文章编号: 1001-0920 (2013) 10-1525-06

直觉模糊时间序列建模及应用

郑寇全[1,2], 雷英杰[1], 王 睿[1], 王艺菲[1]

(1. 空军工程大学 防空反导学院，西安 710051；2. 中国人民解放军 68331 部队，陕西 华阴 714200)

摘 要: 针对模糊时间序列对于预测不确定性的控制、有效的分区间隔和不同分区间隔达到一致的预测准确性方面研究的不足, 构建了直觉模糊时间序列预测模型. 新模型应用直觉模糊 C 均值聚类算法优化序列区间划分, 充分考虑数据点固有的模糊不确定性, 较好地反映了系统的特征分布, 提高了复杂环境中时间序列的预测性能且允许处理多因子预测问题. 最后通过实例验证了所提出方法的有效性和优越性.

关键词: 直觉模糊集；时间序列；直觉模糊逻辑关系；预测

中图分类号: TP18　　**文献标志码:** A

Modeling and application of IFTS

ZHENG Kou-quan[1,2], *LEI Ying-jie*[1], *WANG Rui*[1], *WANG Yi-fei*[1]

(1. Air Defense and Antimissile Institute，Air Force Engineering University，Xi’an 710051，China；2. Unit of 68331，PLA，Huayin 714200，China．Correspondent：ZHENG Kou-quan，E-mail：zhengkouquan0421@163.com)

Abstract: To the limitation of the research of controlling uncertainty in forecasting, effectively partitioning intervals and consistently achieving forecasting accuracy with different interval lengths, the intuitionistic fuzzy time series(IFTS) forecasting model is advanced. The novel forecasting model applies the intuitionistic fuzzy C-means(IFCM) clustering algorithm to optimize interval partitioning, which considers the data point fuzzy uncertainty fully, reflects the character distribution of uncertain system, enhances forecasting functionality in the complex environment better and allows the processing of multiple factors forecasting problems. Finally, a classical instance shows the effectiveness and superiority of the proposed method.

Key words: intuitionistic fuzzy set；time series；intuitionistic fuzzy logic relationship；forecasting

0 引 言

时间序列是指随时间变化的具有随机性且前后相互关联的动态数据序列, 时间序列分析是随机数学一个重要的应用分支, 它利用概率论和数理统计的方法从序列历史数据中提取有关信息, 研究序列数据相互关联的规律, 揭示序列本身的结构特征, 用序列的过去值预测和控制将来值[1]. 传统的时间序列预测依赖大量的历史数据, 而系统的随机性、缺乏相关属性 (参数) 或信息不精确等因素和现实世界中不确定性的广泛存在, 使得历史数据往往是不完整、不准确、非实时性和模糊的, 这便限制了经典时间序列预测模型的应用范围[2].

Song 等[3-4]将序列变量设定为模糊数, 借助模糊逻辑理论能够在不确定环境下描述和处理模糊数据的优势, 分析了时间序列与模糊数学相关的基本特性, 首次提出了模糊时间序列的概念, 标志着时间序列的分析与研究进入了一个崭新的阶段. 模糊时间序列预测的基本思想是利用时变与非时变的模糊关系式描述序列的动态变化趋势, 在历史数据不完整或不合理的情况下, 通过模糊推理获取理想的预测结果. 由于能较好地反映不确定数据的模糊关联本质, 模糊时间序列预测理论一经问世, 便引起学术界和工程技术领域的广泛关注, 并迅速发展成为一个重要而有趣的热门研究课题, 取得了大量有价值的理论研究成果: Chen[5]构建了时间不变性模糊时间序列预测模型; Own 等[6]改进了 Chen 的模型提出了启发式高阶模糊时间序列预测模型; Song[7]通过定义模糊加减算子提出了一种线性模糊时间序列分析方法, 并给出了基于自相关函数的模糊时间序列预测模型; Cheng 等[8]提出了基于模糊熵与梯形模糊数的模糊时

收稿日期: 2012-05-28；**修回日期:** 2012-08-03.

基金项目: 国家自然科学基金项目(60773209, 61272011)；国家重点实验室基金项目(2012ADL-DW0301).

作者简介: 郑寇全(1983−), 男, 博士生, 从事智能信息处理的研究；雷英杰(1956−), 男, 教授, 博士生导师, 从事智能信息处理等研究.

