

基于样本和特征加权 FCM 的交通状态识别^{*}

孙广林, 刘君

(公安部道路交通安全研究中心, 北京 100062)

摘要: 利用交通量、速度、占有率等交通参数进行聚类识别道路交通状态, 参数样本及特征对聚类结果具有不同的作用。为改进传统 FCM 聚类假定数据样本及特征同等重要的缺陷, 选取交通量、平均速度、空间占有率 3 个交通参数, 划分交通状态为自由流、拥堵流和阻塞流, 提出基于样本和特征双加权 FCM 的交通状态识别方法, 采用拉格朗日乘数法动态更新隶属度、样本与特征权值, 进一步设计双加权 FCM 聚类算法。实例分析表明, 与传统 FCM 聚类结果对比, 双加权 FCM 聚类交通状态划分边界更清晰, 样本隶属度函数值接近 0/1 的数量增加 7%, 计算效率提高 1.6 倍, 交通状态识别结果更符合实际运行状态。

关键词: 交通工程; 交通状态识别; 双加权 FCM; 样本加权; 特征加权

中图分类号: U491

文献标志码: A

文章编号: 1671-2668(2020)05-0021-04

道路交通状态是表征交通运行状况的关键指标。准确识别当前交通状态是引导出行者合理选择出行方式及路径, 提高道路交通资源利用效率和防止交通状态持续恶化的重要依据。根据已有研究成果, 交通状态可划分为畅通和拥挤两类, 以此为基础可进一步细分为自由流、密集流、拥堵流和阻塞流等。交通状态识别主要依据交通量、速度、占有率、密度、排队长度、饱和度等基本交通参数, 采用聚类分析方法划分状态类型。实际应用中, 不同交通参数对交通状态聚类的贡献程度不一样, 且样本属性特征间具有不平衡性。传统的聚类分析方法假定所有数据样本及其特征具有一致性, 未体现样本及特征属性对聚类结果的影响。针对传统聚类算法理想化的假设条件, 提出多属性加权的改进聚类算法, 该方法主要采用样本概率分布和最大熵原理确定权重值。多采用信息熵原理确定各交通状态识别指标的权重或根据交通参数间的相似性构建评价函数, 通过最优化方法获得权值。样本与特征双加权聚类算法的提出及在图像识别等领域的应用, 补充了样本及特征对聚类结果的影响, 也为解决传统模糊 C 均值(Fuzzy C-means, FCM)聚类算法在交通状态识别应用中的弊端提供了路径。该文在确定交通状态类别和 FCM 算法的基础上, 利用样本和特征加权 FCM 聚类方法进行交通状态识别, 采用拉格朗日乘数法更新隶属度和簇中心及样本与特征权值, 进一

步设计双加权聚类算法, 通过实例对比分析双加权策略 FCM 与传统 FCM 聚类的效果和性能, 验证双加权策略 FCM 识别交通状态的有效性。

1 交通状态与 FCM 聚类算法

1.1 交通状态划分

传统交通流理论应用交通量、车速和密度描述交通流状态特征, 并将交通流划分为自由流和拥挤流两种状态。这里根据道路交通管理需要及出行者对交通运行的直观感受, 将交通状态划分为自由流、拥堵流和阻塞流 3 种类型。道路交通量、速度、密度是描述交通流基本特征的主要参数, 其中交通密度在实际应用中往往采用容易测量的车道占用率来间接表征, 包括空间拥有率和时间占用率。综合现有研究成果选用交通状态参数的频率及参数样本获取的便利性, 选取交通量、平均速度、空间占用率 3 个交通参数进行聚类分析, 识别交通状态类别。

(1) 自由流。自由流状态下交通流量小、车速高、路段车辆空间占有率低, 驾驶人具有较大的行车自由度, 车辆行驶几乎不受其他车辆影响, 道路交通处于畅通状态。

(2) 拥堵流。拥堵流状态下交通流量接近道路通行能力, 车速显著下降, 路段车辆空间占有率较高, 驾驶人驾车自由度明显受限, 行驶车辆之间相互影响, 道路交通处于不稳定状态, 交通流波动较大。

