第 29 卷 第 2 期 控 制 与 决 策 2014 年 2 月
Vol. 29 No. 2 Control and Decision Feb. 2014

文章编号: 1001-0920 (2014) 02-0251-06 DOI: 10.13195/j.kzyjc.2012.0933

一种关于模糊时间序列模型论域划分的多尺度比率算法

陈 刚, 曲宏巍, 初莹莹

(大连海事大学 数学系，辽宁 大连 116026)

摘 要: 针对模糊时间序列研究中比率划分论域方法存在对非均匀数据划分效果不理想的缺陷, 提出一种基于模糊 C 均值聚类 (FCM) 算法的多尺度比率划分论域的方法. 首先利用 FCM 算法对样本数据进行分类; 然后计算各类数据的平均相对误差, 并基于各类的平均误差划分论域, 产生非等间隔的多尺度论域划分方法; 最后, 通过算例表明了多尺度比率论域划分方法的有效性.

关键词: 模糊时间序列；多尺度比率；模糊 C 均值聚类；有效性

中图分类号: TP273 **文献标志码:** A

A method of multi-scale ratios interval partitioning about fuzzy time series model

CHEN Gang, *QU Hong-wei*, *CHU Ying-ying*

(Department of Mathematics，Dalian Maritime University，Dalian 116026，China．Correspondent：CHEN Gang, E-mail：chengang@dlmu.edu.cn)

Abstract: In view of the shortcoming of Huarng and Yu's ratio method, this paper presents a method of multi-scale ratios interval partitioning. Firstly, the sample data are classified based on the fuzzy C means clustering algorithm. Then, unequal-sized intervals are obtained based on average relative differences of various classes by calculating average relative difference every class. Finally, real examples show the effectiveness of the proposed interval partitioning method.

Key words: fuzzy time series；multi-scale ratios；fuzzy C means；efficiency

0 引 言

传统的时间序列模型能够处理许多预测问题, 在自然灾害、工程控制、经济和金融等领域均取得令人瞩目的成果, 但不能有效解决数据是语言值或是不完整、不确定的预测问题. Song 等[1-3]基于模糊集理论建立了模糊时间序列模型并成功进行了预测, 其预测框架由 4 个步骤构成: 1) 论域的定义和划分; 2) 定义模糊集和模糊时间序列; 3) 建立模糊关系; 4) 去模糊化和预测. 目前, 在模糊时间序列模型的研究中, 大部分研究工作集中于改善 Song 和 Chissom 框架下的步骤1) 和步骤 3). 对于步骤 1), 开始只是局限在等间隔论域的划分方法上[4-5], 之后 Huarng[6]提出了基于分布长度和基于平均长度两种等间隔论域划分方法, 这一研究使人们意识到论域划分的间隔长度对预测结果有极大的影响. 随着研究的深入, 人们开始打破等间隔的论域划分的限制, 提出了非等分论域方法, 如比率论域划分方法[7]、基于模糊 C 均值聚类 (FCM) 算法的论域划分方法[8]和其他非等分论域划分方法[9-13]. 可见, 模糊时间序列模型中论域的非等间隔划分方法引起了人们越来越多的关注. 2006 年, Huarng 等[7]提出了基于比率论域非等分划分方法, 该方法在数据分布较为均匀时效果明显, 提高了预测精度, 但是, 当样本数据在论域上分布不均匀时, 便会在论域划分中出现有的区间数据过分集中、有的区间数据稀少或没有数据的情形, 不利于数据的模糊化, 也不能真实地反映样本数据的分布特征.

基于上述分析, 本文在论域划分上提出基于多尺度比率划分方法. 首先基于 FCM 算法对数据进行分类; 然后求出各类数据的平均相对误差; 基于各类得出的平均误差划分各类, 最终形成多尺度的论域划分方法. 对美国 Aalbama 大学[1,4,7,14-15]的注册人数和库存需求数据进行预测, 实验结果验证了所提出算法的

收稿日期: 2012-06-18；**修回日期:** 2012-11-16.

