

考虑货车因素的高速公路短期交通流风险预测

张兰芳, 赵 焜

(同济大学 道路与交通工程教育部重点试验室, 上海 201804)

摘要: 基于 G15 上海段的交通流数据和交通事故数据, 研究货车比例较高且货车事故率较高的高速公路短期交通流风险预测模型. 分别选取整体交通流参数、货车交通流参数和综合参数作为特征变量, 通过支持向量机进行建模, 运用遗传算法对模型参数进行寻优, 建立不同时间段、不同风险特征变量的分类模型并对比分析. 结果表明事故发生前 5~10 min 的模型预测精度最高. 当加入货车因素时, 总体的预测精度提高 7.1%, 事故预测精度提高 6.1%, 误报率降低 4.7%. 采用平均影响值法进行货车因素对预测结果的影响程度分析, 表明货车因素对于预测模型有较大影响. 该研究模型可用于开发交通安全预警系统, 为高速公路货车安全管理提供理论依据.

关键词: 高速公路; 交通安全; 货车因素; 支持向量机; 风险预测模型

中图分类号: U491

文献标志码: A

Short-term Traffic Flow Risk Prediction on Freeways Based on Truck Factors

ZHANG Lanfang, ZHAO Kun

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

Abstract: Based on the traffic data and crash data collected on G15, a study was made of the short-term traffic flow risk prediction model on freeways with high proportion of trucks and high proportion of truck crashes. The overall traffic flow parameters, the truck traffic flow parameters and the comprehensive parameters were selected as the risk characteristic variables. The support vector machine was adopted for the modeling and genetic algorithm was used to optimize the parameters. Classification models of different time periods, different risk characteristic variables were got and compared. The results show that the model using the data within 5 to 10 minutes before the accident performs the best. When considering truck factors, the overall prediction accuracy improves 7.1%, the crash rate prediction accuracy

improves 6.1% and the false alarm rate is 4.7% lower. Finally, the different importance of characteristic variables was obtained through mean impact value. The results show that truck factors have larger effects on the prediction model. The model in this research can be used to develop early warning system of traffic security and provide theoretical basis of truck safety management on freeways.

Key words: freeway; traffic safety; truck factors; support vector machine; risk prediction model

公安部交通管理局的统计数据显示, 2016 年全国发生货车责任道路交通事故 5.04 万起, 造成 2.5 万人死亡、4.68 万人受伤, 分别占汽车责任事故总量的 30.5%、48.23% 和 27.81%, 远高于货车保有量占汽车总量的比例^[1]. 货车交通事故多发, 已成为严重的交通安全隐患. 因此货车对道路安全的影响研究就显得尤为重要. 鉴于货车事故的严重性, 如果能够采用主动防控技术, 在事故尚未发生、但有一定征兆时, 采取措施来避免事故的发生, 对于改善交通安全具有重要意义.

在货车对交通影响的研究中, Gazis 等^[2]首次提出了“移动瓶颈(moving bottleneck)”的概念, 解释重型车辆对通行能力产生的影响. Middleton 等^[3]针对混行及小客车专用道的事故类型及比例进行调查分析, 结果表明客车专用道的事故率明显低于混行车道事故率, 且混行车道货车与客车发生碰撞的概率极高. Vadlamani 等^[4]发现主要类型的货车交通事故发生的概率与货车比例、交通量、以及天气交通状况等因素有关. Yang 等^[5]对小汽车和货车的混合交通流进行了研究, 发现在不同的小汽车与货车组合中, 小汽车跟随货车的组合对于交通流稳定性影响更大. 在货车安全方面, 货车事故严重程度的影响因素分析也是近年来的研究热点^[6-8].