间序列预测方法; Huarng 等[9]讨论了模糊时间序列与神经网络的融合预测问题; Liu 等[10]基于多属性模糊时间序列建立了自适应期望预测模型; Teoh[11]给出了基于概率和粗糙集规则的模糊时间序列预测模型; Young 等[12]提出了基于层次聚类的模糊时间序列预测方法; Yu 等[2]构建了一种基于 FCM 的模糊时间序列确定性预测模型, 且允许进行两因子预测. 然而, 随着模糊时间序列预测理论日趋成熟, 其局限性也逐渐凸现: 1) 普通模糊集由于隶属度单一, 不能有效描述和处理复杂环境下随机、并发、异步和冲突的不确定性信息, 也不能反映序列数据的实时模糊变化趋势; 2) 等长度划分论域区间无法反映连续数值分布不均匀的特性; 3) 大部分预测模型只能处理单因子问题, 即使少数方法可以进行多因子判别, 但需确定合适的阶数和阶数的冗余. 因此, 模糊时间序列预测理论的拓展研究已成为亟待解决的重要问题.

直觉模糊集 (IFS) 作为 Zadeh 模糊集理论的重要扩充和发展, 其数学描述更加符合客观世界的模糊本质, 为不确定信息的研究和处理提供了新的思路[13]. Oscar 等[14]首次将直觉模糊推理融入时间序列分析, 初步建立了直觉模糊时间序列 (IFTS) 预测理论, 极大地扩展了时间序列对不精确、不完备等粗糙信息的处理功效, 为模糊时间序列预测研究指出了新的突破方向. 然而, 基于 IFTS 预测的理论研究才刚刚起步, 国内外相关学术成果相对较少, 且仅有的研究缺乏理论深度, 没有标准化的定义和模型, 预测精度还有待进一步提高. 鉴于此, 本文首先规范了直觉模糊时间序列的定义, 在基于直觉模糊 C 均值聚类 (IFCM) 算法优化论域区间划分标准的基础上, 提出了一种新的直觉模糊时间序列预测方法. 通过实例验证, 所提出方法有效克服了模糊时间序列预测理论的缺陷, 取得了较好的预测结果.

1 基础理论

定义 1 (直觉模糊时间序列) 令 $\{Y(t), t = 1, 2, \cdots, n\}$ 为论域 U 上的一个时间序列, 给定一个次序分割集合 $\left\{P_i,\ i = 1, 2, \cdots, r;\ \bigcup_{i=1}^{r} P_i = U\right\}$, 其语言变量为 $\{L_i,\ i = 1, 2, \cdots, r\}$. 若在 $\{L_i\}$ 上相对于 $Y(t)$ 的直觉模糊集 $F(t)$ 有隶属度与非隶属度函数对 $\langle\mu_i(Y(t)), \gamma_i(Y(t))\rangle$, 其中 $i \in [1, n]$, $\mu_i(Y(t)), \gamma_i(Y(t)) \in [0, 1]$ 且 $\mu_i(Y(t)) + \gamma_i(Y(t)) \leqslant 1$, 则称 $F(t)$ 为定义在 $Y(t)$ 上的直觉模糊时间序列, 并记为

$$F(t) = \langle\mu_1(Y(t)), \gamma_1(Y(t))\rangle/L_1 + \langle\mu_2(Y(t)), \gamma_2(Y(t))\rangle/L_2 + \cdots + \langle\mu_n(Y(t)), \gamma_n(Y(t))\rangle/L_n. \tag{1}$$

其中: $f_i(t) = \langle\mu_i(Y(t)), \gamma_i(Y(t))\rangle/L_i$ 表示直觉模糊时间序列 $Y(t)$ 相对于语言变量 L_i 及其隶属度与非隶属度的对应关系, $+$ 为连接符号. 若将 $f_i(t)$ 理解为语言变量值, 则 $F(t)$ 作为直觉模糊集的集合即可表示为时间 t 的函数. 由此可见, 与模糊时间序列不同的是, 直觉模糊时间序列的数据集是直觉模糊集.

为了更合理地描述直觉模糊时间序列, 本文利用直觉模糊关系方程作为模型, 对直觉模糊时间序列预测理论进行分析研究. 设论域 U 上的直觉模糊集 $F = \{\langle\mu_i, \gamma_i\rangle\}$, $G = \{\langle\sigma_j, \nu_j\rangle\}$, $i, j = 1, 2, \cdots, r$, 其中 $\langle\mu_i, \gamma_i\rangle$ 和 $\langle\sigma_j, \nu_j\rangle$ 分别为分布于 U 上的隶属度与非隶属度函数对, 则介于 F 与 G 之间的直觉模糊关系 R 可以表示为

$$R = F^{\mathrm{T}} \circ G = [R_{ij}]_{r\times r}. \tag{2}$$

其中: T 为转置符; $\circ$ 为直觉模糊合成运算; $R_{ij} = \langle\mu_{R_{ij}}, \gamma_{R_{ij}}\rangle$ 为介于 F 与 G 之间的直觉模糊关系矩阵元素, 且

$$\mu_{R_{ij}} = \bigvee_{k=1}^{r}\Big(\mu_{ik} \bigwedge \sigma_{kj}\Big),\ \gamma_{R_{ij}} = \bigwedge_{k=1}^{r}\Big(\gamma_{ik} \bigvee \nu_{kj}\Big).$$

