

系统聚类法在机动车司机违法驾驶行为中的应用*

秦鸣, 宁建标, 邓明君

(华东交通大学 交通运输与物流学院, 江西 南昌 330013)

摘要: 以某高速公路交通事故为例, 针对多种机动车驾驶人违法驾驶行为进行定量分析, 考虑事故起数、死亡人数、受伤人数和直接经济损失 4 个评价指标, 通过 6 种不同系统聚类法对违法驾驶行为进行聚类, 使用 MATLAB 软件对数据进行处理, 运用 R 软件得出聚类结果, 在符合实际意义的情况下, 聚类个数 $k=6$ 时结果最优, 类内样本具有相似性; 计算了各类中违法驾驶行为对道路交通安全状况的影响因子 δ , δ 值越大, 则该类中的违法驾驶行为对该高速公路交通安全状况的影响越大, 管理部门需对应加大整治力度。

关键词: 交通安全; 系统聚类法; 机动车司机; 违法驾驶行为

中图分类号: U491.4

文献标志码: A

文章编号: 1671-2668(2017)04-0045-07

机动车的增长导致交通事故增加, 交通事故越来越受到人们的关注, 有些学者运用聚类法对此进行了研究: 王志伟等将模糊聚类分析方法运用于交通事故分析中, 通过对汽车、摩托车、拖拉机、自行车和行人乘车等 5 种交通方式的 4 项评价指标进行模糊聚类分析得出聚类结果; 张璐等以万车事故率作为评价指标, 采用灰色聚类法对某高速公路路段交通事故影响因素进行了聚类评估; 肖慎等运用模糊 c-1 划分空间方法对事故“黑点”成因进行空间划分, 采用模糊聚类 ISODATA 方法确定模糊划分矩阵和成因各特征指标聚类中心, 并进行事故“黑点”成因聚类分析, 利用划分系数 $F_c(U)$ 和划分的平均模糊熵 $H_c(U)$ 对聚类效果进行了评价; Fang C. Y. 等通过实时数据采集, 运用模糊推理网络模型分析了违法驾驶行为; 贾元华等利用数理统计学中的主成分-聚类分析方法, 建立道路交通安全综合评价指标对具体城市近几年交通安全现状进行评价, 客观地反映道路交通事故的特征、变化趋势及事故原因类型等。上述成果采用聚类分析法分析交通事故的影响因素, 但没有对准确的聚类个数进行分析和说明, 聚类结果略显牵强。

在导致交通事故的因素中, 驾驶人驾驶行为等人为因素占 90.3% 的超高比重。该文针对机动车驾驶人违法驾驶行为, 以事故起数、死亡人数、受伤人数和直接经济损失为评价指标, 采用常用的 6 种系统聚类法进行聚类, 分析对比不同聚类法在不同 k

值时其 R^2 值, 确定最优 k 值和聚类结果。

1 系统聚类法

系统聚类法的基本思想: N 个样本, 开始时各自作为一类, 即 N 类, 并规定样本与样本之间的距离和类与类之间的距离, 然后将距离最近的合并成一个新类, 计算新类与其他类的距离; 重复进行两个最近类的合并, 每次减少一类, 直至所有样本合并成一大类。

该文运用最短距离法、最长距离法、中间距离法、类平均法、重心法、离差平方和法等 6 种系统聚类法对机动车驾驶人违法驾驶行为的 4 项评价指标进行聚类评估分析。

(1) 最短距离法(single)。规定样本之间的距离为 d_{ij} ($i, j \in n$), 任意类为 G_j , 合并的类为 G_K 、 G_L 、 G_M 。设 D_{KJ} 为类 G_K 到类 G_j 的距离、 D_{LJ} 为类 G_L 到类 G_j 的距离, 则类 G_M 与类 G_j 的距离 D_{MJ} 为两类最近样本间的距离:

$$D_{MJ} = \min_{\substack{i \in G_M, \\ j \in G_j}} d_{ij} = \min \left\{ \min_{\substack{i \in G_K, \\ j \in G_j}} d_{ij}, \min_{\substack{i \in G_L, \\ j \in G_j}} d_{ij} \right\} = \min \{ D_{KJ}, D_{LJ} \}$$

