

基于 Canopy-K-means 算法的半挂 汽车列车行驶数据分析

李 兵, 屈亚洲, 熊 乐, 王晓亮, 赵晨光

(中汽研汽车检验中心(宁波)有限公司, 浙江 宁波 315336)

摘要:为实时监测车辆行驶状态,建立了基于 Canopy-K-means 算法的车辆行驶安全特征分类模型。采用 Canopy-K-means 聚类算法对车辆行驶数据进行挖掘分析,以欧氏距离大小作为数据集属性间的相似性分类指标,得到表征不同行驶安全特征的离线聚类质心;搭建 TruckSim 与 Simulink 联合仿真平台,设置定半径变车速和方向盘斜阶跃输入仿真工况对车辆行驶状态进行在线识别;同时为验证该方法在实车上的应用效果,设置相同工况对离线聚类质心进行验证分析。仿真和实车结果表明:基于 Canopy-K-means 算法的数据挖掘方法可以对不同行驶状态数据进行分类,得到的表征不同行驶安全特征的聚类质心能在一定程度上对车辆行驶稳定性进行评价,可以作为车辆控制和预警的判定依据。

关键词:行驶稳定性;Canopy-K-means 算法;分类模型;半挂汽车列车;欧氏距离

中图分类号:U461.6 **文献标志码:**A **文章编号:**1671-1807(2021)08-0288-07

近年来,随着物流运输业的快速发展,半挂汽车列车在商用车市场的占比逐年增加。据统计^[1],2013—2017 年国内半挂牵引车年销量整体呈递增趋势,且 2016 和 2017 两年的年销量较上一年同比增长了 55% 和 50%。而半挂汽车列车作为诱发重大交通事故的主要源头之一,其数量的增加势必会给其他交通参与者带来更多的潜在危险,因此,如何提高汽车列车行驶安全性一直是行业的研究热点。

Singh^[2]利用相平面法提出了一定初始条件下半挂汽车列车系统相轨迹趋近和背离平衡点的变化规律,由此判定车辆的稳定和不稳定区域。关志伟等^[3]采用线性系统稳态转向特性、非线性系统稳态转向特性和能量相平面 3 种方法对车辆转向稳定性进行了分析。李西涛^[4]建立了半挂汽车列车非线性动力学模型,分析了其在极限工况下各状态变量平衡点的分岔现象,探讨了极限操纵工况下车辆的横向失稳机制。Ding 等^[5]利用分岔理论分析了恒定运行工况下半挂汽车列车非线性模型的稳定性,采用数值求解得到系统的稳定平衡点和不稳定平衡点,解释了常见的不稳定现象,并对状态空间中

的稳定域进行了划分。陈林^[6]基于卡尔曼滤波算法建立了以半挂牵引车的侧向加速度和横摆角速度为观测变量的状态估计器,并选择双移线和方向盘阶跃工况试验对半挂汽车列车状态进行了估计分析。刘焕^[7]建立了半挂汽车列车三自由度线性和非线性动力学模型,探究了车辆在不同行驶条件下的运动响应和失稳形式,并从能量变化角度对车辆的失稳机理进行解释,为判定车辆失稳提供了新的依据。彭涛等^[8]基于 Lyapunov 法分析了车辆极限状态下的能量和能量变化阈值,求解得到半挂汽车列车高速变道稳定域,为车辆高速转向时的稳定性判定提供依据。综合以上文献可知,现有对半挂汽车列车稳定性分析大多是基于简化动力学模型进行的研究,存在所建模型不精确的问题。

为了解决这一问题,徐强^[9]提出了一种基于 K 均值聚类分析的车辆横向稳定性判定方法,通过对车辆仿真数据进行聚类分析,划分了 4 种不同行驶状态,并设计仿真试验验证了该方法在双半挂汽车列车稳定性判定上的有效性,但是其聚类算法中的 K 值由经验确定,具有一定的主观性,影响对车辆行驶状态的准确划分,同时未进行实车试验验证

