

# 粒子群优化的模糊聚类方法在车辆行驶工况中的应用

石 琴,王楠楠,仇多洋

(合肥工业大学交通运输工程学院,安徽 合肥 230009)

**摘 要:**本文研究了粒子群优化的模糊聚类方法在车辆行驶工况中的应用。采用主成分分析方法将众多反映车辆行驶工况特征的运动学片段特征值进行压缩,用粒子群优化的模糊聚类方法对运动学片段的前三个主成分得分进行聚类,通过 Matlab 编程将上述理论用于合肥市典型道路行驶工况的构建和分析,按时间比例选取合适片段拟合代表性工况,并将代表性工况和采用 K 均值聚类法及模糊 C 均值聚类方法拟合的工况进行对比分析。研究结果表明,将粒子群优化的模糊聚类方法应用到工况的构建中可以有效地提高构建精度。

**关键词:**粒子群;模糊聚类;主成分分析;K 均值聚类;行驶工况

**中图分类号:**U491.2+55 **文献标识码:**A

## 1 引言

随着能源危机和环境污染日益严重,以低能耗、低污染、低排放为基础的“低碳经济”成为全球热点。对于新车型的排放和燃油消耗的检验、车辆设计的动力匹配以及车辆的检测和维护等,需要有一个能够充分代表我国车辆实际行驶状况的行驶工况(Driving Cycle)。

近几十年来,众多学者对行驶工况进行了研究,提出了许多新方法。欧洲、美国和日本根据各国实际的交通状况构建出世界著名的三大工况(ECE15+EUDC、FTP75 和 JPAN10)。Michel 研究了由于车辆状况、驾驶行为以及道路几何条件的不同导致不同城市下的驾驶特性的不同<sup>[1]</sup>。Eva 研究了如何使用城市交通中实际采集到的工况数据来描述车辆行驶工况<sup>[2]</sup>。Lin 和 Dai 都选择使用马尔可夫随机过程方法构建行驶工况<sup>[3-4]</sup>。赵慧等对香港城区汽车行驶工况进行了研究<sup>[5]</sup>。马志雄等研究了动态聚类法在车辆实际行驶工况开发中的应用<sup>[6]</sup>。

目前我国采用欧洲的 ECE 工况,由于交通状况等方面存在较大差距,ECE 工况并不能真实反映我国城市道路车辆实际的行驶状况。所以,亟待研究开发出一个适合我国城市的工况。

本文将主成分分析法与粒子群优化的模糊聚类方法结合起来应用到行驶工况的构建中,并以合肥市典型道路为例进行了实际应用分析。研究表明,采用粒子群优化的模糊聚类方法不仅可以使运动学片段得到清晰的分类,而且构建的工况能够综合反映合肥市的交通状况。

## 2 研究方法

### 2.1 构建方法

在车辆行驶工况的构建中,主成分分析(Principal Component Analysis)和聚类分析(Cluster Analysis)得到普遍的应用。主成分分析可以降低行驶工况中多维数据的维数,减少运动学片段的参数量,聚类分析则可以有效反映不同道路上的行驶特征。常用的聚类方法有包括 K 均值聚类、分层聚类在内的传统聚类和以模糊 C 均值聚类为代表的模糊聚类等。K 均值聚类法对不同的随机初始值可能导致不同的聚类结果,而且它采用爬山式技术寻找最优解,容易陷入局部极小值。模糊 C 均值(Fuzzy C-Means,FCM)聚类算法作为一种局部搜索算法,存在以下缺点:对初始聚类中心和噪声敏感,容易陷入局部极小值,得不到全局最优解。粒子群优化

收稿日期:2010-08-31;修订日期:2010-12-26

基金项目:国家自然科学基金资助项目(70771036);国家自然科学基金资助项目(71071044)

