

DOI:10.11918/j.issn.0367-6234.201805083

结合灰聚簇与 Fisher 变换的城市快速路交通状态判别

宋占国¹,陈 红²,黄 卫¹

(1.东南大学 智能运输系统研究中心,南京 210096;2.长安大学 公路学院,西安 710061)

摘 要:为提高少数据下的城市快速路交通流状态类型判别精度,提出一种结合灰聚簇与 Fisher 变换(GC-Fisher)的组合方法.选择交通量 Q 、速度 v 、占有率 O 作为基础参数,首先经灰聚簇理论将基础参数数据聚簇为4类,其次对分类后的数据构建训练集,训练GC-Fisher模型,获取每一种交通流状态类型的Fisher变换方式及判别函数,最后选择结合 K 均值与多分类支持向量机(K -SVM)的组合方法进行比较.结果表明:在数据量较少条件下对交通流状态类型进行判别,GC-Fisher模型判别率为92%,优于 K -SVM模型的判别率69%,GC-Fisher组合方法在少数据下能够更好地提高交通流状态类型的判别效果.

关键词:交通工程;交通状态判别;灰聚簇;Fisher变换;性能分析

中图分类号:U491 **文献标志码:**A **文章编号:**0367-6234(2019)09-0022-07

Traffic state identification for urban expressway: a combination of gray clustering and fisher transform model

SONG Zhanguo¹, CHEN Hong², HUANG Wei¹

(1. Intelligent Transport System Research Center, Southeast University, Nanjing 210096, China;
2. School of Highway, Chang'an University, Xi'an 710061, China)

Abstract: To improve the accuracy of urban traffic flow state identification under limited data condition, a model combining grey clustering and fisher transform (GC-fisher model) is proposed. First, the parameters of traffic volume (Q), speed (v), and occupation (O) were divided into four categories according to the grey clustering theory. Then training dataset was established by using the classified data, and the fisher transform and discriminant functions of each traffic flow state were obtained by using the GC-Fisher model. Afterwards, the combination of K -means and multi-class support vector machine (K -SVM) model was selected as the comparison model. Results of model discriminant rate and the case study show that the GC-fisher model outperformed K -SVM model under limited data condition: the discriminant rate was 92% by using GC-fisher, while the rate was 69% by using K -SVM. The GC-Fisher model can improve the discrimination accuracy of traffic flow state under limited data condition.

Keywords: traffic engineering; traffic state identification; gray clustering; fisher transform; performance analysis

交通流状态判别作为智能交通系统(ITS)关键技术之一,通过准确的交通流运行状态信息,实时反映道路服务水平,对先进的交通管理系统(ATMS)、先进的出行者信息系统(ATIS)和车辆诱导系统提供策略依据.早期交通流状态判别研究多针对于偶发性交通突发事件的检测,采取的方法有加利福尼亚算法、双指数平滑算法、基于突变理论下的 McMaster 算法等^[1].目前交通流状态判别更多地集中在常发性拥挤方面,常发性拥挤交通流状态判别,主要从状态判别指标、状态判别方法等方面开展.如文献[2]以行程时间参数为基础,构建交通拥挤指数,实现道路拥挤的自动识别;文献[3]利用地点交通流数据和区域

交通流数据,并将二者交通特征融合进行分析,展开对城市交通流状态判别;在状态判别方法方面,提出了多种识别算法,如基于BP神经网络的交通流状态判别方法^[4],基于支持向量机交通流状态模式识别算法^[5],基于聚类方法的交通流状态辨识分析等^[6].

聚类分析方法是目前在状态类型划分过程中应用最为广泛的一种,聚类分析是一种无监督式的机器学习算法,在无任何先验信息情况下实现数据间的模式划分.文献[7]借鉴模糊 C 均值(FCM)聚类算法思想,选择高速公路上收费数据作为分类指标,对交通流状态进行聚类判别;文献[8]以FCM聚类算法思想为基础,通过设置交通参数不同权重情景,提出一种基于参数权重不同条件下的聚类方法,对交通聚类状态进行划分;文献[9-10]以高速公路交通状态聚类划分为研究对象,对比分析3种不同的聚类算法应用性能,由聚类效果判定 K 均值聚类算法分类效果最佳.为更好识别状态分类特征,进一步构建聚类组合

收稿日期:2018-05-17

基金项目:国家自然科学基金(71701046);
江苏省研究生科研与实践创新计划(KYCX18_0151)

作者简介:宋占国(1991—),男,博士研究生;
黄 卫(1961—),男,教授,博士生导师

通信作者:宋占国,zhanguo_song@163.com

模型进行分析.文献[11]将K均值聚类模型与多分类SVM模型相结合,通过聚类算法将高速公路状态进行模糊分类处理,再通过SVM模型对每一特征进行分析识别;文献[12]借助机器学习算法对状态类型进行分类,监督学习算法再以已分类信息作为先验信息进行分类识别;文献[13]将动态聚类方法和投影寻踪技术相结合,建立投影指标函数,采用混合蛙跳算法对投影指标函数不断优化,确定最佳投影方向,作为交通流状态判别阈值;文献[14]提出一种新型聚类方法谱聚类,通过谱图对类型进行划分,在划分结果基础上借助随机子空间集成K最近邻思想,组合一种谱聚类与随机子空间集成K近邻(RS-KNN)模型对快速路交通流状态进行判别模型.