收稿日期:2020-11-11

作者简介:李兵(1989—),男,山东聊城人,中汽研汽车检验中心(宁波)有限公司,工程师,硕士,研究方向为智能网联测试评价方法;通信作者屈亚洲(1992—),男,河南郸城人,中汽研汽车检验中心(宁波)有限公司,助理工程师,硕士,研究方向为车辆可靠性与安全技术。

该算法的有效性。翟德^[10]利用 K 均值聚类算法对半挂汽车列车侧翻和折叠工况进行了研究,进一步验证了该方法在不同失稳形式下应用的有效性,但其缺点与文献[9]类似。基于以上考虑,本文首先采用 Canopy 算法对车辆行驶数据进行“粗”聚类分析,确定 k 值和初始状态聚类质心,然后在此基础上利用 K -means 算法对车辆行驶状态进行精确划分,并将仿真试验和实车试验结果相结合验证该方法的有效性。

1 系统建模

1.1 整车建模

由于本文主要从数据层面对半挂汽车列车行驶稳定性进行分析,需要大量的车辆行驶数据作为研究支撑,因此为了采集多工况下的行驶数据,选择 TruckSim 软件进行车辆动力学建模仿真,仿真

所使用的车辆类型为 3A Cab Over w/3A Euro Trailer。

1.2 仿真工况设计

依据《营运车辆抗侧翻稳定性试验方法 稳态圆周试验》(JT/T 884—2014)和《商用车辆电子稳定性控制系统性能要求及试验方法》(GB/T 8185—2019)试验工况要求进行仿真工况设计,见表 1。

表 1 中,定半径变车速选择速度为 30~70 km/h,以 5 km/h 的间隔依次进行仿真工况设计,这里只以速度为变量,不考虑方向盘转角变化的影响;定车速变转角选择车速为 40~60 km/h,方向盘转角为 30°、60°、90°、120°,速度间隔为 5 km/h 依次进行仿真工况设计;J 转向试验选择速度 30~45 km/h,间隔为 5 km/h,同样不考虑方向盘转角的影响。

表 1 仿真工况设计

试验标准	试验工况	转弯半径/m	速度/(km·h ⁻¹)	速度间隔/(km·h ⁻¹)	方向盘转角/(°)
JT/T 884—2014	定半径变车速	100	30~70	5	—
	定车速变转角	—	40~60	5	30/60/90/120
GB/T 38185—2019	J 转向试验	45.7	30~45	5	—

2 车辆行驶数据分析

2.1 传统的 K-means 算法

K -means 是一种简单的迭代型聚类算法,由于简单易修改、运行速度快,在数据挖掘领域应用非常广泛。其基本原理是:首先给定包含 n 个对象的样本集和聚簇个数 k ,然后从样本集中随机选择 k 个对象作为初始聚类质心,计算剩余样本集与初始聚类质心的欧氏距离,并将它们分配给与其欧氏距离最小的聚簇,再重新计算每个聚簇中各数据对象的均值作为新的聚类质心,不断重复这一过程直到各个聚簇的聚类质心不再变化或满足停止条件,从而实现同一聚簇中的对象相似度较高,而不同聚簇中的对象相似度较小,完成对样本集的划分。

由于 K -means 算法中的聚簇个数 k 必须预先给定,且初始聚类质心的选择是随机的,容易在迭代过程中形成大量冗余,使聚类结果陷入局部最优解^[11]。因此,针对上述问题以及缩短聚类时间周期,采用 Canopy-K-means 聚类算法^[12-13]对车辆行驶数据进行聚类划分。

2.2 基于 Canopy 优化的 K-means 算法

Canopy 算法也是一种常见的聚类算法,是采用快速近似距离度量和两个距离阈值比较方式,实现快速粗聚类^[14],由于其无须提前预设聚簇个数,因

此可以在预处理阶段对车辆行驶数据进行“粗”聚类,确定聚簇个数 k 和初始聚类质心,然后再利用 K -means 算法进行“细”聚类,提高聚类结果的准确性。

Canopy-K-means 算法流程如图 1 所示,可知在聚类工作开始之前,只需确定 Canopy 算法中的两个距离阈值 T_1 、 T_2 ($T_1 > T_2$), K -means 算法的迭代次数和目标函数的终止条件即可。