定义 2 (直觉模糊关系式) 设 $F(t)$ 为给定论域上的直觉模糊时间序列, 假设 $F(t)$ 仅由 $F(t-1)$ 确定, 或仅由 $F(t-2)$ 确定, 或 $\cdots$, 或仅由 $F(t-m)(m > 0)$ 确定, 则一阶直觉模糊时间序列关系式可以表示为

$$F(t) = (F(t-1) \bigcup F(t-2) \bigcup \cdots \bigcup F(t-m)) \circ R(t, t-m). \tag{3}$$

其中: $\bigcup$ 为直觉模糊并运算符, $R(t, t-m)$ 为描述 $F(t-m)$ 与 $F(t)$ 之间的直觉模糊关系矩阵. 若对于任意时间点 t, $R(t, t-m)$ 均与 t 无关, 则称 $F(t)$ 为非时变性直觉模糊时间序列, 否则, 称为时变性直觉模糊时间序列. 由于时变性更能迅速反应序列数据的模糊变动趋势, 本文重点研究时变性直觉模糊时间序列.

若 $F(t)$ 由 $F(t-1), F(t-2), \cdots, F(t-m)(m > 0)$ 同时确定, 则 m 阶直觉模糊时间序列关系式可以表示为

$$F(t) = (F(t-1) \times F(t-2) \times \cdots \times F(t-m)) \circ R_\omega(t, t-m). \tag{4}$$

其中: $\times$ 为笛卡尔积, $R_\omega(t, t-m)$ 为描述 $F(t-1), F(t-2), \cdots, F(t-m)$ 与 $F(t)$ 之间的直觉模糊关系矩阵, ω 为窗口参数.

2 直觉模糊时间序列预测模型

时间序列预测的关键是挖掘历史数据的模糊变化特性, 掌握序列数据的分布规律, 而预测模型构建的根本是在提升预测精度的同时降低系统的复杂性. 本文提出的直觉模糊时间序列预测模型主要包括 3

个基本步骤: 1) 利用 IFCM 聚类算法对每个因子优化区间划分并进行时间序列直觉模糊化处理; 2) 构建直觉模糊关系式; 3) 预测并去模糊化输出. 以下对 3 个步骤分别进行描述.

2.1　基于IFCM的区间划分和时间序列直觉模糊化

IFCM 聚类算法利用 IFS 间的相似性准则, 将聚类归结为一个带约束的线性规划问题, 适用于时间序列等对实时性要求较高的数据分类领域, 能够准确描述和反映序列数据固有的模糊不确定性, 且易于在计算机上实现, 是目前应用最广泛的聚类算法之一[15]. 给定聚类数 c, IFCM 聚类算法根据目标函数 J_m 最小化原则将数据集 $X=\{x_1,x_2,\cdots,x_n\}$ 划分为 c 类, 有

$$J_m(U_\mu,U_\gamma,P)=\sum_{j=1}^{n}\sum_{i=1}^{c}\Big(\frac{(\mu_{ij})^m+(1-\gamma_{ij})^m}{2}\Big)D_\omega(x_j,p_i)^2. \tag{5}$$

其中: $m\in[1,\infty)$ 为平滑参数 (不作特殊要求取 $m=2$); μ_{ij},γ_{ij} 分别为 $x_j=\langle x\mu_{jk},x\gamma_{jk}\rangle$ 属于聚类 i 的隶属度和非隶属度; $p_i=\langle p\mu_i,p\gamma_i\rangle$ 为聚类中心; $D_\omega(x_j,p_i)$ 为样本 x_j 与聚类中心 p_i 之间的距离. 对于第 n 次迭代计算可得

$$p\mu_i=\frac{\sum_{j=1}^{n}\{[(\mu_{ij})^m+(1-\gamma_{ij})^m]/2\}x\mu_j}{\sum_{j=1}^{n}\{[(\mu_{ij})^m+(1-\gamma_{ij})^m]/2\}},$$

$$p\gamma_i=\frac{\sum_{j=1}^{n}\{[(\mu_{ij})^m+(1-\gamma_{ij})^m]/2\}x\gamma_j}{\sum_{j=1}^{n}\{[(\mu_{ij})^m+(1-\gamma_{ij})^m]/2\}}. \tag{6}$$