收稿日期: 2017-05-10

基金项目: 国家自然科学基金(71671126)

第一作者: 张兰芳(1972—), 女, 副教授, 工学博士, 主要研究方向为道路规划与几何设计、道路交通安全. E-mail: zlf2276@tongji.edu.cn

近年来交通智能化发展迅速,交通流实时数据的获取及存储更加高效.越来越多学者开始研究利用实时交通流数据进行风险评估. Oh 等^[9]研究发现 5 min 间隔的速度标准差是检测事故的最佳参数,并采用非参数贝叶斯(non-parametric Bayesian)分类模型分析事故风险. Abdel-Aty 和 Pande^[10]首先将事故风险预测作为一种分类问题进行研究,利用基于贝叶斯判别法的概率神经网络法(Probabilistic Neural Network),最终训练得到 PNN 模型,可以识别 70% 的事故. Pande 等^[11-12]采用多层感知神经网络(Multi-layer Perceptron Neural Network)和正态径向基神经网络(Normalized Radial Basis Function),分别针对追尾事故、侧碰事故进行实时事故风险预测模型建模,达到了理想的分类精度. Yu 等^[13]利用支持向量机(Support Vector Machine)和贝叶斯逻辑回归(Bayesian Logistic Regression)对山区高速公路进行实时交通流风险预测,证明在数据集较小时,支持向量机具有更好的预测性. 孙剑等^[14]应用贝叶斯网络(BN)模型对快速路实时交通流参数与事故风险进行建模,事故预测准确率达到 76.94%.

综上,在货车对于交通的影响的研究中,国内外学者从“移动瓶颈”理论出发,研究微观条件下,货车对交通流的扰动.在研究方法上,多利用仿真软件得到交通流模拟数据,而较少采用实时的交通流数据.在实时交通流风险评估的研究中,使用的模型从传统统计模型逐渐向人工智能模型方面发展.存在的问题是,由于国外货车比例较低,且货车性能与小汽车差别不明显,因此在研究中忽略了货车对交通流风险的影响.针对国内高速公路货车比例及货车事故比例较高的问题,对货车造成的交通流风险还需进行量化研究.通过在特征变量中加入货车因素,提高了高速公路短期交通流风险预测模型的预测精度,为国内货车交通安全研究在事故预测方面提供借鉴.

1 数据准备

1.1 交通流数据与事故数据

本文数据采集自沈海高速公路(G15)上海段,采集时间为 2014 年 1 月至 2015 年 9 月.交通流数据通过线圈检测器进行采集,可得到流量、速度、占有率等交通参数. G15 上海段共布设检测器 21 套,平均间距为 4.5km,检测器数据采集周期为 20 s.交通

流数据采集周期较短会产生数据噪声,因此本文按照 5 min 时间间隔对交通流数据进行集计^[10].提取事故发生前 20 min 的数据,以 5 min 为单位时段,将数据分为事故发生前 0~5 min,5~10 min,10~15 min 和 15~20 min 4 个时间段.

共采集事故数 1 220 起,其中货车相关事故占总事故比例为 61%,如图 1a 所示.统计得到事故发生时货车比例平均为 34%,即本路段货车事故比例明显高于货车比例,表明货车对于本路段的通畅与安全造成了较大的影响.事故形态分布如图 1b 所示.考虑到追尾,碰擦等事故多由交通流拥挤或紊乱等造成,此类事适合使用交通流参数进行预测.并且从图 1b 中可以看出追尾和碰擦两类事故占总事故数的 45%,比例较大,所以事故研究主要为追尾和碰擦事故.

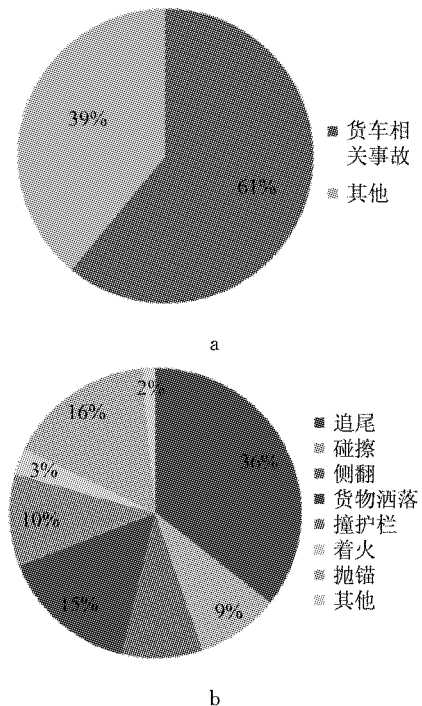


图 1 事故数据分类统计

Fig. 1 Classified statistics of crash data

对事故数据进行筛选.首先将事故发生前 20 min 交通流数据中有异常的事故数据进行剔除.其次由于 G15 布设的线圈检测器并不均匀,当线圈间隔较大时,交通流扰动很难检测到,间隔较小时,所筛选出事故数据样本量较小,预测模型不能得到有效训练,所以筛选上下游检测器间隔在 4 km 内的事故数据,可以保证交通流数据的有效性,且满足模型样本量要求,共筛选到 86 起追尾和碰擦事故.