(2) 最长距离法(complete)。类与类之间的距离定义为两类样本间的最大距离, 即:

$$D_{KL} = \max_{\substack{i \in G_K, \\ j \in G_L}} d_{ij}$$

$$D_{MJ} = \max_{\substack{i \in G_M, \\ j \in G_j}} d_{ij} = \max \left\{ \max_{\substack{i \in G_K, \\ j \in G_j}} d_{ij}, \max_{\substack{i \in G_L, \\ j \in G_j}} d_{ij} \right\}$$

* 基金项目: 江西省交通运输厅科技项目(2015C0012); 江西省青年科学基金项目(20151BAB216024)

$$\max_{\substack{i \in G_L \\ j \in G_J}} d_{ij} = \max\{D_{KJ}, D_{LJ}\}$$

(3) 中间距离法(median)。定义类与类之间的距离为最大距离和最小距离的中间距离。取以 D_{KJ} 、 D_{LJ} 和 D_{KL} 为 3 条边组成的三角形中 D_{KL} 边的中线作为新类 G_M 到任意一类 G_J 的距离 D_{MJ} , 根据三角形初等平面几何理论, 得:

$$D_{MJ}^2 = \frac{1}{2}D_{KJ}^2 + \frac{1}{2}D_{LJ}^2 - \frac{1}{4}D_{KL}^2$$

(4) 类平均法(average)。类平均法采用两种方法定义类与类之间的距离。1) 定义类与类之间的距离为所有样本之间的平均距离[见式(1)], 新类 G_M 到任意一类 G_J 的距离按式(2)计算。2) 定义类与类之间的平方距离为样品对之间平方距离的均值[见式(3)], 新类 G_M 到任意一类 G_J 的距离按式(4)计算。

$$D_{KL} = \frac{1}{n_K n_L} \sum_{\substack{i \in G_K \\ j \in G_L}} d_{ij} \quad (1)$$

$$D_{MJ} = \frac{n_K}{n_M} D_{KJ} + \frac{n_L}{n_M} D_{LJ} \quad (2)$$

$$D_{KL}^2 = \frac{1}{n_K n_L} \sum_{\substack{i \in G_K \\ j \in G_L}} d_{ij}^2 \quad (3)$$

$$D_{MJ}^2 = \frac{n_K}{n_M} D_{KJ}^2 + \frac{n_L}{n_M} D_{LJ}^2 \quad (4)$$

式中: n_K 、 n_L 、 n_M 为类 G_K 、 G_L 、 G_M 中的样本个数。

(5) 重心法(centroid)。类与类之间的距离定义为样本之间的欧式距离[见式(5)], 新类 G_M 到任意一类 G_J 的距离按式(6)计算。

$$D_{KL}^2 = d_{\bar{x}_K \bar{x}_L}^2 = (\bar{x}_K - \bar{x}_L)^T (\bar{x}_K - \bar{x}_L) \quad (5)$$

$$D_{MJ}^2 = \frac{n_K}{n_M} D_{KJ}^2 + \frac{n_L}{n_M} D_{LJ}^2 - \frac{n_K n_L}{n_M^2} D_{KL}^2 \quad (6)$$

式中: \bar{x}_K 为类 G_K 的重心(均值); \bar{x}_L 为类 G_L 的重心(均值)。

(6) 离差平方和法(ward)。该聚类法基于方差分析思想, 如果分类正确, 则类内离差平方和较小, 类间离差平方和较大。设类 G_K 和类 G_L 合并成新类 G_M , G_K 、 G_L 和 G_M 的离差分别为 W_K 、 W_L 、 W_M , 则类 G_K 和类 G_L 之间的距离为 $D_{KL}^2 = W_M - W_K - W_L$, 新类 G_M 到任意一类 G_J 的距离为:

$$D_{MJ}^2 = \frac{n_J + n_K}{n_J + n_M} D_{KJ}^2 + \frac{n_J + n_L}{n_J + n_M} D_{LJ}^2 - \frac{n_J}{n_J + n_M} D_{KL}^2 \quad (7)$$

2 案例分析

以某高速公路为例, 考虑事故起数、死亡人数、受伤人数及直接经济损失 4 项评价指标, 运用系统聚类法对机动车驾驶人违法驾驶行为进行聚类分析, 得出该高速公路机动车司机违法驾驶行为等交通事故影响因素聚类结果和影响程度。数据样本采用 2010 年道路交通事故统计年报中该路段机动车司机违法驾驶行为所引发的交通事故数据, 考虑到不可能将违法驾驶行为全部列出, 仅列出主要 21 项违法驾驶行为, 未列出的(如驾驶时接听电话、未保持安全驾驶距离等)统归到其他违法操作项中。样本数据见表 1, 分析流程见图 1。

