

基于改进狼群算法的模糊时间序列预测模型

鲜思东^{1†}, 李堂金²

(1. 重庆邮电大学 复杂系统智能分析与决策重点实验室, 重庆 400065; 2. 重庆邮电大学 自动化学院, 重庆 400065)

摘要: 当使用模糊时间序列预测模型进行预测时, 模糊区间的不同划分对最后的预测精度有着十分重要的影响. 针对如何更有效的划分模糊区间、进一步提高模糊时间序列的预测精度问题, 本文提出了一种基于改进狼群算法的模糊时间序列预测模型. 为此首先简要介绍了模糊时间序列, 然后阐述了狼群算法并在其游走行为中引入趋向行为和死亡概率对其进行了改进, 最后利用改进狼群算法来划分模糊区间, 建立了一种新的模糊时间序列预测模型. 将Alabama大学入学人数作为实验数据进行实例分析和验证. 通过与现有的一些模型进行对比分析, 本文所提模型具有更高的预测精度, 为模糊时间序列预测提供了新思路.

关键词: 模糊时间序列模型; 狼群算法; 划分; Alabama大学入学人数; 预测

引用格式: 鲜思东, 李堂金. 基于改进狼群算法的模糊时间序列预测模型. 控制理论与应用, 2020, 37(7): 1637 – 1643

DOI: 10.7641/CTA.2019.90662

Fuzzy time series prediction model based on improved wolf pack algorithm

XIAN Si-dong^{1†}, LI Tang-jin²

(1. Key Laboratory of Intelligent Analysis and Decision on Complex Systems,

Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The division of fuzzy interval plays a very important role to the final prediction accuracy when one uses the fuzzy time series prediction model to prediction. In order to solve the problem of how to divide fuzzy interval more effectively and further improve the prediction accuracy, a fuzzy time series prediction model based on improved wolf pack algorithm is proposed in this paper. Firstly, the fuzzy time series is briefly introduced, then the wolf pack algorithm is described and the chemotactic behavior and death probability are introduced into its scouting behavior to improve it. Finally, using the improved wolf pack algorithm to divide the fuzzy interval, a new fuzzy time series prediction model is established. The enrollments of university of Alabama is analyzed and verified as experimental data. By comparing the forecasting results with some existing models, the proposed model has higher prediction accuracy and provides a new idea for fuzzy time series prediction.

Key words: fuzzy time series model; wolf pack algorithm; division; university of Alabama enrollments; forecasting

Citation: XIAN Sidong, LI Tangjin. Fuzzy time series prediction model based on improved wolf pack algorithm. *Control Theory & Applications*, 2020, 37(7): 1637 – 1643

1 引言

以时间轴为依据, 按照一定的时间间隔记录的数据称为时间序列. 利用时间序列进行分析是预测未来结果最古老和最可靠的技术之一. 传统的时间序列预

测方法不能解决时间序列值为“差”、“中等”、“好”等由语言术语描述的问题. 为了克服这一缺陷, Zadeh教授提出了模糊集(fuzzy sets, FS)理论^[1], Song和Chissom将模糊集理论成功引入时间序列模型中, 首次提

收稿日期: 2019-08-09; 录用日期: 2019-12-26.

[†]通信作者. E-mail: sidx@163.com; Tel.: +86 13594608518.

本文责任编辑: 吴敏.

重庆市质量技术监督局重大委托项目(CQZJZD2018001), 重庆市教委研究生教学改革研究项目(YJG183074), 重庆市社会科学规划项目(2018YB SH085), 重庆研究生科研创新项目(CYS18252), 国家自然科学基金项目(11671001)资助.

Supported by the Major Entrustment Projects of the Chongqing Bureau of Quality and Technology Supervision (CQZJZD2018001), the Graduate Teaching Reform Research Program of Chongqing Municipal Education Commission (YJG183074), the Chongqing Social Science Planning Project (2018YB SH085), the Chongqing Research and Innovation Project of Graduate Students (CYS18252) and the National Natural Science Foundation of China (11671001).