1.2 事故对照数据

为了对比事故发生前和正常状态下的交通流动

态特征,交通流数据需要按配对方式选取,采用病例-对照研究方法,其中病例组为交通事故发生前的交通流动态特征,对照组为没有发生事故条件下的交通流动态特征.在交通安全研究领域,病例和对照的比例选择了常用的 1:4^[15-16].

非事故数据选取时,应注意与事故数据在相同地点、相同时间、相同季节、不同日期条件下采集得到,控制道路的几何条件、时段等对事故发生的影响.此外,在提取非事故数据时,要确保取样时段前后 1 h 内没有事故发生,以保证非事故数据能反应正常情况的交通条件.

例如,有一个事故发生在 2014 年 4 月 9 日 17:35,将 17:15~17:35 之间距离事故发生地点最近的检测器数据提取出来,并根据规则,提取 3 月 26 日、4 月 2 日和 4 月 16 日及 4 月 23 日相同时刻的交通流数据.17:30~17:35 的数据为 0~5 min 组,17:25~17:30 的数据为 5~10 min 组,17:20~17:25 的数据为 10~15 min 组,17:15~17:20 的数据为 15~20 min 组.

2 风险预测模型与变量选取

2.1 支持向量机

基于统计学习理论的结构风险最小化原则的 SVM 方法,将实际问题通过非线性变换转换到高维特征空间,在高维空间中构造线性判别函数来实现原空间中的非线性判别问题,其算法复杂度与样本维数无关^[18].并且相比于神经网络,SVM 不易陷入局部最优,更适合于小样本情况.基于交通风险预测问题的特点,本文选用支持向量机进行模型构建.

设有 n 个样本 x_k 及其所属类别 y_k ,表示为 $(x_k, y_k), k=1, 2, 3 \dots N, x_k \in R^N, N$ 表示输入空间的维度, $y_k \in \{1, -1\}$. x_k 指的是第 k 个样本的交通流参数, y_k 是该样本所对应的交通状态:如果是导致事故发生的状态,则 $y_k=1$,如果是非事故发生状态, $y_k=-1$.本文为非线性分类问题,支持向量机通过核函数,将数据映射到高维空间,使其变为线性可分问题,并在这个特征空间中构造最优分类超平面.最终寻优目标函数为式(1),相应的分类函数式为式(2).

$$\max Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (0 \leq \alpha_i \leq C) \quad (1)$$

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i^* y_i K(x_i, x_j) + b^* \right) \quad (2)$$

式中: $K(x_i, x_j)$ 为核函数,可以实现从低维空间向高

维空间的映射; C 为某个指定的常数,起到对错分样本惩罚程度控制的作用; b^* 超平面偏移量最优解, $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_N^*)^T$ 为最优解; α_i 为拉格朗日乘子.通过采用不同的核函数,SVM 算法可以构造相对应的学习机.对于样本量较小的 SVM 模型,径向基函数(Radial-basis function, RBF)作为核函数具有更好的预测性^[13],因此,采用 RBF 作为核函数,式中参数 g 是核函数半径.

$$K(x_i, x_j) = \exp(-g \|x_i - x_j\|^2), g > 0 \quad (3)$$

模型预测精度评价指标一方面是检验 SVM 模型优良程度的标准,另一方面也是 SVM 模型参数交叉验证寻优过程的基本依据.模型分类精度采用的统计指标为总体预测精度、事故预测精度和误报率.总体预测精度是指预测正确的样本数占样本总数的比例,事故预测精度是指预测正确的事故样本数占事故样本总数的比例,误报率是指将非事故样本预测为事故样本的数目占非事故样本总数的比例.