表 1 某高速公路 2010 年机动车司机违法驾驶行为所导致的交通事故统计

机动车违章驾驶行为	事故起数		死亡人数		受伤人数		直接财产损失	
	数量/起	占比	数量/人	占比	数量/人	占比	数量/万元	占比
超速驾驶 G_1	21 754	0.109	9 134	0.152	23 424	0.100	10 869.561 9	0.129
酒后驾驶 G_2	4 378	0.022	1 958	0.033	4 879	0.021	2 038.581 4	0.024
逆向行驶 G_3	9 605	0.048	3 521	0.059	13 434	0.058	5 007.170 0	0.059
疲劳驾驶 G_4	1 890	0.009	1 139	0.019	2 364	0.010	42 369.857 0	0.050
违法变道 G_5	5 464	0.027	1 046	0.017	6 495	0.028	2 421.693 7	0.029
违法超车 G_6	6 850	0.034	2 101	0.035	9 317	0.040	3 044.330 2	0.036
违法侧车 G_7	2 836	0.014	708	0.012	2 350	0.010	476.459 3	0.006
违法调头 G_8	3 028	0.015	341	0.006	3 936	0.017	903.969 2	0.011
违法会车 G_9	8 090	0.041	2 189	0.037	11 647	0.050	2 794.881 7	0.033
违法牵引 G_{10}	51	0.000	29	0.001	49	0.000	21.199 1	0.000
违法抢行 G_{11}	973	0.005	217	0.004	1 091	0.005	259.724 1	0.003
违法上路 G_{12}	3 385	0.006	1 553	0.026	3 857	0.017	2 054.851 2	0.024

续表 1

机动车违章驾驶行为	事故起数		死亡人数		受伤人数		直接财产损失	
	数量/起	占比	数量/人	占比	数量/人	占比	数量/万元	占比
违法停车 G_{13}	1 184	0.006	412	0.007	1 207	0.005	427.071 0	0.005
违法占道 G_{14}	6 665	0.033	2 224	0.037	8 287	0.035	2 652.250 6	0.031
违法装载 G_{15}	1 919	0.010	2 372	0.023	2 126	0.009	15 923.022 2	0.019
违法装运危险品 G_{16}	31	0.000	18	0.000	27	0.000	115.734 3	0.001
违反交通信号 G_{17}	4 862	0.024	1 207	0.020	6 449	0.028	2 322.361 6	0.028
未按规定让行 G_{18}	37 926	0.190	7 788	0.130	42 794	0.183	8 545.092 5	0.101
无证驾驶 G_{19}	12 636	0.063	4 443	0.074	15 539	0.067	2 266.366 4	0.027
违章使用灯光 G_{20}	1 002	0.005	288	0.005	997	0.004	207.455 1	0.002
油门控制不当 G_{21}	64	0.000	22	0.000	108	0.000	46.595 4	0.001
其他违法操作 G_{22}	7 299	0.037	2 186	0.036	7 981	0.034	4 821.321 3	0.057
合计	199 926	1.000	60 019	1.000	233 617	1.000	84 432.882 3	1.000

注:机动车司机在驾驶过程中,不在车道中间行驶,而是靠近车道线甚至压着车道线行驶,但又没有变车道,这种时而正常行驶(车道中间行驶),时而压车道线行驶的驾驶行为称为违法侧车。

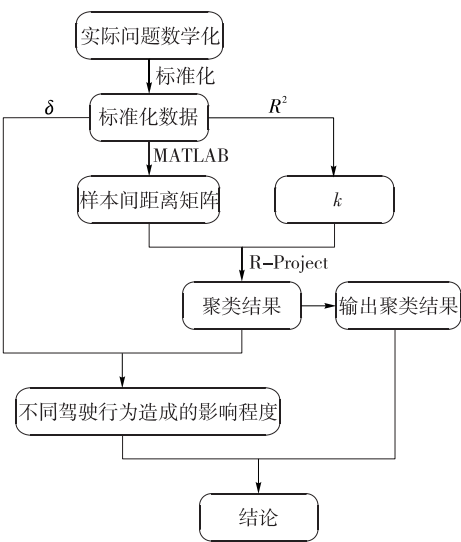


图 1 聚类分析流程

2.1 建立数学模型及数据标准化

将表 1 中 22 项违法驾驶行为看成 22 个样本, 设第 i 项违法驾驶行为为样本 i , 则 $i=1,2,3,\cdots,22$ 。每项违法驾驶行为均考虑 4 项评价指标, 将每项评价指标看成样本的相应特征 $k(k=1,2,3,4)$, x_{ik} 表示样本 i 的第 k 个特征的特征值, 得:

$$x_{ik} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & x_{i3} & x_{i4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{221} & x_{222} & x_{223} & x_{224} \end{bmatrix}$$

样本间的距离采用 $q=2$ 的欧氏距离, 即:

$$d_{ij} = \left[\sum_{k=1}^{22} (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (k=1,2,3,4)$$

式中: x_{ik} 、 x_{jk} 分别为样本 i 和样本 j 的第 k 特征值。

由于 4 项评价指标(变量)的单位不同, 且各指标取值范围相差较大, 计算 Euclidean 距离前先将数据标准化, 使数据属于 $[0,1]$ 。由式(8)、式(9)经 MATLAB 软件编程得出的标准化数据见表 2。

$$\begin{cases} x'_{ik} = \frac{x_{ik} - \bar{x}_k}{\sqrt{S_{kk}}}, & \bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} \\ S_{kk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \end{cases} \quad (8)$$

$$x''_{ik} = \frac{x'_{ik} - x'_{ij\min}}{x'_{ik\max} - x'_{ik\min}} \quad (9)$$

式中: \bar{x}_k 为第 k 个变量的样本均值; S_{kk} 为第 k 个变量的样本方差。

表 2 机动车司机违法驾驶行为所导致交通事故的标准化数据

机动车违章驾驶行为	事故起数	死亡人数	受伤人数	直接经济损失
超速驾驶 G_1	0.883 1	0.000 0	1.000 0	0.121 5
酒后驾驶 G_2	0.828 5	0.000 0	1.000 0	0.027 6
逆向行驶 G_3	0.613 7	0.000 0	1.000 0	0.149 9
疲劳驾驶 G_4	0.018 2	0.000 0	0.029 7	1.000 0

续表 2

机动车违章驾驶行为	事故起数	死亡人数	受伤人数	直接经济损失
违法变道 G_5	0.810 8	0.000 0	1.000 0	0.252 5
违法超车 G_6	0.658 1	0.000 0	1.000 0	0.130 7
违法侧车 G_7	1.000 0	0.098 1	0.794 0	0.000 0
违法调头 G_8	0.747 4	0.000 0	1.000 0	0.156 6
违法会车 G_9	0.623 9	0.000 0	1.000 0	0.064 1
违法牵引 G_{10}	1.000 0	0.261 8	0.932 9	0.000 0
违法抢行 G_{11}	0.865 0	0.000 0	1.000 0	0.048 9
违法上路 G_{12}	0.795 1	0.000 0	1.000 0	0.217 8
违法停车 G_{13}	0.971 1	0.000 0	1.000 0	0.019 0
违法占道 G_{14}	0.732 5	0.000 0	1.000 0	0.070 6
违法装载 G_{15}	0.000 0	0.032 3	0.014 8	1.000 0
违法装运危险品 G_{16}	0.133 0	0.000 0	0.092 1	1.000 0
违反交通信号 G_{17}	0.697 3	0.000 0	1.000 0	0.212 8
未按规定让行 G_{18}	0.860 9	0.000 0	1.000 0	0.021 6
无证驾驶 G_{19}	0.781 3	0.164 0	1.000 0	0.000 0
违章使用灯光 G_{20}	1.000 0	0.101 4	0.993 7	0.000 0
油门控制不当 G_{21}	0.488 4	0.000 0	1.000 0	0.286 0
其他违法操作 G_{22}	0.882 3	0.000 0	1.000 0	0.454 8

2.2 计算样本之间的距离

采用标准化数据计算各违法驾驶行为之间的欧氏距离,运用 MATLAB 软件编程计算得到初始距离矩阵 $D_{(0)22 \times 22}$:

$$D_{(0)22 \times 22} = \begin{bmatrix} 0 & & & & & \\ 1.064\ 4 & 0 & & & & \\ 0.857\ 9 & 0.266\ 2 & 0 & & & \\ \vdots & \vdots & \vdots & & & \\ 1.106\ 6 & 0.181\ 6 & 0.257\ 7 & \cdots & 0.290\ 5 & 0 \end{bmatrix}$$

