

# 中国大陆省会城市客运出租汽车数量管制强度分类<sup>\*</sup>

姚志刚<sup>1</sup>; 程高<sup>2</sup>

(1.长安大学 经济与管理学院, 西安 710064; 2. 长安大学 公路学院, 西安 710064)

**摘要:** 为构建不同管制导向下的客运出租汽车数量调控模型, 运用模糊-*c* 均值聚类法划分中国大陆省会城市客运出租汽车数量管制强度, 选取人口拥有率、建成区面积拥有率、公交客运量比率、GDP 拥有率、道路面积拥有率和对外客运量比率 6 项指标, 采用 Matlab7.0 编程并进行聚类有效性分析和 F-统计量检验。研究表明: 将 31 个样本城市客运出租汽车数量管制强度分为低、较低、中、较高、高 5 个等级最佳, 分类结果为建立出租汽车数量调控模型奠定了基础。

**关键词:** 客运出租汽车; 数量管制强度; 模糊-*c* 均值聚类;

中图分类号: F572 文献标志码: A

## Clustering Taxi Quantity Regulatory Degree of Capital Cities in China

YAO Zhigang<sup>1</sup> CHENG Gao<sup>2</sup>

(1. School of economics and management, Chang'an University, Xi'an 710064 Shaanxi China; 2. School of highway, Chang'an University, Xi'an 710064 Shaanxi China)

**Abstract:** For establishing taxi quantity models under different regulation objectives, a Fuzzy *c*- Means Clustering was presented. Authors identified six primary demand factors such as population, urban built up area, transit volume, GDP, road area and amount of intercity passenger to obtain corresponding indexes of regulatory degree of urban taxi quantity. By compiling a dedicated program and the analysis with software MATLAB 7.0, the clustering effectiveness analysis and F-statistics test revealed that thirty-one capital cities of Mainland China were classified into five categories was the best. The result shows that Fuzzy *c*- Means is effective method to classify the regulatory degree of taxi quantity and the categorization of quantity regulatory degree is the groundwork of modeling taxi license supply.

**Key words:** taxi, quantity regulatory degree, Fuzzy *c*- Means clustering,

## 0 引言

数量管制是城市客运出租汽车监管体系最主要的组成部分。不同强度数量管制政策会对客运出租汽车的运力供给、服务质量、经营收入、经营“牌照”价格产生重要影响<sup>[1]</sup>。目前, 国内外绝大多数城市依然实施不同形式和强度的出租汽车客运数量管制政策。但由于对客运出租汽车的功能定位认识不同, 各城市的客运出租汽车数量管制思路、方法差异较大, 数量管制强度也各不相同<sup>[2]</sup>。纽约、洛杉矶、多伦多、悉尼、巴黎、香港等城市实施特许经营或者“牌照”拍卖制度长期保持客运出租汽车数量基本不变, 伦敦则通过近乎苛刻的驾驶员资格考试间接调控客运出租汽车数量的适度增长<sup>[3]</sup>, 国内大多数

城市则是在行政审批或特许经营制度下, 根据需求变化相应调整客运出租汽车数量<sup>[4]</sup>。通常情况下客运出租汽车数量管制政策根据定性判断或简单依据人口或经济增长情况而制定<sup>[5]</sup>, 到底如何确定客运出租汽车数量控制水平及评价其合理性, 成为政府实施客运出租汽车数量管制面临的难题<sup>[6]</sup>。

千人拥有量和车辆空驶率是两个最常用的客运出租汽车数量配置水平参考指标, 通常被用来进行同类型城市间的客运出租汽车数量管制强度类比。由于不同城市的土地开发与空间布局、经济结构与发展水平、居民出行习惯、交通方式构成、出租车运营模式及时空分布特征等差异较大, 千人拥有量和车辆空驶率指标的适用性受到质疑<sup>[7,8]</sup>, 因此有学者尝试分析客运出租汽车数量与公交客运量、私家车拥有量、居民人口等因素的相关性, 建立统计回

基金项目: 教育部人文社会科学研究青年基金项目 (10YJC790340); 长安大学陕西省哲学社会科学重点研究基地项目 (JD0902), “中央高校基本科研业务费专项资金 (CHDW2011ZD010)