2.2 特征变量选取

提取历史事故数据和交通流原始数据,对数据进行集计,以形成一些容易识别出高事故风险交通状态的特征变量.在短期交通流风险预测模型的研究中,国内外学者提出了很多特征变量,如速度标准差、速度变异系数、不同检测器不同时间间隔的速度、流量、占有率数据^[15-17].为进一步研究货车比例在短期交通流风险预测中的作用,在特征变量选取中考虑了货车因素.为方便命名和编号,事故上游检测器采集到的参数在第一位用 U 标明,事故下游检测器采集到的参数在第一位用 D 标明,特征变量如表 1 所示.

将交通流运行风险分为可能发生事故和不会发生事故两种状态,因此可以将交通流风险预测问题看成一个二分类问题,样本数据中包含事故数据与非事故数据,每个数据包含 A 组、B 组、C 组不同维度的自变量,如果发生事故则将状态变量记为 1,未发生事故则将状态变量记为 -1.将样本数据分为训练样本和测试样本,模型基于训练样本数据对分类器进行训练并建立模型,利用建立的模型对测试数据集进行估计,以此验证模型的精度.

3 模型构建

由于事故样本与非事故样本不均衡,首先对样本进行过抽样,可得到 4 个时间段,3 组特征变量的 12 个数据组.使用 MATLAB 平台 Libsvm 进行

表 1 模型中所用特征变量

Tab.1 Characteristic variables used in the model

A 组整体交通流变量		B 组货车变量	
变量	描述	变量	描述
UV	上游检测器平均速度/(km·h ⁻¹)	UVT	上游检测器货车平均速度/(km·h ⁻¹)
UQ	上游检测器总流量/(辆·5 min ⁻¹)	USVT	上游检测器货车速度标准差/(km·h ⁻¹)
UOC	上游检测器平均占有率/ %	UPT	上游检测器货车比例
USV	上游检测器速度标准差/(km·h ⁻¹)	USPT	上游检测器货车比例标准差
USQ	上游检测器流量标准差/(辆·5 min ⁻¹)	UVD	上游检测器客货速度差/(km·h ⁻¹)
USOC	上游检测器占有率标准差/ %	USVD	上游检测器客货速度差标准差/(km·h ⁻¹)
DV	下游检测器平均速度/(km·h ⁻¹)	DVT	下游检测器货车平均速度/(km·h ⁻¹)
DQ	下游检测器总流量/(辆·5 min ⁻¹)	DSVT	下游检测器货车速度标准差/(km·h ⁻¹)
DOC	下游检测器平均占有率/ %	DPT	下游检测器货车比例
DSV	下游检测器速度标准差/(km·h ⁻¹)	DSPT	下游检测器货车比例标准差
DSQ	下游检测器流量标准差/(辆·5 min ⁻¹)	DVD	下游检测器客货速度差/(km·h ⁻¹)
DSOC	下游检测器占有率标准差/ %	DSVD	下游检测器客货速度差标准差/(km·h ⁻¹)
VD	上下游检测器速度之差/(km·h ⁻¹)	VTD	上下游检测器货车速度之差/(km·h ⁻¹)
QD	上下游检测器流量之差/(辆·5 min ⁻¹)	PTD	上下游检测器货车比例之差
OCD	上下游检测器占有率之差/ %	VDD	上下游检测器客货速度差之差/(km·h ⁻¹)

注:C 组综合变量由 A 组变量与 B 组变量合并形成

SVM 建模,由于惩罚参数 C 和 RBF 核函数参数 g 对 SVM 模型分类会有很大影响,因此,本文采用网格搜索法和遗传算法相结合的方法对参数 C 和 g 进行寻优.得到 12 个 SVM 分类器模型,对分类器效果进行比较分析,可得到预测效果最佳的时间段及特征变量组.

3.1 基于 SMOTE 算法的过抽样方法

在数据选取时,按照病例对照的方法,事故数据与非事故数据采集的比例为 1:4,会出现样本不均衡的问题.为解决非平衡类数据集带来的误分类问题,本文采用 SMOTE 算法改变数据集的样本分布.SMOTE 算法通过人工构造少数类样本来增加正类样本的数量,从而减少数据失衡程度,从而提高模型的预测精度^[19].该算法的基本假设为同类别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:

$$P_j = x_i + \text{rand}(0,1)(y_{ij} - x_i) \quad (4)$$

选取整体交通流特征变量,提取事故前 5~10 min 交通流数据,通过 SMOTE 算法将少数类样本扩大不同的倍数,使用 2 组不同的惩罚因子 C 和核参数 g ,构建 SVM 模型并进行比较分析.为使用同一个分类器模型对不同倍数预测结果进行比较,选取参数时并未使用遗传算法优化参数.结果见表 2.

