

文章编号: 1001-0920(2012)01-0099-05

## 基于证据理论的模糊时间序列预测模型

邱望仁<sup>1,2</sup>, 刘晓东<sup>1</sup>

(1. 大连理工大学 电信学院, 辽宁 大连 116024; 2. 景德镇陶瓷学院 信息工程学院, 江西 景德镇 333403)

**摘要:** 在分析经典模糊时间序列预测模型的基础上, 指出了传统的模型不能处理多因素的情形; 然后分析并改进了证据理论中关于证据合成的方法, 提出了基于证据理论的多因素模糊时间序列预测模型; 最后用1997年~2006年10年间的上海股指数据对所提出的模型进行了实践检验, 实验结果表明该模型是可行的, 其预测效果优于所参照的预测模型.

**关键词:** 模糊理论; 时间序列; 证据理论

**中图分类号:** TP274

**文献标识码:** A

## Fuzzy time series model for forecasting based on Dempster-Shafer theory

QIU Wang-ren<sup>1,2</sup>, LIU Xiao-dong<sup>1</sup>

(1. School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116024, China; 2. School of Computer Engineering, Jingdezhen Ceramic Institute, Jingdezhen 333403, China. Correspondent: QIU Wang-ren, E-mail: qiuone@163.com)

**Abstract:** Based on the analysis of some conventional fuzzy time series forecast models, this paper argues that these conventional models can not deal with multiple-attribute fuzzy time series and proposes a model based on Dempster-Shafer theory to overcome this shortcoming with some revise for combining belief function. The Shanghai stock index over 10 years period 1997~2006 is chosen as the forecasting targets, and the empirical results show that the proposed model outperforms the compared models.

**Key words:** fuzzy theory; time series; Dempster-Shafer theory

### 1 引言

时间序列分析方法不但可以从数量上揭示现象的变化规律, 而且还能预测和控制现象的未来行为和特征. 为解决自然科学与社会科学研究中存在的界限模糊, 认识上存在偏差等不确定性问题, 已诞生了一系列研究不确定性问题的理论, 如: 证据理论、模糊理论、粗糙集等. 这些理论推动了时间序列分析在预测、建模和特征提取等方面的研究和应用.

1993年 Song 和 Chissom<sup>[1]</sup>提出了模糊时间序列模型, 随后该模型广泛地应用于经济、社会生活等领域, 特别是其预测功能常被用于入学人数<sup>[1-8]</sup>、温度<sup>[9]</sup>和股票指数<sup>[6-11]</sup>的预测. 然而, 传统的模糊时间序列模型大多是用单变量来预测下一阶段的值. 事实上, 与该变量相关的因素可能不止一个, 这样会导致人们对传统的模糊时间序列预测模型的应用范围产生质

疑, 而且所得到的结果往往精度不高. 虽然 Kuang<sup>[10]</sup>提出的模型讨论了多因素的情形, 但从其应用上看, 它所考虑的因素是同一个因素的不同表现形式, 故从严格意义上它并不是多因素的模糊时间序列预测.

基于上述分析, 本文结合证据理论能有效处理多种信息的特点, 提出了建立基于证据理论的模糊时间序列预测模型的方法, 并用10年的沪市股指作为实验数据证明了该模型的可靠性和优越性.

### 2 模糊时间序列的基本概念与定义

下面先给出完整的模糊时间序列模型的定义和构建步骤; 然后介绍几个主要的构建模糊关系矩阵的方法, 并对它们在预测中的应用进行分析比较.

**定义 1**  $U$  为论域, 给定  $U$  的一个次序分割集为  $U = \{u_1, u_2, \dots, u_n\}$ , 定义  $A$  为论域  $U$  上的语义变量集 (即模糊集), 并记为

收稿日期: 2010-08-27; 修回日期: 2010-11-10.

基金项目: 国家自然科学基金项目(60875032, 60961003); 江西省自然科学基金项目(2010QJS0127).

作者简介: 邱望仁(1978-), 男, 副教授, 博士, 从事模糊理论、数据挖掘等研究; 刘晓东(1963-), 男, 教授, 博士生导师, 从事模糊理论、数据挖掘等研究.

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

其中:  $f_A$  为定义在  $A$  上的模糊隶属函数,  $f_A: U \rightarrow [0, 1]$ ;  $f_A(u_i)$  为  $u_i$  在模糊集  $A$  上的模糊隶属度的值, 并且  $f_A(u_i) \in [0, 1], 1 \leq i \leq n$ .

