

第七届“泰迪杯”数据挖掘挑战赛——

C 题：运输车辆安全驾驶行为的分析

一、问题背景

车联网是指借助装载在车辆上的电子标签通过无线射频等识别技术，实现在信息网络平台上对所有车辆的属性信息和静、动态信息进行提取和有效利用，并根据不同的功能需求对所有车辆的运行状态进行有效的监管和提供综合服务的系统。当前道路运输行业等相关部门利用车联网等系统数据，开展道路运输过程安全管理的数据分析，以提高运输安全管理水平和运输效率。

某运输企业所辖各车辆均存在常规运输路线与驾驶人员。在驾驶员每次运输过程中，车辆均可自动采集当前驾驶行为下的行车状态信息并上传至车联网系统。驾驶行为可能随气象、路况等因素的变化而变化，进一步影响行车安全、运输效率与节能水平。

请根据该运输企业所采集的数据（见附件 1、附件 2），分析车辆行驶过程中的驾驶行为对行车安全、运输效率与节能情况的影响，运用数据挖掘的方法，建立有效的数学模型进行评价。

二、研究问题

(1) 利用附件 1 所给数据，提取并分析车辆的运输路线以及其在运输过程中的速度、加速度等行车状态。提交附表中 10 辆车每辆车每条线路在经纬度坐标系下的运输线路图及对应的行车的里程、平均行车速度、急加速急减速情况。

(2) 利用附件 1 所给数据，挖掘每辆运输车辆的不良驾驶行为，建立行车安全的评价模型，并给出评价结果。

(3) 综合考虑运输车辆的安全、效率和节能，并结合自然气象条件与道路状况等情况，为运输车辆管理部门建立行车安全的综合评价指标体系与综合评价模型。

附表

序号	1	2	3	4	5	6	7	8	9	10
车牌号	AA00002	AB00006	AD00003	AD00013	AD00053	AD00083	AD00419	AF00098	AF00131	AF00373

三、数据说明

1. 附件 1 给出 450 辆运输车辆的行车轨迹采集数据，由于采集设备精度，实际采集数据可能存在某些异常。

附件 1 的数据说明

序号	指标名称	指标说明	说明
1	vehicleplatenumber	车牌号码	
2	device_num	设备号	
3	direction_angle	方向角	范围：0-359（方向角指从定位点的正北方向起，以顺时针方向至行驶方向间的水平夹角）
4	lng	经度	东经

5	lat	纬度	北纬
6	acc_state	ACC 状态	点火 1/熄火 0
7	right_turn_signals	右转向灯	灭 0/开 1
8	left_turn_signals	左转向灯	灭 0/开 1
9	hand_brake	手刹	灭 0/开 1
10	foot_brake	脚刹	无 0/有 1
11	location_time	采集时间	
12	gps_speed	GPS 速度	单位: km/h
13	mileage	GPS 里程	单位: km

2. 附件 2 给出 2018 年 7 月 30 日至 2018 年 10 月 10 日全国主要城市的自然气象数据。

附件 2 的数据说明

序号	指标名称	指标说明	说明
1	province	省/自治区/直辖市	
2	prefecture_city	地级市	
3	county	县级市/县	
4	wind_direction	风向	
5	wind_power	风力	
6	high_temp	最高温度	
7	low_temp	最低温度	
8	conditions	天气状况	如: 多云、晴、雨、雪
9	relative_humidity	相对湿度	
10	precipitation	降水量	单位: mm
11	record_date	采集日期	

3. 在车辆运输过程中, 不良驾驶行为主要包括疲劳驾驶、急加速、急减速、怠速预热、超长怠速、熄火滑行、超速、急变道等。

附录:

请仔细阅读以下说明:

1、 关于赛题数据

- ① 示例数据: 2019年3月16日随赛题公布。
- ② 答题数据: 2019年4月13日公布。

2、 提交作品

- ① 命名方式: 论文命名为“C题”, 附件请直接命名为“作品附件”。
- ② 论文及附件内请勿出现队号、学校、学院、队员以及指导老师相关任何信息, 否则该作品视为无效作品。
- ③ 请参赛队于2019年4月26日16:00之前在竞赛官网【提交作品】处提交论文(PDF版, 大小不超过50M)及附件(论文正文(Word版)、源数据(组委会提供的源数据除外)、过程数据、程序的压缩包, 大小不超过200M)。