第 18 卷　第 34 期　2018 年 12 月
1671—1815(2018)034-0208-07

科　学　技　术　与　工　程
Science Technology and Engineering

Vol. 18　No. 34　Dec. 2018
ⓒ　2018　Sci. Tech. Engrg.

引用格式:王丝丝, 张敬磊, 王建兴, 等. 基于改进群智能 *K*-means 优化算法的侵犯性驾驶行为异常点检测[J]. 科学技术与工程, 2018, 18(34): 208—214

Wang Sisi, Zhang Jinglei, Wang Jianxin, et al. Outlier detection of reckless driving behaviors based on improved swarm intelligent *K*-means algorithm[J]. Science Technology and Engineering, 2018, 18(34): 208—214

基于改进群智能 *K*-means 优化算法的侵犯性驾驶行为异常点检测

王丝丝[1]　张敬磊[1]*　王建兴[1]　陈　慈[1]　马春杰[2]

(山东理工大学交通与车辆工程学院[1],淄博 255049;北京工业大学城市交通学院[2],北京 100124)

摘　要　如何有效地检测侵犯性驾驶行为异常数据,避免它们对后续建模分析产生不利影响是研究的焦点。在综述侵犯性驾驶行为研究现状基础上,借助驾驶模拟仿真实验系统和 ErgoLAB 人因系统进行侵犯性驾驶实验;并通过不同传感器获取了驾驶行为人车环境高维数据,最后采用优化的 *K*-means 聚类算法对该高维数据进行聚类和异常值检测。结果表明,最佳聚类数为 2;并有效输出了异常点检测结果,为下一步的侵犯性驾驶行为定量分析提供了优质的数据保证。

关键词　侵犯性驾驶行为　群智能算法　*K*-means 聚类　最佳聚类数　异常点检测

中图法分类号　TP391.9　U471.1;　**文献标志码**　A

侵犯性驾驶行为有别于攻击性驾驶行为,两者均是汽车社会发展到一定阶段的产物。攻击性驾驶主要是指驾驶员因为愤怒或者急躁等原因而意图去造成他人心理或身体上伤害的任何驾驶行为,包括直接或间接的言语攻击(如发牢骚、骂脏话、闪前大灯、大声鸣笛等)行为。本文研究的侵犯性驾驶行为主要针对驾驶员操作车辆后的外在表现。其微观过程为车辆启动,驾驶员通过感知外界的道路、交通环境等外界信息;并对其进行综合分析然后决策;最后通过一系列的驾驶操作来改变车辆的运行状况。该行为从产生原因和结果来看,主要是指驾驶员因愤怒或急躁等恶性情绪意图去造成他人心理或身体上伤害的驾驶行为,在交通行为上其主要表现有强行变更车道、强行超车、阻止其他驾驶员超车、违法抢行、占道等,表现出一种带着“路怒症”心理疾病状态下的驾驶行为。

近年来,随着交通拥堵、生活节奏加快等现象越发突出,导致侵犯性驾驶行为的发生日益频繁,逐渐引起了国内外专家学者的重视。相对而言,国外对侵犯性驾驶的研究较早,研究内容主要集中于对侵犯性驾驶的影响因素、产生机制等方面;近年来逐渐有一些学者从车辆运行、地域文化差异等角度对其进行量化研究。例如,Shinar[1] 采用多因子分析方法研究了侵犯性驾驶行为的影响因素;Taubman[2] 通过对比分析驾驶员在不同情感诱导下的侵犯性驾驶意愿情况,表明积极情感能够削弱侵犯性驾驶意愿;并且男女之间在并无显著差异。Bumgarner 等[3] 以及 Danaf 等[4] 研究了愤怒情绪对侵犯性驾驶行为的影响。Bogdan 等[5] 利用元胞分析方法得出愤怒和侵犯性驾驶行为呈正相关关系,此外因驾驶员不同年龄、地区、驾驶经验、性格等这种关系程度会有所不同。Dahlen 等[6] 研究了驾驶员人格、侵犯性驾驶、交通事故之间的关系。结果表明,在驾驶过程中,容易发怒、不随和的驾驶员更容易产生侵犯性驾驶行为,引发交通事故。Wickens 等[7] 研究了性别对侵犯性驾驶与危险驾驶因素之间关系的影响。Schietekat 等[8] 研究了撒哈拉以南地区公共交通中的侵犯性驾驶行为,借助全球定位系统获得车辆的位置和速度,结合三轴加速度计数据,分析安全驾驶和侵犯性驾驶之间的差别,建立了可以判定侵犯性驾驶的车队管理系统。Zeeman 等[9] 基于道路设计标准、车辆运动学及乘客舒适度建立两个侵犯性驾驶行为检测模型。González 等[10] 对智能交通系统中的先进辅助驾驶系统进行了研究;基于具体的速度值、横向加速度值、纵向加速度值建立了检测侵犯性驾驶行为的分类器以便于检测侵犯性驾驶行为。中国的相关研究起步较晚,目前主要从攻击性驾驶