通过表 2 可知,在未进行 SMOTE 过抽样时,事故预测精度均为 0.经过 SMOTE 算法过抽样后,尽

表 2 不同扩大倍数的预测精度

Tab.2 The prediction accuracy of different multiples

模型参数 (C, g)	SMOTE 过 抽样倍数	事故预测 精度/ %	误报率/ %
(5, 1.0)	不扩大	0	7.692 3
	2 倍	33.333 3	20
	4 倍	80	33.846 2
(10, 2.0)	不扩大	0	10.769 2
	2 倍	69.697	24.615 4
	4 倍	87.878 8	30.769 2

管模型参数选择不同,但事故预测精度均得到提升.对比将事故数据扩大 2 倍和 4 倍的结果,扩大 4 倍后,事故样本与非事故样本同样多,不存在数据不均衡性问题,误报率有所降低,但事故预测精度有较大提高.因此,最终选择将少数类样本扩大 4 倍,并对模型参数进行寻优.

3.2 基于遗传算法的 SVM 参数寻优

采用的核函数是径向基核函数,主要的参数是惩罚因子 C 和核参数 g .参数 C 表明了对误差的宽容度, C 越大,容忍出错的程度越小;参数 g 是核函数半径,隐含地决定了数据映射到新的特征空间后的分布.提高 SVM 模型的分类精度就要选择合适的方法对模型参数进行选择.遗传算法属于启发式算法,鲁棒性较强,对函数要求不高,且不容易陷入局部最优^[20],因此,采用遗传算法对 SVM 模型参数进行选择.在用遗传算法对参数进行寻优时,首先要对参数的寻优范围进行设置.采用网格搜索法对模型参数进行粗选,确定参数的大致范围,再用遗传算法对参数进行精选.以 C 组变量为例,首先将参数 C, g 设置为默认值,粗选结果如图 2 所示.

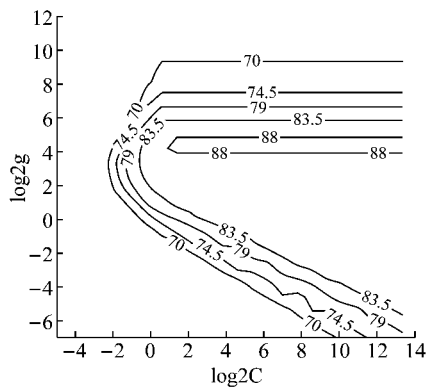
图 2 参数 C 和 g 网格搜索法粗选结果

Fig.2 Rough selecting results of parameters using grid search method

由图 2 可知,预测的分类精度最高为 88% 左右,综合考虑寻优的复杂度及参数寻优质量,得到参数的粗选结果为 $C \in [2^{-1}, 2^5]$, $g \in [2^0, 2^6]$. 得到参数粗选结果后,采用遗传算法对参数进行精选. 对参数进行二进制编码,并将种群初始化,为降低算法的复杂度,初始种群规模不宜太大,因此本文将种群中个体数目定为 20 个. 采用 k -折交叉验证 (k -fold cross validation) 的方法来评价 SVM 分类模型的性能. k -折交叉验证将原始样本分成 k 个互不相交的子集,即 k -折. 然后将每个子集数据分别做一次验证集,其余的 $k-1$ 组子集数据作为训练集,最终,将会得到 k

个模型分类精度 A_1, A_2, \dots, A_k . 用这 k 个分类精度的平均数 A 作为 k -折交叉验证下的性能指标,即

$$A = \frac{1}{k} \sum_{i=1}^k A_i \quad (5)$$

由于遗传算法是随机搜索算法,迭代终止条件并没有统一的方法. 一般采用达到预先设定的代数作为遗传算法迭代终止条件,将迭代次数设为 100,迭代终止后选择最优适应度值作为最后的解. 在运算过程中,可能会有多组的 C 和 g 对应于最高的验证分类准确率,由于惩罚参数 C 取值过大时会造成“过学习”状态的发生,即训练集分类准确率很高而测试集分类准确率很低,使构建的 SVM 模型泛化能力降低. 因此,在处理这种情况时,优先选择参数 C 最小的那组 C 和 g 作为最佳参数,如果参数 C 最小时对应多组参数 g ,则选取搜索到的第 1 组 C 和 g 作为最佳参数.

