

模糊时间序列在中国汽车产业规划中的应用

胡世前¹, 姜倩雯¹, 王 博²

(1. 东北财经大学 公共管理学院, 辽宁 大连 116025; 2. 重庆长安汽车股份有限公司, 重庆 400023)

[摘要]本文分别基于自动聚类算法和FCM 算法的模糊时间序列模型对中国汽车销量进行仿真预测和比较分析, 在此基础上择优为政府制定汽车产业规划和开展汽车污染治理工作提供政策决策工具。首先, 通过自动聚类算法将1995—2014年中国汽车销量历史数据划分成不同长度的区间, 利用单因子三阶模糊逻辑关系对中国汽车销量进行仿真预测。其次, 通过FCM 算法进行区间划分后对中国汽车销量进行仿真预测研究。最后, 对两种算法的仿真预测进行结果精度比较分析, 以检验区间划分对预测结果所产生的影响。研究结果表明: 不同聚类算法的区间划分会对预测结果产生显著影响; 基于FCM 算法的区间划分与自动聚类方法的区间划分相比具有准确度和精度较高的优势, 能有效地反映未来中国汽车市场需求走势, 并帮助政府合理地进行汽车产业规划与科学引导汽车企业的新能源技术研发与应用。

[关键词]汽车产业规划; 汽车污染治理; 模糊时间序列; 自动聚类算法; FCM 算法

中图分类号: F416.471 文献标识码: A 文章编号: 1008-4096(2017)05-0040-06

一、引言

“十三五”是全面建成小康社会、稳增长、促改革、调结构、防风险的关键时期, 复杂的内外部环境在加大产业规划与结构调整工作难度的同时, 也为促进创新驱动夯实中国绿色经济与可持续发展提供了契机。李克强总理在主持召开“十三五”《规划纲要》编制工作会议上重点强调产业发展规划工作应重点围绕贯彻落实创新、协调、绿色、共享等新理念, 依靠大数据等技术平台科学合理地协调产业规划与环境治理之间的矛盾, 并为国家实现增动力、可持续、拓空间等综合目标做出贡献。由于汽车产业具有投资量大、规模经济要求高、中间投入环节多、价值转移效果好、就业面广及消费拉动大等特征, 因而汽车产业的发展早已成为国际社会衡量国家工业化水平、经济实力及科技创新能力的重要标志。

2009年以来, 相对发达国家汽车产销增长相对停滞而言, 发展中国家已经成为世界汽车产销增长的核心, 尤其中国的增长速度最为迅猛。据中国汽车工业协会统计, 2016年中国汽车产销分别完成2 811.90万辆和2 802.80万辆, 比上年同期分别增长14.50%和13.70%, 连续八年蝉联全球第一。另外, 据中国汽车工业协会预测, 到2020年中国汽车总保有量将可能突破1.50亿辆^[1]。中国于2009年步入了初级汽车社会的门槛, 这能从一个侧面体现出改革开放以来中国在经济与社会发展方面所取得的辉煌成果, 但是激增的汽车保有量所产生的汽车污染与汽车垃圾也给生态环境带来了巨大压力。2006年世界卫生组织在联合国大会上就公开提出汽车污染是威胁人类健康、破坏生态环境、引发全球变暖等负面问题的重要因素。2015年11月世界环保组织在

收稿日期: 2017-05-22

基金项目: 辽宁省经济社会发展研究课题“经济新常态下东北老工业基地的环境质量评价研究”(2016lsljdwt-04)

作者简介: 胡世前(1979-), 男, 辽宁丹东人, 讲师, 博士, 主要从事公共政策和政府治理研究。E-mail: hushiqian1979@126.com

姜倩雯(1990-), 女, 辽宁丹东人, 硕士研究生, 主要从事产业经济和政府绩效研究。

王 博(1992-), 男, 重庆人, 硕士研究生, 主要从事数量经济研究。

联合国大会上阐明了全球范围的城市空气污染中50%是由燃油汽车的废气排放引发的观点。伴随着2015年12月12日巴黎气候协议“最终案文”的公布,包括中国政府在内的国际社会势必在开展大气污染治理过程中对汽车产业提出更高的环保标准与合理调整传统能源汽车产业结构的要求^[2]。