作者简介:石琴(1963—),女(汉族),安徽蚌埠人,合肥工业大学交通运输工程学院教授、硕士研究生导师,博士,研究方向:交通环境与安全、车辆现代设计理论与方法。

(Particle Swarm Optimization, PSO) 算法则是一种基于群体智能的全局性随机搜索方法,它容易实现、搜索速度快和搜索范围大<sup>[7]</sup>。粒子群优化的模糊聚类算法(Fuzzy Particle Swarm Optimization, FPSO)能够克服 K 均值聚类和模糊 C 均值聚类的上述缺点,可以实现较好的全局收敛性和较快的收敛速度。同时由于实际的工况数据存在一些模糊性,本文选择使用粒子群优化的模糊聚类方法进行聚类。

## 2.2 模糊聚类算法

模糊聚类算法是将样本空间  $X = \{x_1, x_2, \dots, x_n\}$  中的样本点分为  $c$  类( $1 < c < n$ ),定义目标函数为<sup>[9]</sup>:

$$J_m = \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ij} \quad (1)$$

其中: $n$  为样本数; $m$  为加权系数,且  $m > 1$ ;  $d_{ij}$  为样本点  $x_i$  和中心点  $z_j$  之间的距离;  $u_{ij}$  为第  $j$  个样本属于第  $i$  个聚类中心的适应度,且具有以下性质:

$$(1) \mu_{ij} \in [0, 1], \forall i = 1, 2, \dots, n, \forall j = 1, 2, \dots, c$$

$$(2) \sum_{j=1}^c \mu_{ij} = 1, \forall i = 1, 2, \dots, n$$

$$(3) 0 < \sum_{i=1}^n \mu_{ij} < n, \forall j = 1, 2, \dots, c$$

聚类问题就是求使目标函数(1)最小的适应度矩阵  $U = (\mu_{ij})_{n \times c}$  及聚类中心  $Z = \{z_1, z_2, \dots, z_c\}$ 。通过公式(1)的迭代优化来获取样本空间的模糊分类。

模糊聚类算法的基本步骤如下:

第一步,确定加权系数  $m$ , 初始化适应度函数  $\mu_{ij}$ ,  $i = 1, 2, \dots, n, j = 1, 2, \dots, c$ 。

第二步,计算第  $j$  类的聚类中心  $z_j$ 。

$$z_j = \frac{\sum_{i=1}^n \mu_{ij}^m x_i}{\sum_{i=1}^n \mu_{ij}^m} \quad j = 1, 2, \dots, c \quad (2)$$

第三步,计算欧氏距离  $d_{ij}$ 。

$$d_{ij} = (x_i - z_j) \quad i = 1, 2, \dots, n, j = 1, 2, \dots, c \quad (3)$$

第四步,更新适应度函数  $\mu_{ij}$ 。

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}} \quad (4)$$

第五步,如果没有收敛,返回第二步。

## 2.3 粒子群优化算法

粒子群优化算法是基于群体的算法,将每个个体看成  $D$  维搜索空间中的粒子,在搜索空间中以一

定的速度飞行,并根据本身的飞行经验和同伴的飞行经验动态调整。第  $i$  个粒子表示为  $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$ , 其经历的最好位置为  $p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ , 也称为  $p_{best}$ , 群体所有粒子经历的最好位置为  $g_{best}$ 。  $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$  表示粒子  $i$  的速度。每一代粒子的速度和位置根据如下方程更新<sup>[9]</sup>:

$$v_{ik}(t+1) = \omega v_{ik}(t) + c_1 r_{1k}(t)(p_{ik}(t) - x_{ik}(t)) + c_2 r_{2k}(t) \quad (5)$$

$$(p_{gk}(t) - x_{ik}(t))x_{ik}(t+1) = x_{ik}(t) + v_{ik}(t+1) \quad (6)$$

其中:  $v_{ik}(t)$  为第  $t$  次迭代粒子  $i$  飞行速度矢量的第  $k$  维分量,  $x_{ik}(t)$  为第  $t$  次迭代粒子  $i$  位置矢量的第  $k$  维分量,  $p_{ik}(t)$  为第  $t$  次迭代粒子  $i$  个体最好位置  $p_{best}$  的第  $k$  维分量,  $p_{gk}(t)$  为第  $t$  次迭代群体最好位置  $g_{best}$  的第  $k$  维分量,  $\omega$  为惯性权重,  $c_1$  和  $c_2$  为加速度常数,是调节  $p_{best}$  和  $g_{best}$  相对重要性的参数,  $r_1$  和  $r_2$  为两个在  $[0, 1]$  范围内的随机值。