2018 年 7 月 28 日收到　　山东省自然科学基金(ZR2017LF015)、国家自然科学基金(61573009)和山东省高等学校科技计划(J15LB07)资助

第一作者简介:王丝丝(1992—),女,硕士研究生。研究方向:人-车-环境数据挖掘、驾驶行为及安全。E-mail: wangsis2012@163.com。

* 通信作者简介:张敬磊(1979—),男,山东高密人,副教授。研究方向:智能交通、驾驶行为及安全。E-mail:jinglei@sdut.edu.cn。

行为的产生原因、影响因素等方面进行研究，侵犯性驾驶研究目前较少。研究手段多采取量表测试、主观评价等。比如，李山虎[11]研究了攻击性驾驶的评价方法，用层次分析法等手段分析了车辆运动信息，得出当跟车时距小于 1 s 时会发生侵犯性驾驶。刘志强等[12]通过调查量表对攻击性驾驶的评价体系进行了研究。冯忠祥等[13]建立了愤怒驾驶行为选择多项 Logit 模型；并对其影响因素敏感度进行了分析。雷虎[14]对 456 名驾驶人的调查数据进行了因子分析，得到一份有 21 个项目的驾驶愤怒表现量表；并通过比较月收入不同的驾驶人驾驶愤怒表现得分平均数对量表的结构效度进行了检验。冯庆华[15]从青年性格特性对驾驶行为的影响、青年驾驶人驾驶特性与交通事故相关性入手；并结合青年驾驶人的弯道驾驶经验，分析并提出针对青年驾驶人驾驶特性的道路安全策略。此外，成波等[16]根据驾驶的操作特点提出驾驶状态-驾驶行为-驾驶模式 3 层结构，建立了激进驾驶状态识别算法。张琦等[17]对攻击性驾驶的区域差异进行了研究，研究内容主要集中在产生原因，影响因素，鲜有综合人车环境动态信息对这类危险驾驶行为的定量深入研究。

异常数据的检测，是数据挖掘和统计学方面经典的研究课题；其目的在于分离出数据中较大异于常规数据的内容，继而利用正常的数据进行建模分析以及决策。无线传感器由于其低成本、自组织及感知能力强等特点，已经成为医学、人因工程学等众多领域研究问题所需的重要工具，譬如检测驾驶员的各项生理数据，获取脉搏、血压、心率等指标，同时可能由于传感器节点故障等因素会产生一些对后续分析建模较大影响的异常数据。众所周知，高质量的实验数据是保证分析建模正确性的基础，故而本文重点研究针对这类异常数据的检测并剥离出这些数据。

故而本文综合驾驶过程中人车环境动态信息对侵犯性驾驶行为的影响，主要借助驾驶模拟仿真实验系统和 ErgoLAB 人因系统进行侵犯性驾驶实验，通过不同传感器获取了驾驶行为人车环境高维数据。然后通过改进的蜂群算法及 *K*-means 聚类优化算法对这些数据进行异常值检测，并采用人为加入异常点干扰的方法对模拟产生的数据集应用本算法以及基本 *K*-means 算法进行异常点检测，仿真对比结果以验证本算法的有效性。

