

张 驰,杨绍祥,冯逸伟,等. 基于风险更新理论的高速公路动态风险判别方法[J]. 长安大学学报(自然科学版),2021,41(4):115-126.
ZHANG Chi, YANG Shao-xiang, FENG Yi-wei, et al. Expressway dynamic risk discrimination method based on risk update theory[J]. Journal of Chang'an University(Natural Science Edition), 2021, 41(4): 115-126.

DOI:10.19721/j.cnki.1671-8879.2021.04.011

基于风险更新理论的高速公路 动态风险判别方法

张 驰¹, 杨绍祥², 冯逸伟³, 许 甜⁴, 张昆仑¹

(1. 长安大学 公路学院, 陕西 西安 710064; 2. 云南省交通规划设计研究院有限公司, 云南 昆明 650011;
3. 西安公路研究院, 陕西 西安 710061; 4. 中交第一公路勘察设计研究院有限公司, 陕西 西安 710075)

摘 要: 为了进一步丰富高速公路交通安全保障理论体系, 提出一种包括风险源、影响场、作用对象、测度空间以及时间尺度等风险要素的高速公路动态风险判别方法。首先, 通过美国加州交通管理系统(PeMS)采集了 I80-W 高速公路上 2019 年 1~6 月的交通流数据和事故数据; 其次, 以时序为刻度将风险研究场景分为过去、现在与未来 3 类, 并以判别分析法为测度空间, 建立以 5 种交通流参数为自变量的风险判别函数, 同时以风险更新理论作为时间尺度, 利用判别函数计算风险值 R , 并应用于风险更新算法中的输入场, 完成旧风险模型更新; 最后, 采用统一样本对旧判别函数 R_{1-4} 、新判别函数 R_5 、新旧混合判别函数 R_{1-5} 以及利用风险更新得到的判别函数 R_u 进行精度检验, 其判别精度分别为 67.5%、61%、66%、74%。结果表明: 更新后的风险判别函数应用于高速公路动态风险判别具有更高的预测精度, 基于判别分析法和风险更新算法的高速公路动态风险判别方法可以持续对动态风险评价模型进行更新, 能够应用于未来各时间段的高速公路动态风险判别。

关键词: 交通工程; 动态风险理论; 风险更新理论; 高速公路; 测度空间; 判别分析

中图分类号: U491

文献标志码: A

文章编号: 1671-8879(2021)04-0115-12

Expressway dynamic risk discrimination method based on risk update theory

ZHANG Chi¹, YANG Shao-xiang², FENG Yi-wei³, XU Tian⁴, ZHANG Kun-lun¹

(1. School of Highway, Chang'an University, Xi'an 710064, Shaanxi, China; 2. Broadvision Engineering Consultants, Kunming 650011, Yunnan, China; 3. Xi'an Highway Research Institute, Xi'an 710061, Shaanxi, China; 4. CCCC First Highway Consultants Co., Ltd, Xi'an 710075, Shaanxi, China)

Abstract: In order to further enrich the theoretical system of expressway traffic safety guarantee, a dynamic risk identification method of expressway including risk source, impact field, targets, measurement space and time scale was proposed. Firstly, the traffic flow data and accident data of the I80-W highway from January to June 2019 by the California Traffic Management System

收稿日期: 2021-01-15

基金项目: 国家重点研发计划项目(2017YFC0803906); 陕西省青年科技新星项目(2020KJXX-046)

作者简介: 张 驰(1981-), 男, 四川宜宾人, 教授, 博士研究生导师, E-mail: zhangchi@chd.edu.cn.

(PeMS) were collected. Secondly, the risk research scene into the past, present and future according to the time scale were divided. The discriminant analysis method was used as the measure space to establish the risk discriminant function with five traffic flow parameters as independent variables. At the same time, the risk value R was calculated by using the discriminant function and applied it to the input field of the risk updating algorithm to complete the updating of the old risk model. Finally, the accuracy of the old discriminant function R_{1-4} , the new discriminant function R_5 , the old and new mixed discriminant function R_{1-5} and the discriminant function R_u obtained by risk updating were tested, and the discriminant accuracy was 67.5%, 61%, 66% and 74%, respectively. The results show that the updated discriminant function is applied to expressway dynamic risk discrimination has higher precision of prediction, based on discriminant analysis method and risk updating algorithm of expressway dynamic risk identification method can continuously update dynamic risk evaluation model, can be applied in each period of expressway dynamic risk discriminant in the future. 5 tabs, 9 figs, 32 refs.

Key words: traffic engineering; dynamic risk theory; risk update theory; expressway; measurement space; discriminant analysis

0 引言

近年来,高速公路在中国货运交通及客运交通中起着极为重要的作用,是中国陆路交通的主动脉。随着高速公路建设里程的增加,行车风险的数量和种类均呈增长趋势。各风险因素通过独立作用或叠加效应,对交通安全产生了重大影响。中国目前已采取较多主动安全防控措施以保障高速公路交通安全,如对道路线形、构造物及附属设施等进行安全性评价。研究动态风险对于现有的道路安全治理模式是一个重要的补充,进一步丰富了高速公路交通安全保障理论体系。相较于目前的管理手段,学者们需要在道路智能化设备发展和道路交通数据采集技术逐渐成熟的条件下,思考如何对未来发生的交通安全事件做出预判,与此发展方向密切相关的道路风险研究是未来交通安全领域的研究热点。

高速公路上交通风险的动态变化状况受交通量、车速、车道占有率等交通条件影响。从风险内容上看,动态风险是由交通流等动态因素产生的风险;从风险形式上看,动态风险时刻处于变化过程中,具有较强的时变性。动态风险值 R 作为反映和衡量实时交通条件下危险状况发生可能性的量化指标,对交通安全有重要影响。

目前,关于动态风险的研究主要集中在交通流和事故数据的收集,风险判别模型的预测精度和可移植性等方面,关键词主要分为道路风险评估和道路碰撞预测等。在预测模型研究方面,贾

丰源等通过收集上海市高速公路的交通流数据和事故数据,运用改进的贝叶斯网络(BN)模型,分析了事故发生前 5~10 min 的交通流数据,并与朴素贝叶斯分类、K 近邻、反向传播神经网络等方法进行了对比,结果表明,该方法具有较高的预测精度且相对其他模型预测效果更好^[1-2]。Lin 等将频繁模式(FP)树算法和随机森林算法同时应用于事故数据集的训练与测试,结果表明,基于 FP 树的风险预测模型比随机森林模型具有更低的误报率^[3-4]。Parsa 等研究了支持向量机(SVM)和概率神经网络(PNN)的性能,结果表明,虽然 SVM 模型整体的准确度更高,但 PNN 模型在检测率(DR)方面更好^[5]。Shew 等将基于分类树的神经网络风险预测模型和支持向量机模型分别应用于美国和中国不同的高速公路,结果表明,上述 2 种模型具备较好的可移植性^[6]。上述文献给出多种应用于交通流实时风险判别的数学模型,并证明了其研究中选取模型的适用性和准确度。但风险判别的精度不能仅依据一个数学模型的内部算法确定,风险判别参数的选取、数据收集器与事故发生点的位置关系以及数据选取的时间尺度,甚至天气和大型车比例等都会影响风险判别的效果^[7-8]。