当 $\forall k, D_\omega(x_j,p_k^{(n)})\neq 0$ 时, 有

$$\mu_{ij}^{(n+1)}=\Big[\sum_{k=1}^{c}\Big[\frac{D_\omega(x_j,p_i^{(n)})}{D_\omega(x_j,p_k^{(n)})}\Big]^{2/(m-1)}\Big]^{-1},$$

$$\gamma_{ij}^{(n+1)}=1-\frac{1}{\lambda}\Big[\sum_{k=1}^{c}\Big[\frac{D_\omega(x_j,p_i^{(n)})}{D_\omega(x_j,p_k^{(n)})}\Big]^{2/(m-1)}\Big]^{-1}. \tag{7}$$

当 $\exists k, D_\omega(x_j,p_k^{(n)})=0$ 时, 有

$$\begin{cases}\mu_{ij}=1,\gamma_{ij}=0,\ i=k;\\ \mu_{ij}=0,\gamma_{ij}=1,\ i\neq k.\end{cases} \tag{8}$$

迭代终止条件为 $\|J_m^{(n+1)}-J_m^{(n)}\|\leqslant\varepsilon$. 其中: ε 为迭代停止阈值, $\|\cdot\|$ 为直觉模糊范数运算符. 找到区间 (聚类) 中心后, 将相邻聚类中心的中点作为子区间边界点, 序列数据的极值点作为区间的起始和终结点, 则论域被动态划分为 c 个优化子区间, 其动态语言变量直觉模糊集 $A_i(i=1,2,\cdots,c)$ 定义为

$$A_i=\sum_{j=1}^{c}\langle\mu_{ij},\gamma_{ij}\rangle/p_j, \tag{9}$$

其中 μ_{ij} 和 γ_{ij} 分别为 p_j 属于 A_i 的隶属度和非隶属度, 可按下式计算:

$$\langle\mu_{ij},\gamma_{ij}\rangle=\begin{cases}\langle 1,0\rangle,\ j=i;\\ \Big\langle 1-\frac{p_i-p_j}{\lambda(p_i-q_{i-1})},\frac{p_i-p_j}{p_i-q_{i-1}}\Big\rangle,\ j=i-1;\\ \Big\langle 1-\frac{p_j-p_i}{\lambda(q_j-p_i)},\frac{p_j-p_i}{q_j-p_i}\Big\rangle,\ j=i+1;\\ \langle 0,1\rangle,\ \text{otherwise}.\end{cases} \tag{10}$$

$q_i(i=0,1,\cdots,c)$ 为子区间边界点, $\lambda\leqslant 1$ 为犹豫度调节因子.

为了直觉模糊化原始序列, 数据集 $X=\{x_1,x_2,\cdots,x_n\}$ 中的观测样本 x_i 属于 p_j 的隶属度和非隶属度可以定义为

$$\mu_{ij}=\Big[\sum_{k=1}^{c}\Big[\frac{D_\omega(x_j,p_i^{(n)})}{D_\omega(x_j,p_k^{(n)})}\Big]^{2/(m-1)}\Big]^{-1},$$

$$\gamma_{ij}=1-\frac{1}{\lambda}\Big[\sum_{k=1}^{c}\Big[\frac{D_\omega(x_j,p_i^{(n)})}{D_\omega(x_j,p_k^{(n)})}\Big]^{2/(m-1)}\Big]^{-1}. \tag{11}$$

由此可见, 若序列数据 x_i 直觉模糊化为 A_j, 则 μ_{ij} 取最大值, γ_{ij} 取最小值. 若是多因子预测, 则其他因子均可用同样的方法处理.

2.2　构建直觉模糊关系式

构建直觉模糊关系式是直觉模糊时间序列预测的关键, 基本步骤是首先确定合适的窗口参数 ω, 定义主因子直觉模糊时间序列 $F(t)$ 的标准向量 $C(t)$、关于 $F(t)$ 的操作矩阵 $O^\omega(t)$ 和关于辅助因子直觉模糊时间序列 $G(t)$ 的辅助因子向量 $S(t)$; 然后计算标准向量 $C(t)$、操作矩阵 $O^\omega(t)$ 和辅助因子向量 $S(t)$ 之间的直觉模糊关系矩阵 $R(t)$; 最后根据直觉模糊关系矩阵预测计算序列在时间 t 的输出值.