基金项目: 国家自然科学基金项目(60875032)；中央高校创新团队项目(3132013332).

作者简介: 陈刚(1964−), 男, 教授, 博士生导师, 从事模糊推理、模糊系统优化等研究；曲宏巍(1987−), 女, 硕士生, 从事数据处理与信息提取的研究.

有效性.

1 相关知识

1.1 模糊时间序列概念

令 U 为给定论域, 将 U 划分为 n 个子区间, 即 $U = u_1, u_2, \cdots, u_n$, 则定义在论域 U 中的模糊集合 A 表示为

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \cdots + \frac{f_A(u_n)}{u_n}. \tag{1}$$

其中: $f_{A_i}(\cdot)$ 为模糊集合 A_i 的隶属函数, $f_{A_i}(\cdot): U \to [0,1]$; u_k 为模糊集合 A_i 的一个因素; $f_{A_i}(u_k)$ 为对模糊集合 A_i 的隶属度, $f_{A_i}(u_k) \in [0,1]$, $k = 1,2,\cdots,n$.

定义 1 令 R 中一子集 $Y(t)$ 为给定论域, $f_i(t)(i = 1,2,\cdots)$ 为定义在其上的模糊集合, $F(t)$ 为由 $f_i(t)$ $(i = 1,2,\cdots)$ 组成的集合, 则称 $F(t)$ 为定义在 $Y(t)(t = 0,1,\cdots)$ 上的模糊时间序列[1].

定义 2 假设 $F(t)$ 由 $F(t-1)$ 确定, 即 $F(t-1) \to F(t)$, 可以表示为

$$F(t) = F(t-1) \circ R(t, t-1). \tag{2}$$

式 (2) 称作 $F(t)$ 的一阶模糊时间序列模型, $R(t, t-1)$ 为 $F(t-1)$ 与 $F(t)$ 之间的模糊关系[1].

定义 3 设 $F(t-1) = A_i$, $F_t = A_j$, 则一个模糊逻辑关系可以定义为 $A_i \to A_j$. 其中: A_i 为模糊关系的前件, A_j 为模糊关系的后件[1]. 若有模糊关系重复出现, 则只保留一组.

1.2 FCM 算法

Bezdek 于 1981 年首次提出 FCM 算法, 之后经过人们的不断完善, 目前已成为广泛使用的一种聚类算法. 它通过优化目标函数确定误差最小的函数 $J_m(U, P)$, 优化目标函数为

$$J_m(U,P) = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^2(d_{ik})^2,$$
$$\text{s.t. } U \in M_{fc}. \tag{3}$$

算法步骤如下.

Step 1: 初始化. 首先给定聚类类别数 $c(c \geqslant 2)$、数据个数 n、迭代停止阀值 ε、初始化聚类原型模式 $P^{(0)}$、迭代计数器 $b = 0$.

Step 2: 计算或更新划分矩阵 $U^{(b)}$, 有

$$\mu_{ik}^{(b)} = \Big\{\sum_{j=1}^{c}\Big[\Big(\frac{d_{ik}^{(b)}}{d_{jk}^{(b)}}\Big)^{\frac{2}{m-1}}\Big]\Big\}^{-1}, \tag{4}$$

其中对于 $\forall i,k, \exists\, d_{ik}^{(b)} > 0$. 若 $\exists\, i, r$ 使得 $d_{ik}^{(b)} = 0$, 则 $\mu_{ir}^{(b)} = 1$, 且 $j \neq r$, $\mu_{ir}^{(b)} = 0$.

Step 3: 更新聚类原型模式矩阵 $P^{(b+1)}$, 有

$$P_i^{(b+1)} = \sum_{k=1}^{n}(\mu_{ik}^{b+1})^m x_k \Big/ \sum_{k=1}^{n}(\mu_{ik}^{b+1})^m,$$
$$i = 1,2,\cdots,c. \tag{5}$$

Step 4: 若 $\|P^{(b)} - P^{(b+1)}\| < \varepsilon$, 则算法停止, 输出划分矩阵 U 和聚类原型 P, 否则令 $b = b+1$, 转至 Step 2, 其中 $\|\cdot\|$ 为某种合适的矩阵范数.