然而,上述聚类组合模型对交通流状态的判别在数据量充实条件下效果更好,但由于无检测器装备或检测器设备损坏以致无法获取大量数据时,上述方法对状态类型的判别效果将受到影响.因此,文章提出一种在数据量较少条件下的交通流状态判别方法,即结合灰聚簇与Fisher变换的组合判别模型.灰聚簇与其他聚类模型相比,能够在较少数据下,划分出类内相似度较高,类间相似度较低的类型;而Fisher变换则可利用较少的数据对特征进行辨识,两者结合,能够在数据量较少的条件下对交通流状态特征进行判别.

1 灰聚簇与Fisher变换

1.1 灰聚簇

采用灰聚簇(grey clustering, GC)方法,对城市交通流快速路状态进行类型划分,灰聚簇主要涉及两部分内容,一是计算交通流参数间灰色关联度,作为分类的依据;二是确定灰类划分数量,作为分类的强度^[15].

1.1.1 灰色关联度

灰色关联度可为不同因素之间在数值关系上提供度量标准,在满足整体性、规范性、接近性、偶对称性等4个基础上,按照以下步骤确定参数灰色关联度,其中定义 $X = \{x_i \mid i \in n\}$ 为灰色关联子集, $x_{i,j} \in X$ 为比较序列对, $x_{i,j}(k)$ 为 $x_{i,j}$ 在 k 点数值, $\hat{r}(x_i(k), x_j(k))$ 为任意序列 x_i 对于序列 x_j 在 k 点灰色关联系数, $r(x_i(k), x_j(k))$ 为 m 个灰色关联系数平均值,为灰色关联度.

步骤1 首先对采集的交通流参数数据按照统计的时间间隔进行粒子区间划分,一般选取5 min间隔作为粒子区间的划分标准,划分形式为 $[t_k \ t_{k+1}]$,划分区间的数量 n 作为初步类型数量 n ,且类型中心为其本身.

步骤2 确定数据中两两之间灰色系数^[16]为

$$\hat{r}(x_i(k), x_j(k)) = \frac{\min_{j \neq i} \max_{k=1}^n |x_i(k) - x_j(k)| + \delta \max_{j \neq i} \max_{k=1}^n |x_i(k) - x_j(k)|}{|x_i(k) - x_j(k)| + \delta \max_{j \neq i} \max_{k=1}^n |x_i(k) - x_j(k)|} \quad (1)$$

步骤3 由灰色关联系数的平均值确定各因子灰色关联度为

$$r(x_i(k), x_j(k)) = \sum_{k=1}^m \frac{\hat{r}(x_i(k), x_j(k))}{m} \quad (2)$$

步骤4 进一步确定灰色初始相似度矩阵为

$$R = \begin{Bmatrix} 1 & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & 1 & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,1} & r_{n,2} & \cdots & 1 \end{Bmatrix} \quad (3)$$

步骤5 由灰色关联度计算结果,确定类与类间关联度最为接近的两类,合并为新类型,则总类型数量减少一类.

步骤6 确定新类型与原类型之间关联度为

$$r_{(pq)t} = \frac{r_{pt} + r_{qt}}{2} \quad (4)$$

式中: p, q, t 为原类型, pq 为合并的新类型, r_{pt} 为类型 p, t 之间关联度, r_{qt} 为类型 q, t 之间关联度, $r_{(pq)t}$ 为新类型 pq, t 之间关联度.

步骤7 确定新关联度后继续进行步骤5、步骤6步骤,按照类型数量划分判别标准,获得最佳分类数量.

1.1.2 灰类划分数量

选择统计量RSQ及半偏相关统计量SPRSQ作为灰类型数量划分最佳的标准.统计量RSQ可用来评价类型间合并聚类效果,若 $C_{RSQ} = 1$ 表明所有类型各自为1类,若 $C_{RSQ} = 0$ 表明所有类型合并为1类,这是类型划分的两种极端现象,且认为RSQ值显著减少所对应类型数为最佳分类结果,计算公式为

$$\begin{cases} S_i = \sum_{i=1}^{n_i} (x_{it} - \bar{x}_i)^T (x_{it} - \bar{x}_i), \\ P_k = \sum_{i=1}^k S_i, \\ M = \sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x}) = \sum_{i=1}^n \| (x_i - \bar{x}) \|^2, \\ C_{RSQ} = 1 - \frac{P_k}{M}. \end{cases} \quad (5)$$