**定义 2** 对于任一固定的  $t(t = \cdots, 0, 1, 2, \cdots)$ , 设  $Y(t) \subset R$ , 即为实数域的子集,  $Y(t)$  上定义着一组模糊集  $f_i(t)(i = 1, 2, \cdots)$  且  $F(t) = \{f_1(t), f_2(t), \cdots\}$ , 则称  $F(t)$  为定义在  $Y(t)$  上的模糊时间序列. 其中:  $F(t)$  为语言变量;  $f_i(t)$  为  $F(t)$  中可能的语言值, 即定义 1 中的  $f_A(u_i)$ . 因为对于不同时刻  $t$ ,  $F(t)$  是不同的, 故  $F(t)$  为  $t$  的函数, 且论域可能随  $t$  而变化, 因此使用  $Y(t)$  作为变化的论域.

**定义 3** 设  $R(t, t-1)$  为定义在从  $F(t-1)$  到  $F(t)$  的模糊关系, 若满足  $F(t) = F(t-1) \circ R(t, t-1)$ , 则称  $F(t)$  是由  $F(t-1)$  得到的, 且可以用模糊逻辑关系  $F(t-1) \rightarrow F(t)$  来表示. 其中: “ $\circ$ ”代表合成运算,  $F(t)$  和  $F(t-1)$  都是模糊集, 关系  $R$  称为定义在  $F(t)$  上的一阶模糊关系.

**定义 4** 令  $F(t-1) = A_i, F(t) = A_j$ , 则两个连续的观测值  $F(t)$  和  $F(t-1)$  可以用一模糊逻辑关系表示, 记为  $A_i \rightarrow A_j$ . 其中:  $A_i$  为模糊关系的左件,  $A_j$  为模糊关系的右件.

根据定义 4, 如果  $t = 1, 2, \cdots, N$ , 则有  $N-1$  个模糊关系  $F(i) \rightarrow F(i+1)(i = 1, 2, \cdots, N-1)$ . 用模糊集之间的关系来表示, 它们就是  $N-1$  个  $A_i^t \rightarrow A_j^{t+1}(t = 1, 2, \cdots, N-1)$ , 其中  $A_i^t$  和  $A_j^{t+1}$  分别是  $t$  和  $t+1$  时刻序列值所取的模糊概念.

有了上面的定义, 模糊时间序列预测模型的构建和应用可分为以下几个步骤:

- 1) 根据训练集和模糊隶属函数对论域进行模糊划分;
- 2) 根据训练数据先后的观测值建立模糊关系集合;
- 3) 由 2) 中所有的模糊关系求得模糊关系矩阵;
- 4) 将观测值模糊化, 然后根据 3) 中确定的模糊关系矩阵和给定的预测规则求得预测值.

在第 3) 步利用模糊关系构建模糊关系矩阵时, 不同的模型采用不一样的构建方法, 其中有代表性的是 Song<sup>[1]</sup>, Chen<sup>[2]</sup> 和 Lee<sup>[4]</sup>. 他们的预测规则也有区别, 下面先介绍他们关于模糊关系矩阵的构建方法, 然后对其预测规则和模型进行分析.

在 Song 的模型中, 先定义一个运算“ $\times$ ”来处理模糊关系  $A_i \rightarrow A_j$ , 其中  $A_i$  和  $A_j$  是两个  $n$  维的行向量, 值为相应模糊概念的隶属度, 且  $R^{ij} = A_i^T \times A_j$ . 矩阵

$R^{ij}$  的  $k$  行,  $l$  列的元  $r(k, l)$  被定义为  $r(k, l) = \min(A_i^k, A_j^l)(k, l, i, j = 1, 2, \cdots, n)$ . 其中:  $A_i^k$  和  $A_j^l$  分别是  $A_i$  和  $A_j$  的第  $k$  和  $l$  个元,  $A_i^T$  是  $A_i$  的转置. 然后用取最大值的方法将这些关系矩阵合成为一个模糊关系矩阵, 即  $R = \bigcup_{i,j} R^{ij}$ . 这种方法被称为 max-min 运算.

在 Chen 的模型中, 模糊关系矩阵  $R$  中的元素根据  $A_i \rightarrow A_j$  是否存在来决定. 即如果在训练集中出现了模糊关系  $A_i \rightarrow A_j$ , 则  $R(i, j) = 1$ ; 否则  $R(i, j) = 0$ .

