

Demonstration of A Vehicle Risk Mode Based on On-board Recorder Data

Jinke Liu Xingyu Xiao Zhengquan Xiao

Shenzhen E & D Technology Limited, Shenzhen, Guangdong, 518000, China

Abstract

With the continuous progress of computer technology, more and more on-mounted in-vehicle data recorders (IVDR) have been installed on cars. While obtaining more real-time vehicle operation data, it also provides a richer data basis for the analysis of the causes of automobile accidents. Therefore, this paper will analyze the data of automobile recorder, and combine the driving conditions of various cars and environmental factors, on this basis, proposed a new traffic accident risk factor analysis method. Based on massive information of travel and geographical location, the correlation between traffic risk and accidents, and test the conclusions obtained. After research, we found the influence relationship between vehicle risk factors and accidents, which is specifically manifested as the non-linear relationship between mileage and the probability of accidents. It is hoped that this study can provide some reference value for vehicle risk quantification, transportation planning, personalized auto insurance pricing and other aspects.

Keywords

vehicle recorder data; vehicle risk model; empirical

基于车载记录仪数据的车辆风险模型实证

刘金科 肖星宇 肖政权

深圳市意安迪电子有限公司, 中国 · 广东 深圳 518000

摘要

伴随着计算机技术的不断进步,越来越多的车载记录仪(in-vehicle data recorders,简称IVDR)被安装到了汽车上。在获得了更多的实时车辆运行数据的同时,也为汽车事故的成因的分析提供了更为丰富的数据依据。因此,论文将对汽车记录仪的数据进行分析,并将各种汽车的行驶状况和环境因素相结合,在此基础上,提出了一种新的交通事故风险因子分析方法。论文拟基于海量的出行、地理位置等信息,通过建立多元线性回归模型,分析交通风险与事故的相关性,并对所得结论进行实证检验。经过研究,我们发现了车辆风险因素与事故之间的影响关系,具体表现为:行驶里程与发生事故的概率存在非线性关系等。希望通过本研究,可以为车辆风险量化、交通出行规划、车险个性化定价等方面,提供一定的参考价值。

关键词

车载记录仪数据; 车辆风险模型; 实证

1 引言

当前,以物联网技术为基础的车载记录仪,被越来越多地运用到了车辆的安全驾驶分析上,这为开展车辆危险因素的识别和事故调查,提供了充足的数据支撑。早在 1976 年,外形笨重的驾驶数据记录仪器(磁带机)就已经存在。然而,由于受到仪器设备和分析工具的限制,对驾驶数据的研究和应用造成了很大的限制,直到最近几年,才有了较大的突破。随着现代信息技术的飞速发展,现在,我们可以制造出更好的性能,更小的外形,更适合于各种类型的车辆。与此同时,为进行科学研究提供了大量的与车辆驾驶有关的

数据。与传统的人工记录数据相比,IVDR 所提供的数据具有更高的有效性和实用性。除此之外,IVDR 还可以对一部分人的驾驶行为进行模拟,这就提高了通过总体样本来对车辆危险因素进行分析研究的准确率^[1]。研究人员还将卫星定位和平均车速加入 IVDR 装置中,以使其更符合实际司机的驾驶习惯。新的特性,如视频捕捉等。国外学者 Desyllas 与 Sako 在 2012 年,通过对大量传统文本记录的保险数据的分析,发现行驶里程与事故发生之间存在着正向的关联关系。随着研究数据的不断丰富,本研究从车载记录仪的数据分类汇总入手,对不同类型的车辆风险因素进行了分析比较,并采用多元回归模型,对车辆风险因素与事故进行了相关性研究。最后,将其与实际的出险数据进行比较,以证实上述结论的正确性,并对今后的研究提出一些看法和建议^[2]。

【作者简介】刘金科(1992-),男,中国湖南邵阳人,本科,从事数码电子产品研究。

2 相关研究

在此之前,由于受到数据技术处理的限制,再加上数据收集的复杂性,有关产业组织并未给予足够的重视和重视,对于 IVDR 数据的应用已经进行了数十年的研究,却始终没有什么突破性的进展,直至近年来,随着物联网技术在中国的迅速发展和应用,才有了很大的突破。在对中国这方面的研究进行梳理的过程中,我们可以看到,近几年,IVDR 资料的样本量较大,有较好的应用前景。由于现有的相关理论研究大都是从英文文献中获得,因此论文在变量的选择、指标的设定等方面也是从英文文献中获得的。其中,Jolliffe 等人在 2007 年利用一个基本的线性回归模型,对年度总里程和发生保险损失车辆之间的函数关系进行了回归分析,得到的拟合优度指标较好($R^2=0.82$)。在该研究中,除了使用年度总里程之外,并没有对其他变量进行研究,也没有对样本进行详细分类与描述。尽管这项研究没有给车辆保险产品的创新与个性化带来了新的机遇。