2.3 数据挖掘分析

根据 1.2 节设计的仿真工况进行仿真,得到各个工况下的车辆行驶数据。考虑到实车试验测试设备的可测变量,这里选取牵引车和挂车横摆角速度(ω_1 、 ω_2)、质心侧向加速度(a_{y1} 、 a_{y2})和车身侧倾角(ϕ_1 、 ϕ_2)6 个状态变量作为表征半挂汽车列车行驶稳定性的评价参数。结合各个工况的行驶特性,截取抗侧翻稳定性试验工况稳定时 10 s 的数据,截取 J 转向试验工况 30 s 的数据,采集共 20 500 组数据。为了便于对数据进行批量处理,将各个工况下采集的数据记为同一矩阵 $B = [\omega_{1i}, \omega_{2i}, a_{y1i}, a_{y2i}, \phi_{1i}, \phi_{2i}]$,其中,角标 i 代表矩阵 B 的行数,也即是某运行工况下某一时刻的车辆状态参数。

通过 Canopy 算法对矩阵 B 进行预处理,得到聚簇个数 $k=4$,初始聚类质心

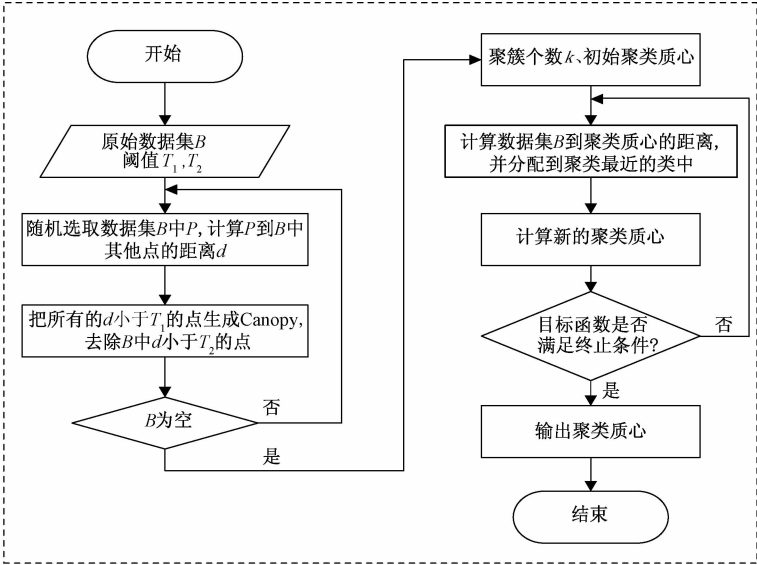


图 1 Canopy-K-means 算法流程

Init_cluster_centr =
$$\begin{bmatrix} 2.969\ 20 & 2.969\ 40 & 0.058\ 70 & 0.058\ 70 & 0.210\ 90 & 0.264\ 50 \\ 10.717\ 60 & 10.949\ 90 & 0.060\ 80 & 0.224\ 20 & 1.221\ 30 & 1.382\ 50 \\ 16.578\ 30 & 15.730\ 60 & 0.474\ 60 & 0.300\ 50 & 2.082\ 80 & 2.387\ 30 \\ 0.001\ 20 & 0.000\ 66 & 0.000\ 74 & 0.000\ 13 & 0.043\ 80 & 0.047\ 70 \end{bmatrix}^{\circ}$$

然后将 Canopy 的聚类结果代入 K-means 算法,得到的聚类结果如图 2 所示,每种聚类质心代表车辆的一种行驶状态。

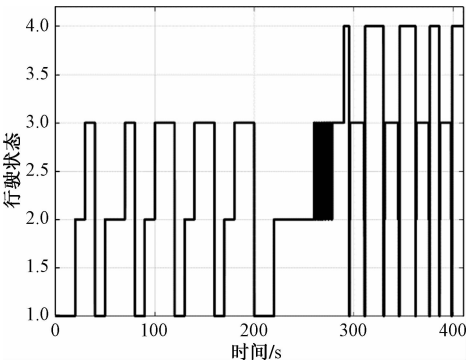


图 2 离线聚类结果

由图 2 可知,经过聚类之后,这些数据被自动化分为 4 个行驶状态,其中 0~200 s 为定车速变转角工况,0~40 s 表示速度为 40 km/h,方向盘转角为 30°、60°、90°、120°时车辆的行驶状态变化,可以发现