为了更好地反映训练集中一些模糊关系出现的频率, 在 Lee 的模型中, 模糊关系矩阵  $R$  的元素用  $A_i \rightarrow A_j$  在训练集中出现的次数所取代.

在应用模糊关系矩阵进行预测时所运用的规则方面, Song 采用的有以下 3 条: 1) 如果观测集所对应各模糊集的隶属度只有一个最大值, 则以该最大值所对应的模糊集的中心值作为预测值; 2) 如果观测集所对应各模糊集的隶属度有多个最大值, 则以最大值所对应模糊集的中心值的平均值作为预测值; 3) 否则, 就用这些模糊隶属度作为权重求得的模糊集中心的加权值作为预测值. 对于 Chen 和 Lee 的预测规则, 可综合为如下表达式:

$$F^o(t+1) = \begin{cases} c_t, & \sum_{j=1}^n R(i_{c_t}, j) = 0; \\ \frac{R(i_{c_t}, :)}{\sum_{j=1}^n R(i_{c_t}, j)} \times C^T, & \text{otherwise.} \end{cases} \quad (2)$$

其中:  $F^o(t+1)$  为  $t$  时刻对  $t+1$  时的预测值,  $c_t$  为  $t$  时刻值所对应隶属度最大的模糊集中心值,  $R(i_{c_t}, :)$  为模糊关系矩阵的模糊集  $c_t$  对应的行,  $C^T$  为  $C$  的转置,  $C = (c_1, \cdots, c_i, c_n)$ ,  $c_i$  为模糊集  $u_i$  的中心值( $i = 1, 2, \cdots, n$ ), “ $\times$ ”表示矩阵的乘运算.

上述 3 种模型按时间顺序先后出现, 它们依次以前面的为基础, 并作了适当的改进. 在 Song 的模型中, 需要通过大量的计算来确定模糊逻辑关系; 当模糊逻辑关系复杂时, max-min 操作会占据大量的计算时间; Chen 的计算方法比 Song 的要简单; Lee 的模型考虑了频数, 从而在预测精度上有了较大的提高. 总之, 这 3 个模型得到的结果不仅越来越好, 而且模型的应用也越来越符合现实. 详细的模型和计算过程可参见文献 [1-2, 4].

然而, 这 3 个模型考虑的都是单变量的情形, 在现实生活中几乎没有一个现象是独立发生的, 与其他因素都有着或多或少的联系. 因此, 如何充分利用这些因素所包含的信息, 建立多因素的模糊时间序列模型是一个值得研究的方向. 下面, 本文将结合证据理论来研究这一问题.

### 3 证据理论

证据理论是由 Dempster<sup>[12]</sup>于 1967 年提出的, 后来 Shafer<sup>[13]</sup>对其进行了推广和完善, 故又称为 Dempster-Shafer 理论 (简称 D-S 理论). 它是一种不确定性推理方法, 允许把整个问题和证据分解为若干个子问题和子证据, 在对子问题、子证据作出相应处理后, 利用 Dempster 合成法则可以得到整个问题的解, 所以证据理论也是一种决策理论, 并已逐渐发展为一种重要的不确定性推理方法. 此外, 它不仅能处理随机性和模糊性所导致的不确定性, 而且可以依靠证据的积累不断地缩小假设集. 正是这些优点使得证据理论的应用越来越广.

下面将基于文献 [12-16] 给出本文所需要的相关定义, 然后提出构建适当信度函数值的取值方法和证据合成公式, 为建立基于证据理论的多因素模糊时间序列模型作准备.

**定义 5** 设某一判决问题, 称其所有可能结果的集合  $\theta$  为识别框架.  $\theta$  在理论上可以是有限集, 也可以是无限集, 但实际处理问题过程中一般为有限集.

**定义 6** 设  $\theta$  为识别框架, 如果集函数  $m: 2^\theta \rightarrow [0, 1]$  ( $2^\theta$  为  $\theta$  的幂集) 满足: 1)  $m(\emptyset) = 0$ ; 2)  $\sum_{A \subseteq \theta} m(A) = 1$ . 则称  $m$  为框架  $\theta$  上的基本可信度分配 (即 mass 函数); 称由  $\text{Bel}(A) = \sum_{B \subseteq A} m(B), \forall A \subseteq \theta$  所定义的函数  $\text{Bel}: 2^\theta \rightarrow [0, 1]$  为  $\theta$  上的信度函数. 如果  $m(A) > 0$ , 则称  $A$  为信度函数  $\text{Bel}$  的焦点; 同时称由  $\text{Pls}(A) = \sum_{B \cap A \neq \emptyset} m(B), \forall A \subseteq \theta$  所定义的函数  $\text{Pls}: 2^\theta \rightarrow [0, 1]$  为  $\theta$  上的似真函数.