近几年围绕着产业经济、民生及社会治理等领域的发展趋势开展预测研究已经成为学术界关注的焦点,而相应领域利用定量分析方法研究时间序列数据已成为常态化。虽然经典的时间序列分析模型和方法已经能处理绝大多数现实问题,但是在时间跨度较大、历史数据缺失或不清晰、信息搜集能力有限等复杂因素影响下,政府部门或主流媒体时常用“激增”、“稳中有降”及“遇冷”等模糊性语言来形容汽车产量、销量及市场需求变化,基于传统时间序列方法的科学的研究将遇到瓶颈。为了有效地使用和挖掘大量模糊不清和碎片化数据的价值,国外学者对可以替代传统时间序列分析模型的新型分析理论与方法产生了强烈需求。Azizzadeh等^[3]初步建立了带有模糊语言变量的模型,Takagi和Sugeno^[4]对模糊模型进行了相应的完善,Song和Chissom^[5]首次将模糊时间序列的预测模型应用到阿拉巴马大学入学人数的研究,后来的Aladag等^[6]的神经网络模型、Egrioglu等^[7]的优化模型、Bai等^[8]的时变模型、Kuo等^[9]的粒子群优化模型及Chen^[10]、Lee等^[11]的模型都以阿拉巴马大学入学人数为例解释说明相应模型的构建原理和应用过程,且取得了良好的研究效果。此后,诸多学者在深入研究模糊时间序列模型基础上,探索性地对传统时间序列分析方法和模糊时间序列分析方法进行比较研究。

由于政府对汽车产业进行规划和开展环境治理工作主要基于汽车污染的流动性、汽车产业的经济地位与产销链较长、新型汽车能源开发利用过程繁琐、汽车市场需求不稳定等复杂因素基础上进行的。因此,在模糊化数据普遍存在的背景下,政府对误差较小、准确率高、普适性广的预测模型产生了强烈需求,目前国内学术界主要是基于时间序列预测法、线性回归模型预测法、非线性回归预测法、灰色系统模型预测法及马尔科夫预测法等定量分析方法对汽车市场需求进行预测研究^[12]。但是基于中国汽车产销数据收集能

力有限、单一化的信息搜集与发布渠道、非线性化的市场需求等现实性因素,在诸多关于中国汽车市场需求或销量预测的定量化研究方法中,模糊时间序列模型具有准确性、实用性与科学性较高的优势^[13]。

二、模型设定及算法

1. 自动聚类算法

自动聚类是一种基于贝叶斯理论的数据聚类算法,通过对数据进行处理,计算出每条数据属于每个类别的几率值,将数据进行聚类,本文应用的自动聚类算法的步骤如下:

第1步,将n个数值数据按升序排列: $x_1, x_2, \dots, x_i, \dots, x_n$ 。其中, x_1 是n个数据里最小的数, x_n 是n个数据里最大的数,且 $1 \leq i \leq n$ 。基于上述递升数列计算聚类的阈值数 τ ,如式(1)所示:

$$\tau = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)}{n-1} \quad (1)$$

第2步,每个数据为单独一类,即 $\{x_1\}, \{x_2\}, \dots, \{x_i\}, \dots, \{x_n\}$ 。其中, x_1 是递升数列里最小的数, x_n 是递升数列里最大的数,且 $1 \leq i \leq n$ 。

第3步,假设有p个类,计算每个聚类 $cluster_j$ 的聚类中心 $center_j$,如式(2)所示:

$$center_j = \frac{\sum_{j=1}^r x_j}{r} \quad (2)$$

其中, x_j 是 $cluster_j$ 中的数据,r表示聚类 $cluster_j$ 中的数据个数,且 $1 \leq j \leq p$ 。计算任意两个相邻聚类中心 $center_j$ 与 $center_{j+1}$ 的距离 $distance_{j,j+1}$,如式(3)所示:

$$distance_{j,j+1} = |center_j - center_{j+1}|, \\ j=1, 2, \dots, p-1 \quad (3)$$

第4步,在所得的距离 $distance_{j,j+1}$ 中找出最小的距离 $smallest_distance$,如式(4)所示:

$$smallest_distance = \min_{j=1,2,\dots,p-1} (distance_{j,j+1}) \quad (4)$$

第5步,如果 $smallest_distance$ 小于阈值 τ ,将其对应的两个聚类合并为一类,转到第3步。否则转到第6步。如果相邻的聚类有相同的最小距离不止一对,那么结合这些聚类成为一个类。例如,假设最小距离位于 $cluster_3$ 跟 $cluster_4$ 之间,同样也位于与 $cluster_4$ 跟 $cluster_5$ 之间,那么将 $cluster_3$ 、 $cluster_4$ 和 $cluster_5$ 归为一类。