3.3 预测结果及分析

以 5 min 为时间间隔选择事故发生前 20 min 的特征变量,用遗传算法对模型参数进行寻优,共建立了 12 个 SVM 分类器,进行多次迭代,如图 3 所示,为 C 组变量迭代的一组实例.

将遗传算法迭代运算后求得的参数组合对训练数据集进行训练,可获得较高预测精度. 将模型应用于测试集,可得到最终的分类预测结果如表 3 所示.

表 3 各组模型参数及预测精度

Tab.3 Parameters and prediction accuracy of models

时间段/min	组	模型参数(C, g)	测试集预测精度/%		
			总体预测精度	事故预测精度	误报率
0~5	A	$C=3.000\ 2\ g=26.794\ 3$	84.782 6	58.064 5	11.639 3
	B	$C=2.068\ 8\ g=17.009\ 4$	91.304 3	83.871	14.918 0
	C	$C=5.001\ 8\ g=3.874\ 1$	88.043 5	70.967 7	8.278 7
5~10	A	$C=3.533\ 7\ g=12.410\ 4$	82.653 1	69.697 0	7.769 2
	B	$C=5.128\ 3\ g=12.529\ 6$	90.816 3	84.848 5	6.153 8
	C	$C=3.645\ 8\ g=5.628\ 9$	89.795 9	75.757 6	3.076 9
10~15	A	$C=7.47\ g=21.284\ 1$	67.777 8	71.875 0	9.448 3
	B	$C=4.689\ g=8.655\ 2$	80.333 3	68.750 0	8.620 7
	C	$C=3.126\ g=4.253\ 8$	75.777 8	78.125	6.896 6
15~20	A	$C=4.729\ 4\ g=4.611\ 8$	52.688 2	63.096 8	25.516 1
	B	$C=7.542\ 9\ g=10.952\ 8$	74.946 2	76.871	14.516 1
	C	$C=5.873\ 3\ g=2.074\ 9$	54.838 7	73.548 4	17.516 1

首先比较不同时间段之间的预测精度. 从表中看出,事故发生前 15~20 min、10~15 min 的总体预测精度、事故预测精度相比于 0~5 min、5~10 min 低,而误报率较高,表明距离事故发生时间较早的交通流数据提供的可用于预测事故的信息量较少,对于交通流风险预测的价值较低. 而本文将事故接警时间等同于事故发生时间,实际事故发生时间要更

早,尽管此处的 0~5 min 时段具有较好的总体预测精度,但此时段可能包含事故发生后的数据. 另外考虑到在实际交通流风险预测的应用中,预测时间应保证一定的提前量,即应当给交通管理部门一定的反应时间来采取措施,因此 0~5 min 时段不作为预测时间段. 综上,5~10 min 时段的预测效果更好.