出了模糊时间序列(fuzzy time series, FTS)模型^[2-4]. 在这种模型中, 观测值是模糊集, 并在当前状态和以前状态之间建立关系.

模糊时间序列的另一个特点是它分析处理小数据集的能力相对于传统的时间序列更好. 20多年以来, 广大学者们在大学招生^[2,5-14]、气候学^[8-9,15]、股票市场^[5,8-10,16-19]、医疗健康^[20]、旅游^[21]等不同的应用领域建立了许多模糊时间序列预测模型.

论域划分是影响模糊时间序列模型预测精度的关键因素之一^[5]. 论域划分方法可以分为两类: 等长度划分和非等长度划分. 早期的研究都采用相等长度的论域划分方法^[6-7], 现在学者们更多地采用了非等长度的论域划分方法. 在基于数学模型的划分方法中, “最小熵原理法”和“梯形模糊化法”^[22]、自动聚类技术^[8,16]、“动态挖掘距离法”^[23]、“基于均值的离散化法”^[9]都表现出了较好的性能. 此外, 为了提高预测精度, “信息粒”^[10,17]和“模糊离散化技术”^[24]也被用来划分区间. 在基于仿生优化算法的论域划分方法中, 粒子群优化算法(particle swarm optimization, PSO)^[11-12,25]、蚁群优化(ant colony optimization, ACO)^[18]、遗传算法(genetic algorithm, GA)^[13,19]、人工鱼群算法(artificial fish swarm algorithm, AFSA)^[14,26]均取得了不错的效果.

为了进一步提高模糊时间序列预测模型的预测精度, 本文将狼群算法(wolf pack algorithm, WPA)进行了改进并用于模糊区间的划分, 提出了一种基于改进的狼群算法(improved wolf pack algorithm, IWPA)的模糊时间序列预测模型(IWPA-FTS). 首先针对基本狼群算法在游走过程中收敛速度慢、容易陷入局部极值的缺陷, 引入趋向行为和死亡概率, 有效提高探狼的收敛速度和跳出局部极值的能力. 然后将改进的狼群算法用于优化论域的划分. 最后以Alabama大学入学人数作为实验数据来验证所提模型的有效性. 通过与现有的一些模型进行对比分析, 本文所提模型具有更高的预测精度.

2 模糊时间序列(FTS)

定义1 设 X 为论域, X 由有限个离散的元素构成, 即 $X = \{x_1, x_2, \dots, x_n\}$. 则定义在论域上的模糊集合可以表示为

$$A = \frac{f_A(x_1)}{x_1} + \frac{f_A(x_2)}{x_2} + \dots + \frac{f_A(x_n)}{x_n}, \quad (1)$$

其中: f_A 是模糊集合 A 的隶属度函数, 并且 $f_A: X \rightarrow [0, 1]$, $f_A(x_i)$ 表示 x_i 属于模糊集 A 的隶属度值.

定义2 设论域 $Y(t)$ ($t = \dots, 0, 1, 2, \dots$)是实数集 $\mathbb{R}$ 上的一个子集, $f_i(t)$ ($i = 1, 2, \dots$)是定义在 $Y(t)$ 上的模糊集. 如果 $F(t)$ 是 $f_1(t), f_2(t), f_3(t), \dots$ 的一个集合, 即 $F(t) = \{f_1(t), f_2(t), f_3(t), \dots\}$, 则把

$F(t)$ 称为定义在 $Y(t)$ 上的模糊时间序列.

定义3 设 $F(t)$ ($t = 1, 2, 3, \dots$)为一个模糊时间序列, 如果 $F(t)$ 由 $F(t-1)$ 确定, $R(t-1, t)$ 为 $F(t)$ 和 $F(t-1)$ 之间的模糊逻辑关系(fuzzy logic relation, FLR), 则

$$F(t) = F(t-1) \circ R(t-1, t).$$

如果 $F(t-1) = A_i$ 且 $F(t) = A_j$, 则这种关系可以表示为 $A_i \rightarrow A_j$, 其中 A_i 和 A_j 分别表示模糊时间序列的当前状态和下一状态.