对于命题  $A$  而言, 信度函数表示对其最低信任程度, 而似真函数则表示对其最高信任程度. 这是两个从不同角度表示对命题  $A$  的信任程度的方法, 本文只选择信度函数  $\text{Bel}$  作为考虑的对象. 证据理论是解决不确定问题的一种好方法, 它能将不同来源的信息进行合成, 得到较为统一的结论. 常用的证据合成方法一般采用下式:

$$m(A) = \frac{1}{1-K} \sum_{X,Y; X \cap Y = A} m_1(X)m_2(Y). \quad (3)$$

其中:  $A \neq \emptyset; K = \sum_{X,Y; X \cap Y = \emptyset} m_1(X)m_2(Y)$ ;  $m_1$  和  $m_2$  为两个信息源对命题  $X$  和  $Y$  的信度函数值;  $X, Y \in 2^\theta$ .

在模糊时间序列预测中, 一般只需预测出数值属于哪一个模糊概念, 因此本文只考虑预测值属于单个模糊概念的程度, 而不必考虑落在哪些模糊概念上. 即只选择单个的模糊概念作为焦点, 并将对

每个焦点的模糊隶属值标准化后作为其信度函数的值. 在进行证据合成时, Dempster-Shafer 的方法 (即式 (3)) 不满足幂等律. 例如: 设  $\theta = \{a, b\}$ . 其中: 一信度函数定义为  $m_1(a) = 0.2, m_1(b) = 0.8$ ; 另一信度函数定义为  $m_2(a) = 0.2, m_2(b) = 0.8$ . 则根据上述合成公式有  $m(a) = 0.06, m(b) = 0.94$ . 为克服这一不足, 本文重新构建证据合成公式为

$$m(A) = \frac{1}{1-K} \sum_{X,Y; X \cap Y = A} \sqrt{m_1(X)m_2(Y)}. \quad (4)$$

其中:  $A \neq \emptyset, K = \sum_{X,Y; X \cap Y = \emptyset} \sqrt{m_1(X)m_2(Y)}$ .

### 4 基于证据理论的模糊时间序列模型

为了将传统的模糊时间序列模型推广到多因素的情形, 首先补充一个定义.

**定义 7** 设一时间序列  $S(t)$  有  $K$  个观测值  $S_j(t)$  (即  $S(t) = (S_1(t), S_2(t), \dots, S_K(t))$ ),  $t$  为时间变量. 需要预测的是属性  $S_1(t)$ , 其他的  $S_j(t) (j \neq 1)$  只是相关的属性. 称  $S_1(t)$  为主属性,  $S_j(t) (j \neq 1)$  为从属性.

根据上述定义, 要预测的是主属性, 而每个属性都可以根据其实际意义用类似于式 (1) 进行模糊划分. 例如令第  $j$  个属性的次序划分为  $U_j = \{u_{j,1}, u_{j,2}, \dots, u_{j,n_j}\}$ , 其对应的模糊集为  $\{A_{j,1}, A_{j,2}, \dots, A_{j,n_j}\}$ ; 然后根据观测值  $S(t)$  求出各属性观测值在各自对应模糊集上的模糊隶属度 (本文采用三角模糊隶属函数), 确定其所在的模糊集. 例如表 1 所示,  $A_j^t$  为时刻  $t$  第  $j$  个属性在模糊化后对应的模糊集,  $A_j^t \in \{A_{j,1}, A_{j,2}, \dots, A_{j,n_j}\}$ .