半偏相关统计量SPRSQ为统计量RSQ的差值,且该统计量值越大表明上一次合并效果更好,计算公式为

$$\begin{cases} S_{p,q,r} = \sum_{i \in G_{p,q,r}} (\mathbf{x}_i - \bar{\mathbf{x}}_{p,q,r})^T (\mathbf{x}_i - \bar{\mathbf{x}}_{p,q,r}), \\ \bar{\mathbf{x}}_r = \frac{n_p \bar{\mathbf{x}}_p + n_q \bar{\mathbf{x}}_q}{n_r} = \frac{n_p \bar{\mathbf{x}}_p + n_q \bar{\mathbf{x}}_q}{n_p + n_q}, \\ W_{pq} = S_r - S_p - S_q = \frac{n_p n_q}{n_r} (\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_q)^T (\bar{\mathbf{x}}_p - \bar{\mathbf{x}}_q), \\ C_{\text{SPRSQ}} = \frac{W_{pq}}{M}. \end{cases} \quad (6)$$

式中: S_i 为类内离差平方和, 此值越小相似度越近; P_k 为所分 k 类离差平方总和; $S_{p,q,r}$ 分别为特征向量 G_p 、 G_q 类型及二者合并为 G_r 类型后的类内离差平方和; W_{pq} 为合并新类型后离差平均和增量。

1.2 Fisher 变换

Fisher 判别^[17]源于 20 世纪 30 年代, 通过将多维指标数据投影至某一方向中, 得到一个由 p 维变量组成的线性函数, 该线性函数按照增大不同类间函数差异值, 缩小同类内各点函数差异值的方式, 确保类与类间尽可能区分, 并由亲疏程度判别未知数据归属。文中采用多总体 Fisher 判别结构, 是在两总体 Fisher 判别结构基础上扩展而来, 计算公式^[18]为

$$y(x) = c_1 x_{1q} + \cdots + c_p x_{pq} = \sum_{i=1}^p c_i x_{iq}. \quad (7)$$

式中: x_{pq} 为在第 p 维变量 q 组数据的观测值, $c_1, \cdots, c_p$ 为 p 维变量所对应的判别系数, $y(x)$ 为多维转变为低维的判别函数值。

进一步可以获得在 $G_1, G_2, \cdots, G_k$ 总体下, 判别函数 $y^{(G_i)}(x)$ 值, 计算公式为

$$y^{(G_i)}(x) = c_1 x_{1q}^{(G_i)} + \cdots + c_p x_{pq}^{(G_i)} = \sum_{i=1}^p c_i x_{iq}^{(G_i)}. \quad (8)$$

为确定 $y^{(G_i)}(x)$ 判别函数的具体形式, 首先确定判定系数 c_i 值, 则

$$I = \frac{\sum_{i=1}^k n_i (\bar{y}^{(G_i)} - \frac{1}{k} \sum_{i=1}^k \bar{y}^{(G_i)})^2}{\sum_{i=1}^k \sum_{q=1}^n (y_q^{(G_i)} - \bar{y}^{(G_i)})^2}. \quad (9)$$

由式(9)按照类间差值越大、类内差值越小的基本原则, 计算 $\max I$ 所对应的特征值, 将特征值对应的特征向量, 作为判别函数的最佳系数 c_i 。结合式(8), 并对式(9)分子分母单独分析, 得到

$$\begin{cases} \mathbf{B} = \sum_{i=1}^k n_i (\bar{\mathbf{x}}^{(G_i)} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}^{(G_i)} - \bar{\mathbf{x}})^T, \\ \mathbf{E} = \sum_{i=1}^k \sum_{q=1}^n \sum_{p=1}^m (\mathbf{x}_{pq}^{(G_i)} - \bar{\mathbf{x}}_p^{(G_i)}) (\mathbf{x}_{pq}^{(G_i)} - \bar{\mathbf{x}}_p^{(G_i)})^T, \\ I = \frac{\mathbf{cBc}^T}{\mathbf{cEc}^T}. \end{cases} \quad (10)$$

对式(10)求系数 $\mathbf{c}$ 的偏导数, 并令偏导数值取 0 所对应的 c_i 为最佳系数, 即

$$\frac{\partial I}{\partial \mathbf{c}} = \partial \left(\frac{\mathbf{cBc}^T}{\mathbf{cEc}^T} \right) / \partial \mathbf{c} = \frac{2\mathbf{Bc}}{\mathbf{cEc}^T} - \frac{2\mathbf{IEc}}{\mathbf{cEc}^T} = 0. \quad (11)$$

由式(11)可知 $\mathbf{Bc} = \mathbf{IEc}$ 中 $\mathbf{I}$ 、 $\mathbf{c}$ 分别作为 $\mathbf{B}$ 、 $\mathbf{E}$ 广义特征根及所对应的特征向量, 从而确定判别函数。但由于所研究 Fisher 判别分类组数为 k , 则仅一个判别函数并不能满足种类判别需求, 因此按照非零特征根的从大到小依次对应的特征向量作为构建的判别函数系数, 构建 m ($m < k$) 个判别函数。其中, 特征根的数量需要满足累计判别效率达到 85% 以上的要求, 特征根累计判别效率的计算为

$$\sum_{i=1}^k \lambda_i / \sum_{i=1}^p \lambda_i \leq 85\%. \quad (12)$$

面对多种类样本数据情景, 对新样本归属类型进行判定, 应遵循规则: 设 $y_i(x)$ 为第 i 个线性判别函数, 计算新样本与以确定样本 $y_i^{(G_k)}(\bar{x}_k)$ 中距离之差, 差值最小者为新样本归属的类型。新样本归属类型计算公式为

$$\begin{cases} d(x, G_k) = \sum_{i=1}^r (y_i(x) - y_i^{(G_k)}(\bar{x}_k))^2; \\ d(x, G_l) = \min_{1 \leq j \leq k} d(x, G_k), \quad x \in G_l. \end{cases} \quad (13)$$