定义4 设有一组一阶模糊模糊逻辑关系为 $A_i \rightarrow A_{k1}, A_i \rightarrow A_{k2}, \dots, A_i \rightarrow A_{km}$, 即它们拥有相同的当前状态, 则可以用一个模糊逻辑关系组(fuzzy logical relation group, FLRG)将它们表示为

$$A_i \rightarrow A_{k1}, A_{k2}, \dots, A_{km}.$$

定义5 设 $F(t)$ 由 $F(t-n), F(t-n+1), \dots, F(t-1)$ 导致, 则 n 阶模糊逻辑关系可以表示为

$$F(t-n), F(t-n+1), \dots, F(t-1) \rightarrow F(t).$$

模糊时间序列的一般预测步骤如下: 1) 确定并划分论域; 2) 模糊化实际数据; 3) 建立FLRs和FLRGs; 4) 去模糊化计算预测值.

3 改进的狼群算法(IWPA)

吴虎胜等人受到自然界狼群狩猎行为和它们的猎物分配模式的启发, 提出了狼群算法(WPA)^[27], 并用它来解决优化问题. WPA中的狼一般分为3类: 头狼、探狼和猛狼, 狼群的整个捕食行为被抽象为3种智能行为, 即游走、召唤和围攻行为, 以及两种智能规则, 即“胜者为王”的头狼生成规则和“强者生存”的狼群更新规则. 下面对这些行为和规则进行详细的说明.

假设人工狼群的猎场空间为 $N \times D$ 维, 其中 N 为人工狼的数量, D 为解空间变量的个数, 则第 i 匹人工狼的位置可以表示为

$$\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iD}), \quad i = 1, 2, \dots, N,$$

x_{id} ($d = 1, 2, \dots, D$)为人工狼 i 的第 d 个待寻优的变量. 位置在 $\mathbf{X}_i$ 的第 i 匹人工狼的适应度值(目标函数值)可定义为 $Y_i = f(\mathbf{X}_i)$.

1) “胜者为王”的头狼产生规则: 在初始解空间中, 选择目标函数值最优的人工狼为头狼. 每次迭代后, 将目标函数值最好的人工狼与上一代头狼进行比较. 如果它的目标函数值更好, 那么上一代头狼将被替换, 头狼的位置将被更新. 头狼不参与执行3种智能行为, 而是直接进入下一次迭代, 直到被更强的狼替换.

2) 游走行为: 除了头狼外, 让目标函数值最佳的 S_{num} 匹人工狼充当探狼. S_{num} 是在区间 $[\frac{N}{(\alpha+1)}, \frac{N}{\alpha}]$

中随机选择的整数. α 为探狼比例因子. 首先, 探狼分别向 h 不同方向迈出一步, 记录每个方向的适应度值. 然后回到原来的位置. 探狼 i 向第 p ($p = 1, 2, \dots, h$) 个方向迈出一步后, 它的位置表示为

$$\mathbf{X}_i^p = \mathbf{X}_i + \text{rand}() \times \text{step}_a, \quad (2)$$

其中: $\text{rand}()$ 是在区间 $[-1, 1]$ 中均匀分布的随机数, step_a 表示游走步长. 探狼 i 向 h 个方向中选择具有最优适应度值且优于原位置的方向前进一步, 更新位置 $\mathbf{X}_i$. 重复上述行为直到某匹人工狼的适应度值优于头狼或者达到最大游走次数 $T_{\max}$.