车辆在方向盘转角为 30°和 60°时为 1 状态,当转角增加到 90°和 120°时,车辆由 1 状态到 2 状态再变为 3 状态;40~200 s 表示速度为 45~60 km/h,在相同方向盘转角变化下,车辆行驶状态变化,其与 40 km/h 的状态变化类似,可知不同行驶状态的危险性高低依次为 3、2、1。200~290 s 是定半径变车速工况,随着车速的增加,车辆侧向加速度越来越大,稳定性降低,此时车辆的行驶状态也呈现 1-2-3 变化。290~410 s 是 J 转向试验工况,表征 30~45 km/h 下车辆的行驶状态,由于截取的是车辆刚进入弯道到驶出弯道的数据,车辆主要在 3 和 4 行驶状态变化,且当车辆在弯道行驶时,此时车辆处于 4 状态。综合以上分析可以得到:车辆不同行驶状态的危险性高低依次为 4、3、2、1。

最后通过分析各聚类质心的特征值大小以及图 2 分析的行驶状态危险性高低,对聚类结果进行排序,结果见表 2。

表 2 基于 Canopy-K-means 算法得到的聚类质心

聚类质心	w_1	w_2	a_{y1}	a_{y2}	ψ_1	ψ_2
1	0.115 00	0.154 00	0.004 00	0.002 00	0.046 00	0.046 00
2	4.175 00	4.261 00	0.103 00	0.089 00	0.414 00	0.490 00
3	7.794 00	7.784 00	0.194 00	0.190 00	1.036 00	1.179 00
4	12.054 00	12.010 00	0.316 00	0.275 00	1.725 00	1.963 00

为了便于后续分析以及考虑到各特征值大小，这里参考文献[15]的方法对各行驶状态概括为 3 种危险等级状态。其中，行驶状态 1 和 2 统称为安全的行驶状态，行驶状态 3 为危险演化阶段，行驶状态 4 为潜在危险。

3 在线仿真平台搭建

3.1 工况设计

2.3 节对车辆行驶数据进行了聚类分析，得到 4 种行驶状态类型，并根据各工况行驶特点，对各行驶状态下车辆的稳定性进行分析，获得各行驶状态的危险性高低依次为 $4>3>2>1$ 。由于上述结论是基于不同工况下的行驶数据离线聚类得出的结果，因此这里选择抗侧翻稳定性和方向盘斜阶跃输入两种试验，来验证聚类结果的有效性，其中抗侧翻稳定性试验工况参数设置见表 3，方向盘斜阶跃输入工况设计如图 3 所示。

表 3 抗侧翻稳定性试验仿真工况设计

试验工况	转弯半径/m	速度/(km·h ⁻¹)
定半径变车速	120	40、45、50、55、60

图 3 为方向盘转角随时间的变化曲线，其中 0~2 s，方向盘转角为 0°，车辆保持直线行驶，在 2~10 s 方向盘转角匀速增加，在第 10 s 时达到 180°，并保持 180°不变，直至仿真结束。

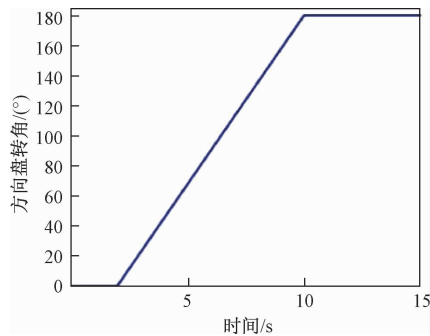


图 3 方向盘斜阶跃输入

3.2 在线识别

利用 TruckSim 和 Simulink 搭建了联合仿真平台，编写了计算欧氏距离的 MATLAB Function 函数，根据表 3 和图 3 所示工况依次进行仿真，其中表 3 所示工况只截取稳态行驶状态下 10 s 内的行驶数据进行分析，共 500 组数据，具体仿真处理结果如图 4、图 5 所示。