2.1 行车里程和车速对汽车危险的影响

“自然驾驶”也被称作“尽可能接近”与“真实”的驾驶数据,它指的是一种以行车数据为基础的、以人们每天的行车习惯为基础的驾驶方式和行为模式的研究。目前已发表的一项研究结果,来自美国高速公路安全局和弗吉尼亚理工大学的交通学院,从 2001 年到 2006 年,对 100 辆汽车进行了为期 5 年的行车数据的跟踪分析。清洗,为研究提供了史无前例的详细测量。Helander 等人在 2010 年对车辆行驶里程和车辆排放等外部性风险变量展开了描述,最后得到了车辆行驶里程和自身性能对于事故风险的正向显著影响,主要分析了速度因素对车辆风险提高正相关,最后得到的结论是驾驶环境对车辆风险的影响明显。论文还提出了行车车速对交通事故危险性的一个重要预测和评价指标^[9]。

2.2 行车环境对交通工具危险性的影响

在对车载记录仪中 GPS 日志信息进行分析的过程中,就可以得到车辆的行为偏向,具体包括了驾驶者的兴趣地点、驾驶习惯、监视信息的实时或历史等。根据行驶地理位置、车主个人信息等隐私数据的暴露水平,结合车载传感装置(包括光照传感器、天气传感器、速度传感器等)、GPS 等设备(地理位置信息等),监测车辆的仪表盘信息、驾驶员的驾驶模式(包括驾驶时段、紧急刹车、突然加速等)等空间信息,在不同路况条件下,可以有效地分析可能产生的危害因素。2010 年,Jun 等通过对亚特兰大公共汽车 IVDR 数据的分析,通过对 167 台安装 IVDR 的车辆的 IVDR 数据的比较,得到了一些有用的信息。通过对两个样本集(高速公路、主干道和城市道路)、行驶时间(行驶时间)下的交通事故的数据进行对比分析,揭示了车速对事故发生风险的影响规律。中国吴义和宁洪等人于 2014 年强调驾驶环境和驾驶行为是事故发生的主要原因,并提出了以车速付费的车

险新模式。

总而言之,汽车风险因素的分析已经逐步从事前的因素转向事后的因素,从静态的风险因素转向动态的风险因素,国内外的学者已经从多个角度以及各种风险因素展开了相应的研究。论文试图利用海量的车载数据、地理位置等信息,加入更多与汽车风险相关的因素,构建多元回归模型,对所得的结果进行分析研究,并对所得的结果进行实证检验。本项目的研究,不但在方法上有较大的创新,而且有望实现对车辆风险的定量分析,并为未来车辆风险的辨识、车险个性化定价等提供借鉴^[4]。

3 资料搜集和整理

本研究采用的是某保险公司 PSYD 类型车辆保险产品,对 IVDR 数据进行测试,在使用的过程中,会对涉及个人隐私的内容展开保护隐藏。当车辆在运行的时候,如果地理位置发生了变化,那么 IVDR 设备将会自动更新一次数据内容,如果地理位置不发生改变,或引擎熄火,则数据的更新将停止。将移动距离与移动时间相结合,求出移动车辆的平均车速和总行驶里程。与 GPS 电子地图的数据相结合,还能得到行驶区域和道路类型的数据。本项目将通过随机取样的方式,以病例-对照的方式,抽取 IVDR 数据库中的研究数据。随机选择了在 2012 年,发生了交通事故的 600 辆肇事车辆,对它们进行了近六个月的 IVDR 数据的采集,并把它们分成了人身伤害事件,致死事件。根据 1/12 的分层抽样,在 50 个交通工具中,平均事故发生率更接近总体抽样的平均值^[9]。论文采用月平均数据进行抽样调查,避免了由于季节变化对产品质量的影响。在采样时,对不含地点信息的事数据进行了剔除。接下来,要做的,就是对开车习惯和交通事故的关系进行研究。参考组是以从 2012 年 7 月到 2012 年 12 月期间,1000 辆没有发生过交通事故的车辆作为参考组。在对数据进行处理的过程中,如果汽车被卷入其他交通事故,则排除在外。在数据记录的过程中出现错误,或者在存储的时候,不能生成日志的车辆数据,GPS 传感器长时间没响应,造成的车辆数据没有地理位置信息和数据更新,GPS 数据不连续,车辆数据不明确,最后,这些研究车辆的数量出现了以下变化:肇事车辆样本组从 600 减少 17 辆,降到 583 辆(下降 2.9%),无肇事车辆样本组从 1000 减少 16 辆,降到 984(下降 1.6%)。由于样本较大,没有把离群点排除在外,所以,我们可以考虑把离群点的影响考虑在内。结果表明,该方法能够有效地提高车辆性能,提高车辆性能,提高车辆性能。其将不同碎片的数据内容整合到了单辆车的特征表现层级中。定义了一个从 N 到 M 的特征矩阵 E , n 行表示 1567 辆车中的第 n 辆车数据, m 列表示了不同行驶里程下的特征值。在矩阵 E 中,聚合算法将会自动地识别出在驾驶里程增加的时候,其他数据所对应的改变值,如图 1 所示。