1　异常值检测概述

无线传感器采集的异常值最早由 Hawkins[18]给出定义，它是指当节点采集的数据包含多个属性维度时，只要在某一个维度上表现为异常，该数据即为异常数据。

异常值产生原因，一方面可能由于传感器节点自身故障或能源耗尽等原因导致其内部结构发生变化，短时内不能人为地去还原或者补充传感器，进而导致产生异常数据；另一方面，当传感器所处环境发生骤变时，也会导致产生大量的异常数据[19]。传感器节点采集到的环境特征点集有别于一般的数据点分布模式，具有正常值数量多、分布范围小、重叠度高、异常值数量明显少于正常值、分布较分散，且离散度较大等特点，许多数据挖掘算法试图将异常点对最终模型的影响减到最小，或在预处理阶段去除它们。这一过程可以描述为从 n 个样本中选 k 个与其他数据显著不同、例外或不一致的样本。目前，异常值检测方法有多种，如基于统计的异常值检测、基于聚类的异常值检测、基于分类的异常值检测、基于邻近模型的异常值检测等[20]。

1.1　*K*-means 聚类算法

聚类分析用于发现局部强相关的对象组，而异常值的一个特点就是与其他对象强相关的对象，因此聚类分析常用来检测异常点。*K*-means 聚类作为一种经典的聚类算法，首先它可以聚类所有的对象，然后评估对象与其簇中心的距离来度量属于簇的程度，判断如果删除一个对象导致该目标显著改进，则可将该对象视为异常点。此外，在处理大规模数据集时该算法是相对可伸缩和高效率的[21]，它将各聚类子集内的全部数据均值作为该聚类的代表点，可以通过迭代使得评价聚类性能的测度函数达到最优从而把数据集划分为不同的类别。

该算法对异常数据检测的主要流程为通过 *K*-means 聚类算法对数据点进行分簇之后，计算各数据点与簇中心点的欧式距离，如大于当前的平均距离，那么将该数据归为疑似异常点；进而将该点到质心的距离与平均距离之差与该簇内全部数据点到质心距离标准差的 1.67 倍（90% 的置信区间）作比较，确定该点是否为异常数据。当计算与簇中心的距离时，按照样本与聚类中心的欧式距离最近的原则进行划分，计算如下：

$$d(\boldsymbol{P},\boldsymbol{Q}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_p - y_p)^2 + \cdots} \tag{1}$$

式(1)中 $\boldsymbol{P} = (x_1, x_2, \cdots, x_n)$、$\boldsymbol{Q} = (y_1, y_2, \cdots, y_n)$ 表示 $n(n \geqslant 2)$ 维空间中的任意两点（x、y 表示各维的坐标值）。需要注意的是，*K*-means 聚类算法对异常点的评估准则函数采用的是均方误差函数。当给定

一个样本数为 n 的数据集 $Y=(y_1, y_2, \cdots, y_n)$，将其划分成 k 类，即 $C_1, C_2, \cdots, C_k$，具体计算方式如下：

$$MSE = \frac{1}{n}\sum_{j=1}^{k}\sum_{y_i \in C_j}\|y_i - z_j\|^2 \tag{2}$$

式(2)中 z_j 表示聚类中心，目标函数值越小，相应的聚类效果越好。

在基本 *K*-means 聚类算法中“k”值是事先给定的，很多时候不知道数据集应该分成多少个类别才合适，目前很多学者[2,7—10]使用经验规则 $k_{opt} \leqslant k_{max} \leqslant \sqrt{n}$（其中 k_{opt} 表示最佳聚类数），并对其合理性进行了证明；也有学者提出一些检验聚类有效性的函数用于计算最佳聚类数 k_{opt}。采用基于聚类分布的有效性度量方法来评价不同聚类数下的聚类效果，以构造聚类评价函数用于寻找最佳聚类数，其定义如下：

$$\mathrm{var}(Y) = \sqrt{\frac{\sum_{i=1}^{n} d^2(y_i, \bar{y})}{n}} \tag{3}$$

式(3)中，$\bar{y} = \sum_{i=1}^{n}\frac{y_i}{n}$；$d(\,)$ 表示统计距离。

聚类密集型计算方法如下：

$$C_{var} = \sum_{i=1}^{k}\frac{\mathrm{var}(C_i)}{\mathrm{var}(Y)}k^{-1} \tag{4}$$

聚类邻近性定义如下：

$$P_r = \frac{1}{k(k-1)}\sum_{i=1}^{k}\sum_{i\neq j, j=1}^{k}\exp\left[-\frac{d^2(C_i, C_j)}{2\sigma^2}\right] \tag{5}$$