作者简介: 姚志刚(1974-), 男, 汉, 陕西澄城人, 副教授, 长安大学工学博士, 研究方向为运输管理。 电子邮件: zhgyao@chd.edu.cn

归模型进行客运出租汽车数量配置合理性评价,发现考虑客运出租汽车功能定位和经济社会条件的客运出租汽车数量回归模型更具有普适性<sup>[9,10]</sup>。基于此,本文以中国大陆省会城市为例,拟用模糊-*c*聚类法对客运出租汽车数量管制强度进行分类,期望以此为基础分别构建不同数量管制强度下的城市客运出租汽车的数量配置模型;同时,客运出租汽车数量管制强度分类结果也可对相关监管政策的制定提供决策参考。

## 1 数量管制强度分类指标

### 1.1 指标选取

客运出租汽车的数量需求客观上会受城市经济发展、社会环境、消费习惯、交通结构及道路建设等多方面影响<sup>[11]</sup>。为克服单一指标表示城市客运出租汽车数量的局限性,文中选取多个指标来表征客运出租汽车数量管制强度。

从《中国城市统计年鉴(2010)》31个大陆省会城市统计资料中选取与出租汽车客运关系密切的社会、经济、交通运输相关统计指标,计算各统计指标与客运出租汽车数量的比率,将其作为客运出租汽车数量管制强度衡量指标;然后,进行各指标间的相关性分析,剔除相关系数高的工业总产值指标,并将公路、铁路、民航客运量合并为对外客运量指标;最后,获得市区人口拥有率(辆/千人)、建成区面积拥有率(辆/平方千米)、公交客运量比率(辆/百万人次)、GDP拥有率(辆/亿元)、道路面积拥有率(辆/万平方米)、城市对外客运量比率(辆/万人次)6项客运出租汽车数量管制强度的特征向量。所选的31个省会城市的分类指标原始数值如表1所示。

虽然城市的游客数量、旅游收入、小汽车拥有量等也对出租汽车数量需求有一定影响,但鉴于这些指标无法从中国城市统计年鉴中获得,文中暂不考虑。

### 1.2 数据标准化

由于客运出租汽车数量管制强度指标的量纲和数量级不同,进行分类前需采用式(1)所示方法对分类指标原始数据进行标准化处理<sup>[12]</sup>。

$$x_{kj} = \frac{x'_{kj} - \min\{x'_{kj}\}}{\max\{x'_{kj}\} - \min\{x'_{kj}\}} \quad (1)$$

式中: $x_{kj}$ 和 $x'_{kj}$ 分别为样本城市 $k(k=1,2,L,n)$ 的第 $j(j=1,2,L,s)$ 个数量管制强度指标标准化值和原始值。

表1 分类指标原始值

Table 1 Original classification indexes

城市	指标	人口拥有率	建成面积拥有率	公交比率	GDP拥有率	道路面积比率	对外客运量比率
北京市		5.752	50.836	14.154	6.455	7.454	3.177
天津市		4.023	49.828	28.483	5.405	3.473	3.650
石家庄市		2.823	35.576	17.851	5.966	2.179	0.676
太原市		2.948	34.840	23.295	6.475	3.562	1.887
呼和浩特市		3.998	30.299	18.151	5.114	3.185	0.881
沈阳市		3.405	46.849	15.620	5.120	3.815	1.570
长春市		5.060	50.939	24.155	8.849	4.181	2.068
哈尔滨市		2.826	39.485	12.689	6.513	4.751	1.311
上海市		3.636	54.243	18.067	3.544	5.496	4.434
南京市		1.876	17.147	9.006	2.945	1.153	0.205
杭州市		2.160	24.978	8.034	2.419	2.153	0.318
合肥市		3.498	25.421	11.961	5.809	1.970	0.495
福州市		2.006	21.158	8.062	3.369	1.886	0.330
南昌市		1.646	33.697	7.023	3.275	2.159	0.631
济南市		2.317	24.896	10.750	3.582	1.373	0.510
郑州市		3.924	33.006	14.923	8.508	3.716	0.445
武汉市		2.369	26.328	8.669	3.578	2.089	0.643
长沙市		3.082	30.222	6.941	3.977	2.212	0.547
广州市		2.920	21.070	7.663	2.494	2.265	0.342
南宁市		1.642	24.201	8.100	4.602	1.741	0.370
海口市		1.260	21.571	0.922	4.429	1.616	0.097
重庆市		0.812	17.590	8.024	3.423	1.491	0.118
成都市		2.496	29.748	11.759	4.649	2.367	0.280
贵阳市		1.417	21.921	5.332	4.820	2.690	0.121
昆明市		2.581	22.655	8.469	5.440	2.929	0.707
拉萨市		6.223	19.661	16.907	46.105	2.667	4.897
西安市		2.141	43.513	8.490	6.137	2.516	0.978
兰州市		2.416	27.727	9.370	6.984	2.134	1.611
西宁市		4.559	77.515	8.707	17.671	6.468	1.469
银川市		5.635	45.099	32.811	14.161	3.881	1.632
乌鲁木齐市		3.494	26.172	10.144	7.866	4.268	2.337