2.3 确定聚类个数

假设总样本数为 n ,所有样本总重心为 \bar{x} ,聚类时把所有样本合并成 k 个类,即 G_1, G_2, \dots, G_k ,类 G_i 的样本数和重心分别为 $n_i, \bar{x}_i (i=1, 2, 3, \dots, k)$,则有:

$$\sum_{i=1}^k n_i = n, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k n_i \bar{x}_i$$

令所有样本总离差平方和为 W ,类 G_i 中样本的类内离差平方和为 W_i , k 个类的类内离差平方和之和为 P_k ,则有:

$$W = \sum_{j=1}^n (x_j - \bar{x})^T (x_j - \bar{x}) = P_k + \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^T (\bar{x}_i - \bar{x})$$

$$W_i = \sum_{j \in G_i} (x_j - \bar{x}_i)^T (x_j - \bar{x}_i)$$

$$P_k = \sum_{i=1}^k W_i$$

令 $R^2 = 1 - P_k/W$,得:

$$R^2 = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^T (\bar{x}_i - \bar{x}) / W$$

P_k/W 的值越小,则 R^2 越大,表明类内离差平方和之和在总离差平方和中所占的比例越小,类内样本越聚集, k 个类分得越开,聚类效果越好。由于 R^2 是一个随着聚类个数增加其值递增的统计量, n 个样本分为一大类时 $R^2=0$,分为 n 时 $R^2=1$,在对实际案例进行聚类时, R^2 值并非越大越好,还要考虑实际情况, R^2 要取适当的值,其值接近零或 1 将失去聚类的意义。

通过对比 6 种系统聚类法在不同聚类个数时的 R^2 值,分析 R^2 值在不同 k 值时的变化,考虑实际的聚类意义,确定最优聚类个数。 R^2 值见表 3。

表 3 各系统聚类法在不同聚类个数时的 R^2 值

k 值	各聚类方法的 R^2 值					
	最短距离法	最长距离法	类平均法	重心法	中间距离法	离差平方和法
1	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0	0.000 0
2	0.841 3	0.841 3	0.841 3	0.841 3	0.841 3	0.841 3
3	0.841 6	0.898 3	0.841 6	0.841 6	0.852 9	0.898 3
4	0.864 3	0.903 6	0.874 6	0.874 6	0.874 6	0.925 9
5	0.885 0	0.918 5	0.915 6	0.915 6	0.895 6	0.945 5
6	0.901 6	0.954 2	0.954 2	0.954 2	0.921 0	0.957 7
...
22	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0

R^2 值相等表示聚类结果相同。从表 3 可见 6 种方法的 R^2 值都是递增的,其中 $k=1, k=22$ 没有实际意义,去除。

令 $\Delta = R_i^2 - R_{i-1}^2 (i=3, 4, 5, \dots, 21)$, 得出当 $i=6$ 时 Δ 最大,最短距离法、最长距离法、类平均法、重心法、中间距离法、离差平方和法的 Δ 值分别为 0.016 6、0.035 7、0.038 6、0.038 6、0.025 4、0.012 2。在聚类个数增加一个或减少一个时 R^2 值变化最大,即类内离差平方和之和在总离差平方和占比中变化最大,由 $i=5 \rightarrow i=6$ 时,聚类效果显著增加。因此,最优聚类个数为 $k=6$ 。