式(5)中 $d(C_i, C_j)$ 表示类 C_i 与 C_j 中心点的距离。

故而，聚类有效性指标 P_C 构造如下：

$$P_C = 1 - [\lambda C_{var} + (1-\lambda)P_r] \tag{6}$$

式(6)中 $\lambda \in [0,1]$，P_C 越大表示聚类效果越佳。

1.2 改进群智能 *K*-means 优化算法

基本 *K*-means 算法具有结构简单、收敛速度快等优点，但聚类结果极易受初始聚类中心的影响陷入局部最优解[22]。人工蜂群算法是一种模仿蜜蜂觅食行为的群智能算法，它作为一种随机寻优算法，具备优越的跳出局部解的性能且不受初始解干扰，可以优化聚类中心，继而克服该聚类算法的缺陷。但传统的蜂群算法具有搜索效率低，收敛速度慢的问题。因此，采用了改进的蜂群算法对 *K*-means 聚类优化。在改进算法中，首先利用粒子群算法中最优解指导搜索方向的思想，对蜂群的搜索方式做出改进，使蜜蜂沿着蜜源质量升高的方向进行搜索，增加寻找到全局最优解的概率；其次，针对领域搜索收敛速度慢的缺点，引入遗传算法中算术交叉操作，具体形式如下：

$$child = \mathrm{rand}\ parent1 + \mathrm{rand}\ parent2 \tag{7}$$

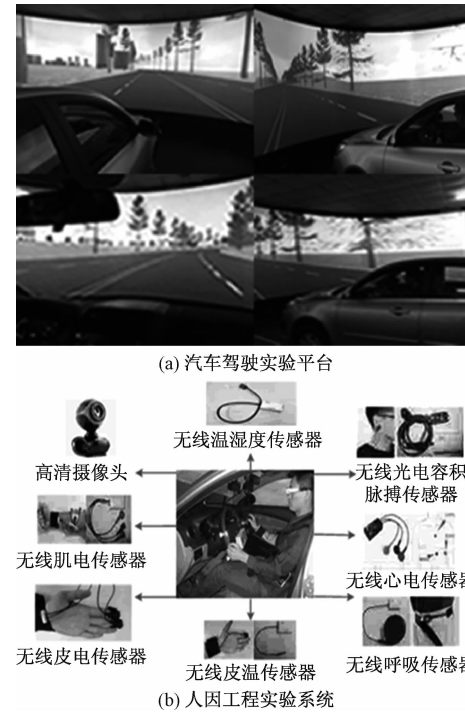
式(7)中，*parent* 表示已存在的解，*child* 表示新的解。对采集数据的所有维数同时进行这种搜索方式能够使得收敛速度更快。利用这样的改进蜂群算法优化 *K*-means 聚类，能够保证在整个搜索空间内，聚类目标函数尽可能小。

2 实验设计及指标标定

2.1 实验设计

2.1.1 实验设备

采用多人多车交互式汽车驾驶模拟器和人因工程实验系统组成的动态人车环境数据采集系统进行实验，所用的主要设备如图 1 所示。实验场景选为典型的城市道路，双向三车道，交通状况为自由流。



(a) 汽车驾驶实验平台

(b) 人因工程实验系统

图 1 人车环境动态数据采集系统

Fig. 1 Dynamic data acquisition system for human-vehicle-environment

2.1.2 实验对象

实验选择不同年龄、驾龄的驾驶员 30 名（18 名男性，12 名女性，平均驾龄在 3 年以上，驾驶里程在 20 000 km 以上），驾驶习惯良好，视觉机能正常，无生理缺陷。通过倾向性心理问卷设定各驾驶员驾驶倾向性类型并对其进行编号，且保证所有被试者没有睡眠不足、过度劳累等情况。实验测试时间在白天，室内温度、天气及道路状况均良好。

2.1.3 实验方案

实验主要基于汽车驾驶仿真实验平台进行，道路环境设定为城市某三车道路段，诱导机制采用周

围车辆反复别车、换道等方式使实验车发生侵犯性驾驶行为。在数据采集前，预先 10 min 用于驾驶员熟悉车辆及驾驶环境。在驾驶过程中，安排 2 名驾驶经验丰富的机动车司机进行跟车，通过主观评价法以 3 min 为时间节点连续标定驾驶行为并记录驾驶员驾驶状态。考虑驾驶状态变化具有连续性，实验过程每 3 min 形成一个样本，连续完成 10 个样本数据的采集。需要注意的是要保证跟车评价者驾驶习惯良好，有较强的观察能力，能够客观公正地评价驾驶行为，最好一年之内无任何交通违法。