Xu 等利用费希尔判别法对风险进行了预判,又研究了天气条件对风险的影响,结果表明,考虑天气变量的模型比不考虑天气变量模型的预测精度提升了 5%^[9-10]。张兰芳等研究了货车因素对支持向量机模型预测精度的影响,结果表明,加入货车因素时总体预测精度提高了 7%^[11]。胡功宏等提出了综合

交通流、天气等条件的高速公路行车风险指数(TRI指标),通过西攀(西昌—攀枝花)高速公路的实时数据验证了模型的有效性^[12]。以上研究证明了风险的诱因是多维度的,把跟风险联系紧密的变量考虑在内对于预测模型精度提升具有较明显的作用。

综上所述,交通流实时风险判别领域的研究大多将风险问题转化成数学模型的风险预测精度问题,以模型输出值体现风险高低,其重点放在了数据、模型和结果分析上,而鲜有研究从风险本质出发,对道路上的动态风险问题进行理论阐述和量化解释。另外,目前该领域所进行的研究大多基于历史事故数据和交通流数据展开,研究过程中往往是将收集的数据简单分为模型的标定数据和验证数据(无时间上的区分),用验证数据来检验标定后模型的准确性,考虑到风险是通过当前的观测对未来的不利事件做出预测,目前少有利用“未来”数据验证现有模型的研究。大多研究对于风险的空间范围有所界定,但缺乏对于交通流动态风险随时间变化这一特点的考虑,且缺乏对于如何完成可持续的动态风险判别的解释。

因此,该领域的动态风险研究应按照风险理论从风险源、影响场、作用对象、测度对象和时间尺度的角度对风险进行不确定意义下的量化分析,并将“未来”数据与目前的观测数据有机结合进行风险预测^[13-19]。鉴于此,本文提出了基于风险源、影响场、作用对象、测度空间和时间尺度的道路动态风险研究方法,并对如何建立各个元素之间的关系进行了量化研究。以判别分析法和线性回归方法为基础,提出量化指标风险值 R ,并以对 R 值的分析结果为依据,进行了道路动态风险等级的划分。在时间尺度的研究方面,提出了基于判别分析的线性关系风险更新算法,这对风险判别模型的更新和精确度提升有着重要的意义。

1 高速公路动态风险研究方法

本文首先对高速公路动态风险进行了量化定义,然后从风险研究的5个基本要素出发,对高速公路动态风险展开分析、判别和更新。

(1)风险分析:为了明确高速公路动态风险中的风险源、影响场和作用对象,作为测度空间和时间尺度的研究前提,采用的理论研究框架在本领域的研究中尚属首次。

(2)风险判别:对五要素中的测度空间进行研究,通过建立实时交通流参数与事故后果的对应关

系,完成当前交通流状况的风险状态判别,是本文进行风险更新的基础性研究,是风险更新研究的基础。

(3)风险更新:对五要素中的时间尺度进行研究,考虑到高速公路动态风险时变性强特点,采用一种较为新颖的研究思想,即对已标定好的风险判别模型结合最新的交通流状况进行更新,是本文的创新点和重点。

本文研究的技术路线如图1所示。

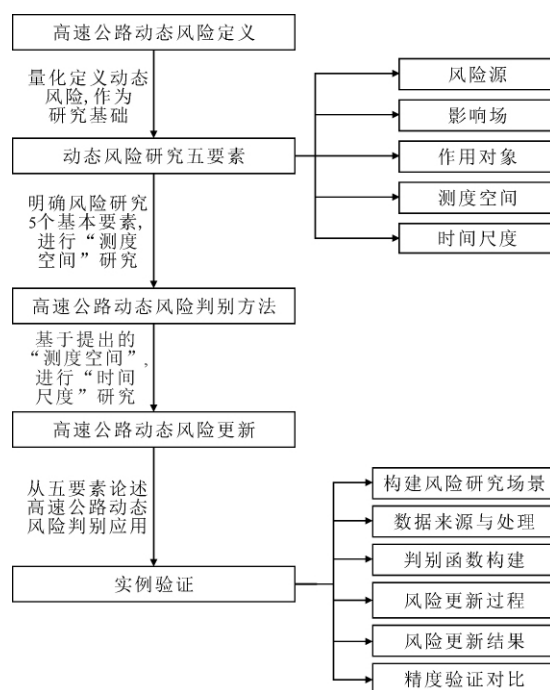


图1 高速公路动态风险研究方法

Fig. 1 Research method of expressway dynamic risk

1.1 动态风险定义

高速公路动态风险理论是为了理解并解释由于实时交通条件引起的风险现象,以时间尺度来讲,高速公路上的动态风险现象介于现在现象与未来现象之间,研究过程是以当前对道路交通条件的观测,推测未来道路交通危险现象发生的可能性。国际风险分析协会(SRA)定义专项委员会指出不同学者的理解角度和研究目标具有差异,因此对于“风险”很难形成统一的定义^[20],在道路交通行业,高速公路动态风险的定义尚不明确,因此下文将给出规定性定义来限定本文的讨论范围。

定义1(抽象化定义):高速公路动态风险是由道路上实时交通流条件引起的,与交通事故或其他交通危险现象相关的一种未来场景。

定义2(操作化定义):由于高速公路动态风险是一个三元集合,因此将高速公路交通安全实时评价指标定义为

$$R=\{S,P,C\} \quad (1)$$

式中: R 为高速公路交通安全实时评价指标,即动态风险值^[21]; S 为风险发生场景; P 为该场景发生的概率; C 为该情景的后果或严重度^[22]。

通过 Fisher 判别算法建立风险判别函数,代入实际情景下的交通流参数得到的 R 值判别可能产生的危险状况。本文对动态风险研究指标的选取主要考虑两方面的因素:交通流指标与事故发生的相关性;数据采集设备的功能及数据精度。本文选取了其中相关性较强、可靠度较高的 5 种交通流参数作为高速公路动态风险场景的数字化表征,即 $S=\{V,Q,p,q,V_{HT}\}$,其中: V 为速度; Q 为流量; p 为车道占有率; q 为大型车比例; V_{HT} 为 5 min 内事故路段上平均交通量与通行时间的乘积,体现交通负荷和交通强度。