由图 4 可知，在 0~10 s，车速为 40 km/h 时，处于第 2 行驶状态，此时车辆为安全状态；在 10~

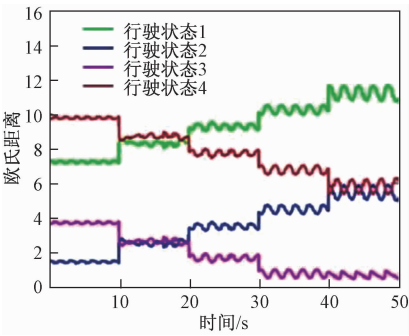


图 4 定半径变车速试验工况的欧氏距离在线识别

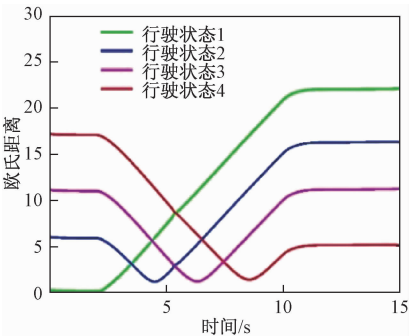


图 5 方向盘斜阶跃输入工况的欧氏距离在线识别

20 s，车速为 45 km/h 时，车辆在第 2 和 3 行驶状态之间变化，表明车辆此时处于安全状态向危险演化状态的过渡阶段；在 20~50 s，速度为 50~60 km/h 时，车辆一直维持在第 3 行驶状态，表明此时车辆处于危险演化状态。

由图 5 可知，在 0~2 s 直线行驶时，处于第 1 行驶状态，此时车辆为安全状态；在 2~10 s 方向盘转角匀速增加时，由第 1、2 行驶状态向 3、4 行驶状态变化，此时车辆逐渐由安全状态向危险演化状态和潜在危险状态过渡；在 10~15 s 保持方向盘转角为 180°不变时，处于第 4 状态，此时车辆处于潜在危险状态。

综合以上两种试验工况的分析结果可知，在已知车辆行驶稳定性变化趋势的情况下，与根据离线聚类结果在线计算欧氏距离判定得到的行驶状态基本一致。因此，利用离线聚类结果对抗侧翻稳定性试验和方向盘斜阶跃输入工况的行驶稳定性判定是有效的。

4 实车验证

为了进一步验证聚类结果的准确性，这里以 3.1 节所列的试验工况进行实车试验，抗侧翻稳定性试验工况下行驶数据的选取方式与 3.2 节相同，同样只截取稳态行驶状态下 10 s 内的行驶数据，计

算其与离线得到的聚类质心的欧氏距离。但考虑到实车试验的安全性以及内容研究的侧重点,半挂汽车列车选择在空载的试验条件下进行,试验样车如图 6 所示。



图 6 试验样车

图 7 所示为根据车辆不同速度下的行驶数据与离线聚类质心计算得到的欧氏距离曲线,从各行驶状态的变化趋势来看,与图 4 所示的仿真结果基本一致,都是在 0~10 s,车速为 40 km/h 时,处于第 2 行驶状态,10~20 s,车速为 45 km/h 时,车辆在第 2 和 3 行驶状态之间变化,20~50 s,速度为 50~60 km/h 时,车辆一直维持在第 3 行驶状态,能明显表现出随着车速增加,车辆由安全向危险演化再向潜在危险状态的过渡过程。

为了更清晰地表征车辆的行驶稳定性变化,绘制了不同速度下牵引车、半挂车侧向加速度及横摆角速度变化曲线,并采用 12 极无阶巴特沃斯滤波器

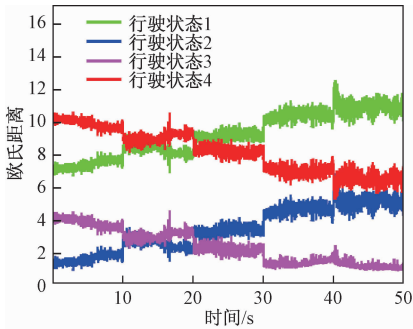


图 7 定半径变车速试验工况的欧氏距离

对原始数据进行滤波处理,如图 8 所示。

图 8(a)和图 8(b)为侧向加速度变化曲线,在 0~10 s,也即为 40 km/h 时,牵引车和半挂车的侧向加速度大小约为 0.1g;在 10~40 s,随着速度的增加,两者的侧向加速度逐渐增大;当在 40~50 s 时,也即速度为 60 km/h 时,半挂车的侧向加速度大小相较于牵引车波动较大,且有明显的甩尾趋势,表明此时车辆的稳定性与之前相比较差。图 8(c)和图 8(d)为横摆角速度变化曲线,其整体变化趋势侧向加速度基本一致,都是随着速度的增加,数值逐渐增大,车辆的稳定性逐渐变差。