FCM 算法也具有另外一种形式, 首先初始化模糊划分矩阵, 利用式 (5) 计算聚类原型 (中心) 矩阵; 然后利用式 (4) 更新模糊分类矩阵, 直到满足停止准则为止.

2 多尺度比率算法

模糊时间序列模型的预测精度主要取决于 3 部分: 论域划分、数据模糊化和构建的模糊关系. 针对目前在模糊时间序列论域划分上存在的问题, 提出基于 FCM 的多尺度比率划分论域方法.

2.1 基于比率的论域划分方法[7]

Step 1: 设训练样本有 n 个数据 $x_1, x_2, \cdots, x_n$, 计算 $x_1, x_2, \cdots, x_n$ 相邻数据间的相对误差为

$$r_t = \frac{x_{t+1} - x_t}{x_t},\ t = 1,2,\cdots,n-1.$$

Step 2: 基于 Step 1 得到的结果描绘出所有 r_1 的分布图[7].

Step 3: 选择适当的比率, 记为 ratio, 一般选 r_1 分布图中 50% 处对应的数据.

Step 4: 给定初值并划分论域. 初值为

$$x_{\text{initial}} = ab' \times 10^z. \tag{6}$$

始于初值, 间隔数通过比率增加, 具体如下所示: 当 $j \geqslant 1$ 时, 有

$$x_j = (1 + \text{ratio})^j x_{\text{initial}},\ u_j = [x_{j-1}, x_j]. \tag{7}$$

其中: $0 \leqslant a \leqslant 9$, $0 \leqslant b \leqslant 9$, $b' = b - 1$. 论域 U 被划分为 $u_1, u_2, \cdots, u_j$, 即 $U = u_1, u_2, \cdots, u_j$. 可以发现, 选择相对误差分布图中 50% 处对应的数据作为比率划分论域, 对样本数据均匀分布效果较好.

2.2 基于多尺度比率的论域划分方法

鉴于比率法的不足, 本文提出首先利用 FCM 算法将数据进行聚类, 然后对每一类采用相同比率进行论域划分, 如此形成多尺度比率的论域划分方法. 该算法可以有效避免上述问题, 具体步骤描述如下.

Step 1: 基于 FCM 算法将 n 个数据分成 $X_1, X_2, \cdots, X_m (m < n)$ 类, 为了方便确定论域划分端点值, 将每一类数据分别按升序排列, 结果如下所示:

$$X_1 = \{d_{11}, d_{12}, \cdots, d_{1n_1}\},$$
$$X_2 = \{d_{21}, d_{22}, \cdots, d_{2n_2}\},$$
$$\vdots$$
$$X_m = \{d_{m1}, d_{m2}, \cdots, d_{mn_m}\}.$$

其中: $1 \leqslant n_k \leqslant n$, $k = 1,2,\cdots,m$.

Step 2: 首先计算 $X_1, X_2, \cdots, X_m$ 类相邻数据的相对误差, 分别为

$$r_{1_i} = \frac{d_{1(i+1)} - d_{1i}}{d_{1i}},\ 1 \leqslant i \leqslant n_1; \tag{8}$$

$$r_{2_i} = \frac{d_{2(i+1)} - d_{2i}}{d_{2i}},\ 1 \leqslant i \leqslant n_2; \tag{9}$$

$$\vdots$$

$$r_{m_i} = \frac{d_{m(i+1)} - d_{mi}}{d_{mi}},\ 1 \leqslant i \leqslant n_m. \tag{10}$$