第6步, $cluster_j$ 的上界 $uBound_j$ 和 $cluster_{j+1}$ 的下界 $lBound_{j+1}$ 的计算如式(5)和式(6)

所示：

$$u\text{Bound}_j = \frac{\text{center}_j + \text{center}_{j+1}}{2} \quad (5)$$

$$l\text{Bound}_{j+1} = u\text{Bound}_j \quad (6)$$

其中， $j=1, 2, \dots, p-1$ 。在第一类之前与最后一类之后都不存在其他类，故第一类的下界 $l\text{Bound}_1$ 与最后一类的上界 $u\text{Bound}_p$ 的计算如式(7)和式(8)所示：

$$l\text{Bound}_1 = \text{center}_1 - (u\text{Bound}_1 - \text{center}_1) \quad (7)$$

$$u\text{Bound}_p = \text{center}_p + (\text{center}_p - l\text{Bound}_p) \quad (8)$$

第7步，让每一聚类 cluster_j 形成一个区间 interval_j ，即聚类 cluster_j 的上界 $u\text{Bound}_j$ 和下界 $l\text{Bound}_j$ 为区间 interval_j 的上界 interval_uBound_j 和下界 interval_lBound_j 。区间 interval_j 的中间值 mid_value_j 计算如式(9)所示：

$$\text{mid_value}_j = \frac{(\text{interval_lBound}_j + \text{interval_uBound}_j)}{2}, \quad j=1, 2, \dots, p \quad (9)$$

2. FCM 算法

聚类就是按照一定的要求和规律对事物进行区分和分类的过程，在这一过程中没有任何关于分类的先验知识，仅靠事物间的相似性作为类属划分的准则，因而属于无监督分类的范畴。聚类分析则是指用数学的方法研究和处理给定对象的分类。“人以群分，物以类聚”，聚类的现象或科学研究伴随着人类社会的产生和发展而不断深化，人类要更深入地认识世界就必须区别不同的事物和甄别事物间的相似性，业内普遍认为最典型的算法是模糊 C-均值聚类（FCM）算法。在 FCM 算法中确定合适的聚类准则是关键步骤之一，本文采用最优聚类的准则就是使得样本数据到聚类中心的距离的平方越小，即表示数据离聚类中心的欧几里得距离越近。聚类数目需要使用者根据需要进行预先设置，对于不同的聚类数目，有的聚类过于分散，而有的分类则过于紧凑。选择合适的聚类数目是最优聚类的前提。FCM 算法就是不断地修正聚类中心 V 和由隶属度值所组成的划分矩阵 U，因而是一个动态聚类过程^[14]。FCM 算法如式(10)所示：

$$\begin{aligned} \text{Min } J_{\text{fcm}}(U, V) &= \sum_{j=1}^p \sum_{i=1}^n u_{ji}^m d_{ji}^2 \\ \text{使得} &\left\{ \begin{array}{l} \sum_{j=1}^p u_{ji} = 1, 1 \leq i \leq n \\ 0 < \sum_{i=1}^n u_{ji} \leq n, 1 \leq j \leq p \\ 0 \leq u_{ji} \leq 1, 1 \leq j \leq p, 1 \leq i \leq n \end{array} \right. \end{aligned} \quad (10)$$

其中， $m>1$ 是模糊系数，本文中取值为 2； $U=u_{ji}$ 是一个 $p \times n$ 的模糊划分矩阵， u_{ji} 是第 i 个样本属于第 j 类的隶属度值，且 $u_{ji} \in [0, 1]$ ； $V=[v_1, v_2, \dots, v_p]$ 是由 p 个聚类中心构成的向量； $d_{ji} = \|x_i - v_j\|$ 表示从样本 x_i 到中心 v_j 的距离。FCM 算法的主要迭代方法是不断地更新隶属度值 u_{ji} 和聚类中心 v_j ，迭代过程截止条件为迭代次数达到 100 或目标函数 $J_{\text{fcm}}(U, V)$ 在两次相邻的迭代过程的最小值提高小于 1×10^{-4} ，具体如式(11)和式(12)所示：

$$v_j = \frac{\sum_{i=1}^n (u_{ji})^m x_i}{\sum_{i=1}^n (u_{ji})^m}, \quad 1 \leq j \leq p \quad (11)$$

$$u_{ji} = \frac{1}{\sum_{k=1}^p \left[\frac{d(x_i, v_j)}{d(x_i, v_k)} \right]^{\frac{2}{m-1}}}, \quad 1 \leq j \leq p, 1 \leq i \leq n \quad (12)$$