图 9 所示为在车辆速度为 40 km/h、方向盘转角斜阶跃输入下的行驶数据与离线聚类质心计算得到的欧氏距离曲线,从各行驶状态的变化趋势来看,与图 5 所示的仿真结果基本一致。在 0~20 s

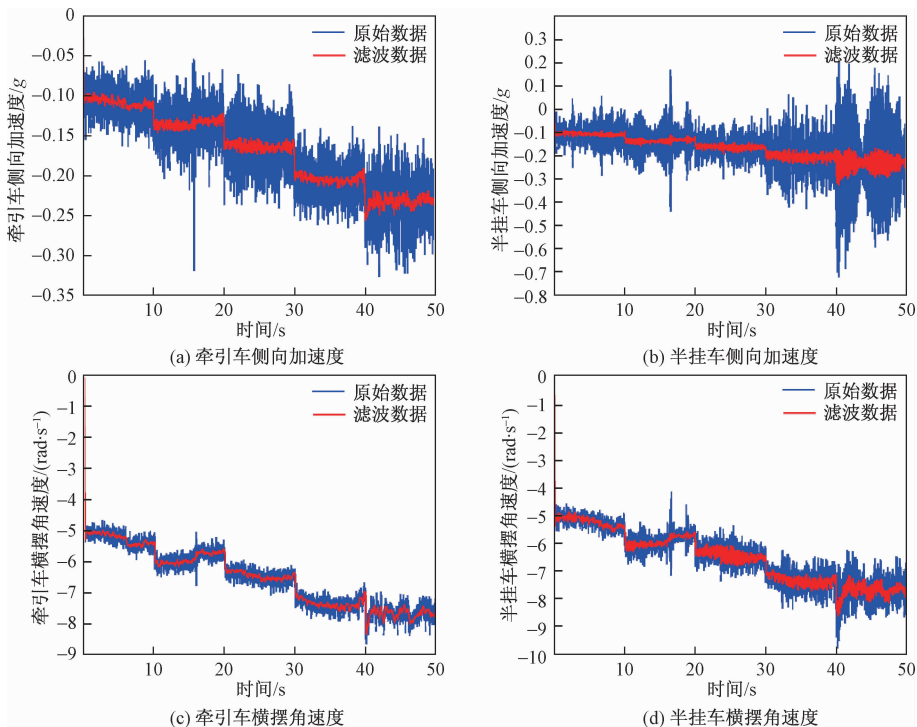


图 8 牵引车和半挂车侧向加速度、横摆角速度变化曲线

直线行驶时,处于第 1 行驶状态,此时车辆为安全状态;在 20~35 s 方向盘转角匀速增加时,由第 1、2 行驶状态向 3、4 行驶状态变化,此时车辆逐渐由安全状态向危险演化和潜在危险状态过渡;在 35~40 s 保持方向盘转角为 180°不变时,处于第 4 行驶状态,此时车辆处于潜在危险状态。

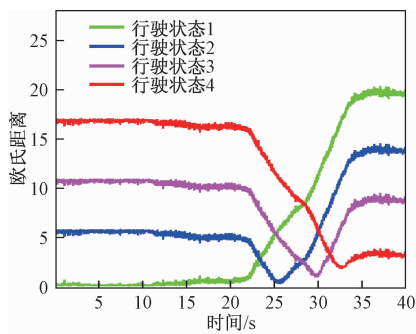


图 9 方向盘斜阶跃输入工况的欧氏距离

为了更清晰地表征车辆的行驶稳定性变化,采用与定半径变车速试验工况的处理方法,绘制了牵引车、半挂车侧向加速度及横摆角速度变化曲线,如图 10 所示。

图 10(a)和图 10(b)为侧向加速度变化曲线,在 0~20 s,车辆以 40 km/h 的速度直线行驶,此时牵引车和半挂车的侧向加速度大小近似为 0;在 20~35 s,随着方向盘转角的匀速增加,两者的侧向加速度逐渐增大;当在 35~40 s 时,也即方向盘转角达