目前交通领域的相关研究常把事故损失或人员伤亡定义为风险后果,而由于在本文研究所收集的事故样本范围内,事故伤亡人数大多为“0”或“1”,如果以伤亡人数作为后果,会导致标定出的风险判别模型在判别的区分度和准确度方面存在不足。另外,PeMS 系统中存在较多无法明确人员伤亡及事故损失参数的事故记录。为了更好地体现风险后果对于高速公路交通状况的影响,本文以事故持续时间作为反映风险后果的参数,而未采用人员伤亡和事故损失参数,即 $C=\{T_s\}$, T_s 为事故持续时间。

中国目前尚未有关于行车风险的相关规范或权威性的指导文件,因此在风险预警指标的提出,风险的量化表征,风险的研究方法,甚至是风险更新等方面,均需要从“源头”开始进行。对于体系性较强的风险研究,只要第一步提出的风险预警指标或定义的量化风险不同,后续的研究方法便截然不同了。本文研究方法是综合考虑风险研究理论、动态风险内容以及研究数据条件三方面的结果。

1.2 风险分析

1.2.1 高速公路动态风险研究五要素

高速公路的风险状况随着时间动态发展,这一发展的过程也是结合道路实时状况逐渐认识风险本质的过程。分析风险就必须明确风险源、影响场、作用对象、测度空间以及时间尺度,从最基本的元素着手,风险源是指导致风险情景产生的源头,高速公路动态风险研究中^[23-24],交通流整体的混乱和个别危险车辆均可以作为风险情景出现的触发因素,即风险源;影响场是指风险源影响范围,通常发生交通事故以后,最直接的影响对象为该路段上的其他正常

行驶车辆;作用对象是指承担风险后果的客体,最直接客体包括车辆和驾驶人;测度空间是指对风险源、影响场和作用对象进行量化描述的数学空间。本文选取了 Fisher 判别算法建立上述风险源、影响场和作用对象的测度空间。时间尺度是风险分析一个相当重要的概念,其原本的定义为“描述风险情景的时间信息”,但这里有必要对其进行补充完善,风险分析的手段是利用现在的信息对未来的风险事件做出预测,那么便产生了一个需要量化的问题,利用现在的信息对于多久以后的情景进行风险预测是有效的?应该如何对一个已经建立的测度空间进行更新使之可以持续应用于未来各个时间的风险预测?为此,本文除了研究建立风险源、影响场和作用对象之间的测度空间外,还研究了测度空间的更新。

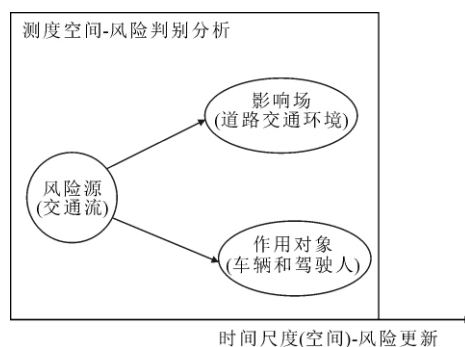


图2 各风险要素关系示意

Fig. 2 Sketch of relationships between risk factors

1.2.2 测度空间研究——基于 Fisher 判别分析算法

将 $S=\{V,Q,p,q,V_{HT}\}$ 定义为量化风险场景,通过 5 种参数的组合反映实时交通流的风险情况,从数学角度看,这 5 种参数分布在多维空间中,直接求解这 5 个参数难度较大,因此,利用 Fisher 判别分析法降维成为了解决此类问题的突破口。

使用数学方法将多维数据降为一维是比较容易实现的,但即使数据在多维空间中集群紧凑分布,投影在线性空间中也会产生重合而难以区分。为了使数据在线性空间容易区分,需要找到一条合适的直线,使各类高维样本投影的集群交互最少,这便是使用 Fisher 判别分析法的主要目的,其数学意义为每个种类内的投影值所形成的类内离差尽可能小,而不同种类内投影值形成的类间离差尽可能大,这条直线的线性表达式为

$$Z=C_1X_1+C_2X_2+\cdots+C_iX_i \quad (2)$$

式中: Z 为函数值; C_i 为系数; X_i 为自变量; i 为自

变量个数。

建立了 Fisher 线性判别函数以后,需要利用其对未知的风险情景进行判别,即将实时的交通流参数代入到已建立的线性判别式中,判断其是否会发生危险状况并对其判别效果进行评价^[25]。

假设存在 2 类后果, A 类和 B 类,按照式(2)分别计算判别函数值 Z ,并进一步求 Z 的 2 类均数和总均数,按下式计算判别界值

$$Z_C = \frac{\overline{Z_A} + \overline{Z_B}}{2} \quad (3)$$

判别规则

$$\begin{cases} Z > Z_C & \text{判为 A 类} \\ Z < Z_C & \text{判为 B 类} \\ Z = Z_C & \text{判为任意一类} \end{cases}$$

式中: Z_A 为 A 类后果; Z_B 为 B 类后果; Z_C 为 A、B 类后果的均数和的平均值(即总均数)。

Fisher 判别分析法要求各判别变量近似服从于正态分布,同时要求样本量是自变量个数的 10~20 倍以上时,建立的判别函数比较稳定;而在 8~10 倍之间时,函数的判别效果比较理想^[26]。本文选取该方法进行动态风险判别的原因主要有两方面:

(1) 相比于目前同类研究采用的支持向量机、贝叶斯网络以及随机森林等非线性模型方法, Fisher 判别法构建的线性模型具有形式简洁,计算速度快等优点,这对于高速公路实时动态风险判别来说尤为重要,假设车辆以 100 km/h 的速度行驶,其通过 2 个感应线圈的时间(研究路段间隔 500~800 m)仅为 20 s,所以从发现风险到风险判别再到风险反馈均对时效性有相当高的要求;

(2) 本文不仅从方法本身角度考虑,还考虑到了下一步需要进行风险更新的因素。风险更新方法属于一种较为新颖的风险研究思路,在交通运输领域还未涉及过具体的应用,因此选择线性关系模型作为风险更新的基础相对于其他参数模型、非参数模型、非线性模型等更为复杂的模型来说,更适合于进行首次理论探索,也较容易实现具体应用和成果落地。

1.2.3 时间尺度研究——基于风险更新理论和算法

(1) 风险更新理论

风险更新是依据最新信息对原有风险判别结果进行修正更改的除旧换新过程,当采用 Fisher 判别分析法度量风险时,风险更新表示着线性组合发生了变化。图 3 是一个三维空间上的线性判别函数示