3) 召唤行为: 头狼通过嚎叫召唤周围的 M_{num} 匹猛狼快速向头狼靠近. 其中 $M_{\text{num}} = N - S_{\text{num}} - 1$. 猛狼以相对较大的步长 step_b 向头狼的位置聚集. 猛狼 i 的位置更新式如下:

$$\mathbf{X}_i^{k+1} = \mathbf{X}_i^k + \text{step}_b \times \frac{(\mathbf{G}^k - \mathbf{X}_i^k)}{|\mathbf{G}^k - \mathbf{X}_i^k|}, \quad (3)$$

其中 $\mathbf{G}^k$ 是第 k 代头狼的位置. 在奔袭的过程中, 如果猛狼 i 的适应度值 Y_i 比头狼的适应度值 Y_{lead} 更好, 该猛狼将代替头狼, 同时 $Y_{\text{lead}} = Y_i$. 然后由新的头狼发起召唤行为. 否则, 猛狼 i 继续向头狼的位置靠拢直到 $d(i, \text{lead}) < d_{\text{near}}$, 然后等待进入围攻行为. $d(i, \text{lead})$ 表示猛狼 i 和头狼之间的距离; d_{near} 表示判定距离. 它决定了猛狼 i 是否由奔袭转为围攻行为. d_{near} 由式(4)计算得到

$$d_{\text{near}} = \frac{1}{D \cdot \varpi} \cdot \sum_{d=1}^D |\max_d - \min_d|, \quad (4)$$

其中: ϖ 是距离判定因子, $[\min_d, \max_d]$ 是第 d 个变量的取值范围.

4) 围攻行为: 通过向头狼的位置聚集后, 猛狼离头狼的位置都比较近了. 这时将头狼的位置视为猎物的位置, 猛狼将和探狼一起进行围攻猎物. 围攻过程中人工狼 i 的位置更新式如下:

$$\mathbf{X}_i^{k+1} = \mathbf{X}_i^k + \lambda \times \text{step}_c \times |\mathbf{G}^k - \mathbf{X}_i^k|, \quad (5)$$

其中: λ 是在区间 $[-1, 1]$ 均匀分布的随机数; step_c 表示围攻步长. 如果人工狼 i 执行围攻行为后, 新位置的适应度值优于原位置的值, 则更新该狼的位置 $\mathbf{X}_i$. 否则不变.

在3种智能行为中涉及的步长 step_a , step_b 和 step_c 之间具有如下的关系:

$$\text{step}_a = \frac{\text{step}_b}{2} = 2 \cdot \text{step}_c = \frac{|\max_d - \min_d|}{S}, \quad (6)$$

其中 S 表示步长因子, 它决定了人工狼在解空间寻找最优解的精细程度.

5) “强者生存”的狼群更新规则: 这个规则模拟达尔文提出的自然选择法则. 为了保障种族的延续, 食物按照先强后弱的顺序进行分配, 这会导致一些弱

小的狼因为分配不到食物而饿死. 在实际算法中表现为淘汰 m 匹具有最差适应度值的人工狼, 同时在解空间内随机生成 m 匹新的人工狼. m 是在区间 $[\frac{N}{2\beta}, \frac{N}{\beta}]$ 随机选择的一个整数. β 为群体更新比例因子.

WPA算法具有良好的全局收敛能力, 同时也有一些缺点, 如收敛速度慢、容易陷入局部极值等. 本文对WPA的游走行为进行改进, 提出了改进的狼群算法(IWPA).

在狼群算法的游走行为中, 当 h 个方向中具有最优适应度值的位置比原位置好的时候, 人工狼仅仅向该方向前进一步, 然后重新选择 h 个方向重复前面的行为, 这导致了狼群算法的收敛速度过慢. 事实上, 具有最优适应度值的方向是最可能改善适应度值的方向. 因此, 为了加速狼群算法的收敛速度, 本文在这里引入一个趋向行为. 当人工狼朝着改善适应度值的方向移动一步后, 人工狼将继续沿同一方向前进几步, 直到人工狼的适应度值不再改善或移动步数(N_s)达到了最大值($N_{s \max}$). 在前进的过程中, 人工狼的位置更新式如下:

$$\mathbf{X}_{i, \text{next}} = \mathbf{X}_i + \text{rand} \times \text{step}_a \times \Phi(i), \quad (7)$$

其中: rand 是在区间 $[0, 1]$ 之间均匀分布的随机数,

$$\Phi(i) = \frac{\mathbf{X}_j - \mathbf{X}_i}{|\mathbf{X}_j - \mathbf{X}_i|}$$

为适应度改善的方向, $\mathbf{X}_j$ 为具有更优适应度值的位置向量.