将 3 组特征变量事故发生前 5~10 min 的预测

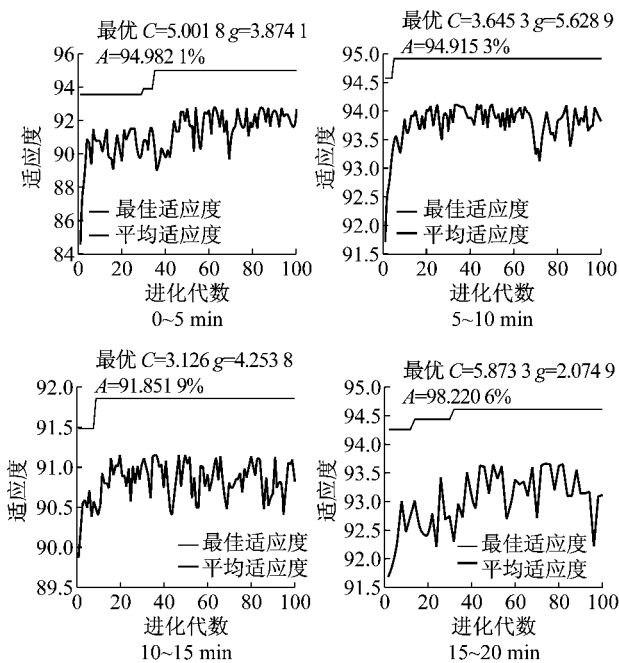


图 3 SVM 模型参数寻优图

Fig.3 parameter optimization of SVM models

结果进行比较,结果发现:与 A 组变量相比,C 组变量总体预测精度与事故预测精度均得到提高,且误差率降低至 3.076 9%,说明加入货车相关变量,可提高交通流风险预测精度. B 组变量即仅采用货车特征变量时,总体预测精度和事故预测精度相比于 C 组有所提高,但误报率有所升高. B 组和 C 组变量均比 A 组变量有更好的预测效果,即本路段在进行交通流风险预测时,应当将货车变量作为风险特征变量.

3.4 基于 MIV 值的敏感性分析

考虑到模型中风险特征变量较多,因此采用平均影响值法(MIV 法)研究在 C 组变量中,货车因素对预测结果的影响程度,同时对特征变量进行分析和筛选. MIV 方法在神经网络中是非常有效的一种变量筛选方法,在模型训练终止后,将训练样本中每一特征变量在其原值的基础上分别加减 10% 构成两个新的训练样本,再利用已建成的模型得到两个新的结果,新结果之间的差值即为变动该特征变量对预测结果产生的影响变化值. 结合前文中对所建立的 SVM 预测模型比较,选择事故发生前 5~10 min 的 C 组特征变量的预测模型,计算各特征变量的 MIV 值,取绝对值后进行排序,得到的结果如表 4 所示.

在 30 个特征变量的前 10 个主要影响变量中,下游占有率标准差对预测模型的影响最大,即当下游占有率发生变化时,由于交通流的传递性,上游交

表 4 C 组变量模型的 MIV 值统计表

Tab.4 Statistical table of MIV values of Group C model

特征变量	MIV 值	排序
下游占有率标准差	0.023 1	1
下游货车速度标准差	0.017	2
上游速度	0.009 7	3
下游货车比例	0.009	4
上游货车平均速度	0.008 6	5
上游货车比例	0.008 6	6
下游流量	0.008 4	7
上游流量	0.007 5	8
上游货车比例标准差	0.006 4	9
上下游货车速度之差	0.006 2	10

通流受到干扰,从而增大事故发生的可能性. MIV 值排序第 2 的变量为下游货车速度标准差,即下游货车速度变化值对交通流有较大影响,此结论与 Oh 等学者^[9]研究结论一致,只是其研究中并未对客货车进行区分. 可以看到,在前 10 个主要特征变量中,有 6 个为货车相关变量,而且上、下游货车比例均在其中,表明在模型中考虑货车相关变量可以提高预测准确率. 在货车比例及货车事故率较高路段进行主动安全管理时,应当加强对货车速度以及货车比例的等因素实时监控,从而提高预测的可靠性.

4 结语

将交通流风险预测问题看成是一个二分类问题,基于支持向量机,建立了高速公路交通流风险预测模型. 分别选取断面整体交通流参数、货车交通流参数和综合参数作为风险特征变量,以 5 min 为间隔选择事故发生前 20 min 的特征变量,用遗传算法对模型参数进行寻优. 结果表明事故发生前 5~10 min 的模型预测精度最高,加入货车因素的预测模型比仅考虑交通流参数的模型精度更高,因此可用货车交通流参数作为预测模型的特征变量. 在综合变量预测模型中,采用平均影响值法进行分析,结果表明货车因素对于预测模型有较大影响. 该模型可用于开发交通安全预警系统,为高速公路安全运营管理提供依据. 但由于仅为 G15 上海段建立了模型,接下来研究应在搜集不同道路条件以及不同货车比例的基础上进行模型优化,提高模型适用性.