意,假设总体分布为真实分布,符合道路动态风险的本质规律。在利用交通流参数进行风险分析时,需要充分运用有限度的观察样本来估算出整个样本总体的分布。从线性组合空间形态来看,对比基于旧样本估算得到的总体分布,将新样本数据与旧样本数据混合后的样本对总体估计的结果更加精确^[27]。

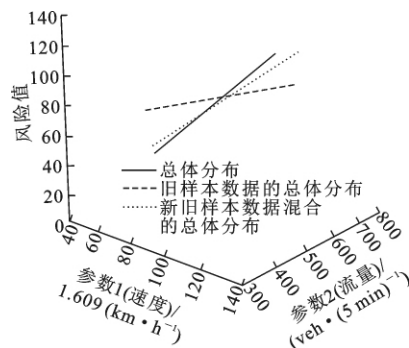


图3 三维线性函数更新示意

Fig. 3 3D linear function updated sketch

从旧的样本估计结果变化到新的样本估计结果转换为风险的过程即为风险的更新。风险是一个未来的现象,风险的更新也是必然发生的,其根本目的就是为了利用最新的信息来修正自己原有的认识,减少因为自己原有的认识不足而造成的主观不确定性和客观不确定性。

(2) 风险更新算法

风险更新算法的思路是将新的风险及旧的风险投影到背景同为属性空间 H 的一个数学空间中,风险更新的原理如图 4 所示。

风险更新的原理为基于数据形态方面的特征,以过去相似形态为背景,创建新旧信息的形态发生系统,工具选用背景的投影算子,采用形态计算技术,通过记录与重现形态来估计其内在的结构,再根据形态与背景的相关性来确定加权值,最后计算更新的风险。

2 算例

算例中将详细介绍如何以 1 个月为时间尺度,对道路风险评价模型进行更新,并对其风险判别精度进行验证。

2.1 构建风险研究场景

本文的风险研究场景构建分为空间场景构建和时间场景构建^[28]。在空间场景方面,本文选取了美国 I80-W 高速公路 0~25 × 1.609 km/h 范围作为研究路段,研究对象是在该路段上行驶的交通流,研究方法是通过数据监测断面收集车流实时状态参数

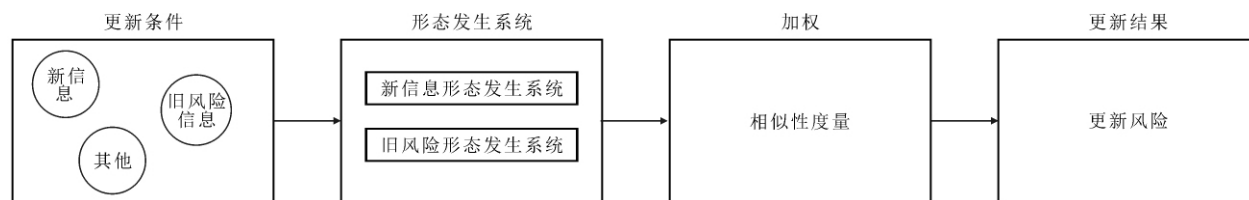


图4 风险更新算法流程

Fig. 4 Processes of risk updating algorithm

信息,采用 Fisher 判别分析法判别其风险严重程度,研究目的是在产生风险时给予驾驶人提醒,避免风险的进一步发展。

时间场景构建的目的是把风险研究在时间轴上“点”的形式扩展为“线”的形式,基于上述提到的未来数据收集方面的问题,本文提出采用假设法构建动态风险研究的时间场景。假定现在的时间是

2019 年 6 月 1 日,当前选取的研究路段上风险判别模型是依据 2019 年 1~4 月的事事故数据和交通流参数标定的,现在需要利用 5 月收集到的数据更新风险判别模型,同时,将 6 月的数据作为验证数据对风险模型进行精度验证。将 1~4 月的数据定义为过去数据,5 月的数据定义为现在数据,6 月的数据定义为未来数据,如图 5 所示。



图5 信息结构

Fig. 5 Information structure

2.2 数据收集

PeMS 系统通过在美国加州范围内所有高速公路上间隔 500~800 m 布设交通流检测线圈,收集全天不同时段交通流数据,从 PeMS 获取到的数据时间精度为 30 s,本文参照已有研究,以 5 min 为时间间隔处理原始数据^[29]。根据对风险场景的量化定义,收集了 I80-W 高速公路(事故数较高)2019 年 1~6 月 616 组事故数据(包含事故发生时间、位置、持续时间、伤亡等参数),按 5:1 的比例将数据分为训练样

本和验证样本,其中训练样本分为 1~4 月的 404 例旧数据模型事故数据和 5 月的 112 例新数据模型事故数据,验证样本包含 6 月的 100 例验证数据。并对事故发生前 5 min 以内事故点上游断面(距事故点 800 m 以内)的平均速度、流量、车道占有率、大型车比例以及 V_{HT} 等参数进行收集,见表 1,表 2。

上述参数中,将收集到的事故数据按照事故持续时间排序后,发现在 120 min 内事故的分布较为密集,见图 6,因此将事故后果分为 3 个类型:0 表示

表1 样本收集情况

Tab. 1 Samples collection

数据样本类型	数据时间范围	数据量/例	事故组与对照组比例
旧探测器数据样本——标定旧模型 R_{1-4}	2019 年 1~4 月	404	1:3(对照组与事故组时间地点均相同,但日期与事故组相邻且未发生交通事故)
新探测器数据样本——更新旧模型与标定新模型 R_5	2019 年 5 月	112	
验证数据	2019 年 6 月	100	

表2 单组数据形式

Tab. 2 Single-arm data form

事故 后果	事故信息			上游断面参数收集信息					
	发生时间	持续时间/ min	事故桩号 (Abs pm)	检测断面桩号 (Abs pm)	流量/ (veh · (5 min) ⁻¹)	车道占有 率/%	V_{HT} / (veh · h ⁻¹)	大车 比例/%	速度/ (1.609 km · h ⁻¹)
1	2019-01-31,15:36	111	4.1	3.3	388	24.2	45.8	2.6	26.1
0	2019-02-01,15:36				455	5.6	22.7	4.0	61.8