^{*} 基金项目: 公安部技术研究计划项目(2018JSYJB05)

(3) 阻塞流。阻塞流状态下交通流量呈断崖式下降,车辆间歇性停驶,车头时距趋于稳定,路段车辆空间占有率非常高,整个路段车辆处于同步跟驰状态,相邻车道车速基本一致。

1.2 FCM 聚类算法

FCM 聚类是基于目标优化的聚类方法。设给定的数据集 $X = \{X_1, X_2, \dots, X_n\}$ 包含 n 个样本,将样本分为 c 类, $1 < c < n$, 对任意 $X_i \in X$, 其特征矢量 $X_i = \{x_{i1}, x_{i2}, \dots, x_{ir}\}$, x_{ir} 表示第 i 个样本的第 r 个特征属性。应用隶属度函数及样本到簇中心的距离定义目标函数为:

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (u_{ij})^m (d_{ij})^2 \quad (1)$$

$$\text{s.t.} \begin{cases} 0 \leq u_{ij} \leq 1 \\ \sum_{i=1}^c u_{ij} = 1 \end{cases} \quad (2)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^c u_{ij} = 1 \end{cases} \quad (3)$$

式中: $J(U, V)$ 为样本到簇中心的距离加权值; $U = \{u_{ij}\}$, 为 $c \times n$ 阶模糊分类矩阵; $V = [v_1, v_2, \dots, v_c]$, 为 $p \times c$ 阶簇中心矩阵; v_i 为簇中心; u_{ij} 为第 j 个样本隶属于第 i 个簇的程度, $1 \leq i \leq n, 1 \leq j \leq c$; m 为加权指数; $(d_{ij})^2 = \|x_j - v_i\|^2$, 为样本到簇中心的欧式距离。

由样本距离模糊簇中心最小准则,可获得聚类目标表达式:

$$\min \{J(U, V)\} = \min \left\{ \sum_{i=1}^c \sum_{j=1}^n (u_{ij})^m (d_{ij})^2 \right\} \quad (4)$$

拉格朗日乘数法是求解变量在多个条件限制下多元函数极值的方法,具有较高的通用性和普适性。因此,通过构建拉格朗日函数,采用拉格朗日乘数法求解隶属度 u_{ij} 及簇中心 v_i 。

2 加权 FCM 聚类算法

基于式(1)~(4)FCM 聚类算法,为表征交通状态识别指标样本对聚类结果的影响及指标特征的作用,同时引入样本权值 α_j ($j = 1, 2, \dots, n$) 和特征权值 β_k ($k = 1, 2, \dots, c$), α_j 和 β_k 分别满足约束条件 $\sum_{j=1}^n \alpha_j = 1, \sum_{k=1}^c \beta_k = 1$ 。 p 和 q 为不等于 1 的整数。双加权 FCM 聚类方法表示为:

$$\min_{U, V} J(U, V) = \sum_{i=1}^c \sum_{j=1}^n \alpha_j^p u_{ij}^m \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2 \quad (5)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^c u_{ij} = 1 & (j = 1, 2, \dots, n) \end{cases} \quad (6)$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^n \alpha_j = 1 \end{cases} \quad (7)$$

$$\text{s.t.} \begin{cases} \sum_{k=1}^c \beta_k = 1 \end{cases} \quad (8)$$

2.1 隶属度与权值

2.1.1 隶属度的确定

由式(5)~(8),采用 Lagrange 方法确定隶属度。定义 f, g, h 为 Lagrange 乘子,其中 f 是由 Lagrange 乘子构成的向量, $f = (f_1, f_2, \dots, n)$ 。Lagrange 函数 L 表示为:

$$L(U, V; f, g, h) = \sum_{i=1}^c \sum_{j=1}^n \alpha_j^p u_{ij}^m \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2 - \sum_{j=1}^n f_j \left(\sum_{i=1}^c u_{ij} - 1 \right) - g \left(\sum_{j=1}^n \alpha_j - 1 \right) - h \left(\sum_{k=1}^c \beta_k - 1 \right) \quad (9)$$