粒子的速度更新公式(5)包含三部分:第一部分为粒子先前的速度;第二部分为粒子当前位置与自己最好位置之间的距离;第三部分为当前位置与群体最好位置之间的距离。粒子  $i$  通过公式(5)和(6)来决定下一步的运动位置。粒子群优化算法通过粒子在解空间内不断跟踪个体最好位置与全局最好位置进行搜索,直至达到规定的迭代次数或满足预置的误差为止。

## 2.4 粒子群优化的模糊聚类算法

本文利用粒子群优化的模糊聚类方法全局优化性能好的特点,首先对样本聚类中心进行初始化以获得其全局最优簇中心,然后再充分发挥 FCM 聚类方法局部搜索精度高的特点,达到最好的聚类效果。粒子群优化的模糊聚类算法流程如图 1 所示。

## 3 道路行驶工况的构建

### 3.1 数据采集和运动状态的划分

本文以合肥市典型道路为采集对象,选取屯溪路、胜利路等 5 条典型道路进行样本采集。每条测试路段连续采样 14 天(包括周六和周日)。采样时间为每天的 7:30~11:30 和 13:00~18:00,包括了车流量的高峰期与非高峰期以及工作日与非工作日的各个时段。驾驶员采用固定路线跟车法,跟随测试路段上的车流行驶进行数据采集。

将采集到的 18000 多组数据进行片段划分,按照以下定义划分了 88 组运动学片段:

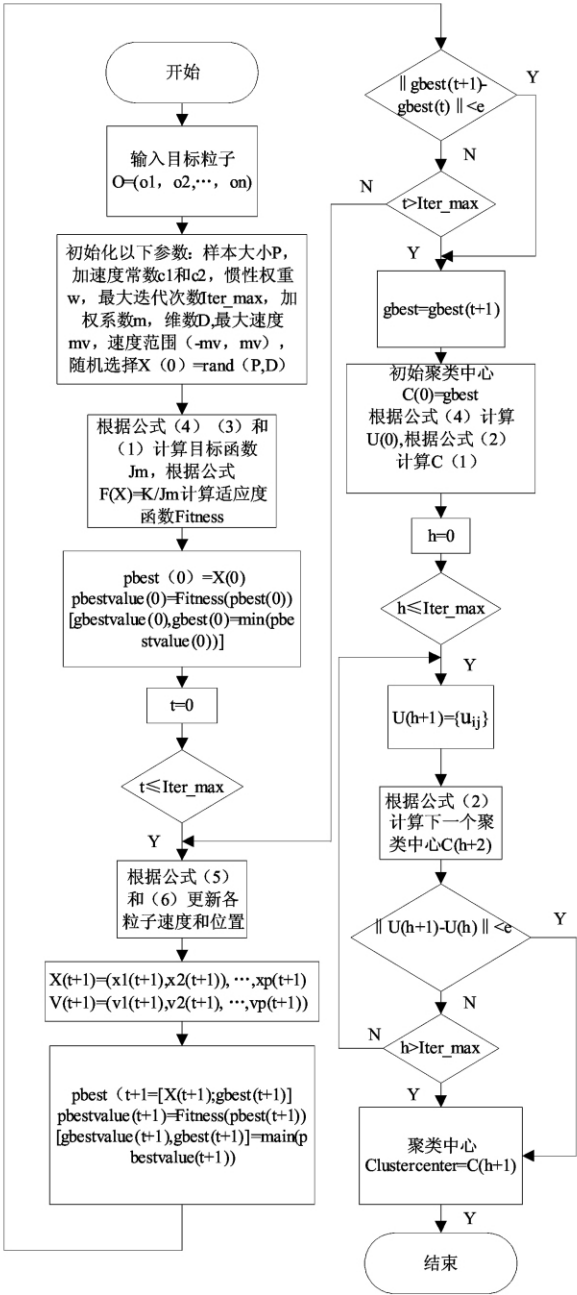


图 1 粒子群优化的模糊聚类算法流程图

- (1)怠速工况:发动机工作且  $V=0$ ;
- (2)加速工况:车辆  $a \geq 0.36\text{m/s}^2$  且  $V \neq 0$ ;
- (3)减速工况:车辆  $a \leq -0.36\text{m/s}^2$  且  $V \neq 0$ ;
- (4)匀速工况:车辆  $|a| < 0.36\text{m/s}^2$  且  $V \neq 0$ 。

为了便于主成分分析,本文选择了怠速比例  $P_i$ 、加速比例  $P_a$ 、减速比例  $P_d$ 、匀速比例  $P_c$ 、平均速度  $V_m$ 、平均运行速度  $V_{mr}$ 、最大加速度  $a_{max}$ 、最小加速度  $a_{min}$ 、平均加速度  $a_m$ 、速度标准偏差  $vsd$  和加速度标准偏差  $asd$  共 11 个统计学参数来描述和评价运动学片的特征,并计算出每个运动学片的特征值,如表 1 和表 2 所示。

表 1 各运动学片段及其特征参数

序号	特征参数	意义	单位
1	$P_i$	怠速比例	%
2	$P_a$	加速比例	%
3	$P_d$	减速比例	%
4	$P_c$	匀速比例	%
5	$V_m$	平均速度	km/h
6	$V_{mr}$	平均运行速度	km/h
7	$a_{max}$	最大加速度	$\text{m/s}^2$
8	$a_{min}$	最小加速度	$\text{m/s}^2$
9	$a_m$	平均加速度	$\text{m/s}^2$
10	$vsd$	速度标准偏差	km/h
11	$asd$	加速度标准偏差	$\text{m/s}^2$

表 2 各运动学片段的特征参数值

片段	$P_i$	$P_a$	$P_d$	$P_c$	...	$vsd$	$asd$
1	0.023	0.224	0.213	0.540	...	11.619	0.519
2	0.064	0.211	0.183	0.541	...	16.009	0.561
3	0.337	0.176	0.214	0.273	...	19.581	0.569
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
87	0.009	0.205	0.199	0.588	...	30.071	0.540
88	0.182	0.255	0.236	0.327	...	7.330	0.627

3.2 主成分分析和粒子群优化的模糊聚类分析

分别求出 88 组运动学片的特征参数,利用 SPSS 软件进行主成分分析,得到各主成分(分别用  $M_i, i=1 \sim 11$  表示)的特征值和贡献率,见表 3。

表 3 各主成分贡献率和累积率

	特征值	贡献率(%)	累积率(%)
$M_1$	4.857	44.154	44.154
$M_2$	1.862	16.931	61.084
$M_3$	1.542	14.021	75.105
$M_4$	0.032	9.386	84.491
$M_5$	0.852	7.747	92.238
$M_6$	0.456	4.145	96.383
$M_7$	0.220	1.997	98.380
$M_8$	0.110	0.003	99.383
$M_9$	0.049	0.447	99.830
$M_{10}$	0.018	0.164	99.994
$M_{11}$	0.001	0.006	100.000

通过表 3 可以看出,前四个主成分特征值的累积率达到了 84.491%,在进行主成分分析时只需选取累积率超过 80%的前几个主成分。为了便于聚类分析,本文选取前三个主成分进行粒子群优化的模糊聚类分析。通过粒子群优化的模糊聚类分析将样本数据分为两类。同时,采用 FCM 聚类方法和 K 均值聚类法对数据进行聚类,与粒子群优化的模糊聚类结果进行比较。聚类结果如图 2、图 3 和表 4。