2.4 R 软件运行代码

确定最优聚类个数后,为了得出聚类结果,使用 R 软件进行编程。R 软件运行程序如下:

```
x = c(0.883 1, 0.828 5, 0.613 7, 0.018 2,
0.810 8, 0.658 1, 1, 0.747 4, 0.623 9, 1, 0.865 0,
0.795 1, 0.971 1, 0.732 5, 0, 0.133 0, 0.697 3,
0.860 9, 0.781 3, 1, 0.488 4, 0.882 3, 0, 0, 0, 0,
0, 0.098 1, 0, 0, 0.261 8, 0, 0, 0, 0, 0.032 3, 0, 0, 0,
0.164 0, 0.101 4, 0, 0, 1, 1, 1, 0.029 7, 1, 1,
0.794 0, 1, 1, 0.932 9, 1, 1, 1, 1, 0.014 8, 0.092 1,
1, 1, 1, 0.993 7, 1, 1, 0.121 5, 0.027 6, 0.149 9, 1,
0.252 5, 0.130 7, 0, 0.156 6, 0.064 1, 0, 0.048 9,
0.217 8, 0.019 0, 0.070 6, 1, 1, 0.212 8, 0.021 6, 0,
0, 0.286 0, 0.454 8);
dim(x)=c(22,4);d=dist(x)
h=hclust(d,"single");
h=hclust(d,"complete");
h=hclust(d,"average");
h=hclust(d^2,method="centroid");
h$height=sqrt(h$height);
```

```
cbind(h$merge,h$height);
h=hclust(d^2,method="median");
h$height=sqrt(h$height);
cbind(h$merge,h$height);
h=hclust(d^2,method="ward");
h$height=sqrt(h$height);
cbind(h$merge,h$height);
opar=par(mfrow=c(2,3))
plot(h,hang=-1,main="聚类谱系图—最短距离法");
rect.hclust(h,k=6,border='red');
plot(h,hang=-1,main="聚类谱系图—最长距离法");
rect.hclust(h,k=6,border='red');
plot(h,hang=-1,main="聚类谱系图—类平均法");
rect.hclust(h,k=6,border='red');
plot(h,hang=-1,main="聚类谱系图—重心法");
rect.hclust(h,k=6,border='red');
plot(h,hang=-1,main="聚类谱系图—中间距离法");
rect.hclust(h,k=6,border='red');
plot(h,hang=-1,main="聚类谱系图—离差平方和法");
rect.hclust(h,k=6,border='red');
par(opar)
```

