

基于 GPS 轨迹数据实现营运客车风险驾驶行为分析

梁珂宪

北京工业大学 北京 100124

[摘要]随着互联网技术的发展和智能手机的普及，越来越多的人开始利用手机记录生活中的点点滴滴，尤其是对于营运客车驾驶人员来说，通过 GPS 轨迹数据来记录营运客车驾驶员的行为特征，形成一种以行为特征为基础，以行为特征来识别风险驾驶行为的方法。

[关键词]GPS；数据；分析

[中图分类号] G641 **[文献标识码]** A **[文章编号]** 1647-9325 (2023)-0065-07 **[收稿日期]** 2022-07-19

首先根据营运客车驾驶员 GPS 轨迹数据特征提取出特定驾驶员风险驾驶行为特征，然后利用机器学习算法建立风险驾驶行为模型进行分析。该模型能够很好地识别营运客车驾驶员在道路上行驶过程中的驾驶行为风险，并可以通过建立风险指标来有效评估交通事故发生概率和严重程度。研究数据来源为吉林省境内某客运公司提供的 GPS 轨迹数据。根据吉林交通安全研究所对营运客车 GPS 轨迹数据处理结果分析得出以下结论：1）营运客车驾驶员在道路上行驶时存在危险驾驶行为，有较高的道路交通安全事故发生率；2）危险驾驶行为是营运客车驾驶员不良驾驶行为的集中体现，从时间分布上来看，主要集中在凌晨、傍晚以及夜间；3）危险驾驶行为发生在城市道路、高速公路和乡村道路等不同道路环境下。

1. 轨迹数据预处理

车辆 GPS 轨迹数据是一种特殊的时空序列数据，在构建风险驾驶行为模型时，需要对 GPS 轨迹数据进行预处理。GPS 轨迹数

据预处理的主要目的是去除非运动状态数据、补充缺失的轨迹点信息，以及对数据进行离群点剔除等。利用交通事故发生率等数据进行训练建立风险驾驶行为模型时，需要将非运动状态下的异常点剔除，即剔除出现非运动状态的车辆，避免其对模型的准确性造成影响。GPS 轨迹数据预处理方法主要包括：（1）对缺失值进行填充；（2）对非运动状态下的异常点进行剔除；（3）对数据进行清洗。其中，缺失值填充方法有两种，分别是分段填充法和累加填充法。分段填充法可以将所有时间点的轨迹都填充完整；累加填充法可以将部分时间点的轨迹按照一定的比例进行累加，使其形成完整的连续轨迹。本文采用分段填充法对缺失值进行填充；（4）对于非运动状态下异常点，需考虑该异常点与正常点之间是否存在时间关系或空间关系等因素，判断其是否是由于该异常点导致的非运动状态下数据异常。

2. 数据特征分析

根据吉林交通安全研究所对营运客车

GPS 轨迹数据处理结果,提取出营运客车驾驶员在道路上行驶时存在的危险驾驶行为特征,主要包括以下三个方面。(1)驾驶员存在危险驾驶行为是导致交通事故的主要原因。通过对驾驶员危险驾驶行为特征进行分析,可以有效识别出危险驾驶行为,从而可以更好地预防和控制交通事故的发生,保障车辆道路运输安全。(2)驾驶员危险驾驶行为具有时间特性。通过对营运客车驾驶员 GPS 轨迹数据分析,可以发现驾驶员在道路上行驶时存在以下两个方面的特征:①驾驶员行驶速度较快;②驾驶时间较长。营运客车驾驶员存在的两个方面特征都说明了其存在危险驾驶行为的可能性较大,并且根据事故调查数据分析发现:在发生道路交通安全事故的驾驶员中存在非常高比例的营运客车驾驶员存在危险驾驶行为。

3. 风险驾驶行为的识别

风险驾驶行为是指驾驶员在道路上行驶时存在的交通安全风险。对于营运客车驾驶员来说,危险驾驶行为主要包括闯红灯、超速行驶、进入高速公路行驶等。从时间分布上来看,这些行为发生的时间通常是在夜间或者凌晨,这说明这些行为是由不良的驾驶习惯所导致的。根据这些行为发生的时间段,可以将风险驾驶行为分为“危险驾驶”和“不危险驾驶”两种类型。以闯红灯为例,从时间分布上来看,这些行为一般在清晨、傍晚、夜间和凌晨发生较为频繁,主要是由于这些时间段没有足够多的交通参与者进行活动,因此在该时间段发生闯红灯等危