然后计算 $X_1, X_2, \cdots, X_m$ 类相邻数据的平均相对误差, 分别为

$$\mathrm{ratio}_1 = \frac{\sum_{i=1}^{n_1} r_{1_i}}{n_1}, \tag{11}$$

$$\mathrm{ratio}_2 = \frac{\sum_{i=1}^{n_2} r_{2_i}}{n_2}, \tag{12}$$

$$\vdots$$

$$\mathrm{ratio}_m = \frac{\sum_{i=1}^{n_m} r_{m_i}}{n_m}. \tag{13}$$

Step 3: 每一类根据各自的比率进行划分, 将前一类中根据比率划分的最后一个小区间的右端点作为下一类开始划分的左端点, 如此进行划分, 可以将整个论域分成不同小区间. 通过上述方法定义论域 $U_1, U_2, \cdots, U_m$ 分别为

$$U_1 = [d_{11} - \sigma_{11}, u_a], \tag{14}$$

$$U_2 = [u_a, u_b], \tag{15}$$

$$\vdots$$

$$U_m = [u_q, d_{mn_m} + \sigma_{m2}]. \tag{16}$$

其中: σ_{11}, σ_{m2} 为适当的正数; u_a, u_b, u_q 为每一类中通过各自比率划分后的最后一个小区间的右端点值. 此外, 定义整个论域 U 为

$$U = [d_{11} - \sigma_{11}, d_{mn_m} + \sigma_{m2}]. \tag{17}$$

Step 4: 给定初值并划分论域. 设初值 x_{initial} 为

$$x_{\mathrm{initial}} = d_{11} - \sigma_{11}. \tag{18}$$

始于初值, 间隔数通过比率增加, 具体如下所示: 当 $j \geqslant 1$ 时, 有

$$x_j = (1 + \mathrm{ratio})^j x_{\mathrm{initial}},\ u_j = [x_{j-1}, x_j]. \tag{19}$$

将论域 U_1 划分为 $i+1$ 个区间, 有

$$u_{1_1} = [d_{11} - \sigma_{11}, (1 + \mathrm{ratio}_1)(d_{11} - \sigma_{11})], \tag{20}$$

$$u_{1_2} = [(1+\mathrm{ratio}_1)(d_{11}-\sigma_{11}), (1+\mathrm{ratio}_1)^2(d_{11}-\sigma_{11})], \tag{21}$$

$$\vdots$$

$$u_{1_i} = [(1 + \mathrm{ratio}_1)^{i-1}(d_{11} - \sigma_{11}), u_a], \tag{22}$$

此时 $u_a = (1 + \mathrm{ratio}_1)^i(d_{11} - \sigma_{11})$. 将论域 U_2 划分为 $k+1$ 个区间, 有

$$u_{2_1} = [u_a, (1 + \mathrm{ratio}_2)u_a], \tag{23}$$

$$u_{2_2} = [(1 + \mathrm{ratio}_2)u_a, (1 + \mathrm{ratio}_2)^2 u_a], \tag{24}$$

$$\vdots$$

$$u_{2_k} = [(1 + \mathrm{ratio}_2)^{k-1} u_a, u_b], \tag{25}$$

此时 $u_b = (1 + \mathrm{ratio}_2)^k u_a$. 将论域 U_m 划分为 $s+1$ 个区间, 有

$$u_{m_1} = [u_q, (1 + \mathrm{ratio}_m)u_q], \tag{26}$$

$$u_{m_2} = [(1 + \mathrm{ratio}_m)u_q, (1 + \mathrm{ratio}_m)^2 u_q], \tag{27}$$

$$\vdots$$

$$u_{m_s} = [(1 + \mathrm{ratio}_m)^{s-1} u_q, d_{mn_m} + \sigma_{m2}]. \tag{28}$$

Step 5: 论域 U 被划分成的区间为 $u_{1_1}, u_{1_2}, \cdots, u_{1_i}, u_{2_1}, u_{2_2}, \cdots, u_{2_k}, \cdots, u_{m_1}, u_{m_2}, \cdots, u_{m_s}$.