直觉模糊时间序列 $F(t)$ 的标准向量 $C(t)$ 和操作矩阵 $O^\omega(t)$ 分别为

$$C(t)=f(t-1)=[C_1\ \ C_2\ \ \cdots\ \ C_m], \tag{12}$$

$$O^\omega(t)=\begin{bmatrix}f(t-2)\\ \vdots\\ f(t-\omega)\end{bmatrix}=\begin{bmatrix}O_{11}&\cdots&O_{1m}\\ \vdots&\ddots&\vdots\\ O_{(\omega-1)1}&\cdots&O_{(\omega-1)m}\end{bmatrix}. \tag{13}$$

其中: $f(t-1)$ 为主因子直觉模糊时间序列 $F(t)$ 在 $t-1$ 和 $t-2$ 之间的模糊化变动; m 为论域中元素的个数; C_j 和 O_{ij} 为精确值且 $C_j,O_{ij}\in[0,1],i\in[1,\omega-1],j\in[1,m]$. 由此可见, 标准向量 $C(t)$ 和操作矩阵 $O^\omega(t)$ 都是由主因子时间序列推演出来的.

辅助因子向量 $S(t)$ 的表达式为

$$S(t)=g(t-1)=[S_1\ \ S_2\ \ \cdots\ \ S_m]. \tag{14}$$

其中: $S(t)$ 为在 $t-1$ 时刻的辅助因子向量; $g(t-1)$ 为辅助因子直觉模糊时间序列 $G(t)$ 在 $t-1$ 时刻的模糊化数据; $S_i \in [0,1], i \in [1,m]$.

单因子时变性直觉模糊时间序列预测模型的直觉模糊关系是标准矩阵和操作矩阵之间的关系, 而在两因子时变性直觉模糊时间序列预测时, 必须考虑标准矩阵、操作矩阵和辅助因子向量间的直觉模糊关系. 因此, 两因子预测模型的直觉模糊关系矩阵可以表示为

$$R(t)=O^{\omega}(t)\otimes S(t)\otimes C(t)=[R_{ij}]_{(\omega-1)\times m}. \quad (15)$$

其中: 矩阵元素 $R_{ij}=O_{ij}\bigcap S_j\bigcap C_j$, $\bigcap$ 为直觉模糊交运算符, $i\in[1,\omega-1]$, $j\in[1,m]$. 因此, 依此类推便可将两因子直觉模糊关系矩阵推广到多因子时变性直觉模糊时间序列预测中.

2.3 预测和去模糊化预测输出

根据直觉模糊关系矩阵 $R(t)$ 可以得到直觉模糊时间序列 $F(t)$ 的模糊化预测变动为

$$U(t)=\Big[\bigvee_{i=1}^{\omega-1}(R_{i1})\ \ \bigvee_{i=1}^{\omega-1}(R_{i2})\ \ \cdots\ \ \bigvee_{i=1}^{\omega-1}(R_{im})\Big]. \quad (16)$$

最后的去模糊化输出是获取预测结果 A_j 的精确值, 而 A_j 由属于 p_j 的最大隶属度和最小非隶属度确定, 因此, 在充分研究数据分布模糊趋势的前提下, 本文采用直觉模糊区间重心法进行预测结果的去模糊化计算, 表达式为

$$\mathrm{def}(A_j)=\sum_{j=1}^{c}\tau_j m_j\Big/\sum_{j=1}^{c}\tau_j. \quad (17)$$

其中: $\tau_j=(\mu_j+(1-\gamma_j))/2$ 为模糊趋势逼近因子, m_j 为划分区间的中心位置.

3 算法应用

3.1 实例设计

为了验证预测模型的有效性, 本文在实际两因子数据集上进行实验. 测试数据采用某市 2000 年 6 月的日平均气温和云密度, 利用直觉模糊时间序列预测模型对其进行研究, 结果如表 1 所示, 输入的时间序列可以表示为 $Y(t)=(26.1,36),(27.8,23),\cdots,(30.2,19)$.

Step 1: 运用 IFCM 算法进行论域区间划分.

选取与文献 [16] 相同的区间划分值 ($c_{\text{main favtor}}=9, c_{\text{second factor}}=7$), 直觉模糊 C 均值聚类算法单独应用于主因子和辅助因子的论域区间划分, 其聚类中心的结果集分别为