表 1 观测序列模糊化

$t$	$S_1$	$S_2$	$\dots$	$S_j$	$\dots$	$S_K$
$S(1)$	$A_1^1$	$A_2^1$	$\dots$	$A_j^1$	$\dots$	$A_K^1$
$S(2)$	$A_1^2$	$A_2^2$	$\dots$	$A_j^2$	$\dots$	$A_K^2$
$\vdots$						
$S(t)$	$A_1^t$	$A_2^t$	$\dots$	$A_j^t$	$\dots$	$A_K^t$
$\vdots$						

模糊化以后, 对于第  $j$  个属性根据训练集得到一系列的模糊关系  $A_j^t \rightarrow A_j^{t+1} (t = 1, 2, \dots)$ . 注意, 这里  $A_j^t$  是对应于第  $j$  个属性的某个模糊集, 不同的属性其模糊集及其个数都可能不一样 (即上面的  $n_j$  随  $j$  的取值不同而变化), 这样对序列属性的要求降低了, 从而拓宽了模型的适用性. 下面利用这些模糊关系构建模糊关系矩阵. 根据前面第 2 节的分析, 本文采用文献 [4] 中的方法来构建模糊关系矩阵  $R^j$ . 与经典模型不同的是  $R^j$  不一定是方阵, 其行和列数分别为第  $j$  和第  $l$  个属性的模糊集的个数. 这样可得到  $K$  个模糊关系矩阵  $R^1, R^2, \dots, R^K$ .

下面讨论如何利用这些模糊关系矩阵进行预测. 设  $R^1, R^2, \dots, R^K$  为根据训练样本得到的模糊关系矩阵,  $S(t) = (S_1(t), S_2(t), \dots, S_K(t))$  为  $t$  时刻的观测值, 根据给定的模糊隶属函数求得这  $K$  个观测值所对应的模糊集分别为  $\{A_{t_1}, A_{t_2}, \dots, A_{t_K}\}$ , 其中  $A_{t_j}$  为观测值  $S_j(t)$  模糊化后对应的模糊集, 且  $A_{t_j} \in \{A_{j,1}, A_{j,2}, \dots, A_{j,n_j}\}$ . 由此, 将模糊关系矩阵  $R^j$  的模糊集  $A_{t_j}$  所对应的行归一化, 作为它对  $t+1$  点预测值中  $S_1(t+1)$  取值为  $\{A_{1,1}, A_{1,2}, \dots, A_{1,n_1}\}$  的信度函数值. 即

$$m_j(A_{1,i}) = \frac{R^j(t_j, i)}{\sum_i R^j(t_j, i)}, \quad i = 1, 2, \dots, n_1. \quad (5)$$

式(5)的含义是, 当序列  $S(t)$  的第  $j$  个值是模糊概念  $A_{t_j}$  时, 它预测  $S(t+1)$  主属性为模糊集  $A_{1,i}$  的信度值为  $m_j(A_{1,i})$ . 这样, 可得到由每个属性和相应观测值所提供的预测值取各个模糊集的信度值; 然后根据证据理论中证据合成方法得到一个综合的信度值向量, 不妨记为  $m = (m_t^1, m_t^2, \dots, m_t^{n_1})$ ; 最后根据如下预测公式进行模糊时间序列的预测:

$$F^o(t+1) = (m_t^1 \ m_t^2 \ \dots \ m_t^{n_1}) \times C^T. \quad (6)$$

其中:  $F^o(t+1)$  为序列  $S(t)$  在  $t+1$  时主属性的预测值;  $(m_t^1, m_t^2, \dots, m_t^{n_1})$  为  $t$  时刻的综合信度值向量;  $C^T$  为  $C$  的转置,  $C = (c_1, \dots, c_i, c_{n_1})$ ,  $c_i$  为主属性的模糊集  $u_i$  的中心值 ( $i = 1, 2, \dots, n_1$ ).

在第 2 节已经对模糊时间序列分析的 3 个经典模型进行了对比分析, 并根据原文献的预测规则和思想归纳总结出了预测公式(2), 即本文提出的式(2)是对 Chen 和 Lee 模型的精确概括和数学表达. 式(6)是基于证据理论的模糊时间序列预测的数学表达式, 是式(2)在多因素模糊时间序列预测方面的推广. 此外, 本文提出的模型在预测步骤方面主要有以下 3 点与传统的模型不同: 1) 建立模糊关系时每个属性与主属性之间的模糊关系都要考虑, 有多少个属性便有多少个模糊关系矩阵; 2) 用来预测的观测值要根据属性求得其相应模糊关系矩阵中对应的预测规则; 3) 不同属性之间的预测规则利用第 3 节提出的证据合成公式(4)进行融合, 从而求出式(6)中的  $(m_t^1, m_t^2, \dots, m_t^{n_1})$ . 其他步骤不变.