3　新方法下的模糊时间序列模型

在多尺度比率划分论域方法下, 建立模糊时间序列模型的过程如下.

Step 1: 定义论域和划分区间. 论域 $U = [D_{\min} - D_1, D_{\max} + D_2]$. 其中: $D_{\min}$ 和 $D_{\max}$ 分别为样本数据中的最小值和最大值, D_1 和 D_2 为适当的正值. 基于第 2.2 节中提出的多尺度比率算法, 将论域 U 划分为长度不等的区间 $u_1, u_2, \cdots, u_n$.

Step 2: 定义模糊集和模糊数据. 定义模糊集合为

$$A_k = \frac{f_{A_k}(u_1)}{u_1} + \frac{f_{A_k}(u_2)}{u_2} + \cdots + \frac{f_{A_k}(u_n)}{u_n}.$$

通过区间对数据进行模糊化, 数据模糊化规则为: 若 x_k 是样本中任意一个数据点, 则该数据属于 $u_k (1 \leqslant k \leqslant n)$, 模糊集 A_k 的最大隶属度对应区间 u_k, 这样即可将该数据模糊化成 A_k. 将所有样本按上述方法模糊化后得到对应的模糊时间序列.

Step 3: 建立模糊关系为 $A_j \to A_q, A_j \to A_q, \cdots$; 建立模糊关系组为 $A_j \to A_q, A_r, \cdots$.

Step 4: 添加启发式知识并建立启发式模糊逻辑关系组[6]. 若模糊关系组为 $A_j \to A_q, A_r, \cdots$, 则一个启发式模糊关系组为 $A_j \to A_{p_1}, A_{p_2}, \cdots, A_{p_k}$, 其中 $A_{p_1}, A_{p_2}, \cdots, A_{p_k}$ 选自 $A_q, A_r, \cdots$.

Step 5: 去模糊化和预测. 利用平均值去模糊化法得到预测结果.

4　新算法有效性验证

为了验证多尺度算法的有效性, 本文选文献 [7] 中一组 Alabama 大学 1971 ∼ 1992 年的注册人数作为

样本数据进行比较研究. 对于另一组库存需求数据, 限于篇幅, 这里仅给出计算结果.

对于Alabama大学1971～1992年的注册人数, 首先计算两相邻年间的差值(后一年人数减去前一年人数), 正值表示注册人数增加, 负值表示注册人数减少, 得到21个变化值数据如表1所示.

表 1 实际注册人数和数据变化值

年份	实际人数	变化值
1971	13 055	
1972	13 563	+508
1973	13 867	+304
1974	14 696	+829
1975	15 460	+764
1976	15 311	−149
1977	15 603	+292
1978	15 861	+258
1979	16 807	+946
1980	16 919	+112
1981	16 388	−531
1982	15 433	−955
1983	15 497	+64
1984	15 145	−352
1985	15 163	+18
1986	15 984	+821
1987	16 859	+875
1988	18 150	+1291
1989	18 970	+820
1990	19 328	+358
1991	19 337	+9
1992	18 876	−461

选取聚类数 $c = 3$, 根据第3节描述的建模步骤, 给出具体实例的预测过程($c = 3$是随机选取的, 仅用来描述本文的算法).

Step 1: 定义论域和划分区间.

Step 1.1: 设训练样本有22个数据, 利用FCM算法将22个数据分成 X_1, X_2, X_3 类, 结果如下所示:

$$X_1 = \{13\,055, 13\,563, 13\,867\},$$

$$X_2 = \{14\,696, 15\,145, 15\,163, 15\,311, 15\,433, 15\,460, 15\,497, 15\,603, 15\,861, 15\,984, 16\,388, 16\,807, 16\,859, 16\,919\},$$