$$M_{\text{main factor}}=\{27.4,27.8,28.3,28.6,29.1,29.4,29.5,30.1,30.8\},$$

$$M_{\text{second factor}}=\{13.8,22.7,29.2,31.0,44.9,55.6,63.1\}.$$

表 1 某市历史的和模糊化的日平均气温、云密度、预测结果和误差

日期	气温	模糊化气温	云密度	模糊化云密度	预测状态	解模糊	误差
1	26.1	A_1	36	B_4	–	–	–
2	27.8	A_2	23	B_2	A_2	27.9	−0.1
3	29.0	A_5	23	B_2	A_5	29.1	−0.1
4	30.5	A_9	10	B_1	A_9	30.3	0.2
5	30.0	A_8	13	B_1	A_8	30.1	−0.1
6	29.5	A_7	30	B_3	A_7	29.6	−0.1
7	29.7	A_7	45	B_5	A_7	29.6	0.1
8	29.4	A_6	35	B_4	A_6	29.4	0
9	28.8	A_4	26	B_3	A_4	28.7	0.1
10	29.4	A_6	21	B_2	A_6	29.4	0
11	29.3	A_6	43	B_5	A_6	29.4	−0.1
12	28.5	A_4	40	B_5	A_4	28.7	−0.2
13	28.7	A_4	30	B_3	A_4	28.7	0
14	27.5	A_1	29	B_3	A_1	27.4	0.1
15	29.5	A_7	30	B_3	A_7	29.6	−0.1
16	28.8	A_4	46	B_5	A_4	28.7	0.1
17	29.0	A_5	55	B_6	A_5	29.1	−0.1
18	30.3	A_8	19	B_2	A_8	30.1	0.2
19	30.2	A_8	15	B_1	A_8	30.1	0.1
20	30.9	A_9	56	B_6	A_9	30.3	0.6
21	30.8	A_9	60	B_7	A_9	30.3	0.5
22	28.7	A_4	96	B_7	A_4	28.7	0
23	27.8	A_2	63	B_7	A_2	27.9	−0.1
24	27.4	A_1	28	B_3	A_1	27.4	0
25	27.7	A_2	14	B_1	A_2	27.9	−0.2
26	27.1	A_1	25	B_2	A_1	27.3	−0.2
27	28.4	A_3	29	B_3	A_3	28.3	0.1
28	27.8	A_2	55	B_6	A_2	27.9	−0.1
29	29.1	A_5	29	B_3	A_5	29.1	0
30	30.2	A_8	19	B_2	A_8	30.1	0.1

两个因子的直觉模糊集 $A_i(i\in[1,9])$ 和 $B_j(j\in[1,7])$ 相应地由式 (9) 和 (10) 确定. 由于样本数据的分布是不均匀的, 各划分子区间的长度各异. 可见, 利用 IFCM 聚类算法优化论域区间划分比常用的等区间划分更能反映不确定数据的分布特性.

Step 2: 构建语言变量直觉模糊集并直觉模糊化历史数据.

设定主因子语言变量直觉模糊集为

$$A_1=\{特低\},\ A_2=\{很低\},\ A_3=\{低\},$$

$$A_4=\{较低\},\ A_5=\{适中\},\ A_6=\{较高\},$$

$$A_7=\{高\},\ A_8=\{很高\},\ A_9=\{特高\}.$$

其中: $A_i=\displaystyle\sum_{j=1}^{c}\langle\mu_{ij},\gamma_{ij}\rangle/p_j$, p_j 为聚类中心, μ_{ij} 和 γ_{ij} 分别为 p_j 属于语言值 A_i 的隶属度和非隶属度. 辅助因子语言变量直觉模糊集定义为

$$B_1=\{很小\},\ B_2=\{小\},\ B_3=\{较小\},\ B_4=\{持平\},$$

$$B_5=\{较大\},\ B_6=\{大\},\ B_7=\{很大\}.$$

根据式 (9) ∼ (11), 取 $\lambda=0.95$, 计算主因子和辅助因子直觉模糊集隶属度和非隶属度函数值分别如表 2

表 2　主因子直觉模糊集隶属度与非隶属度函数取值

主因子	隶属函数取值
A_1	$\langle 1,0\rangle\langle 0.35,0.62\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle$
A_2	$\langle 0.75,0.24\rangle\langle 1,0\rangle\langle 0.19,0.77\rangle\ \langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle$
A_3	$\langle 0,1\rangle\ \langle 0.25,0.71\rangle\ \langle 1,0\rangle\ \langle 0.42,0.55\rangle\ \langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle$
A_4	$\langle 0,1\rangle\langle 0,1\rangle\langle 0.42,0.55\rangle\langle 1,0\rangle\ \langle 0.19,0.77\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle$
A_5	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0.19,0,77\rangle\ \langle 1,0\rangle\ \langle 0.10,0.86\rangle\langle 0,1\rangle\ \langle 0,1\rangle\ \langle 0,1\rangle$
A_6	$\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0.42,0.55\rangle\langle 1,0\rangle\langle 0.74,0.25\rangle\langle 0,1\rangle\langle 0,1\rangle$
A_7	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0.58,0.40\rangle\langle 1,0\rangle\langle 0.34,0.63\rangle\ \langle 0,1\rangle$
A_8	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle\langle 0.03,0.92\rangle\ \langle 1,0\rangle\ \langle 0.07,0.88\rangle$
A_9	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0.26,0.70\rangle\langle 1,0\rangle$

和表 3 所示. 因此, 时间序列 $Y(t)$ 的直觉模糊结果可以表示为 $F=(A_1,B_4)(A_2,B_2)(A_5,B_2)\cdots(A_8,B_2)$.