## 5 模型的应用

将本文模型与 Song, Chen 和 Lee 的单变量模型的预测结果进行比较, 同时为了突现该方法的有效性, 还列出了 Huarng 的基于 2-型模糊集时间序列模型的结果. 与大多数文献一样, 本文也采用均方根误差(见式(7), 其中  $F_f(t)$  和  $F_r(t)$  分别代表  $t$  时刻的预测值和真实值)来衡量预测精度, 即

$$\text{RMSE} = \sqrt{\left(\sum_{t=1}^n (F_f(t) - F_r(t))^2\right) / n}. \quad (7)$$

首先考虑 1997 年 1 月 2 日~1997 年 12 月 31 日这 243 个交易日的数据, 以开盘价为主属性, 最高价和最低价为从属性. 由于这 3 个价相差不大, 可以用相同的模糊集来表示, 且这一年的股价波动在 800~1600 之间, 故将论域  $U$  分为 8 个区间, 分别是  $u_1 = [800, 900]$ ,  $u_2 = [900, 1000]$ ,  $\dots$ ,  $u_8 = [1500, 1600]$ . 对应的模糊集记为  $A_1, A_2, \dots, A_8$ .

以数据中的前 190 个作为训练样本, 后 53 个数据作为测试样本, 采用本文提出的模型可依次得到 3 个模糊关系矩阵, 最后得到的预测值如图 1 所示. 为了比较模型的优势, 图 1 中还列出了 Lee 的模型和 Huarng 的 2-型模糊集模型的预测结果. 由于 Lee 的模型明显优于 Song 和 Chen 的, 为保持图的简洁性, 不影响图示的对比效果, 在此略去了后两个模型的预测结果.

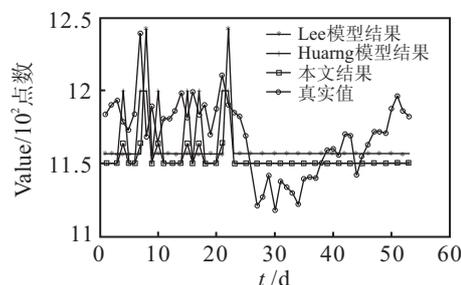


图 1 1997 年沪指预测结果比较对照

从图 1 中可以看出, 本文提出的模型和 Huarng 模型的结果都明显优于 Lee 模型; 本文的结果与 Huarng 的非常相似, 但稍微有些预测的点不同. 为了进一步比较结果, 下面列出了从 1997 年 1 月 2 日~2006 年 12 月 29 日沪市这 10 年的股指交易预测误差. 由于这里只是比较不同模型的预测精度, 在做实验时将训练样本和测试样本取每年的交易数据, 数据的结构与前面一样, 都是 3 个属性, 分别是开盘值、最高值和最低值. 模糊集及其个数是根据该年股指的波动范围来决定, 每波动 100 个点作为一个区间, 预测的方法与上例中的相同, 预测结果列于表 2 中.

从表 2 可以看出, Kuarnng 和本文方法明显优于 Song, Chen 和 Lee 模型, 且本文方法的预测平均误差最小, 这充分说明了本文提出的多因素预测模型在预测精度上要高于单变量预测模型. 此外, 表 2 的倒数第 2 列的预测误差是本文模型应用经典的证据合成方法得到的(即利用式(3)), 最后一例是利用式(4)得到的. 最后一列的数值明显小于倒数第 2 列所对应的值, 整体上也优于 Kuarnng 模型, 这表明本文对证据合成公式的改进是很有必要的.

表 2 1997 年~2006 年沪市股指预测均方根误差值对照

年份	Song <sup>[1]</sup>	Chen <sup>[2]</sup>	Lee <sup>[4]</sup>	Kuarng <sup>[10]</sup>	本文 1	本文 2
1997	50.12	38.07	36.57	32.03	33.81	30.84
1998	44.19	34.51	32.02	24.80	31.28	25.60
1999	38.38	37.47	35.81	25.23	32.55	27.74
2000	46.16	42.15	40.33	30.26	34.82	30.07
2001	42.32	37.23	36.76	27.49	33.39	27.49
2002	38.80	47.19	35.59	39.90	33.18	27.65
2003	32.71	36.57	30.98	29.90	31.08	25.02
2004	38.66	35.98	33.66	26.61	30.36	24.82
2005	31.02	36.51	33.01	31.71	30.75	25.52
2006	41.31	40.57	35.42	35.68	32.53	29.08
平均	40.37	38.63	35.01	30.36	32.37	27.38