$$X_3 = \{18\,150, 18\,876, 18\,970, 19\,328, 19\,337\}.$$

Step 1.2: 计算 X_1, X_2, X_3 类相邻数据的相对误差分别为

$$r_{1_2} = 0.038\,9,\ r_{1_3} = 0.022\,4,$$
$$r_{2_2} = 0.030\,6,\ r_{2_3} = 0.001\,2,$$
$$r_{2_4} = 0.009\,8,\ r_{2_5} = 0.008\,0,$$
$$r_{2_6} = 0.001\,7,\ r_{2_7} = 0.002\,4,$$
$$r_{2_8} = 0.006\,8,\ r_{2_9} = 0.016\,5,$$
$$r_{2_{10}} = 0.007\,8,\ r_{2_{11}} = 0.025\,3,$$
$$r_{2_{12}} = 0.025\,6,\ r_{2_{13}} = 0.003\,1,$$
$$r_{2_{14}} = 0.003\,6,\ r_{3_2} = 0.040\,0,$$
$$r_{3_3} = 0.005\,0,\ r_{3_4} = 0.018\,9,$$
$$r_{3_5} = 0.000\,5.$$

计算 X_1, X_2, X_3 类相邻数据的平均相对误差分别为

$$\mathrm{ratio}_1 = 0.030\,7,\ \mathrm{ratio}_2 = 0.010\,9,\ \mathrm{ratio}_3 = 0.016\,1.$$

Step 1.3: 根据 X_1, X_2, X_3 类数据, 分别定义论域 U_1, U_2, U_3 为

$$U_1 = [13\,000, 14\,233],$$
$$U_2 = [14\,233, 16\,940],$$
$$U_3 = [16\,940, 20\,000].$$

其中: $\sigma_{11} = 55$, $\sigma_{m1} = 663$. 定义论域 U 为 $U = [13\,000, 20\,000]$.

Step 1.4: 给定初值并划分论域, 有

$$x_{\mathrm{initial}} = d_1 - \sigma_{11},$$
$$u_{1_1} = [13\,000, 13\,399],$$
$$u_{1_2} = [13\,399, 13\,810], \cdots, u_{3_{10}} = [19\,555, 20\,000].$$

Step 2: 定义模糊集和模糊数据. 根据Step 1.1得到29个区间, 定义模糊集为

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_{28}} + \frac{0}{u_{29}},$$
$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \cdots + \frac{0}{u_{28}} + \frac{0}{u_{29}},$$
$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \cdots + \frac{0}{u_{28}} + \frac{0}{u_{29}},$$
$$\vdots$$
$$A_{29} = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \cdots + \frac{0.5}{u_{28}} + \frac{1}{u_{29}}.$$

依照第2.2节建模步骤中描述的数据模糊化规则进行数据模糊化, 结果如表2所示.

Step 3: 建立模糊关系为

$$A_1 \to A_2,\ A_2 \to A_3, A_3 \to A_6,$$
$$A_6 \to A_{11},\ A_{11} \to A_{10},\ A_{10} \to A_{12},$$
$$A_{12} \to A_{13},\ A_{13} \to A_{19},\ A_{19} \to A_{19},$$
$$A_{19} \to A_{16},\ A_{16} \to A_{11},\ A_{11} \to A_{11},$$
$$A_{11} \to A_9,\ A_9 \to A_9,\ A_9 \to A_{14},$$
$$A_{14} \to A_{19},\ A_{19} \to A_{24},\ A_{24} \to A_{27},$$
$$A_{27} \to A_{28},\ A_{28} \to A_{28},\ A_{28} \to A_{26}.$$

建立模糊关系组为

$$A_1 \to A_2,\ A_2 \to A_3,\ A_3 \to A_6,\ A_6 \to A_{11},$$

$A_{11} \to A_9, A_{10}, A_{11}$, $A_{10} \to A_{12}$, $A_{12} \to A_{13}$, $A_{13} \to A_{19}$, $A_{19} \to A_{16}, A_{19}, A_{24}$, $A_{16} \to A_{11}$, $A_9 \to A_9, A_{14}$, $A_{14} \to A_{19}$, $A_{24} \to A_{27}$, $A_{27} \to A_{28}$, $A_{28} \to A_{28}, A_{26}$.