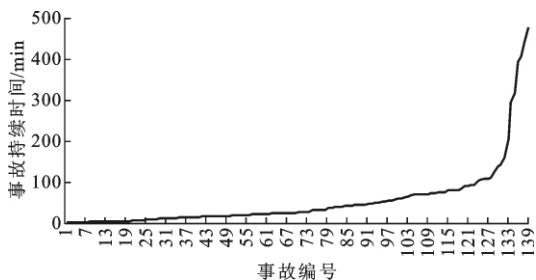


图 6 事故持续时间分布

Fig. 6 Distributions of accidents duration

不存在事故发生;1 表示事故持续时间在 120 min 以内;2 表示事故持续时间在 120 min 以上(因事故后果 2 数据量过少,表 2 未列出)。表 2 中流量为事故发生 5 min 内上游断面通过的车辆数;车道占有率为观测时间 5 min 内,车辆通过某一断面的累计时间所占单位观测时间的百分比;速度为 5 min 以内通过该断面所有车辆的平均速度。

2.3 模型计算与结果分析

本研究中将风险值定义为 R ,通过将实际条件下的交通流参数代入动态风险判别函数得到,作为本研究中评价高速公路动态风险的量化指标,依据图 7 所示计算步骤进行风险更新研究。

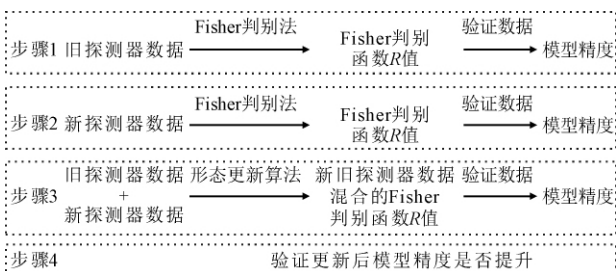


图 7 计算方法

Fig. 7 Method of calculation

2.3.1 Fisher 判别结果

利用 SPSS 中的判别分析功能对新旧样本分别进行 Fisher 判别后,得到未标准化典则判别函数系数,如表 3 所示。

表 3 利用新旧样本判别函数系数

Tab. 3 Discriminant function coefficients using new and old samples

数据	函数	常数	V	Q	p	q	V _{HT}
旧数据样本判别函数 R_{1-4}	1	-5.364	0.001	0.059	-0.073	0.065	0.082
	2	0.789	-0.003	0.114	0.018	-0.056	-0.018
新数据样本判别函数 R_5	1	-2.401	-0.014	0.565	-0.029	-0.034	0.086
	2	1.184	0.007	-0.129	0.056	0.065	-0.084

Fisher 判别会把空间高维数据投影到 2 条直线上,形成 2 个判别函数,而本文需要找到区分度更高,投影重叠面积更少的函数作为 Fisher 判别式系数。图 8、图 9 分别为旧数据样本判别函数 R_{1-4} 和新数据样本判别函数 R_5 的典则判别函数图。

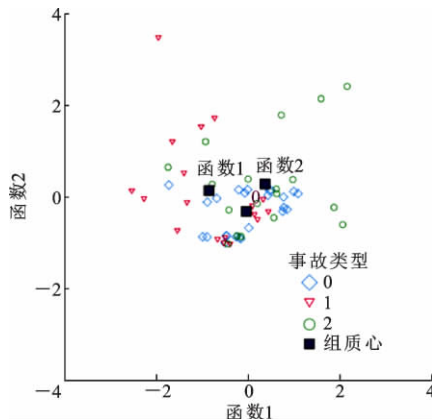


图 8 旧数据模型典则判别函数

Fig. 8 Canonical discriminant function of old data model

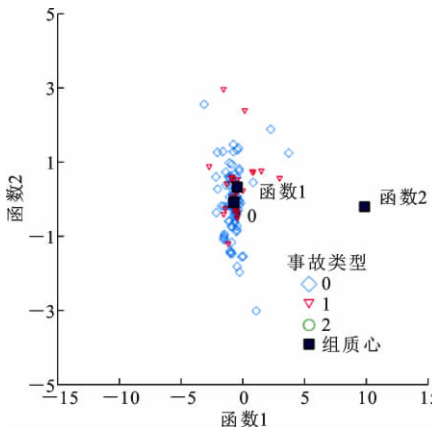


图 9 新数据模型典则判别函数

Fig. 9 Canonical discriminant function of new data model

从图 8 和图 9 可以看出:由旧数据标定的判别模型在函数 1 的方向上质间距更大,因此投影在函数 1 上的重叠面积更少,所以将函数 1 作为旧数据模型的判别函数;同理也选择函数 1 作为新数据模型的判别函数。可以得到函数关系式如下

$$R_{1-4} = -5.364 + 0.001Q + 0.059p - 0.073V_{HT} + 0.065q + 0.082V \quad (4)$$

$$R_5 = -2.401 - 0.014Q + 0.565p - 0.029V_{HT} - 0.034q + 0.086V \quad (5)$$

式中: R_{1-4} 为 2019 年 1~4 月数据的 Fisher 判别函数; R_5 为 2019 年 5 月数据的 Fisher 判别函数。

2.3.2 风险更新算法

旧探测器数据包含更多的风险结构信息,新探测器数据包含对风险的最新认识。理论上讲将两

者结合就可以获得更多对风险的认识,但并不是简单地将 2 组数据叠加。在旧的大样本数据中,也存在一些不能准确反映风险本质或者不符合当前风险状况的信息,本文需要利用新的样本数据完成除旧换新的过程,此即为应用风险更新算法的目的。

风险更新算法分为 4 步:①估计新探测器数据模型的分布,建立背景;②建立新探测器数据模型的形态发生系统,计算其内源值和夹角余弦 $\cos(\gamma_n)$,由内源值估计总体参数 θ_n ;③建立旧探测器数据模型的形态发生系统,计算其内源值 S_0 和夹角余弦 $\cos(\gamma_0)$,由内源值估计总体参数 θ_0 ;④令 $Z = \cos(\gamma_n)/(\cos(\gamma_n) + \cos(\gamma_0))$,由 $\theta_n = Z\theta_0 + (1-Z)\theta_0$ 计算更新参数。

本文通过建立交通流参数与事故持续时间之间的线性关系模型判别动态风险,因此通过建立线性背景来估计参数,由旧探测器数据模型建立的形态发生系统以 6 个变换值 y 为对象,总体构成了 7 维对象空间。以对应的风险值 R 为输入场,建立基本属性为: $x=1$ 和 $x=y_i, i=1,2,\dots,6$ 的 6 维空间,得到的形态信息见表 4。

表 4 形态信息
Tab. 4 Form information

信息	H_1	H_2	H_3	H_4	H_5	H_6	X
O_1	1	502	10.1	26.4	5.4	58.6	-0.92
O_2	1	511	8.7	24.4	5.3	64.0	-0.36
O_3	1	522	12.7	30.5	5.6	52.7	-1.51
O_4	1	427	18.5	3.4	1.2	26.6	-1.73
O_5	1	323	7.0	2.6	5.9	30.6	-1.85
O_6	1	451	22.1	4.6	2.4	24.7	-1.65