表 3　辅助因子直觉模糊集隶属度与非隶属度函数取值

辅助因子	隶属函数取值
B_1	$\langle 1,0\rangle\langle 0.23,0.73\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle$
B_2	$\langle 0.26,0.70\rangle\langle 1,0\rangle\langle 0.07,0.88\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle$
B_3	$\langle 0,1\rangle\langle 0.37,0.60\rangle\langle 1,0\rangle\ \langle 0.78,0.21\rangle\langle 0,1\rangle\ \langle 0,1\rangle\langle 0,1\rangle$
B_4	$\langle 0,1\rangle\langle 0,1\rangle\langle 0.62,0.36\rangle\langle 1,0\rangle\ \langle 0.24,0.72\rangle\langle 0,1\rangle\langle 0,1\rangle$
B_5	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0.01,0,94\rangle\ \langle 1,0\rangle\langle 0.22,0.74\rangle\ \langle 0,1\rangle$
B_6	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\ \langle 0.36,0.61\rangle\langle 1,0\rangle\langle 0.79,0.20\rangle$
B_7	$\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0,1\rangle\langle 0.37,0.60\rangle\langle 1,0\rangle$

Step 3: 设定阶次, 根据主因子和辅助因子的直觉模糊结果以及原始序列构建两因子直觉模糊推理关系式.

本文利用直觉模糊合成运算, 阶次的设定必须在 2 阶以上才能运算, 如以预测 15 日气温为例, 当阶次数为 5 时, 基准矩阵由 14 日直觉模糊化气温数据组成, 而运算矩阵由 9 ∼ 13 日的直觉模糊化气温数据组成; 辅助因子向量由 14 日直觉模糊化云密度数据组成. 套用式 (12) ∼ (15) 即可求得直觉模糊关系矩阵 $R(15)$. 因此, 如表 1 所示, 根据式 (16) 即可预测序列 $F(15)$ 的值.

Step 4: 去模糊化.

利用直觉模糊区间重心法计算式 (17) 对预测结果进行去模糊化处理得到预测结果, 明确数值如表 1 所示.

3.2　算法性能评估与比较

根据度量标准中的均方差 (MSE) 和平均预测误差率 (AFER)[2]将本文提出的预测方法与几种常用的经典模糊时间序列预测方法进行性能比较, 结果如表 4 所示. 常用的模糊时间序列预测方法中预测结果均方差和平均预测误差率最低分别达到 2.67 和 0.902 1, 本文预测结果较其分别降低了 23.2 和 14.8 个百分点, 相对于 Song 等最初建立的模糊时间序列预测模型则分别降低了 47.6 和 55.3 个百分点. 由此可见, 直觉模糊时间序列预测方法充分考虑了序列数据的模糊波动趋势, 并采用辅助预测因子较好地反映了模糊数据的不确定性本质, 获得了较高的预测精度.

表 4　不同预测方法的预测误差比较

度量标准	Avg.AFER	Avg.MSE
Song-Chissom 模型[3-4]	3.91	1.719
Chen 模型[5]	3.72	1.565
Song 模型[7]	3.31	1.316
Huarng 模型[9]	3.13	1.23
Yu 模型[2]	2.98	1.102
Young 模型[12]	2.67	0.9021
IFTS 预测模型	2.05	0.769

3.3　算法的复杂度分析

算法复杂度是考量算法性能的重要指标, 分为时间复杂度和空间复杂度. 时间复杂度是指算法的时间耗费, 它是算法所求解问题规模 n 的函数, 通常用 $O(n)$ 表示; 空间复杂度是指完成一个算法所需要占用的空间. 当评价算法性能优劣时, 算法的时间复杂度是重要的参照标准. 由本文模型的预测步骤可以看出, 算法的时间复杂度为 $O(mp\omega n)$. 其中: m 为因子数 (本文取 $m=2$), p 为子区间分组数, ω 为阶次数, n 为数据笔数. 如表 5 所示, 算法的运算时间与序列维数正相关. 空间复杂度则依赖于输入数据的维数, 由于额外空间相对于输入数据量而言是常数, 除特别说明外, 均按最坏情况分析. 因此, 本文算法的空间复杂度与常用的模糊时间序列预测算法基本相当. 根据表 5 数据结合各参数的取值可知, 在提高算法预测精确度的前提下, 直觉模糊时间序列预测模型达到了较低的算法复杂度.