2 GC-Fisher 模型构建框架

将灰聚簇分类结果采用多总体 Fisher 变换分析, 构建 GC-Fisher 判别模型, 该模型流程分为以下步骤。

步骤 1 选取交通状态参数中交通量 Q 、速度 v 、占有率 O 三个参数作为判别的特征量, 采集相关数据并进行统一的标准化处理, 处理方式

$$x'_{ij} = \frac{x_{ij} - x_{j,\min}}{x_{j,\max} - x_{j,\min}}. \quad (14)$$

式中: $x_{j,\max}, x_{j,\min}$ 分别表示特征量 j 在采集序列中的最大值和最小值, x'_{ij} 为归一化后的数值。

步骤 2 由参数之间灰色关联度确定类型间合并依据, 由统计量 RSQ、SPRSQ 值, 确定状态划分的最佳分类数量 k , 可设置相对应标签加以区分, 如状态类型 1, 2, $\cdots, k$ 等。

步骤 3 通过灰聚簇已确定总体样本类型数量为 k , 选择标准化后的交通量 Q 、速度 v 、占有率 O 作为三维指标, 每一维指标选取 n 个样本数目, 构造 Fisher 变换训练集, 训练集输出多总体下的 Fisher 线性判别映射函数。

步骤 4 采用实验数据训练上述 GC-Fisher 模型, 输入新样本类型的三维指标数据, 通过应用多总

体下的 Fisher 线性判别映射函数,由亲疏远近程度差异输出最终样本数据状态归属类型.

3 实例分析

3.1 数据描述

选择长沙市湘府路作为研究对象,采集交通流量 Q 、速度 v 、占有率 O 连续 5 d 工作日(2016 年 11 月 14 日至 2016 年 11 月 18 日)17:00—20:00 数据,其中设置连续前 4 d 的数据作为训练数据,第 5 d 数据作为测试数据进行分析. 数据采集间隔为 5 min,则连续的 3 h 数据共划分 36 个研究单元,统计结构形式为 $[t_k \quad t_{k+1}]$,其中 $k \in [0, 36]$.

$$\begin{cases} q_i = \sum_{j=1}^n q_{ij}, \\ v_i = \left(\sum_{j=1}^n q_{ij} v_{ij} \right) / \sum_{j=1}^n q_{ij}, \\ O_i = \left(\sum_{j=1}^n q_{ij} o_{ij} \right) / \sum_{j=1}^n q_{ij}. \end{cases} \quad (15)$$

式中: q_i 、 v_i 、 O_i 分别代表断面 i 上的流量、速度、占有率, j 为断面车道编号, n 为断面车道数量.

通过式(15)可将采集的车道交通流数据转化为断面交通流数据.

3.2 基于灰聚簇的状态类型划分

对采集的交通流量 Q 、速度 v 、占有率 O 数据进行

标准化处理,在处理结果基础上获得对应参数的灰色关联度矩阵. 灰色关联矩阵 R_Q 、 R_v 、 R_O 分别为

$$\begin{aligned} R_Q &= \begin{bmatrix} 1 & 0.65 & \cdots & 0.44 \\ 0.65 & 1 & \cdots & 0.58 \\ \vdots & \vdots & \ddots & \vdots \\ 0.44 & 0.58 & \cdots & 1 \end{bmatrix}, \\ R_v &= \begin{bmatrix} 1 & 0.59 & \cdots & 0.33 \\ 0.59 & 1 & \cdots & 0.45 \\ \vdots & \vdots & \ddots & \vdots \\ 0.33 & 0.45 & \cdots & 1 \end{bmatrix}, \\ R_O &= \begin{bmatrix} 1 & 0.55 & \cdots & 0.51 \\ 0.55 & 1 & \cdots & 0.61 \\ \vdots & \vdots & \ddots & \vdots \\ 0.51 & 0.61 & \cdots & 1 \end{bmatrix} \end{aligned}$$

计算统计量 RSQ 及半偏相关统计量 SPRSQ,并观察其变化趋势,确定各参数最佳分类数量.

图 1 为 Q 、 v 、 O 统计量 RSQ 及半偏相关统计量 SPRSQ 分类变化信息,由统计量判定标准,确定最佳分类结果. 其中,统计量 RSQ 判定标准为该值显著减少时所对应的分类数即是分类数的最佳结果;而统计量 SPRSQ 则选择该值最大时的上一次合并类型数目作为最佳分类结果. 结合图 1 统计量 RSQ 和 SPRSQ 值变化,及二者判定标准,确定交通流状态类型最佳划分数量为 4.