注: $H_1 \sim H_6$ 为 H 空间中的属性,其中 H_1 为 $x=1$ 时的系数, $H_2 \sim H_6$ 分别为流量 ($\text{veh} \cdot (5 \text{ min})^{-1}$)、车道占有率(%)、 V_{HT} ($\text{veh} \cdot \text{h}^{-1}$)、大型车比例(%)、速度 ($1.609 \text{ km} \cdot \text{h}^{-1}$); $O_1 \sim O_6$ 分别为 6 个月的故事; X 为输入场,即风险值 R 。

通过形态计算得到内源、投影以及夹角余弦,由写运算得外源值 S_E 为

$$S_E = H^T X =$$

$$\begin{bmatrix} -0.008 & -3.514 & -0.113 & -0.097 & -0.032 & -0.299 \end{bmatrix}^T$$

$$H = \begin{bmatrix} 1 & 502 & 10.1 & 26.4 & 5.4 & 58.6 \\ 1 & 511 & 8.7 & 24.4 & 5.3 & 64.0 \\ 1 & 522 & 12.7 & 30.5 & 5.6 & 52.7 \\ 1 & 427 & 18.5 & 3.4 & 1.2 & 26.6 \\ 1 & 323 & 7.0 & 2.6 & 5.9 & 30.6 \\ 1 & 451 & 22.1 & 4.6 & 2.4 & 24.7 \end{bmatrix}$$

$$X = \begin{bmatrix} -0.92 & -0.36 & -1.51 & -1.73 & -1.85 & -1.65 \end{bmatrix}^T \quad (6)$$

内源值 S_I 为

$$S_I = (H^T H)^{-1} S_E =$$

$$\begin{bmatrix} -6.130 & 0.005 & 0.030 & -0.086 & 0.080 & 0.073 \end{bmatrix}^T \quad (7)$$

由内源知 Fisher 判别函数参数 $\beta_i \sim \beta_6$ 为

$$\begin{cases} \beta_1 = -6.130 \\ \beta_2 = 0.005 \\ \beta_3 = 0.030 \\ \beta_4 = -0.086 \\ \beta_5 = 0.080 \\ \beta_6 = 0.073 \end{cases} \quad (8)$$

由读运算得投影 Y 为

$$Y = HS_I =$$

$$\begin{bmatrix} -0.92 & -0.36 & -1.51 & -1.73 & -1.85 & -1.65 \end{bmatrix}^T \quad (9)$$

其中夹角余弦 $\cos(\alpha) = XY / (\|X\| \|Y\|) = 1$, 估计误差的模为 $\|e\| = \|X - Y\| = 0$, 这说明输入场与背景完全匹配。

新探测器数据模型的形态发生过程有 2 个重要信息将会被应用到更新模型中:①对背景的估计;②新探测器数据模型的参数 $\theta_n = (0.443, -0.001, -0.007, 0.142, 0.210, -0.149)$ 和夹角余弦值 $\cos(\gamma_n) = 1$ 。

以夹角余弦作为加权依据进行更新, β_i 更新后参数 β_{ui} , 更新过程为

$$\beta_{ui} = \beta_i \frac{\cos(\gamma_0)}{\cos(\gamma_0) + \cos(\gamma_n)} +$$

$$\beta_n \frac{\cos(\gamma_n)}{\cos(\gamma_0) + \cos(\gamma_n)} \quad (10)$$

$$\begin{cases} \beta_{u1} = -2.843 \\ \beta_{u2} = 0.004 \\ \beta_{u3} = 0.015 \\ \beta_{u4} = 0.028 \\ \beta_{u5} = 0.145 \\ \beta_{u6} = -0.076 \end{cases} \quad (11)$$

更新后的风险值 R_u 为

$$R_u = \beta_{u1} + \beta_{u2}Q + \beta_{u3}p + \beta_{u4}V_{HT} + \beta_{u5}q + \beta_{u6}V =$$

$$-2.843 + 0.004Q + 0.015p +$$

$$0.028V_{HT} + 0.145q - 0.076V \quad (12)$$

利用更新后的风险表达式[式(12)]判别风险类型,还需要利用旧模型 R_{1-4} 和新模型 R_5 对其组质心进行更新,更新后结果为 $(\alpha_0, \alpha_1, \alpha_2) = (-0.045, 0.709, 3.925)$, 其中 $\alpha_0, \alpha_1, \alpha_2$ 分别为事故类型为 0、1、2 的组质心在函数 R_u 上的投影值。

2.3.3 精度验证和对比

对 R_{1-4} 、 R_5 、 R_u 这 3 种判别函数分别进行判别结果准确性检验,为了说明风险更新原理和算法的

有效性,在结果对比中加入了 R_{1-5} ,即将 1~4 月的数据样本和 5 月的数据样本直接混合后标定的判别函数。

检验过程为将检验样本代入到各组判别函数中,距离哪一类事故后果的质心距离最近,则被归为相应的类别。这里需要注意的是,在 Xu 等的研究中也做过类似的正确性检验,但其采用了同一时间段的数据分别去标定和检验模型^[9]。考虑到风险判别应用当前的观测去预测未来发生的事件,风险判别模型的检验数据和标定数据应有明确的时间界

限,使用未来的数据去检验当前模型的判别准确性更为合理。因此,本文利用 2019 年 6 月交通流数据对各组风险判别函数分别进行判别精度检验,结果如表 5 所示。

由表 5 可见,通过对旧判别函数 R_{1-4} 、新判别函数 R_5 、新旧混合判别函数 R_{1-5} 以及利用风险更新得到的判别函数 R_u ,采用统一样本进行精度检验后,判别精度分别为 67.5%、61%、66%、74%。这表明,更新后的风险判别模型对于 2019 年 6 月的交通流参数具有更高的预测精度。

表 5 判别精度检验

Tab. 5 Accuracy test

判别函数	R_{1-4}			R_5			R_{1-5}			R_u		
实际事故类型分类	0	1	2	0	1	2	0	1	2	0	1	2
正确预测样本数	104	22	9	104	19	8	103	20	9	114	24	10
总样本数	150	36	14	150	36	14	150	36	14	150	36	14
预测分类精度/%	69.3	61.1	64.3	69.3	52.8	57.1	68.7	55.6	64.3	76.0	66.6	71.4
合计/%	64.9			61.0			66.0			74.0		