2.5 聚类结果及安全影响分析

通过运行上述代码($k=6$),得出各系统聚类法的聚类结果(见图 2)。由图 2 可知:最长距离法、类平均法、重心法的聚类结果完全相同;中间距离法、

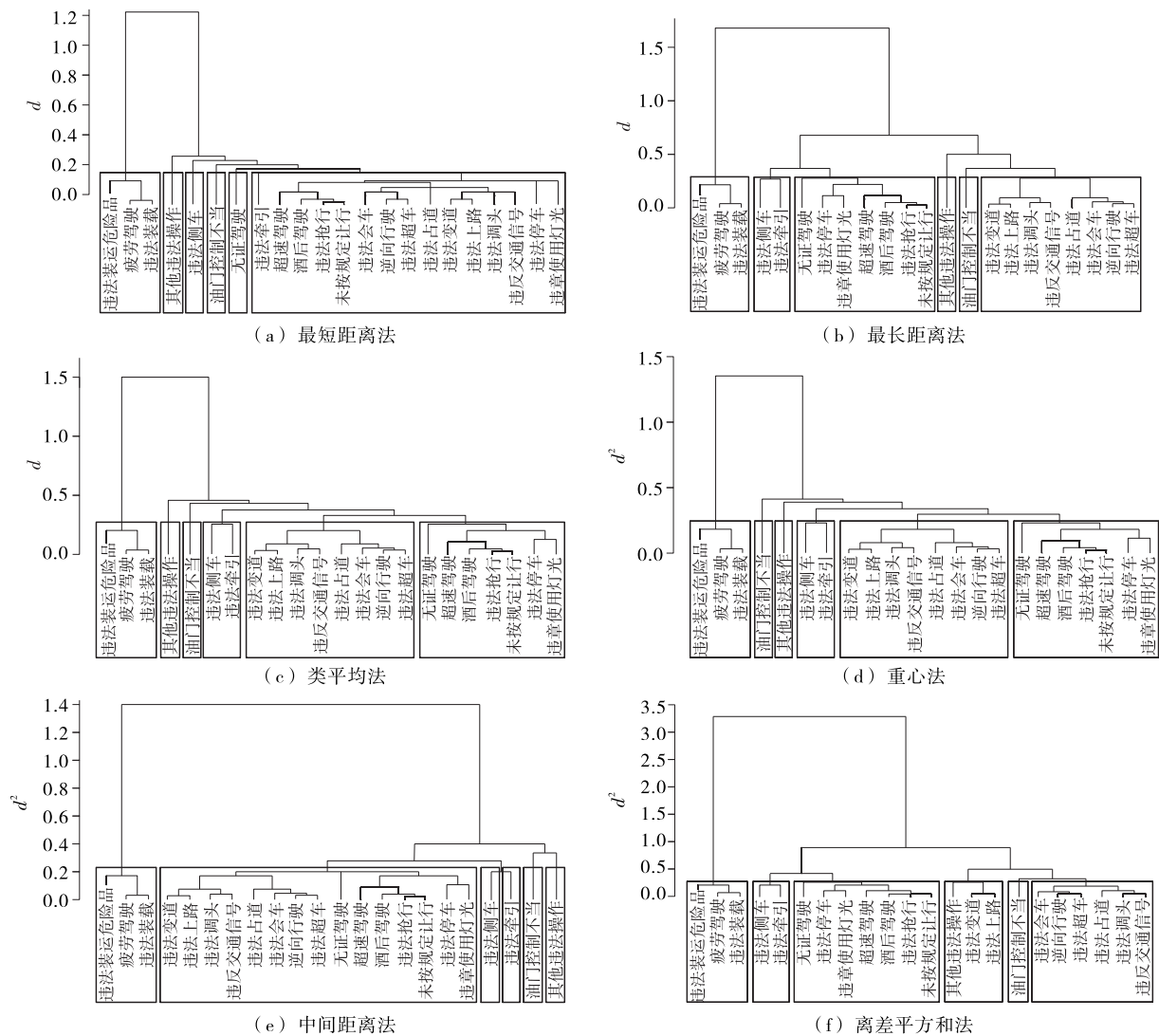


图2 聚类结果

离差平方和法略有差异,只有最短距离法的结果出入较大。考虑到实际的聚类意义,22项违法驾驶行为总体可分为六类(见表4)。

引入各类中违法驾驶行为对道路交通安全的影响因子 δ ,令:

$$\begin{cases} \delta_j = \frac{1}{n_j} \sum_{i=1}^{n_j} \eta_i & (j=1,2,\dots,6) \\ \eta_i = \frac{1}{4} \sum_{k=1}^4 |x'_{ik}| \end{cases} \quad (10)$$

式中: δ_j 为第 j 类对道路交通安全的影响因子; n_j 为第 j 类中包含的违法驾驶行为项数; η_i 为第 j 类中第 i 项违法驾驶行为对道路的影响程度; x'_{ik} 为对应 η_i 的违法驾驶行为标准化的评价指标值。

通过式(10)计算得到各类对道路安全状况的影响因子为 $\delta_1=0.785\ 2$ 、 $\delta_2=0.713\ 0$ 、 $\delta_3=0.847\ 5$ 、 δ_4

表4 最终聚类结果($k=6$)

类别	聚类结果
第一类	其他违法操作 G_{22}
第二类	油门控制不当 G_{21}
第三类	违法侧车 G_7 、违法牵引 G_{10}
第四类	疲劳驾驶 G_4 、违法装载 G_{15} 、违法装运危险品 G_{16}
第五类	超速驾驶 G_1 、酒后驾驶 G_2 、违法抢行 G_{11} 、违法停车 G_{13} 、未按规定让行 G_{18} 、无证驾驶 G_{19} 、违章使用灯光 G_{20}
第六类	逆向行驶 G_3 、违法变道 G_5 、违法超车 G_6 、违法调头 G_8 、违法会车 G_9 、违法上路 G_{12} 、违法占道 G_{14} 、违反交通信号 G_{17}

$=0.748\ 0$ 、 $\delta_5=0.858\ 5$ 、 $\delta_6=0.826\ 1$,对道路安全的影响程度为第五类>第三类>第六类>第一类>第

四类>第二类。

上述聚类法得出的结果与实际情况较相符。如第二类油门控制不当(操作不当)单独作为一类,以区别于其他违法驾驶行为的类;第六类中违法变道、违法超车、违法调头、逆向行驶等都与违法使用车道有关。计算合并为不同类的违法驾驶行为的道路交通安全影响因子,从而得出不同违法驾驶行为对道路交通安全的影响程度。

3 结语

通过将具有潜在联系的违法驾驶行为合并为一类,得出不同类机动车违法驾驶行为等道路交通安全影响因素对交通安全的影响程度。结果表明目标高速公路上超速驾驶、酒后驾驶等第五类中的违法驾驶行为对其安全的影响程度最大,应重点加强整治;其次为违法牵引、违法侧车等,也应重点预防。通过聚类分析,找出重点,对症下药,有助于提升道路交通管理水平,防范交通事故的发生,减少生命、财产损失。