## 2 数量管制强度分类方法

### 2.1 基本原理

选用模糊-*c*均值聚类进行客运出租汽车数量管制强度分类,主要因为该方法在其他领域的分类研究中有很好普适性<sup>[13]</sup>。设待分类城市样本集为 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, L, \mathbf{x}_n\}$ ,  $\mathbf{x}_k = (x_{k1}, x_{k2}, L, x_{ks})^T$ 表示城市样本 $\mathbf{x}_k$ 为*s*维特征向量。采用模糊-*c*均值聚类进行客运出租汽车数量管制强度分类,即根据分类指标将城市样本集分为*c* ( $2 \leq c \leq \sqrt{n}$ )个模糊组,并求每组聚类中心,当各聚类中心加权距离平方和的目标函数最小时获得最佳分类方案,目标函数可表示为:

$$\min \{J_m(U, P)\} = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m d_{ik}^2 \quad (2)$$

$$\sum_{i=1}^c \mu_{ik} = 1$$

$$0 \leq \mu_{ik} \leq 1, 1 \leq i \leq c$$

式中： $U$ 为隶属度矩阵或分类矩阵， $\mu_{ik}$ 为 $U$ 的第*i*行第*k*列，表示城市*k*属于聚类中心*i*的隶属度； $P$ 为聚类中心矩阵； $d_{ik}$ 为聚类中心*i*与城市样本*k*的欧氏距离； $m \in (1, \infty)$ 为模糊加权指数。

## 2.2 数量管制强度聚类算法

根据模糊- $c$ 均值聚类原理，城市客运出租汽车数量管制强度分类算法如下：

**Step 1:** 初始化，给定客运出租汽车数量管制强度聚类中心数 $c$ 和模糊加权指数 $m$ ，设定迭代停止阈值 $\varepsilon$ ，初始化聚类中心 $P^{(0)}$ ，设置迭代计数器 $b = 0$ 。

**Step 2:** 用式(3)计算隶属度矩阵 $U^{(b)}$ 。对于 $\forall i, k$ ，如果 $\exists d_{ik}^{(b)} > 0$ ，则有

$$\mu_{ik}^{(b)} = \frac{1}{\sum_{j=1}^c \left( \frac{d_{ik}^{(b)}}{d_{ij}^{(b)}} \right)^{\frac{2}{m-1}}} \quad (3)$$

**Step 3:** 用式(4)更新聚类中心矩阵 $P^{(b+1)}$

$$p_i^{(b+1)} = \frac{\sum_{k=1}^n (\mu_{ik}^{(b+1)})^m \cdot \mathbf{x}_k}{\sum_{k=1}^n (\mu_{ik}^{(b+1)})^m} \quad (4)$$

**Step 4:** 如果 $\|P^{(b)} - P^{(b+1)}\| < \varepsilon$ ，则算法停止并输出隶属度矩阵 $U$ 和聚类中心矩阵 $P$ ，否则令 $b = b + 1$ ，转向Step 2继续迭代，直至算法停止。

## 3 分类结果

用 Matlab7.0 对上述数量管制强度聚类算法编程，给定模糊加权指数 $m$ 和聚类数 $c$ 情况下将标准化后的城市客运出租汽车数量管制强度分类指标输入程序，运行后可得到分类结果。

### 3.1 分类有效性判别

由于聚类过程中模糊加权指数 $m$ 和聚类数 $c$ 的赋值很大程度上影响着分类结果，因此需要分析不同 $m$ 或 $c$ 取值情况下的分类结果的有效性。通常

采用基于可能性分布的聚类有效性函数 $P(U, c)$ 来判别分类效果的优劣， $P(U, c)$ 由式(5)计算

$$P(U, c) = \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^c u_{ik} - \frac{1}{c} \sum_{i=1}^c \left( \frac{\sum_{k=1}^n u_{ik}^2}{\sum_{k=1}^n u_{ik}} \right) \quad (5)$$