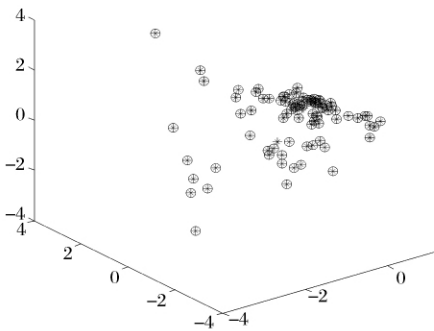


图 2 PSO 模糊聚类结果

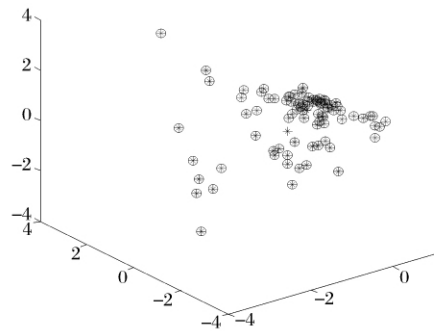


图 3 FCM 聚类结果

表 4 K 均值聚类结果

K 均值聚类中心			
1	0.25689	-0.2082	-0.3274
2	-0.87341	0.7078	1.1130

通过图 2 可以清晰看到,采用粒子群优化的模糊聚类方法将 88 组数据精确分为两类,第一类有 23 个片段,第二类有 65 个片段,根据相关系数大小按各类的时间比例从第一类中选取片段 70,81,从第二类中选取片段 67,6,83,26,拟合工况 1 如图 4。同时,采用 FCM 聚类方法进行聚类,得到两类的分类结果,第一类有 36 个片段,第二类有 52 个片段,根据相关系数大小按时间比例从第一类中选取片段 6,83,26,从第二类中选取片段 67,63,17,64,78,拟合工况 2 如图 5。同理,采用 K 均值聚类方法得到第一类有 68 个片段,第二类有 20 个片段,从第一类中选取片段 67,83,26,51,从第二类中选取片段 70,81,拟合工况如图 6。

为了验证粒子群优化的模糊聚类方法在道路行驶工况中应用效果,本文将实验数据的特征参数值与拟合工况 1 和工况 2、3 的数据参数值分别进行比较,并计算出者的误差值,如表 5 所示。