2.2　指标标定

侵犯性驾驶行为的传感数据包括驾驶人驾驶过程中的生理-心理人因指标、车辆运行指标两大类数据。车辆运行数据反映车辆的运动状态，是从车辆运动的角度去量化驾驶行为的表现，包括车辆的速度、加速度、档位、油门、刹车、离合等指标[23,24]。驾驶人的生理-心理指标反映的是驾驶员的状态，研究的是从人因工程学角度量化驾驶人内在表现的微观状态，包括驾驶员的呼吸、脉搏、心电、皮电、肌电等，具体定义如下：d 为档位，无量纲；l、s、y 分别表示离合、刹车和油门，指踏板踩踏深度，范围是 0 ~ 1，其中 0 表示未踩，1 表示全部踩下；v 为速度，m/s；a 为加速度，m/s^2；z 为行驶里程；c 为车道线偏距，指车辆中心线与车道中线的横向偏距，由系统自动测量，m；q 表示前轮转角，是指车辆转向时，前轮转向的角度，左为负，右为正，(°)；b 为呼吸，rp；m 为脉搏，bp；x 为心电，μV；p 为皮电，μS；j 为肌电，μV。

2.3　数据采集

实验的传感数据采集处理包括驾驶员生理-心理状态的人因数据、车辆运行等共计 15 项指标，均通过传感器等输出在 PC 机上。人因数据借助 Ergo-LAB 同步平台通过各类便携式传感器（图 1）分时段、分事件场景获取脉搏、肌电、心电、皮电、呼吸等信息。实验记录了 30 位实验者的侵犯性驾驶人车环境指标数值集，分别记为数据集 1 ~ 30。需要注意的是，在异常值检测之前对原始数据采用了标准差标准化方法实现各指标间的无量纲化。计算过程为每一变量与其平均值之差除以该变量的标准差 σ_q，标准化后数据集的平均值为 0，标准差为 1，进而消除量纲和数量级的影响。这里以前轮转角为例，计算方法如下：

$$E_q = \frac{1}{n}\sum_{i=1}^{n} q_i \tag{8}$$

$$Z_q = \frac{q_i - E_q}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(q_i - E_q)^2}} \tag{9}$$

式中，E_q 表示前轮转角平均数，q_i 表示 i 时刻的前轮转角，Z_q 表示无量纲化的前轮转角；$\sigma_q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(q_i - E_q)^2}$。选择数据集 1 的二维可视化图像代替表格展示初始数据，这些数据无量纲化后截取的 60 s 片段数据如图 2 所示，其余研究对象进行同样处理。

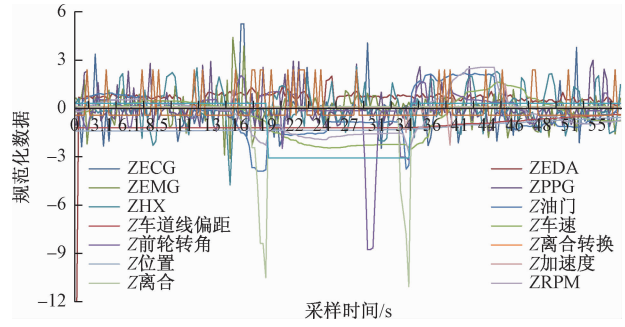


图 2　部分侵犯性驾驶行为采样情况

Fig. 2　Samples of some reckless driving behaviors

可以看出有些数据较大的偏离了常规运行情况，出现跳跃式增长，这些样本点极可能是异常值。

3　侵犯性驾驶行为异常值检测

3.1　改进群智能算法基本思想

K-means 作为典型基于距离的非层次聚类算法，首先依据给定的“k”值按照最小化误差函数的原则聚类所有的对象，并将其划分 k 个不同簇，然后评估对象与其聚类中心的距离来度量属于簇的程度，判断如果删除一个对象导致该目标显著改进从而区分异常点对象。因此聚类数和聚类中心的选择尤其重要。采用一种改进的群智能算法优化其聚类中心（基于遗传算法中的算术交叉操作与粒子群算法中最优解理论改进蜂群算法去优化聚类中心）。在此基础上，添加了聚类准则函数对聚类数进行评估求解，得到最佳聚类数。最后依据距离评估数传感数据是否为异常点。