将 Lagrange 函数 L 分别对 u_{ij} 和 f_j 求偏导,令 $\partial L / \partial u_{ij} = 0, \partial L / \partial f_j = 0$, 则:

$$m \alpha_j^p u_{ij}^{m-1} \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2 - f_j = 0 \quad (10)$$

$$\sum_{i=1}^c u_{ij} = 1 \quad (11)$$

由式(10)计算隶属度 u_{ij} , 得:

$$u_{ij} = \left[\frac{f_j}{m \alpha_j^p \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2} \right]^{1/(m-1)} \quad (12)$$

将 Lagrange 函数 L 对簇中心 v_{ik} 求导,由 $\partial L / \partial v_{ik} = 0$ 得到簇中心值:

$$v_{ik} = \frac{\sum_{j=1}^n \alpha_j^p u_{ij}^m x_{jk}}{\sum_{j=1}^n \alpha_j^p u_{ij}^m} \quad (13)$$

联合式(11)、式(12)求解隶属度 u_{ij} , 得:

$$u_{ij} = \frac{1}{\sum_{i=1}^c \left[\frac{\sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2}{\sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2} \right]^{1/(m-1)}} \quad (14)$$

2.1.2 权值的确定

将 Lagrange 函数 L 分别对样本权值 α_j 、特征权值 β_k 求导,由 $\partial L / \partial \alpha_j = 0, \partial L / \partial \beta_k = 0$, 得样本与特征权值分别为:

$$\alpha_j = \frac{1}{\sum_{i=1}^n \left[\frac{\sum_{k=1}^c u_{ij}^m \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2}{\sum_{k=1}^c u_{ij}^m \sum_{k=1}^c \beta_k^q (x_{jk} - v_{ik})^2} \right]^{1/(p-1)}} \quad (15)$$

$$\beta_k = \frac{1}{\sum_{l=1}^n \left[\frac{\sum_{j=1}^c \alpha_j^p u_{lj}^m (x_{jk} - v_{ik})^2}{\sum_{j=1}^c \alpha_j^p u_{lj}^m (x_{jk} - v_{ik})^2} \right]^{1/(q-1)}} \quad (16)$$

2.2 算法实现

根据式(5)~(16),设计双加权 FCM 聚类算法(DFCM)如下:

(1) 初始化 m 、 p 、 q 、最大迭代次数 $Iter$ 、迭代终止条件 ϵ 、 u_{ij} ($1 \leq i \leq n, 1 \leq j \leq c$)、 α_j ($j = 1, 2, \dots, n$)、 β_k ($k = 1, 2, \dots, c$),对数据进行标准化处理。

(2) 若 $j < Iter$,由式(13)更新簇中心,由式(14)更新隶属度,由式(15)更新样本权值,由式(16)更新特征权值。计算相邻两次迭代簇中心向量差的范数值 $Err = \|v^j - v^{j-1}\|$,若 $Err < \epsilon$,则停止计算,否则 $j = j + 1$ 。算法结束。

3 实例分析

3.1 数据准备与初始化

3.1.1 数据准备

选取北京市西五环晋元桥—石景山路出口段南向交通视频数据作为聚类分析的数据源,该路段长 1 km,单向三车道,最高限速 80 km/h。视频数据记录时间为 2019 年 3 月 13 日 7:00—19:00,未发生交通事故或其他突发事件。以 5 min 为单位(共 145 个样本)统计该路段交通量、平均速度、空间占有率时间分布情况,结果见图 1~3。