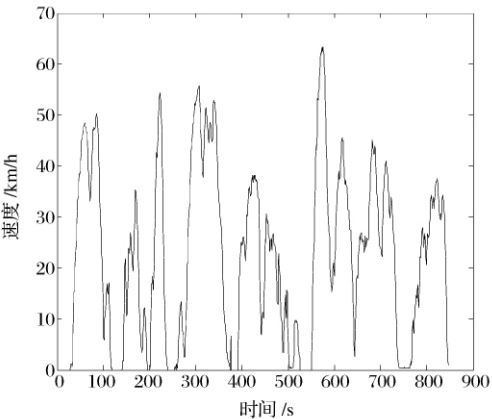


图 4 PSO 模糊聚类拟合工况图

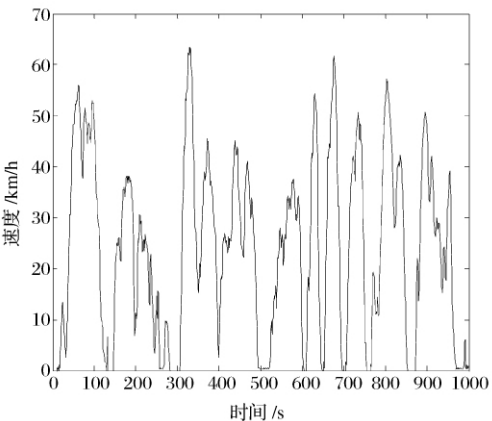


图 5 FCM 聚类拟合工况图

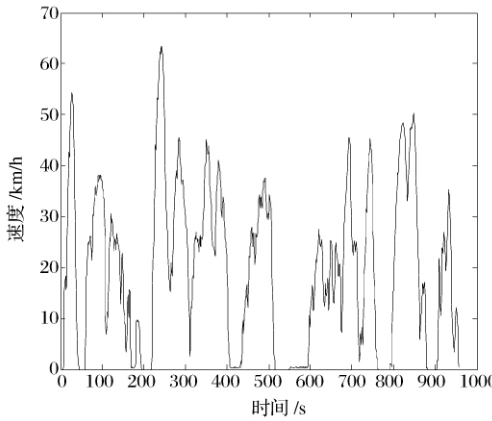


图 6 K 均值聚类拟合工况图

通过表 5 可以计算出拟合工况 1 与实验数据的特征参数的平均相对误差为 4.57992%,小于 5%,说明将粒子群优化的模糊聚类方法应用到行驶工况的构建中是有效可行的。同时,可以计算出拟合工况 2 与实验数据的特征参数的平均相对误差小于拟合工况 1 与实验数据的特征参数的平均相对误差。而且,工况 1 和工况 2 与实验数据的特征参数的平

均相对误差小于工况 3。所以,基于粒子群优化的模糊聚类方法的聚类效果较 FCM 聚类方法和 K 均值聚类方法好。

表 5 实验数据与拟合工况 1、工况 2 和工况 3 的比较

特征参数	工况 1	工况 2	工况 3	实验数据	误差 1	误差 2	误差 3(%)
<i>Pi</i>	0. 12541	0. 12022	0. 13480	0. 11259	4. 31709	6. 3467	12. 122
<i>Pa</i>	0. 21495	0. 20943	0. 18391	0. 22900	2. 63573	9. 3444	12. 184
<i>Pd</i>	0. 19550	0. 19101	0. 17555	0. 20203	2. 35066	5. 76933	8. 097
<i>Pc</i>	0. 46450	0. 48202	0. 50679	0. 45932	3. 68657	4. 70935	5. 139
<i>Vm</i>	24. 15312	25. 27942	21. 53996	23. 80612	4. 45540	5. 82806	14. 793
<i>Vmr</i>	7. 56361	28. 73391	25. 58423	27. 52757	4. 07289	4. 19831	10. 962
<i>Amax</i>	2. 27889	2. 20278	2. 1388	2. 20001	3. 45518	0. 12575	2. 9
<i>Amin</i>	-2. 17300	-2. 26667	-2. 2	-2. 23278	4. 13249	14. 9514	2. 941
<i>a</i>	0. 00043	0. 00047	0. 0006	0. 00042	8. 51064	10. 6383	26. 081
<i>vsd</i>	23. 31063	24. 50501	19. 7077	23. 41319	4. 87402	4. 4555	19. 577
<i>asd</i>	0. 51857	0. 56222	0. 54947	0. 60050	7. 76386	6. 80908	2. 267

4 结语

(1)本文以合肥市典型道路中采集的实验数据为基础,定义了能够描述和评价运动学片段特征的 11 个统计学参数,运用主成分分析法选出前三个主成分得分进行粒子群优化的模糊聚类,经过聚类选取合适的片段拟合工况。

(2)将粒子群优化的模糊聚类方法构建的工况与实验数据相比,误差小于 5%,能够综合反映合肥市实际的交通状况,且为工况数据的聚类分析提供了一种新的方法,也可以广泛应用于图像分割、多目标优化、信号处理、模式识别等多个科学和工程领域。

(3)采用 FCM 聚类方法和 K 均值聚类法拟合工况,并与粒子群优化的模糊聚类方法拟合的工况进行比较。结果表明,粒子群优化的模糊聚类方法的聚类效果较 FCM 聚类方法和 K 均值聚类法好,其优化能力强,搜索精度高。