3.2　主要步骤

算法的具体步骤如下。

（1）确定聚类范围。令初始聚类数 $k = 2$，聚类数 k 的搜索范围 $2 \leqslant k \leqslant \sqrt{n}$，设定聚类的最大阈值为 C_{max}。

（2）计算当前最优解。按照已确定的聚类数完成食物源初始化，计算该条件下的目标函数评估该点是否为异常点，并得出当下最优解 G_{best}。

（3）计算聚类有效性指标 P_C。按照改进的蜂群算法进行循环搜索，输出循环结束后的最优聚类结果，并按照聚类准则函数计算聚类有效性指标 P_C。

(4)令 $k = k + 1$，如若 $k < C_{max}$，则转到步骤(2)，否则转到下一步。

(5)比较聚类有效性指标 P_C，得出最佳的聚类数，并输出其聚类结果及异常点编号。

3.3 检测结果及分析

通过实验记录各位实验者的侵犯性驾驶行为人车环境指标数据，即数据集 1 ~ 数据集 30。每个数据集包括 3 min 的实验片段，平均 1 s 采集 3 个样本点，因此每个数据集包含 540 个样本点，共计 16 200 条。仿真分析过程中，设置最大迭代次数为 500 次。限于篇幅，选取数据集 1 的检测结果如下，最佳聚类数为 2，聚类记录见图 3 所示。

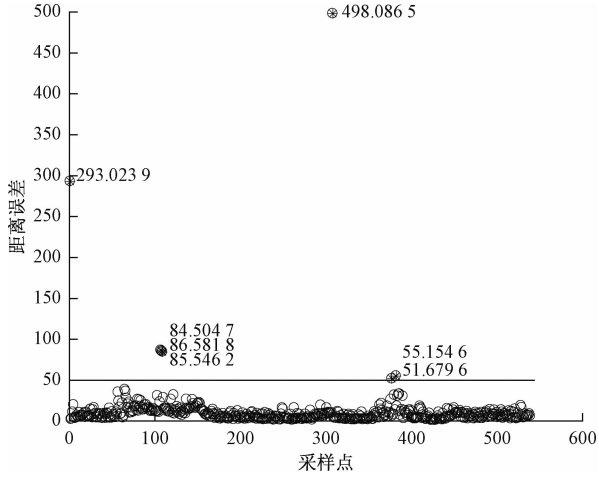


图 3 数据集 1 驾驶行为记录

Fig. 3 Data set 1 driving behaviors clustering

检测到数据集 1 的第 1、107、108、109、308、377、382 等采样点已经明显超过正常的容错范围，会对后续基于这些数据的建模结果产生较大影响，故而将其视为异常点剔除。

3.4 仿真验证

由于侵犯性驾驶行为的人车环境数据维数较高，不便于直观观察本算法的有效性。故而利用模拟的人工二维数据集以及人为加入异常点作为仿真数据源，并对比基本 *K*-means 聚类的异常点检测结果以验证算法的优越性。数据产生的方法是随机产生 900 个符合正态分布 $N(5,1)$ 的点，以及 900 个随机均匀分布($x^2 + y^2 \leqslant 1$)的点，并加入 10 个异常数在空白区域，数据集分布概率情况如图 4 所示。检测结果见表 1 所示。

从检测结果看，运用提出的优化群智能 *K*-means 聚类优化算法能够直接准确给出最佳聚类数为 8，且检测准确率为 100%，相对比基本 *K*-means 聚类算法，需要根据先验知识给定聚类数，由于已经由本算法得出最佳聚类数为8，因此这里分别将聚类数设为 6 ~ 9，检测结果如表 2 中所示，都不能准确地检测出异常点数据，且当聚类数为 8 时，异常点检测结果为 70%，进一步验证了本算法的优越性，同时验证了侵犯性驾驶异常数据点检测的有效性。