从图 1~3 可以看出:测试路段 7:25—8:25、17:25—18:25 出现显著的流量高峰,流量波动幅度较大,交通运行处于拥堵流状态;中午时段流量相对

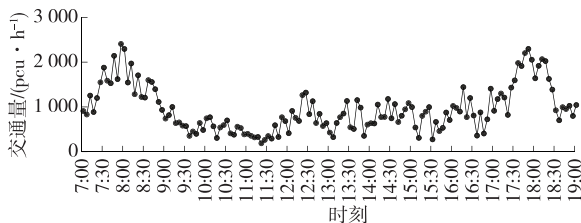


图 1 测试路段交通量时间分布

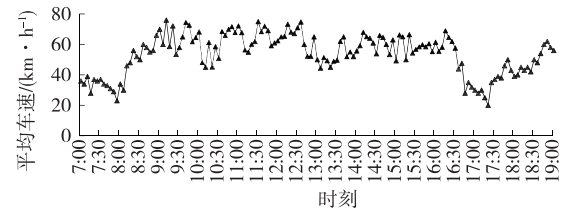


图 2 测试路段平均速度时间分布

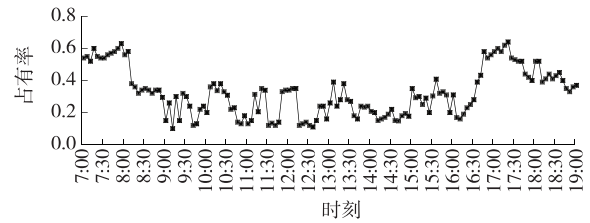


图 3 测试路段空间占有率时间分布

较低,对应的平均车速较高、道路空间占有率较低,交通运行处于自由流状态;17:20 平均速度降至最低 20 km/h,道路空间占有率达到最高 0.64,同时流量显著下降,交通运行处于阻塞流状态。测试路段 7:00—19:00 时段内具有不同的交通运行状态特征分布,源数据完整可用并具有良好的适用性。

3.1.2 参数初始化

根据路段交通状态参数的性质,设定双加权聚类算法参数初始值,隶属度矩阵指数 $m = 2.25$ 、最大迭代次数 $Iter = 100$ 、迭代终止条件 $\epsilon = 1 \times 10^{-5}$ 、 $p = 2$ 、 $q = 3$ 。采用 MATLAB 软件进行计算,随机生成隶属度矩阵 U 、样本权值 α_j ($j = 1, 2, \dots, 145$)、特征权值 β_k ($k = 1, 2, 3$),并进行标准化处理,令其满足 $\sum_{i=1}^3 u_{ij} = 1$ ($j = 1, 2, \dots, 145$)、 $\sum_{j=1}^{145} \alpha_j = 1$ 、 $\sum_{k=1}^3 \beta_k = 1$ 。

3.2 聚类效果分析

输入测试路段源数据及初始化参数值,采用 MATLAB 模糊聚类工具箱输出传统 FCM 聚类结果,对比分析 3 类交通状态隶属度函数值、聚类目标函数值,比较两类算法的聚类效果与计算效率。

3.2.1 簇中心矩阵

FCM 和 DFCM 聚类算法簇中心矩阵 V_1 、 V_2 为:

$$V_1 = \begin{bmatrix} 1020.9 & 54.5 & 0.3 \\ 1911.1 & 41.1 & 0.5 \\ 510.7 & 59.8 & 0.3 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} 963.3 & 53.9 & 0.3 \\ 989.9 & 53.7 & 0.3 \\ 1083.4 & 53.0 & 0.4 \end{bmatrix}$$

V_1 与 V_2 中,行表示交通状态,列表示状态识别指标,各类交通状态识别指标值均在交通量阈值[189,2 047]、平均速度阈值[20,76]、占有率阈值[0.1,0.64]内,簇中心可用于聚类计算。

3.2.2 隶属度值

聚类算法样本隶属度值接近 0/1 的数量可用来衡量聚类算法效果的优劣,隶属度值越接近 0/1,类内样本的相似度越高,不同类间的差异越显著,聚类效果愈显著。FCM 和 DFCM 聚类算法 3 类交通状态隶属度函数值见图 4~6。