到 180°并保持不变时,半挂车的侧向加速度大小相较于牵引车波动较大,表明此时车辆的稳定性与之前相比较差。图 10(c)和图 10(d)为横摆角速度变化曲线,其整体变化趋势与侧向加速度基本一致,且都是随着方向盘转角的匀速增加,数值逐渐增大,车辆稳定性变差。

综合以上两种试验工况的分析结果可知,根据实车采集的行驶数据和离线聚类质心进行计算欧氏距离,判定得到的行驶状态与表征车辆行驶稳定性指标的变化基本一致。因此,从实车试验层面,验证了利用离线聚类结果对抗侧翻稳定性试验和方向盘斜阶跃输入工况的行驶稳定性判定也是有效的。

5 结语

选用 TruckSim 构建了半挂汽车列车整车模型,依据 JT/T 884—2014 和 GB/T 38185—2019 试验标准设计仿真工况并采集了车辆行驶数据,基于 Canopy-K-means 算法对行驶数据进行聚类分析,并搭建了在线联合仿真平台和实车试验验证平台,从仿真和实车试验两个层面验证该算法在车辆稳定性判定领域的可行性。结果表明,采用 Canopy-K-means 聚类算法可以对不同行驶状态数据进行分类,能够获得表征不同行驶安全特征的聚类质心,通过将定半径变车速、方向盘斜阶跃输入两种试验工况采集的行驶数据和聚类质心进行欧氏距

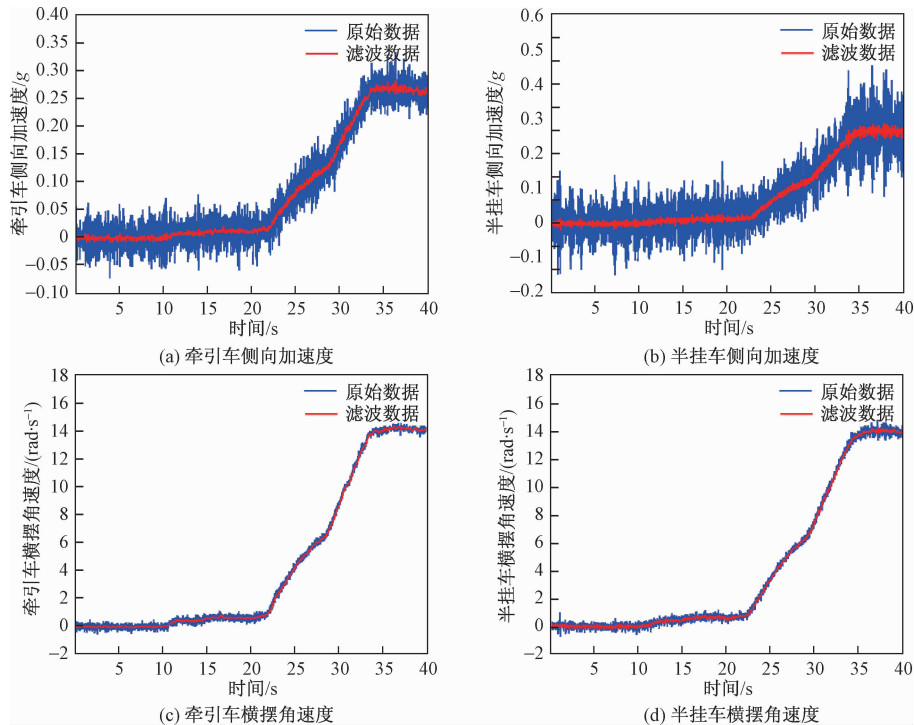


图 10 牵引车和半挂车侧向加速度、横摆角速度变化曲线

离计算,判定得到的行驶状态与车辆所处工况下的行驶稳定性基本一致,能在一定程度上对车辆稳定性进行评价,因此可以使用表征不同行驶安全特征的聚类质心为车辆的控制和预警提供判定依据。

参考文献

[1] 屈亚洲. 基于磁流变阻尼器的半挂汽车列车悬架控制策略研究[D]. 长春: 吉林大学, 2019.