表 2　实际注册人数、模糊化数据和变化值

年份	实际人数	模糊集	变化值
1971	13 055	A_1	
1972	13 563	A_2	+ 508
1973	13 867	A_3	+304
1974	14 696	A_6	+829
1975	15 460	A_{11}	+764
1976	15 311	A_{10}	−149
1977	15 603	A_{12}	+292
1978	15 861	A_{13}	+258
1979	16 807	A_{19}	+946
1980	16 919	A_{19}	+112
1981	16 388	A_{16}	−531
1982	15 433	A_{11}	−955
1983	15 497	A_{11}	+64
1984	15 145	A_9	−352
1985	15 163	A_9	+18
1986	15 984	A_{14}	+821
1987	16 859	A_{19}	+875
1988	18 150	A_{24}	+1291
1989	18 970	A_{27}	+820
1990	19 328	A_{28}	+358
1991	19 337	A_{28}	+9
1992	18 876	A_{26}	−461

计算 RMSE 如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(X_{\text{actual}}(t) - X_{\text{forecasted}}(t))^2}{n}}. \tag{29}$$

计算 RMSE 如下:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(X_{\text{actual}}(t) - X_{\text{forecasted}}(t))^2}{n}. \tag{30}$$

Step 4: 添加启发式知识并建立启发式模糊逻辑关系组. 建立启发式模糊逻辑关系组, 具体结果如表 3 所示.

Step 5: 去模糊化和预测. 通过启发式模糊关系组预测 $F(t)$, 预测结果如表 4 所示. 相对误差 Error 计算如下:

$$\text{Error} = \frac{X_{\text{forecasted}}(t) - X_{\text{actual}}(t)}{X_{\text{actual}}(t)} \times 100\%.$$

两种方法 RMSE 的比较结果如表 5 所示. 由表 5 两种方法预测结果的比较可见, 本文提出的多尺度方法明显优于 Huarng 的比率法[7]. 目前, 较好的预测结果是 Egrioglu 等[21]的 MSE 值 60 140, 而本文得到的结果更优, 因此验证了多尺度划分论域方法的有效性, 具体结果如表 6 所示.

表 3　启发式模糊逻辑关系组

模糊逻辑关系组	启发式	启发式模糊关系组
$A_1 \to A_2$	↑	$A_1 \to A_2$
$A_2 \to A_3$	↑	$A_2 \to A_3$
$A_3 \to A_6$	↑	$A_3 \to A_6$
$A_6 \to A_{11}$	↑	$A_6 \to A_{11}$
$A_{11} \to A_9, A_{10}, A_{11}$	↑	$A_{11} \to A_{11}$
$A_{11} \to A_9, A_{10}, A_{11}$	↓	$A_{11} \to A_9, A_{10}$
$A_{10} \to A_{12}$	↑	$A_{10} \to A_{12}$
$A_{12} \to A_{13}$	↑	$A_{12} \to A_{13}$
$A_{13} \to A_{19}$	↑	$A_{13} \to A_{19}$
$A_{19} \to A_{16}, A_{19}, A_{24}$	↑	$A_{19} \to A_{19}, A_{24}$
$A_{19} \to A_{16}, A_{19}, A_{24}$	↓	$A_{19} \to A_{16}, A_{19}$
$A_{14} \to A_{19}$	↑	$A_{14} \to A_{19}$
$A_{24} \to A_{27}$	↑	$A_{24} \to A_{27}$
$A_{27} \to A_{28}$	↑	$A_{27} \to A_{28}$
$A_{28} \to A_{28}, A_{26}$	↑	$A_{28} \to A_{28}$
$A_{28} \to A_{28}, A_{26}$	↓	$A_{28} \to A_{28}, A_{26}$