参考文献:

- [1] Klauter S G, Dingus T A, Neale V L, et al. Comparing real-world behaviors of drivers with high versus low rates of crashes and near crashes[R]. National Highway Traffic Safety Administration, 2009.
- [2] 王伟志.模糊聚类分析在交通事故中的应用[J].辽宁工学院学报, 2007, 27(4).
- [3] 张璐.灰色聚类法在交通事故影响因素分析中的应用

[J].公路与汽运, 2011(2).

- [4] 肖慎, 过秀成.模糊聚类法在公路交通事故黑点成因分析中的应用[J].交通运输系统工程与信息, 2002, 2(3).
- [5] Fang C Y, Hsueh H L, Chen S W. Dangerous driving event analysis system by a cascaded fuzzy reasoning petri net[A]. Intelligent Transportation Systems Conference 2006[C]. 2006.
- [6] 贾元华, 邬万江, 常胜, 等.主成分-聚类分析法在道路交通安全评价中的应用[J].交通标准化, 2010(19).
- [7] 田兵.系统聚类法及其应用研究[J].阴山学刊, 2014, 28(2).
- [8] 公安部交通管理局.中华人民共和国道路交通事故统计年报(2010)[R].北京:公安部交通管理科学研究所, 2011.
- [9] 陈明, 郑彩云, 张铮. MATLAB 函数与实例速查手册[M].北京:人民邮电出版社, 2014.
- [10] 李倩星. R 语言实战:编程基础、统计分析与数据挖掘宝典[M].北京:电子工业出版社, 2016.
- [11] 喻伟, 方磊.城市轨道交通与常规公交协调的灰色聚类评价[J].公路与汽运, 2011(4).
- [12] 袁伟, 付锐, 郭应时, 等.驾驶员视觉搜索模式模糊聚类评价方法[J].中国公路学报, 2011, 24(1).
- [13] 赵海存.道路交通事故数据统计分析系统研究[D].西安:长安大学, 2005.
- [14] 陈玉玲.基于实例的系统聚类分析法在环境空气质量评价中的应用[J].环境科学与管理, 2010, 35(8).
- [15] 黎美清, 杜岩, 罗义学, 等.机动车驾驶员道路交通违法行为危险因素分析[J].应用预防医学, 2010, 16(3).

收稿日期:2017-01-14

(上接第44页)

- [7] 张弛, 孟良, 汪双杰, 等.高速公路曲线路段小客车制动行为侧滑风险仿真分析[J].中国公路学报, 2015, 28(12).
- [8] 何杰, 刘霞, 陈一锴, 等.恶劣天气路面条件对行车安全的影响[J].交通运输工程学报, 2011, 11(1).
- [9] 刘晨.基于Carsim仿真软件的公路线形安全评价方法研究[D].西安:长安大学, 2014.
- [10] Jangyeol Yoon, Dongshin Kim, Kyongsu Yi. Design of a rollover index-based vehicle stability control scheme[J]. Vehicle System Dynamics, 2007, 45(5).
- [11] Jung Jongchul, Taehyun Shim, Jamie Gertsch. A vehicle roll-stability indicator incorporating roll center movements[J]. IEEE Transactions on Vehicular Technology, 2009, 58(8).

- [12] JTG B01-2014, 公路工程技术标准[S].
- [13] JTG D20-2006, 公路路线设计规范[S].
- [14] 唐歌腾, 任春晓, 李臣, 等.基于Trucksim的不同弯道半径安全车速确定方法[J].公路交通科技, 2016, 33(6).
- [15] 陈冉, 辜志强, 范李, 等.基于仿真的某型商用车弯道行驶安全速度模型研究[J].公路与汽运, 2016(1).
- [16] 刘晨.基于Carsim仿真软件的公路线形安全评价方法研究[D].西安:长安大学, 2014.
- [17] 王慧丽, 史忠科.山区道路车辆侧翻模型与安全分析[J].交通信息与安全, 2013, 31(3).
- [18] 孙川, 吴超仲, 褚端峰, 等.基于车辆侧向稳定性分析的弯道行驶安全评价[J].交通信息与安全, 2014, 32(6).

收稿日期:2017-03-17