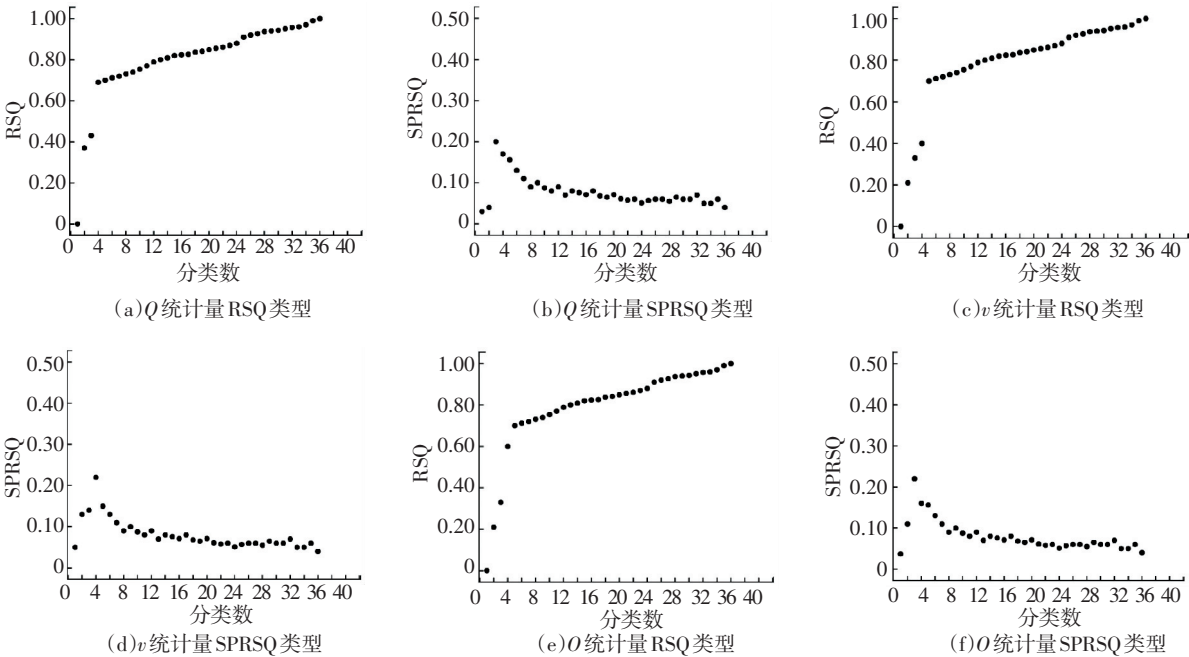


图 1 分类结果图

Fig.1 Diagram of classification result

在确定 4 种状态聚类类型的基础上,借助 R 语言软件平台,获取采集数据 Q 、 v 、 O 在每一类型下的聚类源,其中表 1 表示 4 种不同状态下 Q 、 v 、 O 的聚

类源. 进一步可由 4 种状态聚类结果,获得每一状态下参数之间对应变化关系,主要包括 Q - v 、 Q - O 及 v - O 等变化关系,如图 2 所示,根据变化特征

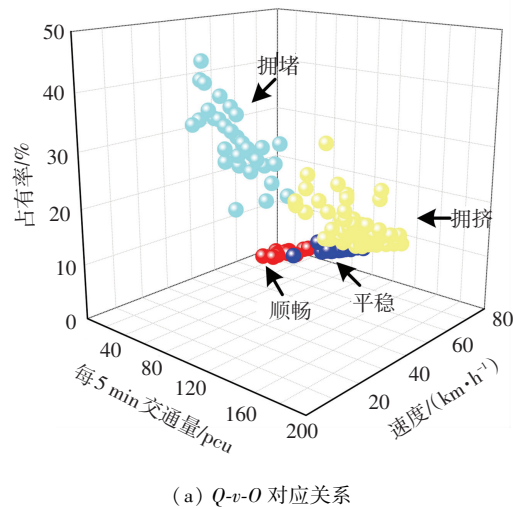
可明显判断聚类簇与簇之间界限特征.

表 1 不同状态下各参数聚类中心

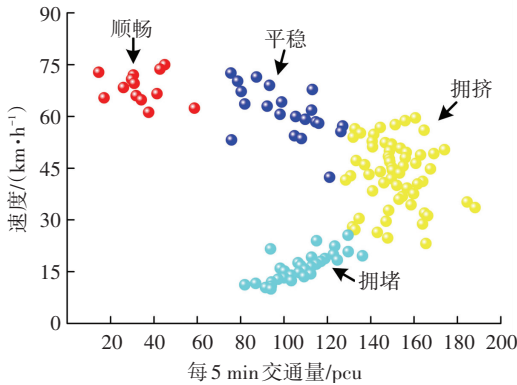
Tab.1 Clustering centers of various parameters under different conditions

状态类型	状态特征	交通量/ (pcu · min ⁻¹)	速度/ (km · h ⁻¹)	占有率/%
1	顺畅	7	68.4	5.61
2	平稳	20	54.8	15.83
3	拥挤	26	40.7	27.17
4	拥堵	32	18.2	37.89