## 6 结 论

本文在分析传统模糊时间序列预测模型的基础上,着重介绍了 Song, Chen 和 Lee 三个经典的模糊时间序列模型在构建模糊关系矩阵和预测规则上的不同,指出了传统的模型不能处理多因素的情形;然后分析并提出用证据理论中合成证据的方法来处理多种信息(即多因素)预测问题;为了使证据理论更好地适用于模型时间序列模型,本文还对证据合成的方法进行了修改,进而提出了基于证据理论的模糊时间序列预测模型;最后用 1997 年~2006 年上海股指数据对所提出的模型进行了实践检验,所得结果证明了本文方法不仅比传统的单变量模糊时间序列预测模型的预测误差小,而且还比 Kuarng 的基于 2-型模糊集的预测模型要好。

需要指出的是,本文只是从利用证据理论的角度提出了构建解决多因素的模糊时间序列预测模型的方法;从预测精度的角度,关于改进证据理论的合成规则和证据的利用方面,已经有了大量的研究<sup>[14-16]</sup>;此外,众多学者<sup>[4-6,10-11]</sup>证明了模糊区间的划分方法对预测精度有很大的影响.因此,在提高模型的预测准确率方面还有很多工作可以做。

## 参考文献(References)

- [1] Song Q, Chissom B S. Fuzzy time series and its models[J]. Fuzzy Sets and Systems, 1993, 54(3): 269-277.
- [2] Chen S M. Forecasting enrollments based on fuzzy time series[J]. Fuzzy Sets and Systems, 1996, 81(3): 311-319.
- [3] Lee M H, Efendi R, Ismail Z. Modified weighted for enrollment forecasting based on fuzzy time series[J]. Matematika, 2009, 25(1): 67-78.
- [4] Ufuk Yolcu, Erol Egrioglu, Vedide R Uslu, et al. A new approach for determining the length of intervals for fuzzy time series[J]. Applied Soft Computing, 2009, 9(2): 647-651.
- [5] Huarng K H. Effective lengths of intervals to improve forecasting in fuzzy time series[J]. Fuzzy Sets and Systems, 2001, 123(3): 387-394.
- [6] Huarng K H. Ratio-based lengths of intervals to improve fuzzy time series forecasting[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B, 2006, 36(2): 328-340.
- [7] Cheng C H, Chen T L, Chiang C H. Trend-weighted fuzzy time-series model for TAIEX forecasting[J]. Lecture Notes in Computer Science, 2006, 4234(3): 469-477.
- [8] Yu H K. Weighted fuzzy time series model for TAIEX forecasting[J]. Physica A, 2005, 349(3/4): 609-624.
- [9] Wang N Y, Chen S M. Temperature prediction and TAIEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series[J]. Expert Systems with Applications, 2009, 36(2): 2143-2154.
- [10] Huarng K H, Yu H K. A type 2 fuzzy time series model for stock index forecasting[J]. Physica A: Statistical Mechanics and Its Applications, 2005, 353(1-4): 445-462.
- [11] Leu Y H, Lee C P, Jou Y Z. A distance-based fuzzy time series model for exchange rates forecasting[J]. Expert Systems with Applications, 2009, 36(4): 8107-8114.
- [12] Shafer G. A mathematical theory of evidence[M]. Princeton: Princeton University Press, 1976: 5-31.
- [13] Dempster A P. New methods for reasoning towards posterior distributions based on sample data[J]. Annals of Mathematical Statistics, 1966, 37(2): 355-374.
- [14] Philippe Smets. Analyzing the combination of conflicting belief functions[J]. Information Fusion, 2007, 8(4): 387-412.
- [15] Liu W. Analyzing the degree of conflict among belief functions[J]. Artificial Intelligence, 2006, 170(11): 909-924.
- [16] 龚本刚. 基于证据理论的不完全信息多属性决策方法研究[D]. 合肥: 中国科学技术大学管理学院, 2007. (Gong B G. Methods based on Dempster-Shafer theory for multi-attribute decision making with incomplete information[D]. Hefei: School of Management, University of Science and Technology of China, 2007.)