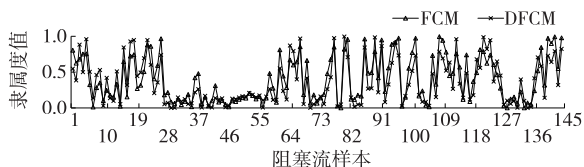


图4 FCM与DFCM阻塞流样本隶属度

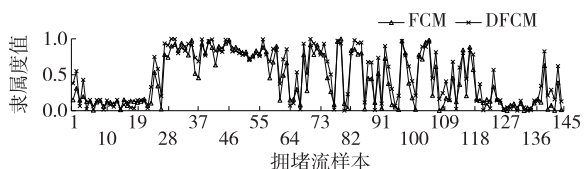


图5 FCM与DFCM拥堵流样本隶属度

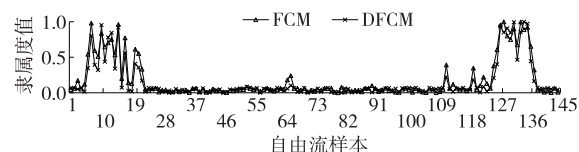


图6 FCM与DFCM自由流样本隶属度

3.2.3 目标函数

传统 FCM 和 DFCM 聚类算法目标函数值的变化见图 7。

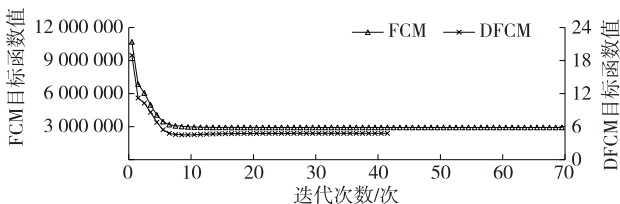


图7 FCM与DFCM的目标函数值

由图 7 可知:DFCM 聚类算法的目标函数值远小于 FCM 聚类算法,DFCM 聚类算法经过 43 次迭代即满足迭代终止条件,而 FCM 聚类算法经过 70 次迭代才满足迭代终止条件,DFCM 聚类算法的计算效率比 FCM 聚类算法提高 1.6 倍。

4 结论

引入样本权值和特征权值,提出双加权 FCM

的交通状态识别方法。针对实测路段交通视频数据,分别采用传统 FCM 与双加权 FCM 进行计算,并对聚类结果进行对比分析。结果表明:双加权 FCM 聚类算法获得的交通状态识别效果、相同状态内样本相似度、不同状态间的差异性及计算性能均优于传统 FCM 聚类算法,能有效体现数据样本和特征属性对交通状态聚类结果的影响。

参考文献:

- [1] CHRISTIANE S, THOMAS A. Classification and prediction of road traffic using application-specific fuzzy clustering[J]. IEEE Transactions on Fuzzy Systems, 2002, 10(3): 297-308.
- [2] LINDLEY J. Urban freeway congestion: Quantification of the problem and effectiveness of potential solution [J]. ITE Journal, 1987, 57(1): 27-32.
- [3] 陈利正, 吴聪. 多变量聚类分析的高速公路交通流状态实时评估[J]. 交通运输系统工程与信息, 2018, 18(3): 225-233.
- [4] 张亮亮, 贾元华, 牛忠海, 等. 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): 147-151.
- [5] 林甲祥, 吴丽萍, 巫建伟, 等. 基于样本与特征双加权的自适应 FCM 聚类算法[J]. 黑龙江大学自然科学学报, 2018, 35(2): 244-252.
- [6] 任福田, 刘小明, 荣建, 等. 交通工程学[M]. 北京: 人民交通出版社, 2003: 6.
- [7] 李凯, 高岩, 曹喆. 自动调整样本和特征权值的模糊聚类算法[J]. 哈尔滨工程大学学报, 2018, 39(9): 1554-1560.
- [8] 吴启顺, 蔡晓禹, 蔡明. 基于 FCM 快速路交通状态判别加权指数研究[J]. 科学技术与工程, 2017, 17(6): 289-295.
- [9] 曾真, 穆程. 城市道路交通瓶颈识别方法研究[J]. 公路与汽运, 2018(4): 17-20.
- [10] 刘小明, 王灏, 李颖宏, 等. 基于可拓决策理论的城市交通状态识别研究[J]. 武汉理工大学学报(交通科学与工程版), 2011, 35(2): 337-340.
- [11] 王宇俊, 田锋, 叶道均, 等. 改进 FCM 的交通状态判别算法[C]//中国城市规划设计研究院城市交通专业研究院. 2017 年中国城市交通规划年会论文集. 北京: 中国城市规划设计研究院城市交通专业研究院, 2017.
- [12] 梅朵, 杨庆芳, 鄂旭. 基于 MR-FCM 的区域交通状态识别方法[J]. 武汉理工大学学报(交通科学与工程版), 2018, 42(4): 564-568.