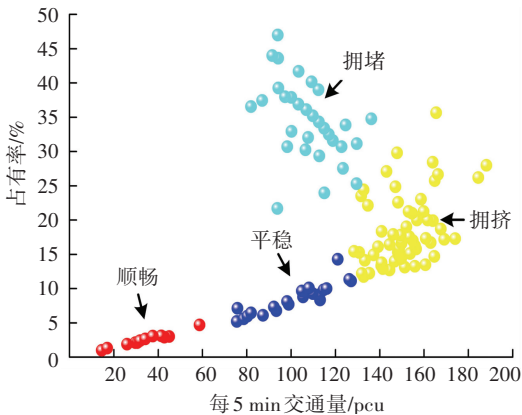
图 2 信息能够表明簇与簇间分类特征较为明显,其中顺畅状态下,交通量总体水平较少,车辆间干扰能力较弱,占有率低,车辆运行处于自由行驶状态;平稳状态下,随着交通量数量增加,车辆间干扰程度加深,占有率和速度分别呈现升高和下降趋势;拥挤状态下,车辆间干扰程度进一步加剧,交通量增加速度缓慢,占有率上升和速度下降的趋势明显;拥堵状态下,车辆间运行受到严重程度干扰,交通量减小,占有率急剧升高,速度急剧下降,交通流运行呈现极不稳定状态. 因此,通过灰聚簇划分的状态类型能够很好表征交通流在不同阶段下运行特征.



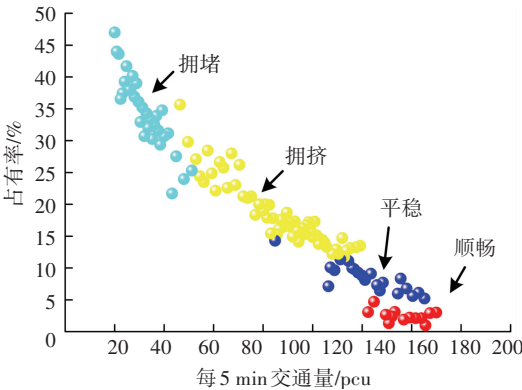
(a) Q - v - O 对应关系



(b) Q - v 对应关系



(c) Q - O 对应关系



(d) v - O 对应关系

图 2 参数变化特征

Fig.2 Parameter change characteristics

将每一时段对应状态类型归类之后,则可判断连续 5 d 在 17:00—20:00 时间段中交通流状态变化趋势. 分别定义 3 个坐标轴代表信息, x 轴表示一天之内状态变化趋势,选择每 5 min 作为一个单元,将 17:00—20:00 划分为 1-36 个整数单元; y 轴表示一周中对应的日期; z 轴表示 4 种状态类型,用编号 1、2、3、4 表示. 由图 3 连续 5 d 状态类型变化趋势,sl 可发现变化特征具有一定相似性,整体呈现“凸型”变化,从状态 2 逐渐过渡为状态 4,再由状态 4 逐渐变为状态 1,变化过程中虽存在不同程度波动,但整体较为稳定,符合交通流变化规律特征,因此,采用灰聚簇对状态类型划分具有一定有效性.

3.3 GC-Fisher 训练结果

选择前 4 d 状态数据作为基础构建训练集,对状态类型的判别展开分析. 根据训练数据集状态类型聚簇结果,确定样本总体类型为 4,指标维度 p 为 3,为交通量 Q 、速度 v 、占有率 O ,构建 Fisher(4,3) 转换模型的判别函数,其中训练集分类结果见表 2.

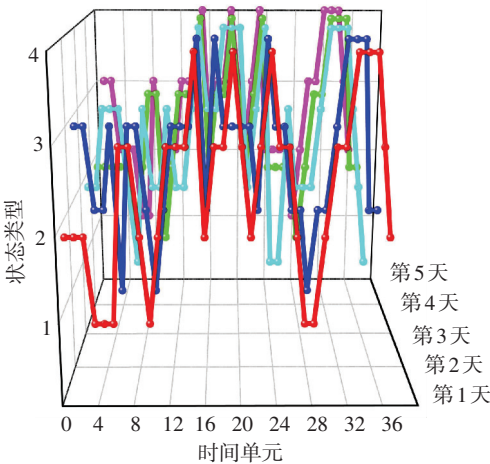


图 3 每天交通流状态变化趋势图

Fig.3 Diagram of daily traffic flow status trend

表 2 训练集样本数据类统计量(标准化后)

Tab.2 Statistical data of training set sample (after standardization)

状态类型	参数	均值	标准差
1	Q	0.110	0.067
	v	0.898	0.067
	O	0.033	0.021
2	Q	0.788	0.074
	v	0.514	0.146
	O	0.380	0.113
3	Q	0.538	0.076
	v	0.097	0.064
	O	0.727	0.125
4	Q	0.498	0.095
	v	0.786	0.112
	O	0.164	0.049