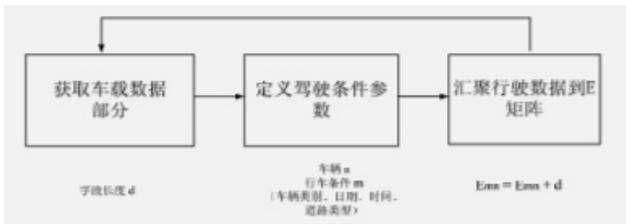


图 1 数据聚合过程

4 建模和修正

在此基础上，采用多变量 Logistic 回归模型，对实验数据和实际事件的误差进行系统分析。

$$P(\text{casdx}) = 1 - P(+x) = 1 + e^8 \quad (1)$$

采用最大似然方法，以风作为截短系数，以 A 系数作为预测概率与实际观测数据的误差（0~1）。其中重点介绍了如下几个方面的内容：首先，选取合适的约束条件，然后对矩阵 E 的确定进行了分析。

- ①一天中几点钟；
- ②一星期中哪个日子；
- ③开车时最多走哪条路；
- ④平均行车车速；
- ⑤在影响因子总数中，该模式的适用性如何。

将不同行车状况下的数据进行归类汇总，需对一段连续的离散时间、速度变量进行研究。在时间变量的处理上，以一小时为一次，以一天为一段；公路类型可划分为城市道路、城郊道路、高速公路 3 类；每增加 30km/h 为一段，发展了 120km/h 和 120km/h 以上的区间。若将这些条件综合考虑，则 E 矩阵中的组合条件可达到 2520 个之多。然而，由于某些组合变量无法达到模型化的要求，论文对这些组合变量做了适当的调整，使模型中各栏的数量降至 39 个，如表 1 所示。

表 1 矩阵 E 的聚合结构分析表

Time of day	Day of week	Road type	Velocity interval	Σ
24 类	7 类	3 类	5 类	39 类

对于一个初始模型，我们将整个变量集包含在风险矩阵 E 中。我们预期模型中可能会包含 30~40 个变量，这些变量包括：每天的时间间隔、每周的日期、月平均里程，道路类型（城市，市郊，高速公路），车速间隔 0~30km/h，60~90m/h 等。然而，该模型有很大的误差（如时间变量 >2800），这很有可能与其高的共线性有关。为此，论文提出了一种新的改进方法，即融合多个邻近特征，以减少该方法所引起的误差。

5 结语

论文以构建多元回归的模型来对 IVDR 数据进行处理，对车辆事故和危险因素之间的联系进行了分析。在 1567 辆行驶地理位置发生改变的车辆 IDVR 数据中，在此基础上，本项目拟采用主成分分析法，优化多项共线性问题以减少影响因子；在此基础上，通过个人比较研究，区别不同类型的肇事车与非肇事车的差异，探索时间点、日期、路况类型、车速等多种可变程度下，基于多元回归模型得到的结果。可以看到各风险因素与环境变量对风险事故存在的显著影响。此外，当车辆行驶的其他条件固定不变时，在行车里程和风险事故之间，存在着一种偏离线性函数的关系。论文还对技术应用这一部分的潜在应用可能性进行了进一步的探讨。

参考文献

- [1] Huang H. META: A mobility model of METropolitan Taxis extracted from GPS traces. Proc. Of the China IEEE.2010.
- [2] Bolderdijk JW. Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: Results of a Dutchfield experiment. Accid. Anal. Prev. 2011, 43 (3): 1181-1186.
- [3] 华荣晖.美国强制车险费率制度的特点与启示——以马萨诸塞州为例[J].上海金融学院学报,2009(1).
- [4] 黄永波.车险费率市场化改革及中小保险公司应对策略[J].金融会计,2011(2).
- [5] 吴义,宁洪.车载导航大数据在车险行业的应用[J].计算机光盘软件与应用,2014(7):23-24+27.