表 5　不同预测方法时间复杂度比较

预测模型	$O(n)$
Song-Chissom 模型[3-4]	$O(kn)$
Chen 模型[5]	$O(n^2)$
Song 模型[7]	$O(n^2)$
Huarng 模型[9]	$O((k+1)n)$
Yu 模型[2]	$O(kn^2)$
Young 模型[12]	$O(2p\omega n)$
IFTS 预测模型	$O(2p\omega n)$

4 结 论

本文针对模糊时间序列预测理论的不足, 从优化区间划分和考虑多因素影响的角度出发, 建立了新的直觉模糊时间序列预测模型. 新模型充分利用了直觉模糊集对不确定性数据集描述和推理方面的优势, 较大程度地提高了系统的预测性能. 采用IFCM聚类算法动态划分论域区间, 反映了样本数据分布的不均匀性; 对直觉模糊关系转换矩阵进行改进, 使得新模型能够处理多因子预测问题; 基于数据分布模糊趋势的去模糊化处理, 更加接近不确定数据分布的实际. 分析实验结果可知, 直觉模糊时间序列预测模型在预测准确度和性能方面均优于常用的模糊时间序列预测模型. 如何使预测模型进一步推广用于处理更多因子和更长期时间的预测问题将是下一步的研究重点.

参考文献(References)

[1] Rafiei D. On similarity-based queries for time series data[C]. Proc of the 15th Int Conf on Data Engineering. Sydney, 1999: 410-417.

[2] Yu W L, Fang J W, Liao J P. A novel FCM-based deterministic forecasting model for fuzzy time series[J]. Computer Engineering and Science, 2010, 32(7): 112-116.

[3] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series: Part I[J]. Fuzzy Sets and Systems, 1993, 54(1): 1-9.

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

[5] Chen S M. Forecasting enrollments based on high-order fuzzy time series[J]. Cybernetics and Systems, 2002, 33(1): 1-16.

[6] Own C M, Yu P T. Forecasting fuzzy time series on a heuristic high-order model[J]. Cybernetics and Systems, 2005, 36(7): 705-717.

[7] Song Q. A note on fuzzy time series model relation with sample autocorrelation functions[J]. Int J of Cybernetics and Systems, 2003, 34(2): 93-107.

[8] Cheng C H, Chang J R, Yeh C A. Entropy-based and trapezoid fuzzification based fuzzy time series approaches for forecasting IT project cost[J]. Technological Forecasting and Social Change, 2006, 73(5): 524-542.

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

[10] Liu Jing-wei, Chen Tai-liang, Cheng Ching-hsue, et al. Adaptive-expectation based multi-attribute FTS model for forecasting TAIEX[J]. Computers and Mathematics with Applications, 2010, 59(2): 795-802.

[11] Teoh Hia Jong, Cheng Ching-hsue, Chu Hsing-hui, et al. Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets[J]. Data & Knowledge Engineering, 2008, 67(1): 103-117.

[12] Young Keun Bang, Lee Chul-heui. Fuzzy time series prediction using hierarchical clustering algorithms[J]. Expert Systems with Applications, 2011, 38(4): 4312-4325.

[13] 雷英杰, 王宝树, 苗启广. 直觉模糊关系及其合成运算[J]. 系统工程理论与实践, 2005, 25(2): 113-118.
(Lei Y J, Wang B S, Miao Q G. On the intuitionistic fuzzy relations with compositional operations[J]. Systems Engineering Theory and Practice, 2005, 25(2): 113-118.)

[14] Oscar Castillo, Arnulfo Alanis, Mario Garcia. Hector arias an intuitionistic fuzzy system for time series analysis in plant monitoring and diagnosis[J]. Applied Soft Computing, 2007, 7(4): 1227-1233.

[15] 申晓勇, 雷英杰, 李进, 等. 基于目标函数的直觉模糊集合数据的聚类方法[J]. 系统工程与电子技术, 2009, 31(11): 2732-2735.
(Shen X Y, Lei Y J, LI J, et al. Clustering technique to intuitionistic fuzzy sets data based on objective function[J]. Systems Engineering and Electronics, 2009, 31(11): 2732-2735.)

[16] Lee L W, Wang H, Chen S M, et al. Handling forecasting problems based on two-factors high-order fuzzy time series[J]. IEEE Trans on Fuzzy Systems, 2006, 14(3): 468-477.