[2] SINGH M. Stability bounds of a tractor-semitrailer [J]. Vehicle System Dynamics, 1980, 9(2): 69—86.

[3] 关志伟, 王新建, 郑明锋. 牵引车—半挂车列车转向稳定性分析[J]. 汽车技术, 2011 (6): 41—45.

[4] 要李西涛. 半挂汽车列车操纵稳定性分析与评价研究[D]. 昆明: 昆明理工大学, 2013.

[5] DING N, SHI X, ZHANG Y, et al. Analysis of bifurcation and stability for a tractor semi-trailer in planar motion[J]. Vehicle System Dynamics, 2014, 52(12): 1729—1751.

[6] 陈林. 基于卡尔曼滤波的半挂汽车列车状态估计[D]. 沈阳: 东北大学, 2014.

[7] 刘焕. 半挂汽车列车三自由度平面运动稳定性分析[D]. 长春: 吉林大学, 2016.

[8] 彭涛, 关志伟, 张荣辉, 等. 半挂汽车列车高速变道稳定域估计[J]. 交通运输工程学报, 2018, 18(4): 90—102.

[9] 徐强. 基于 K 均值聚类分析的车辆横向稳定性判定与控制策略研究[D]. 长春: 吉林大学, 2018.

[10] 翟德. 半挂汽车列车侧向稳定性判定及控制研究[D]. 长春: 吉林大学, 2019.

[11] 郭卫霞, 薛涛, 李婷. 基于 Hadoop 的 Canopy-K-means 并行算法的学生成绩与毕业流向关系分析[J]. 西安工程大学学报, 2018, 32(6): 705—712.

[12] 高榕, 李晶, 肖雅夫, 等. 基于云环境 K-means 聚类的并行算法[J]. 武汉大学学报(理学版), 2015, 61(4): 368—374.

[13] HE H, GUO L, GENG Y. The optimization of cmac neural network structure based on Canopy-K-means algorithm [J]. International Journal of Advancements in Computing Technology, 2012, 4(22): 641—647.

[14] 王海燕, 崔文超, 许佩迪, 等. Canopy 在划分聚类算法中对 K 选取的优化[J]. 吉林大学学报(理学版), 2020, 58(3): 634—638.

[15] 刘宏飞, 徐强, 许洪国, 等. 基于 K 均值聚类分析的车辆横向稳定性判定方法[J]. 湖南大学学报(自然科学版), 2018, 45(8): 48—53.

Driving Data Analysis of Tractor Semi-trailer Based on Canopy-K-means Algorithm

LI Bing, QU Ya-zhou, XIONG Le, WANG Xiao-liang, ZHAO Chen-guang
(CATRA Automotive Test Center (Ningbo) Co., Ltd., Ningbo Zhejiang 315336, China)

Abstract: In order to monitor the driving status of vehicles in real time, a classification model of vehicle driving safety features based on Canopy-K-means algorithm was established. Using Canopy-K-means clustering algorithm to mine and analyze vehicle driving data, using Euclidean distance as the similarity classification index between attributes of data set, to obtain offline clustering centroids representing different driving safety features; building TruckSim and Simulink joint simulation Platform, set the fixed radius variable speed and steering wheel oblique step input simulation working conditions to identify the vehicle's driving status online; at the same time, in order to verify the application effect of the method on the real vehicle, the same working conditions were set to verify the offline clustering centroid analysis. The simulation and real vehicle results show that the data mining method based on the Canopy-K-means algorithm can classify different driving state data, and the obtained cluster centroids representing different driving safety characteristics can evaluate the driving stability of the vehicle to a certain extent, which can be used as the judgment basis for vehicle control and early warning.

Key words: driving stability; Canopy-K-means algorithm; classification model; semi-trailer; Euclidean distance