3 结果与讨论

3.1 指标选取和数据精度

由于中国在高速公路路网数据信息采集技术方面尚未成熟,本文仅收集了美国加州 PeMS 交通管理系统中 I80-W 高速公路的相关数据。在建立本文的判别函数时,通过多因素 F 检验选取了相关性较强的一组参数,但是其中缺少类似于速度标准差、车头时距等较为普遍应用和认可的指标^[30-31]。因此,对比该高速公路实际事故数据,风险判别模型的预测精度仅维持在 70% 左右。

本文以 5 min 为时间间隔对原始交通流数据进行处理,研究事故发生前 5 min 的交通流状况和事故后果的关系,而对于车速较高的高速公路来说,车辆通过 2 个探测器之间仅需几秒,与事故相关性最强的数据应是事故前几秒的交通流参数。因此,数据的精度也是制约模型精确度的主要因素之一,但这类问题均会随着道路信息采集设备的完善和技术的发展逐渐得到解决。

3.2 模型建立

采用 Fisher 判别分析法建立交通流参数与事故持续时间之间的线性函数关系,并验证了判别函数检验的精确度。

但此结果并不代表交通流参数与事故后果间仅存在线性关系,目前很多学者采用随机森林方法,贝叶斯神经网络和机器学习等方法研究了此类的问题,

均取得了不错的预测效果^[1]。本文选取此方法的一个突破点在于判别分析的结果可以得到一个关于交通流参数的综合风险指数,该参数可用于风险更新算法中背景 H 的建立。

本文中提到的风险更新算法,其基础理论来源于形态发生更新。该理论和方法最大的作用是解决“缺失原始资料条件下如何用新观测资料直接对旧风险进行更新”这一科学问题,该方法对于新建公路或新建风险评价体系的公路具有重要意义。

3.3 结果分析

采用了同一组验证数据对 4 种不同参数和方法得到的判别函数精度进行了检验,而在检验时没有采用常规的 McNemar-Test 检验方法^[32]。因为研究假设目前的时间点为 2019 年 6 月 1 日,而采用的验证数据(整个 6 月的交通流数据)对应的事故后果是已知的,直接将判别函数的计算值与各事故组质心进行比对具有更高的精度检验效果。通过旧判别函数 R_{1-4} 、新判别函数 R_5 、新旧混合判别函数 R_{1-5} 和利用风险更新得到的判别函数 R_u ,分别得到了 67.5%、61%、66%、74% 的预测精度。由表 5 可知, R_{1-4} 准确率高于 R_5 ,旧探测器数据量更大,更能通过数据反映出风险的本质,旧探测器数据的样本量约为新探测器数据的 4 倍,但在判别准确率上仅比后者高出 6.5%。值得注意的是,道路交通环境是不断变化的,新探测器数据为更新后数据,更符合目前的道路风险状况。 R_{1-5} 精度略低于 R_{1-4} 。 R_{1-5} 虽然

具有更大的样本量,但 Fisher 判别样本在自变量的 8~10 倍之间时判别效果最为理想,更大的样本量并不能带来更好的预测精度; R_{1-5} 精度远低于更新后的判别函数 R_u ,说明样本的混合不仅是符合风险本质信息的叠加,也是干扰因素的叠加。

3.4 风险更新的时间尺度

以 1 个月为时间尺度,使用新收集的探测器数据对旧的模型进行更新,关于确定更新时机方面,其理论依据为数据收集的样本量,当新收集的样本组数达到自变量个数的 10 倍时,可以建立效果理想的判别函数,此时便可对原有判别模型进行更新。但在实际情况中,收集到 50 组样本的时间取决于数据收集方式和交通状况。本文以 1 个月为间隔是较为保守的估计,如果具有秒级精度的交通流参数数据,风险更新的尺度也可以缩减为 1 h 甚至更小。随着以后探测器布设密度和收集数据精度的提高,可以实现更短周期的风险更新。

3.5 方法适用性及优缺点分析

本文采用了美国 PeMS 数据库数据进行了该方法的可行性研究,因此也产生了难以避免的方法适用性问题,对此笔者做出如下解释:一项利用美国数据产生的研究成果应用于中国的高速公路时,由于中美两国的高速公路在线形设计特点,驾驶方式以及交通特点方面的不同,其具体的结论性成果不宜直接应用。但本文方法在利用美国 PeMS 数据进行验证后应用于中国高速公路,后续研究可借鉴本文方法,同时结合中国道路的数据资源和信息,形成直接应用于中国高速公路的理论成果。

以风险源为核心,对风险源进行深度分析后再选取完全适用于参数特点的风险判别模型,在数据资源稀缺的条件下,会受到数据库不适配或难以开展现场数据收集等制约。因此,本文以完善包括风险分析、风险判别和风险更新的整套风险研究体系为目的。在研究过程中也存在一些不足之处,包括风险源中的参数并不完全相互独立,部分参数间存在线性相关性且各自存在时空特性,后续研究中可采用非线性模型考虑参数时空特征的变化。

4 结 语

(1)通过风险分析、风险判别和风险更新,较为系统地阐述了高速公路动态风险理论体系。通过对公路动态风险赋予参数化定义,从风险分析的基本要素入手判别公路动态风险,利用不同交通流参数组合条件下的综合风险指数 R ,并引入了与动态风

险密切相关的风险更新理论和算法,得到更新后的风险判别函数 R_u ,其相较于函数 R_{1-4} 、 R_5 、 R_{1-5} 的判别精度分别提升了 6.5%、13%、8%。

(2)结合 2 种算法在高速公路动态风险研究,主要说明了如何结合新数据和旧数据对公路动态风险评价模型进行不断地更新,提高其判别精度。

(3)结合本文的研究内容和公路动态风险内涵,为今后高速公路动态风险研究提供依据:动态风险研究方面应对多参数共同作用下的事故发生的必然条件做出界定;指标选取方面,需要找出最适合于动态风险判别的一组参数;实时风险变化对于数据的连续性要求较高,不断提升数据采集的精度也十分必要。

参考文献:

References:

- [1] 贾丰源,孙 杰,孙 剑.快速路交通流运行安全关键参数识别与评估[J]. 同济大学学报(自然科学版), 2015,43(2):221-225,324.
JIA Feng-yuan, SUN Jie, SUN Jian. Key variables identification and proactive assessment of real-time traffic flow accident risk on urban expressway[J]. Journal of Tongji University (Natural Science), 2015, 43(2):221-225,324.
- [2] 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):873-879.
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):873-879.
- [3] LIN L, WANG Q, SADEK A W. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction[J]. Transportation Research Part C, 2015, 55:444-459.
- [4] 付存勇,王俊骅.基于监控数据的高速公路实时事故风险模型[J]. 交通信息与安全, 2017, 35(5):11-17,36.
FU Cun-yong, WANG Jun-hua. A real-time accident risk model on freeways based on monitoring data[J]. Journal of Transport Information and Safety, 2017, 35(5):11-17,36.
- [5] PARSA A B, TAGHIPOUR H, DERRIBLE S, et al. Real-time accident detection: Coping with imbalanced data[J]. Accident Analysis & Prevention, 2019, 129:

- 202-210.
- [6] SHEW C, PANDE A, NUWORSOO C. Transferability and robustness of real-time freeway crash risk assessment[J]. *Journal of Safety Research*, 2013, 46: 83-90.
- [7] 游锦明, 王俊骅, 唐 棠, 等. 基于支持向量机的高速公路实时事故风险研判[J]. *同济大学学报(自然科学版)*, 2017, 45(3): 355-361.
- YOU Jin-ming, WANG Jun-hua, TANG Tang, et al. Support vector machines approach for predicting real-time rear-end crash risk on freeways[J]. *Journal of Tongji University (Natural Science)*, 2017, 45(3): 355-361.
- [8] 游锦明, 方守恩, 张兰芳, 等. 高速公路实时事故风险研判模型及可移植性[J]. *同济大学学报(自然科学版)*, 2019, 47(3): 347-352.
- YOU Jin-ming, FANG Shou-en, ZHANG Lan-fang, et al. Real-time crash risk prediction models and transferability analysis on freeways[J]. *Journal of Tongji University (Natural Science)*, 2019, 47(3): 347-352.
- [9] XU C, LIU P, WANG W, et al. Development of a crash risk index to identify real time crash risks on freeways[J]. *KSCE Journal of Civil Engineering*, 2013, 17(7): 1788-1797.
- [10] 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, 43(1): 68-73.
- XU Cheng-cheng, 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): 68-73.
- [11] 张兰芳, 赵 焜. 考虑货车因素的高速公路短期交通流风险预测[J]. *同济大学学报(自然科学版)*, 2018, 46(2): 208-214.
- ZHANG Lan-fang, ZHAO Kun. Short-term traffic flow risk prediction on freeways based on truck factors[J]. *Journal of Tongji University (Natural Science)*, 2018, 46(2): 208-214.
- [12] 胡功宏, 林 雨, 高建平. 高速公路交通流实时安全性评价[J]. *安全与环境学报*, 2015, 15(1): 57-63.
- HU Gong-hong, LIN Yu, GAO Jian-ping. Real-time safety probability evaluation of the expressway traffic flow[J]. *Journal of Safety and Environment*, 2015, 15(1): 57-63.
- [13] LI M, LIU Z, LI X, et al. Dynamic risk assessment in health care based on Bayesian approach[J]. *Reliability Engineering and System Safety*, 2019, 189: 327-334.
- [14] LEITE F L, SCHNEIDER D, ADLER R. Dynamic risk management for cooperative autonomous medical cyber-physical systems[C]//GALLINA B. *Proceedings of International Conference on Computer Safety, Reliability and Security*. Cham: Springer, 2018: 126-138.
- [15] ZHANG Q, CUI L, ZHANG J, et al. Grid based dynamic risk assessment for grassland fire disaster in Hulunbuir[J]. *Stochastic Environmental Research and Risk Assessment*, 2015, 29: 589-598.
- [16] 张明媛, 袁永博, 周 晶. 城市自然灾害风险分析新方法[J]. *大连理工大学学报*, 2010, 50(5): 706-711.
- ZHANG Ming-yuan, YUAN Yong-bo, ZHOU Jing. A new method of urban natural disaster risk analysis[J]. *Journal of Dalian University of Technology*, 2010, 50(5): 706-711.
- [17] HARTIGAN A, THRASHER D, ADLAM R. Real time decision making-incorporating dynamic risk management[C]//Society of Petroleum Engineers. *Proceedings of Intelligent Energy International Conference and Exhibition*. Aberdeen: SPE program committee, 2016: 1-11.
- [18] 黄崇福. 风险分析基本方法探讨[J]. *自然灾害学报*, 2011, 20(5): 1-10.
- HUANG Chong-fu. Discussion on basic methods for risk analysis[J]. *Journal of Natural Disasters*, 2011, 20(5): 1-10.
- [19] 黄崇福. 自然灾害动态风险分析的一个形式化模型[J]. *灾害学*, 2015, 30(3): 1-9, 53.
- HUANG Chong-fu. A formal model for dynamic risk analysis in natural disasters[J]. *Journal of Catastrophology*, 2015, 30(3): 1-9, 53.
- [20] THOMPSON K M, DEISLER P F, SCHWING R C. Interdisciplinary vision: The first 25 years of the society for risk analysis (SRA), 1980-2005 [J]. *Risk Analysis*, 2005, 25(6): 1333-1386.
- [21] 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, 42(3): 555-559.
- XU Cheng-cheng, 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, 42(3): 555-559.
- [22] 田 静. 形态发生估计及其在概率风险更新中的应用[D]. 北京, 北京师范大学, 2012.
- TIAN Jing. Morphogenetic estimation and its application in updating probabilistic risk[D]. Beijing: Beijing Normal University, 2012.
- [23] 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, 2005(1908): 51-58.
- [24] ABDEL-ATY M, PANDE A. Identifying crash propensity using specific traffic speed conditions[J]. Journal of Safety Research, 2005, 36: 97-108.
- [25] SUN W, XIE S, HAN N. Robust discriminant analysis with adaptive locality preserving[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(10): 2791-2804.
- [26] ABDULLAH A S, SELVAKUMAR S. Assessment of the risk factors for type II diabetes using an improved combination of particle swarm optimization and decision trees by evaluation with Fisher's linear discriminant analysis[J]. Soft Computing, 2019, 23(20): 9995-10017.
- [27] ZHENG J, WANG H, SONG Z, et al. Ensemble semi-supervised Fisher discriminant analysis model for fault classification in industrial processes[J]. ISA Transactions, 2019, 92: 109-117.
- [28] ZHONG K, HAN M, QIU T, et al. Fault diagnosis of complex processes using sparse kernel local Fisher discriminant analysis[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(5): 1581-1591.
- [29] 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): 358-363.
- [30] ABDEL-ATY M, UDDIN N, ABDALLA F, et al. Predicting freeway crashes based on loop detector data using matched case-control logistic regression[J]. Transportation Research Record, 2004(1897): 88-95.
- [31] ZHENG Z, AHN S, MONSERE C M. Impact of traffic oscillations on freeway crash occurrences[J]. Accident Analysis & Prevention, 2010, 42(2): 626-636.
- [32] LENSBERG T, EILIFSEN A, MCKEE T E. Bankruptcy theory development and classification via genetic programming[J]. European Journal of Operational Research, 2006, 169(2): 677-697.