表 4　实际值、预测值和相对误差

年份	实际人数	预测值	相对误差
1971	13 055		
1972	13 563	13 604	+ 508
1973	13 867	14 022	+304
1974	14 696	14 626	+829
1975	15 460	15 444	+764
1976	15 311	15 277	−149
1977	15 603	15 613	+292
1978	15 861	15 783	+258
1979	16 807	16 849	+946
1980	16 919	16 849	+112
1981	16 388	16 578	−531
1982	15 433	15 444	−955
1983	15 497	15 444	+64
1984	15 145	15 277	−352
1985	15 163	15 534	+18
1986	15 984	15 534	+821
1987	16 859	16 849	+875
1988	18 150	18 201	+1291
1989	18 970	19 094	+820
1990	19 328	19 401	+358
1991	19 337	19 401	+9
1992	18 876	19 097	−461

表 5　两种方法预测结果 RMSE 比较

预测数据	比率法	新方法
Alabama 大学注册人数	396.1	135.7
库存需求数据 (19～24)	15.1	8.1

表 6　不同方法预测结果比较

方法	Huarng 法	Cheng 法	Egioglu 法	新方法
MSE	226 611	261 142	60 140	47 143

5 结 论

针对Huarng比率法[7]存在的不足, 利用FCM算法建立了一种多尺度比率划分论域的方法. 通过Alabama大学和库存需求数据的对比研究充分表明了多尺度比率论域划分方法的有效性, 从而验证了所提出方法对于样本数据多样性的适应能力.

参考文献(References)

[1] Song Q, Chissom B S. Fuzzy time series and its models[J]. Fuzzy Sets and Systems, 1993, 54(1): 269-277.

[2] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series: Part I[J]. Fuzzy Sets and Systems, 1993, 54(1): 1-10.

[3] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series: Part II[J]. Fuzzy Sets and Systems, 1994, 62(1): 1-8.

[4] Chen S M. Forecasting enrollments based on fuzzy time-series[J]. Fuzzy Sets and Systems, 1996, 81(3): 311-319.

[5] Hwang Jeng-ren, Chen Shyi-ming, Lee Chia-hoang. Handling forecasting problems using fuzzy time series[J]. Fuzzy Sets and Systems, 1998, 100(2): 217-228.

[6] Huarng K. Heuristic models of fuzzy time series for forecasting[J]. Fuzzy Sets and Systems, 2001, 123(3): 369-386.

[7] Huarng K, Yu T H K. Ratio-based lengths of intervals to improve fuzzy time series forecasting[J]. IEEE Trans on Systems, Man and Cybernetics, Part B: Cybernetics, 2006, 36(2): 328-340.

[8] Li S T, Cheng Y C, Lin S Y. A FCM-based deterministic forecasting model for fuzzy time series[J]. Computers and Mathematics with Applications, 2008, 56(12): 3052-3063.

[9] Huarng K, Yu T H K. The application of neural networks to forecast fuzzy time series[J]. Physica A, 2006, 363(2): 481-491.

[10] Cheng C H, Chang R J, Yeh C A. Entropy-based and trapezoidal Fuzzification based fuzzy time series approach for forecasting IT project cost[J]. Technological Forecasting and Social Change, 2006, 73(5): 524-544.

[11] Chen S M, Chung N Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms[J]. Int J of Intellegent Systems, 2006, 21(5): 485-501.

[12] Chen S M, Chung N Y. Forecasting enrollments of students by using fuzzy time series and genetic algorithms[J]. Information and Management Sciences, 2006, 17(3): 1-17.

[13] Kuo I H, Horng S J, Kao T W, et al. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization[J]. Expert Systems with Applications, 2009, 36(3): 6108-6117.

[14] Sullivan J, Woodall W H. A comparison of fuzzy forecasting and Markov modeling[J]. Fuzzy Sets and Systems, 1994, 64(3): 279-293.

[15] Aladag C H, Basaran M A, Egrioglu E, et al. Forecasting in high order fuzzy time series by using neural networks to define fuzzy relations[J]. Expert Systems with Applications, 2009, 36(3): 4228-4231.

(责任编辑: 郑晓蕾)