由表 2 分别获得 4 组样本类型的均值和标准差,经 Fisher 转化得到每一样本类型下的 $y_i^{(G_k)}(x_k)$ 值,其中转换方式为

$$\begin{cases} y_i^{(G_1)} = -1.10Q_i + 213.01v_i + 205.89O_i - 100.42, \\ y_i^{(G_2)} = 109.36Q_i + 187.03v_i + 199.92O_i - 130.44, \\ y_i^{(G_3)} = 60.45Q_i + 173.26v_i + 226.01O_i - 108.24, \\ y_i^{(G_4)} = 62.7Q_i + 208.08v_i + 205.46O_i - 115.62. \end{cases}$$

(16)

转换方式确定之后,需进一步确定判别函数.训练数据集中分类组数 k 为 4,因此需要构建多个判别函数才能保证样本类型准确识别.共确定相对应的非零特征根共 3 个,根值分别为 6.117、4.26、2.629,其相对应的方差为 47.03%、32.75%、20.21%,满足 85%以上累计判别效率要求.由此,需要构建 3 个判别函数数量,由特征根值得到标准化典型判别

式函数系数,判别函数为

$$\begin{cases} F1 = 0.894Q_i + 0.49v_i - 0.031O_i, \\ F2 = -0.421Q_i + 0.487v_i + 1.676O_i, \\ F3 = -0.040Q_i - 0.495v_i + 1.746O_i. \end{cases} \quad (17)$$

测验数据由式 (17) 得到 3 种判别函数值,与 4 种样本类型的转换平均值 $\bar{y}^{(G_1)}$ 、 $\bar{y}^{(G_2)}$ 、 $\bar{y}^{(G_3)}$ 、 $\bar{y}^{(G_4)}$ 处理得到距离差,选择距离差值的和最小作为归属类型.

3.4 模型性能测试

选择第 5 d 数据进行分析判断构建测验集,测验 GC-Fisher 模型对交通流状态类型判别性能效果.本文针对模型性能测试采用对比分析方式,选择 K-SVM 组合判别模型作为比较, K-SVM 模型 RBF 核函数参数定值取 2.2,惩罚系数确定取 10.5^[7].以矩阵形式展示判别效果信息,其中主对角线表示正确判别类别信息,非主对角线表示误判类别信息,数字表示判别的个数,百分比指该个体数量占总体数量比重,灰色方格中上面百分比为判别率,下面百分比为误判率.

结合图 4 提供的判别效果比较信息,两种判别模型均存在误判情况.当所采集的数据量较少时,采用 GC-Fisher 模型对交通流状态判别率为 92%,而 K-SVM 模型的交通流状态判别率为 69%.因此,当数据量较少时,可采用 GC-Fisher 组合判别模型对交通流状态类型进行判别.

9	1	0	0	90%
25%	3%	0%	0%	10%
0	9	1	0	90%
0%	25%	3%	0%	10%
0	1	4	0	80%
0%	3%	11%	0%	20%
0	0	0	11	1
0%	0%	0%	31%	0
1	82%	80%	1	92%
0	8%	20%	0	8%

目标交通状态
(a) GC-Fisher

7	2	1	0	70%
19%	6%	3%	0%	30%
1	6	2	0	67%
3%	17%	6%	0%	33%
0	0	3	1	75%
0%	0%	8%	3%	25%
0	2	2	9	69%
0%	6%	6%	25%	31%
88%	60%	38%	90%	69%
12%	40%	62%	10%	31%

目标交通状态
(b) K-SVM

图 4 判别效果比较图

Fig.4 Comparison of discriminant effect

4 结 论

1) 选择交通量 Q 、速度 v 、占有率 O 作为参数指标, 并利用灰聚簇理论中灰色关联度方法, 结合统计量 RSQ 及半偏相关统计量 SPRSQ 提供的灰类划分标准, 确定交通流的 4 种状态类型, 并给出 4 种状态的聚类源。

2) 以 4 种交通流状态类型作为总体, 提出了一种基于灰聚簇与 Fisher 变换 (GC-Fisher) 相结合的状态类型判别模型。

3) 选择“K-SVM”作为比较模型, 通过构造数据训练集和测验集, 对判别性能进行分析。结果表明 GC-Fisher 模型的判别率为 92%, 优于 K-SVM 模型的判别率 69%。因此在数据量较少情况下, GC-Fisher 模型对状态类型判别具有更高的判别精度。

4) 仅选择长沙市湘府路作为研究对象较为单一, 下一阶段研究仍需从道路空间特征方面着手, 选取不同快速路验证模型的普适性, 同时仍需进一步设置在不同数据样本量情景下, 对 GC-Fisher 模型的数据样本量敏感程度进行分析。