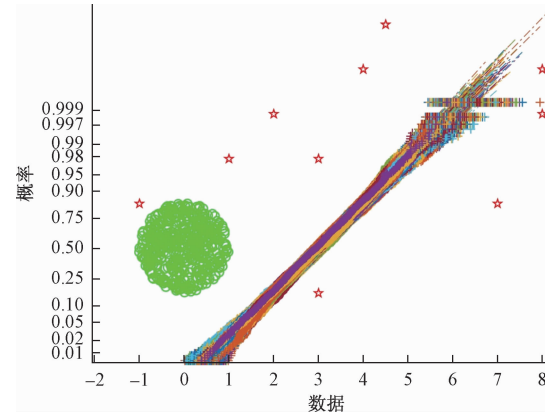


图 4 仿真模拟数据集

Fig. 4 Artificial simulation datasets

表 1 数据集异常点检测结果

Table 1 Outlier detection result of datasets

检测算法	*k* 值	异常点个数
本算法	8	10
K-means 算法	6	5
	7	13
	8	7
	9	12

4 结论

侵犯性驾驶行为作为一种由恶劣情绪引发的危险驾驶行为，近年来随着生活节奏的加快以及交通拥堵的日益突出，该行为引发的交通事故越发频繁，对侵犯性驾驶行为的研究越发引起重视。在综述侵犯性驾驶行为研究情况的基础上，综合考虑驾驶人生理-心理、车辆运行以及交通环境的影响，对侵犯性驾驶行为进行定量研究，借助汽车仿真驾驶模拟系统以及 ErgoLAB 人因系统等设备进行侵犯性驾驶实验，获取人车环境数据集。考虑异常数据对后续建模研究的影响，因此利用改进群智能的 *K*-means 聚类优化算法进而对实验数据集进行异常值检测。结果显示，数据最佳聚类数为 2，并析别出了与正常值相差较大的异常点数据，为下一步的分析建模消除了异常点的影响。

参 考 文 献

1 Shinar D. Aggressive driving: The contribution of the drivers and the situation. Transportation Research Part F: Traffic Psychology & Behaviour, 1998;1(2):137—160

2 Taubman B A O. The effects of positive emotion priming on self-re-

ported reckless driving. Accident Analysis & Prevention, 2012;45(2):718—725

3 Bumgarner D J, Webb J R, Dula C S. Forgiveness and adverse driving outcomes within the past five years: Driving anger, driving anger expression, and aggressive driving behaviors as mediators. Transportation Research Part F: Traffic Psychology & Behaviour, 2016; 42: 317—331

4 Danaf M, Abouzeid M, Kaysi I. Modeling anger and aggressive driving behavior in a dynamic choice—Latent variable model. Accident Analysis & Prevention, 2015;75:105—118

5 Bogdan S R, Măirean C, Havarneanu C E. A meta-analysis of the association between anger and aggressive driving. Transportation Research Part F Traffic Psychology & Behaviour, 2016; 42: 350—364

6 Dahlen E R, Edwards B D, Tubré T, *et al.* Taking a look behind the wheel: An investigation into the personality predictors of aggressive driving. Accident Analysis & Prevention, 2012; 45(1): 1—9

7 Wickens C M, Mann R E, Stoduto G, *et al.* Does gender moderate the relationship between driver aggression and its risk factors? Accident Analysis and Prevention, 2012; 45:10—18

8 Schietekat J M, Booysen M J. Detection of reckless driving in the sub-saharan informal public transportation system using acceleration-sensing telematics. EuroCon. Zagreb: IEEE, 2013:597—601

9 Zeeman A S, Booysen M J. Combining speed and acceleration to detect reckless driving in the informal public transport industry. 16th International IEEE Conference on Intelligent Transportation Systems. Hague: IEEE, 2013:756—761

10 González A B R, Wilby M R, Díaz J J V, *et al.* Modeling and detecting aggressiveness from driving signals. IEEE Transactions on Intelligent Transportation Systems, 2014;15(4):1419—1428

11 李山虎. 攻击性驾驶行为评价方法研究. 西安：长安大学, 2011
Li Shanhu. The research on evaluation method of aggressive driving. Xi'an: Chang'an University, 2011

12 刘志强，李云霞. 攻击性驾驶行为的 FA 和 AHP 评价系统研究. 重庆理工大学学报(自然科学), 2015; 29(11): 11—16
Liu Zhiqiang, Li Yunxia. Study of aggressive driving behavior evaluation system based on factor analysis and AHP. Journal of Chongqing Institute of Technology (Natural Science), 2015; 29(11): 11—16

13 冯忠祥，雷叶维，周穆雄，等. 修订版驾驶愤怒量表在中国驾驶人中的效信度分析. 重庆交通大学学报(自然科学版), 2016;35(4):141—145
Feng Zhongxiang, Lei Yewei, Zhou Muxiong, *et al.* Validity and reliability of the revised driving anger scale for Chinese drivers. Journal of Chongqing Jiaotong University (Natural Science), 2016; 35(4): 141—145