式中 $P(U, c)$ 的数值随 $m$ 和 $c$ 的赋值不同而变化， $P(U, c)$ 值越小说明聚类效果越好。

由于选取31个省会城市分类，则有 $2 \leq c \leq 5$ 。分类有效性判别可以根据函数 $P(U, c)$ 的取值确定出 $c$ 位于 $[2, 5]$ 、 $m$ 位于 $[1.2, 2.8]$ 情况下的最优分类数，所对应的聚类有效性函数 $P(U, c)$ 值变化曲线如图1所示。图1显示 $m$ 值所对应的 $P(U, c)$ 最小值大多数情况下是在 $c = 5$ 时获得，即所选样本城市客运出租汽车的数量管制强度分为5类时最佳。

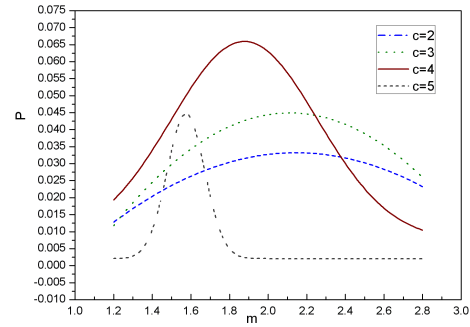


图1 有效性函数曲线

Fig.1 Validity function curves

考察采用模糊- $c$ 均值聚类算法进行城市客运出租汽车的数量管制强度分类的运算效率，聚类迭代过程目标函数变化如图2所示，可以看出，算法迭代25次或更少次数时能使目标函数达到最小，具有较好收敛性。

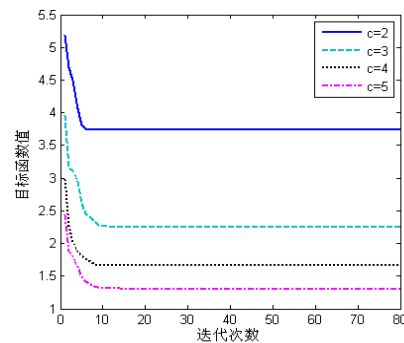


图2 目标函数迭代曲线

Fig.2 Iterative curves of objective function

### 3.2 结果分析

根据上述最佳分类数和应用模糊- $c$ 均值聚类法时的模糊加权指数选取原则,可将 $m=2$ 、 $c=5$ 对应的客运出租汽车数量管制强度分类结果视为最佳,具体如表2所示。同时,本文用 $F$ -统计量法验证分为5类时的合理性,计算95%置信水平的 $F$ 值为13.02、 $F_{0.05}$ 值为2.74,  $F > F_{0.05}(4,27)$ 表明分5类时各类之间存在显著差异,即分类合理。而且,表2列出 $c=2,3,4$ 时的分类结果以便于对比。

根据最佳分类方案,将31个省会城市的客运出租汽车数量管制强度分别称为I(低)、II(较低)、III(中)、IV(较高)、V(高)管制强度5个等级:

I级管制强度的城市有天津、长春、拉萨、西宁、银川,其中天津、长春和银川的公交比率远高于全国省会城市平均水平,而拉萨的人口拥有率、GDP拥有率和对外客运比率以及西宁市的建成区面积拥有率居全国省会城市首位;

II级管制强度的城市有北京和上海,除上海的GDP拥有率指标外,北京和上海的其他数量管制强度指标均高于全国省会城市平均水平,且北京的道路面积拥有率居全国省会城市首位;

表2 分类结果

Table 2 Classification results

城市 \ 分类	$c=2$	$c=3$	$c=4$	$c=5$
北京市	I	I	I	II
天津市	I	I	I	I
石家庄市	II	II	III	IV
太原市	I	II	II	III
呼和浩特市	II	II	III	III
沈阳市	I	II	II	III
长春市	I	I	I	I
哈尔滨市	I	II	II	III
上海市	I	I	I	II
南京市	II	III	IV	V
杭州市	II	III	IV	V
合肥市	II	III	III	IV
福州市	II	III	IV	V
南昌市	II	III	IV	V
济南市	II	III	IV	V
郑州市	II	II	III	III
武汉市	II	III	IV	IV
长沙市	II	III	III	IV
广州市	II	III	IV	IV
南宁市	II	III	IV	V
海口市	II	III	IV	V
重庆市	II	III	IV	V
成都市	II	III	IV	IV
贵阳市	II	III	IV	V
昆明市	II	III	III	IV
拉萨市	I	I	I	I
西安市	II	III	III	IV
兰州市	II	III	III	IV
西宁市	I	I	I	I

银川市	I	I	I	I
乌鲁木齐市	I	II	II	III

III级管制强度的城市有太原、呼和浩特、沈阳、哈尔滨、郑州、乌鲁木齐,这六个城市的数量管制强度指标均接近或略高于全国省会城市平均水平;