参考文献

- [1] 姜桂艳. 道路交通状态判别技术[M]. 北京: 人民交通出版社, 2004: 66
JIANG Guiyan. Road traffic condition discrimination technology [M]. Beijing: China Communications Press, 2004: 66
- [2] 杨聚芬, 姜桂艳, 李琦. 基于收费数据的高速公路交通拥挤自动判别方法[J]. 哈尔滨工业大学学报, 2014, 46(12): 108
YANG Jufen, JIANG Guiyan, LI Qi. The automatic traffic congestion identification of freeway based on charging data[J]. Journal of Harbin Institute of Technology, 2014, 46(12): 108
- [3] BHASKAR A, TSUBOTA T, KIEU L M. Urban traffic state estimation: fusing point and zone based data[J]. Transportation Record Part C: Emerging Technologies, 2014, 48: 120
- [4] 巫威眺, 靳文舟, 林培群. 基于 BP 神经网络的道路交通状态判别方法研究[J]. 交通信息与安全, 2011, 29(4): 71
WU Weitiao, Jin Wenzhou, LIN Peiqun. Research on road traffic condition discrimination based on BP neural network[J]. Traffic Information and Security, 2011, 29(4): 71
- [5] 于荣, 王国祥, 郑继媛. 基于支持向量机的城市道路交通状态模式识别研究[J]. 交通运输系统工程与信息, 2013(1): 130
YU Rong, WANG Guoxiang, ZHENG Jiyuan. Research on urban road traffic pattern recognition based on support vector machines[J]. Transportation Systems Engineering and Information, 2013(1): 130
- [6] 殷俊杰, 丁宏飞, 薄雾, 等. 基于模糊聚类的城市快速路交通流状态划分[J]. 重庆交通大学学报(自然科学版), 2013, 32(4): 652
YIN Junjie, DING Hongfei, BO Wu, et al. Classification of urban expressway traffic flow based on fuzzy clustering[J]. Journal of Chongqing Jiaotong University (Natural Science Edition), 2013, 32

- (4): 652
- [7] 杨庆芳, 马明辉, 梁士栋. 基于收费数据的高速公路交通状态判别方法[J]. 华南理工大学学报(自然科学版), 2014, 42(12): 51
YANG Qingfang, MA Minghui, LIANG Shidong. Distinguishing method of freeway traffic status based on toll data[J]. Journal of South China University of Technology (Natural Science Edition), 2014, 42(12): 51
- [8] 张亮亮, 贾元华, 牛忠海. 交通状态划分的参数权重聚类方法研究[J]. 交通运输系统工程与信息, 2014, 14(6): 147
ZHANG Liangliang, JIA Yuanhua, NIU Zhonghai. Clustering method of parameter weights for traffic conditioning[J]. Journal of Transportation Systems Engineering and Information, 2014, 14(6): 147
- [9] AZIMI M, ZHANG Y L. Categorizing freeway flow conditions by using clustering methods[J]. Transportation Research Record: Journal of the Transportation Research Board, 2010, 2010(2173): 105
- [10] MONTAZERI-GH M, FOTOUHI A. Traffic condition recognition using the k-means clustering method[J]. Scientia Iranica, 2011, 18(4): 930
- [11] DENG C, WANG F, SHI H M. Real-time freeway traffic state estimation based on cluster analysis and multiclass support vector machine[C]//Proceedings of International Workshop on Intelligent Systems and Applications (ISA). Wuhan: IEEE, 2009
- [12] ANTONIOU C, KOUTSOPOULOS H N, YANNIS G. Dynamic data-driven local traffic state estimation and prediction[J]. Transportation Research Part C: Emerging Technologies, 2013, 34: 89
- [13] 邴其春, 龚勃文, 杨兆升, 等. 基于投影寻踪动态聚类的快速路交通状态判别[J]. 西南交通大学学报, 2015, 50(6): 1164
BING Qichun, GONG Bowen, YANG Zhaosheng, et al. Expressway traffic state discrimination based on projection pursuit dynamic clustering[J]. Journal of Southwest Jiaotong University, 2015, 50(6): 1164
- [14] 商强, 林赐云, 杨兆升, 等. 基于谱聚类与 RS-KNN 的城市快速路交通状态判别[J]. 华南理工大学学报(自然科学版), 2017, 45(6): 52
SHANG Qiang, LIN Ciyun, YANG Zhaosheng, et al. Identification of urban expressway traffic State based on spectral clustering and RS-KNN[J]. Journal of South China University of Technology (Natural Science Edition), 2017, 45(6): 52
- [15] 田民, 刘思峰, 卜志坤. 灰色关联度算法模型的研究综述[J]. 统计与决策, 2008(1): 24
TIAN Min, LIU Sifeng, BU Zhikun. A review of research on grey correlation algorithm model[J]. Statistics and Decision, 2008(1): 24
- [16] 李善梅, 徐肖豪, 王超, 等. 基于灰色聚类的交叉航路拥挤识别方法[J]. 西南交通大学学报, 2015, 50(1): 189
LI Shanmei, XU Xiaohao, WANG Chao, et al. Congestion identification of crossing air routes based on grey clustering method[J]. Journal of Southwest Jiaotong University, 2015, 50(1): 189
- [17] TANG W J, GONG F, DONG R R, et al. Face recognition based on the fusion of wavelet packet sub-images and fisher linear discriminant[J]. Multimedia Tools and Applications, 2017, 76: 22725
- [18] YE H S, LI Y J, CHEN C. Fast Fisher discriminant analysis with randomized algorithms[J]. Pattern Recognition, 2017, 72: 82

(编辑 魏希柱)