3. 基于模糊逻辑关系的预测研究

基于模糊逻辑关系和聚类技术的预测步骤如下：

第1步，应用聚类算法将数据划分为不同长度的区间。

第2步，假设第1步得出 p 个区间 u_1, u_2, \dots, u_p ，用语言项 A_1, A_2, \dots, A_p 表示模糊集，如式(13)所示：

$$\left\{ \begin{array}{l} A_1 = a_{11}/u_1 + a_{12}/u_2 + \dots + a_{1p}/u_p \\ A_2 = a_{21}/u_1 + a_{22}/u_2 + \dots + a_{2p}/u_p \\ \dots \\ A_p = a_{p1}/u_1 + a_{p2}/u_2 + \dots + a_{pp}/u_p \end{array} \right. \quad (13)$$

其中， $a_{ij} \in [0, 1]$ ， $1 \leq i \leq p$ 且 $1 \leq j \leq p$ ， a_{ij} 为 u_j 在模糊集 A_i 上的隶属度值。

第3步，将每个数据用一个模糊集来表示。如果数据属于 u_j ，那么该数据被模糊为 A_j ，且 $1 \leq j \leq p$ 。

第4步，在第3步被模糊化的数据集的基础上创建模糊逻辑关系。基于模糊逻辑的当前状态，让拥有相同当前状态的模糊逻辑关系形成一个模糊逻辑关系组^[15-16]。

三、中国汽车销量聚类分析与仿真预测

1. 基于自动聚类算法和模糊逻辑关系的中国汽车销量预测

本文利用 1995—2014 年中国汽车销量历史数据，根据自动聚类算法进行区间划分，建立模糊逻辑关系后构建模糊预测模型。如表 1 第 2 列所示，1995—1999 年中国汽车销量增速平稳，

2000—2009 年中国汽车销量增速迅猛,由于受到 2008 年美国次贷危机和中国进入汽车初级社会的双重影响,2010—2014 年中国汽车销量增

幅较小。因此,1995—2014 年中国的汽车销售呈现出非线性变化趋势。

表 1 1995—2014 年中国汽车销量及预测销量 单位:万辆

年份	汽车销量	基于自动聚类算法的预测销量	FCM9 的预测销量	FCM10 的预测销量
1995	144.18	-	-	-
1996	145.87	-	-	-
1997	156.59	-	-	-
1998	160.31	207.85	198.20	175.85
1999	183.30	207.85	198.20	175.85
2000	207.84	207.85	198.20	175.85
2001	237.11	207.85	198.20	218.76
2002	324.85	207.85	198.20	324.99
2003	439.26	515.32	510.02	452.04
2004	507.16	515.32	510.02	452.04
2005	575.82	515.32	510.02	565.27
2006	721.60	714.81	717.68	721.67
2007	879.15	975.82	878.69	909.30
2008	938.05	975.82	937.79	909.30
2009	1 364.48	1 375.01	1 364.73	1 364.75
2010	1 806.19	1 821.95	1 861.24	1 861.24
2011	1 850.51	1 821.95	1 861.24	1 861.24
2012	1 930.64	1 821.95	1 861.24	1 861.24
2013	2 198.41	2 152.12	2 276.21	2 276.20
2014	2 349.19	2 349.19	2 276.21	2 276.20

本部分采用自动聚类算法对中国汽车销量数据进行区间划分,然后基于模糊逻辑关系对中国汽车销量进行仿真预测,具体步骤如下:

第 1 步,对历史数据进行排序处理。如表 1 第 2 列所示,1995—2014 年中国汽车销量数据是递升数列,计算聚类的阈值数 τ 为 116.05。

第 2 步,运用自动聚类算法的计算步骤在得出聚类结果的基础上,求出各聚类的中心值。

第 3 步,根据自动聚类算法,将处理得到的结果转化为区间,如表 2 所示。

表 2 基于自动聚类算法的区间划分

区间	聚类中心	区间下界	中间值	区间上界
u ₁	176.46	102.26	176.46	250.66
u ₂	324.85	250.66	333.40	416.13
u ₃	507.41	416.13	515.32	614.51
u ₄	721.60	614.51	714.81	815.10
u ₅	908.60	815.10	975.82	1 136.54
u ₆	1 364.48	1 136.54	1 375.01	1 613.47
u ₇	1 862.45	1 613.47	1 821.95	2 030.43
u ₈	2 198.41	2 030.43	2 152.12	2 273.80
u ₉	2 349.19	2 273.80	2 349.19	2 424.58

第 4 步,定义“汽车销量”语言项的表示 A_1, A_2, \dots, A_9 ,如下所示:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_9$$

$$A_2 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_9$$

...

$$A_9 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_9$$

第 5 步,将历史数据进行模糊化处理,若数据属于 u_i ,则数据被模糊化为 A_i , $1 \leq i \leq 9$ 。1995—2014 年中国汽车销量历史数据将被模糊化为语言值:1995—2001 年每年销量被模糊化为 A_1 ,2002 年的销量被模糊化为 A_2 ,2003—2005 年每年销量被模糊化为 A_3 ,2006 年的销量被模糊化为 A_4 ,2007—2008 年每年销量被模糊化为 A_5 ,2009 年的销量被模糊化为 A_6 ,2010—2012 年每年销量被模糊化为 A_7 ,2013 年的销量被模糊化为 A_8 ,2014 年的销量被模糊化为 A_9 。

第 6 步,构建模糊逻辑关系。构建单因子三阶模糊时间序列的模糊逻辑关系,如表 3 所示。例如,2001、2002、2003 年和 2004 年模糊化的汽车销量分别是 A_1 、 A_2 、 A_3 和 A_4 ,本文将这四个连续数据之间的模糊逻辑关系表示为“ $A_1, A_2, A_3 \rightarrow A_4$ ”,其中,“ A_1, A_2, A_3 ”和“ A_3 ”分别被

称之为该模糊逻辑关系下的当前状态和下一状态。在此基础上,将拥有相同当前状态的模糊逻辑关系构成一个模糊逻辑关系组。

表3 基于自动聚类算法的单因子三阶

模糊逻辑关系

$A_1, A_1, A_1 \rightarrow A_1$	$A_3, A_3, A_4 \rightarrow A_5$
$A_1, A_1, A_1 \rightarrow A_1$	$A_3, A_4, A_5 \rightarrow A_5$
$A_1, A_1, A_1 \rightarrow A_1$	$A_4, A_5, A_5 \rightarrow A_6$
$A_1, A_1, A_1 \rightarrow A_1$	$A_5, A_5, A_6 \rightarrow A_7$
$A_1, A_1, A_1 \rightarrow A_2$	$A_5, A_6, A_7 \rightarrow A_7$
$A_1, A_1, A_2 \rightarrow A_3$	$A_6, A_7, A_7 \rightarrow A_7$
$A_1, A_2, A_3 \rightarrow A_3$	$A_7, A_7, A_7 \rightarrow A_8$
$A_2, A_3, A_3 \rightarrow A_3$	$A_7, A_7, A_8 \rightarrow A_9$
$A_3, A_3, A_3 \rightarrow A_4$	

第7步,运用模糊预测原则计算出预测值。对2004年和2011年的汽车销量进行预测时,2001—2003年模糊化的汽车销量分别是 A_1 、 A_2 和 A_3 。由表3可知,第三组得模糊逻辑关系是“ $A_1, A_2, A_3 \rightarrow A_3$ ”。因此,2004年汽车销量预测值为 u_3 的中间值,由于 u_3 的中间值为515.32,即2004年中国汽车销量预测值为515.32万辆。同理,2011年中国汽车销量的预测值为1821.95万辆。基于此,通过单因子三阶模糊时间序列对1995—2014中国汽车销量进行预测,如表1第3列所示。

2. 基于FCM算法和模糊逻辑关系的中国汽车销量预测

本部分采用FCM算法对中国汽车销量数据进行区间划分,然后基于模糊逻辑关系对中国汽车销量进行仿真预测。将聚类数分别设定为9和10,以聚类数为9的分析步骤如下:

第1步,通过运用Matlab自带的FCM聚类函数,得出各聚类的中心值。

第2步,根据FCM算法,将相应结果转化为区间,如表4所示。

表4 基于FCM算法的区间划分

区间	聚类中心	区间下界	中间值	区间上界
u_1	168.49	144.18	212.58	280.98
u_2	317.03	280.98	331.52	382.06
u_3	510.02	382.06	515.39	648.71
u_4	717.68	648.71	724.55	800.38
u_5	878.69	800.38	854.49	908.60
u_6	937.79	908.6	1 029.94	1 151.27
u_7	1 364.73	1 151.27	1 368.31	1 585.34
u_8	1 861.24	1 585.34	1 824.94	2 064.53
u_9	2 276.21	2 064.53	2 206.86	2 349.19

第3步,定义“汽车销量”的语言项 A_1 , A_2 ,..., A_9 ,如下所示:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \dots + 0/u_9$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \dots + 0/u_9$$

...

$$A_9 = 0/u_1 + 0/u_2 + 0/u_3 + \dots + 0/u_9$$

第4步,将历史销量数据进行模糊化处理,1995—2014年中国汽车销量历史数据将被模糊化为语言值:1995—2001年每年销量被模糊化为 A_1 ,2002年的销量被模糊化为 A_2 ,2003—2005年每年销量被模糊化为 A_3 ,2006年的销量被模糊化为 A_4 ,2007年的销量被模糊化为 A_5 ,2008年的销量被模糊化为 A_6 ,2009年的销量被模糊化为 A_7 ,2010—2012年每年销量被模糊化为 A_8 ,2013—2014年每年销量被模糊化为 A_9 。

第5步,构建单因子三阶模糊逻辑关系,如表5所示。

表5 基于FCM算法的单因子三阶
模糊逻辑关系

$A_1, A_1, A_1 \rightarrow A_1$	$A_3, A_3, A_4 \rightarrow A_5$
$A_1, A_1, A_1 \rightarrow A_1$	$A_3, A_4, A_5 \rightarrow A_6$
$A_1, A_1, A_1 \rightarrow A_1$	$A_4, A_5, A_6 \rightarrow A_7$
$A_1, A_1, A_1 \rightarrow A_1$	$A_5, A_6, A_7 \rightarrow A_8$
$A_1, A_1, A_1 \rightarrow A_2$	$A_6, A_7, A_8 \rightarrow A_8$
$A_1, A_1, A_2 \rightarrow A_3$	$A_7, A_8, A_8 \rightarrow A_8$
$A_1, A_2, A_3 \rightarrow A_3$	$A_8, A_8, A_8 \rightarrow A_9$
$A_2, A_3, A_3 \rightarrow A_3$	$A_8, A_8, A_9 \rightarrow A_9$
$A_3, A_3, A_3 \rightarrow A_4$	

第6步,基于模糊逻辑关系对1995—2014年中国汽车销量进行仿真预测。同样地,对聚类数为10时也可进行如上的分析,得到的预测销量分别如表1第4列和第5列所示。

四、中国汽车销量预测结果比较

本文通过均方误差(MSE)和平均预测误差率(AFER)两个方面对基于自动聚类算法和FCM算法的预测结果进行比较。

表6 基于自动聚类算法和FCM算法的
预测结果比较

算 法	MSE	AFER
自动聚类算法	3 109.90	0.09
FCM 9	2 823.90	0.09
FCM 10	1 532.10	0.04

如表6所示,基于自动聚类算法的预测值与实际值间的偏差较基于FCM算法的预测结果都要大。另外,基于自动聚类算法所预测的数值在均方误差和平均预测误差率方面都比FCM算法的预测结果要大。虽然较自动聚类算法而言,FCM算法在聚类数确定环节存在需要事先人工设定聚类数目,且区间取值与划分存在着一定主观性等,但是从预测精度和误差等方面来看,FCM算法在科学性和可操作性方面都具有优势。从FCM算法的聚9类和聚10类的计算结果来看,验证了不同区间划分会对预测结果产生重要影响的假设。

