.95% 7/359	71.33% 520/729
总体	84.85% 1 042/1 228	0.33% 4/1 210	92.21% 2 248/2 438	85.19% 1 053/1 236	0.92% 11/1 200	92.04% 2 242/2 436	83.44% 1 018/1 220	0.58% 7/1 210	91.40% 2 221/2 430

由表 5 可以发现,从总体样本的分类效果上看,基于 5~10 min 构建的支持向量机分类器有着最高的总体预测精度 92.21%,其事故预测精度 84.85%与基于 10~15 min 构建的支持向量机分类器的事 故预测精度 85.19%相近,说明该两种分类器对于事 故均有着较优的分类效果,但从误报率上来看,由于 在实际交通管理过程中,误报率需要保持可接受范 围内且尽可能低以减少错误预警发生的频率,因此 5 ~10 min 的预测器(0.33%)要优于 10~15 min 的 预测器(0.92%);就测试集的分类效果而言,5~10 min 预测器具备着最高的事故预测精度 50.53%,误 报率也同时最低 1.13%。所以,为了保证模型的可靠 性及在实际管理中的可应用性,最终选用 5~10 min 的支持向量机分类器作为最优分类器。

4 结语

采用 G60 高速上海段的单流量检测器采集的交 通流数据,通过配对案例对照的方法对追尾事故进 行预测并建立了支持向量机模型,结果表明基于事 故前 5~10 min 的交通流数据构建的支持向量机分 类器能对事故进行最佳预测,总体误报率为0.33%, 表明了该 SVM 分类器在实际应用中具备潜在的应 用价值,同时也表明了基于单流量检测器的交通流 数据对事故进行预测的可靠性。

从单检测器数据入手,采用车道级别的交通流

数据构建实时的追尾事故风险预测模型,一方面初 步解决了现阶段低信息化高速公路管理过程中滞后 性问题,提高交通安全调度管理的前瞻性和主动性; 另一方面通过对已有的交通运营数据进行挖掘,为 高速公路管理部门提供实时可靠风险预警的理论依 据。

参考文献:

[1] Madanat S, Liu PC. A prototype system for real-time incident likelihood prediction[C]//IDEA Project Final Report(ITS-2): Washington D C: Transportation Research Board of the National Academies, 1995 [2016-02-08]. [http://ntl.bts.gov/lib/jpodocs/repts\\_te/4111.pdf](http://ntl.bts.gov/lib/jpodocs/repts_te/4111.pdf).  
[2] Lee C, Saccomanno F, Hellinga B. Analysis of crash precursors on instrumented freeways [J]. Transportation Research Record:Journal of the Transportation Research Board, 2002, 1784(1):1.  
[3] Abdel-Aty M, Uddin N, Pande A, et al. Predicting freeway crashes from loop detector data by matched case-control logistic regression[J]. Transportation Research Record:Journal of the Transportation Research Board, 2004, 1897(1):88.  
[4] Golob T F, Recker W W, Alvarez V M. Freeway safety as a function of traffic flow[J]. Accident Analysis and Prevention, 2004, 36:933.  
[5] Oh J S, Oh C, Ritchie S G, et al. Real-time estimation of accident likelihood for safety enhancement [J]. Journal of Transportation Engineering, 2005, 131(5):358.  
[6] Pande A, Abdel-Aty M. Comprehensive analysis of the relationship between real-time traffic surveillance data and

- rear-end crashes on freeways [J]. Transportation Research Record: Journal of the Transportation Research Board, 2006, 1953(1):31.
- [7] Golob T F, Recker W, Pavlis Y. Probabilistic models of freeway safety performance using traffic flow data as predictors [J]. Safety Science, 2008, 46:1306.
- [8] Hossain M, Muromachi Y. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways [J]. Accident Analysis and Prevention, 2012, 45:373.
- [9] Xu C, Liu P, Wang W, *et al.* Evaluation of the impacts of traffic states on crash risks on freeways [J]. Accident Analysis & Prevention, 2012, 47:162.
- [10] Ahmed M M, Abdel-Aty M, Yu R. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data [J]. Transportation Research Record: Journal of the Transportation Research Board, 2012, 2280(1):60.
- [11] Ahmed M, Abdel-Aty M. A data fusion framework for real-time risk assessment on freeways [J]. Transportation Research Part C: Emerging Technologies, 2013, 26:203.
- [12] 徐铖铖, 刘攀, 王炜, 等. 基于判别分析的高速公路交通安全实时评价指标 [J]. 东南大学学报: 自然科学版, 2012(3):555.  
XU Chengcheng, LIU Pan, WANG Wei, *et al.* Discriminant analysis-based method to develop real-time crash indicator for evaluating freeway safety [J]. Journal of Southeast University: Natural Science Edition, 2012(3):555.
- [13] 徐铖铖, 刘攀, 王炜, 等. 恶劣天气下高速公路实时事故风险预测模型 [J]. 吉林大学学报: 工学版, 2013(1):68.  
XU Chengcheng, LIU Pan, WANG Wei, *et al.* Real time crash risk prediction model on freeways under nasty weather conditions [J]. Journal of Jilin University: Engineering and Technology Edition, 2013(1):68.
- [14] 李志斌, 刘攀, 金茂菁, 等. 高速公路常发拥堵路段追尾事故风险实时预测 [J]. 吉林大学学报: 工学版, 2013(6):1482.  
LI Zhibin, LIU Pan, JIN Maojing, *et al.* Real-time prediction of rear-end crashes near freeway recurrent bottlenecks [J]. Journal of Jilin University: Engineering and Technology Edition, 2013(6):1482.
- [15] Pande A, Abdel-Aty M, Hsia L. Spatiotemporal variation of risk preceding crashes on freeways [J]. Transportation Research Record: Journal of the Transportation Research Board, 2005, 1908:26.
- [16] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000(1):36.  
ZHANG Xuegong. Introduction to statistical learning theory and support vector machines [J]. Acta Automatica Sinica, 2000 (1):36.
- [17] Ghosh M, Chen MH. Bayesian inference for matched case-control studies [J]. Sankhya: The Indian Journal of Statistics, Series B, 2002, 64:107.
- [18] 陶新民, 刘福荣, 杜宝祥. 不平衡数据 SVM 分类算法及其应用 [M]. 哈尔滨: 黑龙江科学技术出版社, 2011.  
TAO Xinmin, LIU Furong, DU Baoxiang. Imbalanced data SVM classifier algorithm and application [M]. Harbin: Heilongjiang Science & Technology Press, 2011.
- [19] Chawla N V, Bowyer K W, Hall L O, *et al.* SMOTE: synthetic minority over-sampling technique [J]. Journal of Artificial Intelligence Research, 2011, 16(1):321.