(4)粒子群优化算法是一种方兴未艾的研究领域,许多理论和应用领域还需更深入地研究,而模糊聚类分析也是聚类分析中新的发展趋势,如何将两者的最新理论结合起来并应用于行驶工况的构建将是工况研究的一个新方向。

参考文献:

[1] Andre,M. The ARTEMIS European driving cycles for measuring car pollutant emissions[D],2004.  
[2] Ericsson, E. Variability in urban driving patterns[J]. Transportation Research,2000,337-354.  
[3] Lin, J. A Markov Proccess Approach to Driving Cycle Development[D]. California University Doctor Paper,

2002.

[4] Dai, Z. Driving Cycles: A New Cycle—Building Method that Better Represents Real—World Emissions [D]. Department of Civil and Environmental Engineering University of California,2008.  
[5] 赵慧,张镇顺,熊永达. 香港城区汽车行驶工况的研究[J]. 环境科学学报, 2000,20(3): 12-16.  
[6] 马志雄, 朱西产, 李孟良. 动态聚类法在车辆实际行驶工况开发中的应用[J]. 武汉理工大学学报,2005,27(11): 69-71.  
[7] 李帅,王新军,高丹丹. 基于内部空间特性的 PSO 聚类算法[J]. 计算机工程,2009,35(5): 197-199.  
[8] 邹刚,孙即祥,敖永红. 粒子群模糊聚类方法在病理图像分类中的应用[J]. 计算机工程与设计,2009,30(22): 5155-5157.  
[9] Izakian, H. Abraham, A. Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization[J]. World Congresson Nature and Biologically Inspired Computing, 2009: 1690-1694.  
[10] 冯征,阎敏,张智峰. 一种基于 PSO 的模糊聚类算法[J]. 计算机工程与应用,2006,27(2): 150-151.  
[11] 许磊,张凤鸣. 基于 PSO 的模糊聚类算法[J]. 计算机工程与科技,2006,27(21): 4128-4129.  
[12] 王玲,贺兴时. 基于 PSO 的模糊 C 均值聚类算法[J]. 甘肃联合大学学报,2008,22(2): 78-81.  
[13] 艾国和,乔维高,李孟良. 车辆运动学参数构成分析[J]. 公路交通科技,2006, 23(2):154-157.  
[14] 潘姝月,周雅夫. 城市公交车行驶工况的研究[J]. 机械设计与制造,2010,2(6): 263-265.  
[15] 朱西产,李孟良,马志雄. 车辆行驶工况开发方法[J]. 江苏大学学报,2005,26(2): 110-113.  
[16] Shi, S Zou, G. Liu, L. . Study on the fuzzy clustering method of the microtrips for passenger car driving cycle in changchun[J]. Vehile power and propulsion conference,2009,1555-1560.

Application of Fuzzy Clustering Based on Particle Swarm Optimization in Vehicle Driving Cycle

SHI Qin, WANG Nan-nan, QIU Duo-yang

(School of transportation Engineering, Hefei University of Technology, Hefei 230009, China)

**Abstract:** Fuzzy clustering based on particle swarm optimization method in vehicle driving cycle is tested in this paper. Principal component analysis is used to reduce the characteristics of the whole kinematic segments, which represents road running characteristics. The scores of the first three principal components of the kinematic segments are classified by using fuzzy clustering based on particle swarm optimization method. Programming with Matlab, the above theory is used to construct and analyze the typical roads in Hefei, and the representative driving cycle is obtained by selecting proper segments according to the ratio of time. The representative driving cycle and driving cycle obtained from k-means clustering and fuzzy c-means clustering method are compared with the experimental data respectively. The research results shaws that fuzzy clustering based on particle swarm optimization method which is used to construct driving cycle can improve construction precision effectively.

**Key words:** particle swarm; fuzzy clustering; principal component analysis; k-means clustering; driving cycle