IV级管制强度的城市有石家庄、合肥、武汉、长沙、广州、成都、昆明、西安、兰州,这九个城市的数量管制强度指标均接近或略低于全国省会城市平均水平;

V级管制强度的城市有南京、杭州、福州、南昌、济南、南宁、海口、重庆、贵阳,其中南京的建成区面积比率和道路面积比率、杭州的GDP拥有率、海口的公交比率和对外客运比率以及重庆的人口拥有率均居全国省会城市末位。

此外,从表2可以看出,不同分类数情况下始终有保持稳定的五个城市组合: {天津、长春、拉萨、西宁、银川}、{北京、上海}、{太原、沈阳、哈尔滨、乌鲁木齐}、{合肥、长沙、昆明、西安、兰州}、{南京、杭州、福州、南昌、济南、南宁、海口、重庆、贵阳},表明同一组内城市的客运出租汽车数量管制强度相似度更高。

### 4 结语

从《中国城市统计年鉴(2010)》统计资料中选取与出租汽车客运关系密切的社会、经济、交通运输等统计指标并计算,确定出人口拥有率、建成区面积拥有率、公交客运量比率、GDP产值拥有率、道路面积拥有率、对外客运量比率6项指标作为城市客运出租汽车数量管制强度表征向量;用模糊- $c$ 均值聚类进行城市客运出租汽车数量管制强度划分,给出聚类算法,及确定最佳分类数、最佳分类结果的方法;分类结果显示,31个中国大陆省会城市客运出租汽车数量管制强度分为I(低)、II(较低)、III(中)、IV(较高)、V(高)5个等级最佳,其中天津、长春、拉萨、西宁、银川5个城市数量管制强度最低,北京和上海2个城市次之,太原、呼和浩特、沈阳、哈尔滨、郑州、乌鲁木齐6个城市数量管制强度居中,石家庄、合肥、武汉、长沙、广州、成都、昆明、西安、兰州9个城市的数量管制强度较高,而南京、杭州、福州、南昌、济南、南宁、海口、重庆、贵阳9个城市的数量管制最严格;要得到不同管制强度类型对应的客运出租汽车数量配置模型,还需在分类的基础上对样本城市的社会、经济指标进行统计分析后才能获得。

---

## 参 考 文 献

1. Moore, AT, Balaker.T. Do Economists Reach a Conclusion on Taxi Deregulation? [J], *Economic Journal*, 2006,3(1):109-132.
2. Transport Research Centre. Taxi Services: Competition and Regulation[R]. Roundtable on Taxi Services Regulation and Competition. Paris, OECD, 2007.
3. Walter Skok, Juan Antonio Martinez. An International Taxi Cab Evaluation [J], *Knowledge and Process Management*, 2010,17(3):145-153.
4. 陆建,王炜.城市出租车拥有量确定方法[J].*交通运输工程学报*, 2004,4(1): 92-95.
5. Bruce Schaller. Entry controls in taxi regulation: Implications of US and Canadian experience for taxi regulation and deregulation[J]. *Transport Policy*, 2007,14 (6) :490-506.
6. Tamer Çetin, Kadir Yasin Eryigit. Estimating the effects of entry regulation in the Istanbul taxicab market[J]. *Transportation Research Part A: Policy and Practice*, 2011,45(6):476-484.
7. 翁剑成,刘文韬,陈智宏等. 基于浮动车数据的出租车运营管理研究[J]. *北京工业大学学报*,2010, 36(6): 779- 784.
8. Bekken, J.T.. Experiences with (de-)regulation in the European taxi industry[R]. (De)Regulation of the Taxi Industry Roundtable 133, Paris,OECD, 2007:31-58.
9. Bruce Schaller. A Regression Model of the Number of Taxicabs in U.S. Cities[J]. *Journal of Public Transportation*, 2005,8(5):63-78.
10. Kattan, K. Barros, A. Wirasinghe, S.C. Analysis of Work Trips Made by Taxi in Canadian Cities[J]. *Journal of Advanced Transportation*, 2010,44(5): 11-18.
11. 唐炉亮,常晓猛,李清泉等. 基于蚁群优化算法与出租车 GPS 数据的公众出行路径优化[J]. *中国公路学报*, 2011,24 (2) :89-95.
12. Hesam Izakian, Ajith Abraham. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem[J]. *Expert Systems with Applications*, 2011,38(3): 1835 -1838.
13. 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): 103- 108.