另外, 在游走过程中, 当 h 方向中的最优适应度比初始位置差时, 算法可能已经陷入局部极值. 在这种情况下, 为了提高人工狼跳出局部极值的能力, 本文做出如下改进: 对于上述情况的人工狼, 本文设置一个死亡概率 P_e , 当人工狼满足死亡概率时, 该人工狼死亡, 并随机地在解空间的任何地方生成一个新的个体, 否则位置不变. 改进后的人工狼游走行为的伪代码如算法1所示.

算法1 改进后的游走行为.

```

for  $t = 1 : T_{\max}$  do
  for  $i = 1 : S_{\text{num}}$  do
    for  $j = 1 : h$  do
      根据式(2)生成位置向量;
      计算该位置的适应度值;
    end
    找出最优适应度值 $Y_1$ 所对应的位置 $\mathbf{X}_1$ ;
    if  $Y_1 < Y_i$  then
       $\mathbf{X}_i = \mathbf{X}_1, Y_i = Y_1$ ;
    While  $N_s < N_{s \max}$  do
      根据式(7)生成位置向量 $\mathbf{X}_2$ ;

```

```

    计算位置 $\mathbf{X}_2$ 处的适应度值 $Y_2$ ;
    if  $Y_2 < Y_i$  then
         $\mathbf{X}_i = \mathbf{X}_2, Y_i = Y_2$ ;
    else
        set  $N_s = N_{s \max}$ ;
    end
end
else if 满足死亡概率 then
    在解空间中随机生成一个新的个体;
end
end
找出最优适应度值 $Y_3$ 所对应的位置 $\mathbf{X}_3$ ;
if  $Y_3 < Y_{\text{lead}}$  then
     $Y_{\text{lead}} = Y_3$ ;
    更新头狼的位置;
    break;
end
end

```

4 IWPA-FTS预测模型

为了进一步提高模糊时间序列模型的预测精度, 本文将IWPA算法用于模糊区间的划分, 建立了IWPA-FTS预测模型. 利用IWPA寻找最优的区间划分点的具体步骤如下:

步骤1 初始化参数.

确定论域 $U = [x_{\min} - l_1, x_{\max} + l_2]$, 其中: $x_{\max}$ 和 $x_{\min}$ 分别是实际数据集的最大值和最小值, l_1 和 l_2 是两个适当的正数. 如果区间划分个数为 n , 那么划分点的个数 $D = n - 1$. 设置人工狼的数量为 N , 探狼游走方向个数 h , 算法最大循环次数 M 以及算法涉及的其他相关参数. 创建一个 $N \times D$ 维狼群位置矩阵, 其中的每一个元素都均匀分布在论域范围内. 按升序排列每匹狼的位置向量.

步骤2 计算每匹人工狼在当前位置的适应度值.

步骤2.1 模糊化每一个历史数据.

在模糊时间序列中, 模糊集通常由下面的公式定义:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + \cdots + 0/u_{n-1} + 0/u_n, \quad (8a)$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + \cdots + 0/u_{n-1} + 0/u_n, \quad (8b)$$

$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + \cdots + 0/u_{n-1} + 0/u_n, \quad (8c)$$

$\vdots$

$$A_{n-1} = 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 1/u_{n-1} + 0.5/u_n, \quad (8d)$$

$$A_n = 0/u_1 + 0/u_2 + 0/u_3 + \cdots + 0.5/u_{n-1} + 1/u_n, \quad (8e)$$

其中 $u_1, u_2, u_3, \cdots, u_n$ 表示区间.

确定好模糊集后, 就可以将每个数据模糊化. 为了确定数据 x 所对应的模糊集, 本文首先要确定数据 x 所属的模糊区间 u_i , 然后计算数据 x 属于区间 u_i 的隶属度值, 接着找出最大隶属度值所对应的模糊集 A_i . 最后, 数据 x 就被模糊化为 A_i .

步骤2.2 建立模糊关系和模糊关系组.