14 雷　虎. 愤怒情绪下的汽车驾驶行为特征及其对交通安全的影响研究. 武汉：武汉理工大学, 2011
Lei Hu. The characteristics of angry driving behaviors and its effects on traffic safety. Wuhan: Wuhan University of Technology, 2011

15 冯庆华. 青年驾驶人攻击性驾驶行为分析. 吉林交通科技, 2013;(3):45—47
Feng Qinghua. Analysis of aggressive driving behaviors of young drivers. Jilin Science & Technology of Communications, 2013;(3): 45—47

16 成　波，徐少兵，王文军，等. 激进驾驶状态识别算法研究. 汽车工程, 2014;35(7):785—789
Cheng Bo, Xu Shaobing, Wang Wenjun, *et al.* A study on the recognition algorithm of aggressive driving state. Automotive Engineering, 2014; 35(7): 785—789

17 张　琦，吴超仲，张　晖. 我国攻击性驾驶行为的区域差异研究. 交通信息与安全, 2016;(6): 914—929
Zhang Qi, Wu Chaozhong, Zhang Hui. Regional differences of aggressive driving behaviors in China. Journal of Transport Information and Safety, 2016;(6): 914—929

18 Hawkins D. Identification of outliers. London: Chapman and Hall, 1980

19 Zhang Y, Hamm N A S, Meratnia N, *et al.* Statistics-based outlier detection for wireless sensor networks. International Journal of Geographical Information Science, 2012; 26(8): 1373—1392

20 贾瑞玉，宋建林. 基于聚类中心优化的 *K*-means 最佳聚类数确定方法. 微电子学与计算机, 2016; 33(5): 62—66
Jia Ruiyu, Song Jianlin. *K*-means optimal clustering number determination method based on clustering center optimization. Micro Electronics & Computer, 2016; 33(5): 62—66

21 费　欢，李光辉. 基于 *K*-means 聚类的 WSN 异常数据检测算法. 计算机工程, 2015; 41(7):124—128
Fei Huan, Li Guanghui. Abnormal data detection algorithm for WSN based on *K*-means clustering. Computer Engineering, 2015; 41(7): 124—128

22 Wang X M. Improved artificial bee colony clustering algorithm based on *K*-means. Applied Mechanics & Materials, 2014; 556—562

23 张敬磊，王晓原，王梦莎，等. 三车道动态环境下汽车驾驶倾向性的转移概率及其预测. 交通运输系统工程与信息, 2017; 17(1): 82—90
Zhang Jinglei, Wang Xiaoyuan, Wang Mengsha, *et al.* Transition probability and prediction of automobile driver's propensity under three-lane dynamic conditions. Journal of Transportation Systems Engineering and Information Technology, 2017;17(1): 82—90

24 张敬磊，王晓原，王云云，等. 驾驶任务缓急与汽车驾驶倾向性的相关性. 深圳大学学报(理工版), 2017;34(2): 195—203
Zhang Jinglei, Wang Xiaoyuan, Wang Yunyun, *et al.* Correlation between task urgency and driver's propensity. Journal of Shenzhen University (Science and Engineering), 2017;34(2): 195—203

Outlier Detection of Reckless Driving Behaviors Based on Improved Swarm Intelligent *K*-means Algorithm

WANG Si-si[1], ZHANG Jing-lei[1*], WANG Jian-xing[1], CHEN Ci[1], MA Chun-jie[2]
(School of Transportation and Vehicle Engineering, Shandong University of Technology[1], Zibo 255049, China;
College of Metropolitan Transportation, Beijing University of Technology[2], Beijing 100124, China)

[**Abstract**] How to detect abnormal data effectively and avoid its adverse effects on further study is focused. First, reckless driving experiments are carried out using the simulation driving experiment system and the ErgoLAB human factor system based on reviewing its progresses, and the high dimensional data of the human-vehicle-environment's are obtained by sensors. And then, the optimized *K*-means clustering algorithm is used to cluster and outlier detection. The results show, the optimal number is 2, and the outlier detection are effectively output, which provides high quality data guarantee for the next further study.

[**Key words**] reckless driving swarm intelligence algorithm *K*-means clustering optimal clustering number outlier detection