参考文献:

[1] 中华人民共和国公安部交通管理局. 部分货车车型违规生产问题突出,货车运输安全隐患严重[EB/OL]. [2017-05-03]. <http://www.mps.gov.cn/n2255040/n2255043/c5609936/>

- content.html.
- Traffic Management Bureau of the Public Security Ministry of the People's Republic of China. Some trucks have severe problems of illegal production, and the security problem of trucking is serious [EB/OL]. [2017-05-03]. <http://www.mps.gov.cn/n2255040/n2255043/c5609936/content.html>.
- [2] GAZIS D C, HERMAN R. The moving and "phantom" bottleneck [J]. *Transportation Science*, 1992, 26(3): 223.
- [3] MIDDLETON D, VENGLAR S, QUIROGA C, *et al.* Strategies for separating trucks from passenger vehicles: final report [R]. Texas: Texas Transportation Institute, 2006.
- [4] VADLAMANI S, CHEN E, AHN S, *et al.* Identifying large truck hot spots using crash counts and PDOEs [J]. *Journal of Transportation Engineering*, 2011, 137(1): 11.
- [5] YANG D, JIN P, PU Y, *et al.* Stability analysis of the mixed traffic flow of cars and trucks using heterogeneous optimal velocity car-following model [J]. *Physica A: Statistical Mechanics and Its Applications*, 2014, 395(4): 371.
- [6] ZOU W, WANG X, ZHANG D. Truck crash severity in New York city: an investigation of the spatial and the time of day effects [J]. *Accident Analysis & Prevention*, 2017, 99: 249.
- [7] NAIK B, TUNG L W, ZHAO S. Weather impacts on single-vehicle truck crash injury severity [J]. *Journal of Safety Research*, 2016, 58: 57.
- [8] OSMAN M, PALETI R, MISHRA S, *et al.* Analysis of injury severity of large truck crashes in work zones [J]. *Accident Analysis & Prevention*, 2016, 97: 261.
- [9] OH C, OH J S, RITCHIE S G, *et al.* Real-time estimation of freeway accident likelihood [C/CD]//80th Annual Meeting of the Transportation Research Board. Washington D C: Transportation Research Board, 2001.
- [10] ABDEL-ATY M, PANDE A. Classification of real-time traffic speed patterns to predict crashes on freeways [C/CD]//83rd Annual Meeting of the Transportation Research Board. Washington D C: Transportation Research Board, 2004.
- [11] PANDE A, ABDEL-ATY M. Comprehensive analysis of 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.
- [12] PANDE A, ABDEL-ATY M. Assessment of freeway traffic parameters leading to lane-change related collisions [J]. *Accident Analysis & Prevention*, 2006, 38(5): 936.
- [13] YU R, ABDEL-ATY M. Utilizing support vector machine in real-time crash risk evaluation [J]. *Accident Analysis & Prevention*, 2013, 51(2): 252.
- [14] 孙剑, 孙杰. 城市快速路实时交通流运行安全主动风险评估 [J]. *同济大学学报(自然科学版)*, 2014, 42(6): 873.
- SUN Jian, SUN Jie. Proactive assessment of real-time traffic flow accident risk on urban expressway [J]. *Journal of Tongji University: Natural Science*, 2014, 42(6): 873.
- [15] ABDEL-ATY M, UDDIN N, PANDE A, *et al.* Predicting freeway crashes based on loop detector data using matched case-control logistic regression [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2004, 1897(1): 88.
- [16] ABDEL-ATY M, UDDIN N, PANDE A. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways [J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2005, 1908(1): 51.
- [17] 徐铨铨, 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, 43(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, 43(1): 68.
- [18] VAPNIK V N. 统计学习理论 [M]. 许建华, 张学工, 译. 北京: 电子工业出版社, 2004.
- VAPNIK V N. *Statistical learning theory* [M]. Translated by XU Jianhua, ZHANG Xuegong. Beijing: Publishing House of Electronics Industry, 2004.
- [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.
- [20] 付阳, 李昆仑. 支持向量机模型参数选择方法综述 [J]. *电脑知识与技术*, 2010, 6(28): 8080.
- FU Yang, LI Kunlun. A survey of model parameters selection method for support vector machines [J]. *Computer Knowledge and Technology*, 2010, 6(28): 8080.