五、结 论

本文以夯实中国绿色经济基础与通过有效治理汽车污染以推动社会可持续发展为主线,基于汽车污染的流动性、汽车产业链较长及经济贡献大的特征,为政府制定和调整汽车产业政策提供决策工具,结合前期相关研究,使用模糊逻辑关系和聚类算法的模糊时间序列分别对中国汽车销量进行仿真预测,在比较分析后择优选择预测精度高和误差小的研究方法。首先,通过自动聚类算法将1995—2014年中国汽车销量历史数据划分成不同长度的区间,利用单因子三阶模糊逻辑关系对中国汽车销量进行仿真预测。其次,通过FCM算法进行区间划分后对中国汽车销量进行仿真预测研究。最后,对两种算法的仿真预测进行结果精度比较分析,以检验区间划分对预测结果所产生的影响。

研究结果表明:第一,不同聚类算法的区间划分会对预测结果产生显著影响。第二,基于FCM算法的区间划分和自动聚类方法的区间划分相比具有准确度和精度较高的优势,能准确有效地反映未来中国汽车市场需求走势,并帮助政府合理地进行汽车产业规划与科学引导汽车企业的新能源技术研发与应用。

参考文献:

- [1] 中国汽车工业协会.2016年中国汽车产销量均超2800万辆[N].第一财经日报,2017-01-13.
- [2] 胡世前.激发东北老工业基地全面振兴活力[N].人民日报(理论版),2015-11-19.
- [3] Azizzadeh,L., Zahed, L., Zadeh, L. A. Fuzzy Sets, Information and Control [J]. Information & Control,

1965,8(3):383-353.

- [4] Takagi, T. , Sugeno, M. Fuzzy Identification of Systems and Its Applications to Modeling and Control[J]. IEEE Transaction on SMC,1985,15(1):116-132.
- [5] Song, Q. , Chissom, B. S. Fuzzy Time Series and Its Model[J]. Fuzzy Sets & System, 1993, 54 (3) : 269 -277.
- [6] Aladag, C. H. , Basaran, M. A. , Egrioglu, E. , Yolcu, U. , Uslu, V. R. Forecasting in High Order Fuzzy Times Series by Using Neural Networks to Define Fuzzy Relations[J]. Expert Systems with Applications,2009, 36 (3):4228-4231.
- [7] Egrioglu, E. , Aladag, C. H. , Yolcu, U. , Uslu, V. R. , Erilli, N. A. Fuzzy Time Series Forecasting Method Based on Gustafson – Kessel Fuzzy Clustering [J]. Expert Systems with Applications, 2011,38(8):10355 -10357.
- [8] Bai, E. , Wong, W. K. , Chu, W. C. , Xia, M. , Pan, F. A Heuristic Time – Invariant Model for Fuzzy Time Series Forecasting [J]. Expert Systems with Applications,2011,38(3):2701-2707.
- [9] Kuo, I. H. , Horng, S. J. , Chen, Y. H. , Run, R. S. , Kao, T. W. , Chen, R. J. , Lai, J. L. , Lin, T. L. Forecasting TAIFEX Based on Fuzzy Time Series and Particle Swarm Optimization [J]. Expert Systems with Applications, 2010,37(2):1494-1502.
- [10] Chen,S. M. Forecasting Enrollments Based on Fuzzy Time Series[J]. Fuzzy Sets & Systems,1996,81(3) : 311-319.
- [11] Lee, M. H. , Efendi, R. , Ismail, Z. Modified Weighted for Enrollment Forecasting Based on Fuzzy Time Series [J]. MATEMATIKA, 2009, 25 (1) : 67 -78.
- [12] 约翰·T·门泽尔,卡罗尔·C·贝恩斯多克.销售预测:方法系统管理[M].罗玉波等译,北京:人民邮电出版社,2004.9-11.
- [13] 王维国,殷亮.半参数趋势阈值面板模型及其参数估计[J].数量经济技术经济研究,2014,(9):124-137.
- [14] 高铭.模糊时间序列在汽车销售中的应用[D].大连:大连理工大学硕士学位论文,2015.25-27.
- [15] 邱望仁,刘晓东.模糊时间序列模型研究综述[J].模糊系统与数学,2014,(3):173-181.
- [16] 曲宏巍.模糊时间序列模型相关理论的研究[D].大连:大连海事大学硕士学位论文,2012.14-15.

(责任编辑:尚培培)