险驾驶行为的可能性较大。如图 3 所示为某客运公司提供的 GPS 数据处理结果。从时间分布上来看,该客运公司所属路段中存在较多闯红灯的行为发生。从不同时段发生的事件数量来看,分别出现了早高峰和晚高峰以及中午和下午的两个高峰时段。

4. 风险评价指标构建

根据营运客车驾驶员在道路上行驶时的行为特征,并结合营运客车驾驶员的基本信息,我们构建了营运客车驾驶员风险评价指标体系,如图 1 所示。选取风险驾驶行为特征,建立营运客车驾驶员风险评价模型。将营运客车驾驶员的基本信息作为输入,选取各道路环境下危险驾驶行为特征作为输出,风险评价指标作为输出。具体如下:其中 C、D、E、F 分别代表城市道路、高速公路、乡村道路和其他道路环境,P 代表危险驾驶行为在上述 4 种道路环境中的概率值。在城市道路下的风险评价指标为 $C=2.30+0.43+0.18+0.11+0.14$;在高速公路下的风险评价指标为 $C=2.24+1.11+0.04+0.06=4.45$;在乡村道路下的风险评价指标为 $C=2.15+0.24+0.03=4.56$;在其他道路环境下的风险评价指标为 $C=2.11+0.04=4.60$ 。其中,模型参数 β_1 、 β_3 分别表示营运客车驾驶员的风险驾驶行为特征以及道路环境下危险驾驶行为特征。

5. 基于机器学习算法的风险驾驶行为识别模型建立

目前用于机器学习的算法包括随机森林

算法、梯度提升决策树算法、支持向量机等,随机森林算法是最早用于分类识别的机器学习模型之一,其原理是对数据进行多维空间划分,再根据划分后的空间信息构建决策树模型来实现分类识别。梯度提升决策树算法则是根据训练样本集构造的一个逻辑树来实现对样本进行分类,其原理是先通过对每个样本进行特征选择,再通过不断的学习来逐渐学习出一个最优解。机器学习模型需要大量的样本数据进行训练才能提高预测精度和泛化能力。针对营运客车驾驶员驾驶行为风险识别,选择随机森林算法进行训练并建立模型,对营运客车驾驶员驾驶行为数据进行训练和识别。首先通过对随机挑选出的训练样本集进行随机森林机器学习,得到分类结果如图 2 所示,在模型建立过程中需要先将样本按照出现频率由高到低进行排序,然后依次对这些样本集中进行训练建立风险驾驶行为分类模型。

6. 风险评估应用

为了更好地评价风险驾驶行为对交通事故的影响,利用风险驾驶行为评估模型,首先对营运客车驾驶员的风险驾驶行为进行分析,将营运客车驾驶员的风险驾驶行为分为 1—5 分四个等级;然后对每个风险等级的车辆风险水平进行计算,以某一年发生的所

有事故数作为基础,将营运客车驾驶员的行驶行为分为驾驶员驾驶能力、行驶速度、危险因素三个方面分别进行计算,然后将所有计算结果相乘得到交通事故发生概率;最后通过与实际发生交通事故进行对比,计算出营运客车驾驶员风险驾驶行为对交通事故发生概率的影响程度。通过计算得出每辆车驾驶员的平均行驶速度(km/h)、平均危险因素(道路状况、天气因素、车辆技术)三个方面分别为 0.29、0.21 和 0.14,可见,平均危险因素对交通事故的影响较大。对于一辆营运客车驾驶员来说,当其存在较高的平均危险因素时,即有较高的交通事故发生概率,因此应该对营运客车驾驶员进行严格管理以降低交通事故发生率。同样地,当其存在较低的平均风险因素时则可以认为其不存在较高的交通事故发生概率。

参考文献

- [1]陈明华.自动紧急刹车系统避撞策略的研究[D].2019.
- [2]刘康,陈智勇,黄庆鹏,等.基于 GPS 大数据的租车人群出行行为分析[J].电脑知识与技术.2020,(20).
- [3]陈东攀.基于北斗车辆定位数据的驾驶行为分析[D].2020.

Risk driving behavior analysis of operating passenger cars is realized
based on GPS trajectory data

Liang section Xi'an

Beijing University of Technology, Beijing 100124

Abstract: with the development of Internet technology and the popularity of

smart phones, more and more people began to use mobile phone record dribs and drabs of life, especially for operating bus drivers, through the GPS trajectory data to record the behavior characteristics of operating bus drivers, forming a based on the behavior characteristics, behavior characteristics to identify the risk driving behavior method.

Key words: GPS; data; analysis