通过定义3-4建立FLRs和FLRGs. 高阶FLRs可根据定义5建立.

步骤2.3 去模糊化并计算预测值.

预测值通过两种不同的方法计算: 对训练阶段的数据采用基于下一状态嵌入式贝叶斯网络(embedded Bayesian network, EBN)的方法, 对测试阶段的数据采用主投票(master voting, MV)方法进行计算.

Kuo等^[28]提出了EBN算法. 其基本思想是: 对每一个FLRG中的所有下一状态所对应的模糊区间等长度划分为3个小区间, 然后根据下面的表达式计算出预测值:

$$\text{预测值} = \frac{\sum_{k=1}^n \frac{\text{submid}_k + \text{mid}_k}{2}}{n}, \quad (9)$$

其中: n 表示在同一FLRG中下一状态的总数, mid_k ($1 \leq k \leq n$)表示下一状态对应的模糊区间的中点值, submid_k 表示第 k 个下一状态所对应的实际数据在3个小区间中所属的那个小区间的中点值. 为便于理解, 本文举个简单的例子进行说明: 假设在某一FLRG中的其中一个下一状态对应的实际数据为1550并且该状态对应的模糊区间为 $u_i = [1000, 1600]$, 则这3个小区间分别为 $u_{i1} = [1000, 1200]$, $u_{i2} = (1200, 1400]$, $u_{i3} = (1400, 1600]$. 因此,

$$\text{mid}_k = 1300, \text{submid}_k = \frac{1400 + 1600}{2} = 1500.$$

在测试阶段, 本文采用Kuo等^[11]提出的MV方法. 其计算式如下:

$$\text{预测值} = \frac{(m_{t_1} \times W_h) + m_{t_2} + \cdots + m_{t_\lambda}}{W_h + (\lambda - 1)}, \quad (10)$$

其中 λ 代表预测时需要用的与预测时刻相邻的前面时刻数据的个数, m_{t_i} ($i = 1, 2, \cdots, \lambda$)为前面第 i 时刻的数据所属模糊区间的中点值, W_h 表示与预测时刻相邻时刻的数据对预测值的影响程度. 为了方便比较, 本文取 $W_h = 15$.

步骤2.4 计算适应度值.

本文把预测值与真实值之间的均方根误差(root mean square error, RMSE)作为人工狼的适应度值, 其计算式如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\text{预测值} - \text{真实值})^2}{N}}. \quad (11)$$

比较所有人工狼的初试适应度值, 选取拥有最优适应度值(最小RMSE值)的人工狼作为头狼。

步骤3 选择人工狼的行为。

步骤3.1 除头狼外, 选择 S_{num} 匹具有最佳适应度值的狼作为探狼, 执行改进的游走行为, 直到某匹狼的适应度值优于头狼或者达到最大游走次数 T_{max} , 然后转向步骤3.2。

步骤3.2 头狼通过嚎叫执行召唤行为。剩余的 M_{num} 匹猛狼快速向猎物靠近。在奔袭的过程中, 如果某匹猛狼的适应度值优于头狼, 则该猛狼将代替头狼并发起召唤行为, 否则, 猛狼继续向头狼靠拢直到与头狼的距离 $d(i, \text{lead}) < d_{\text{near}}$, 然后转入步骤3.3。

步骤3.3 探狼和猛狼一起执行围攻行为。

步骤3.4 根据“强者生存”的狼群更新原则更新狼群。

步骤4 判断结束条件。

如果算法达到最大迭代次数 M 或者已经达到预设精度, 则输出最后一次迭代后头狼的适应度值和相应的位置向量, 同时得到的还有模糊时间序列、FLRs、FLRGs和训练阶段的预测值。否则, 重复步骤3—4。

5 实例分析

在本节中, 利用了Alabama大学1971年至1992年的实际入学人数(表1第2列)来证明本文所提模型的有效性。本实验使用MATLAB 2016a编写仿真程序, 在Windows 10操作系统, 英特尔i5-6500 3.19 GHz处理器, 16 G内存的PC机上实现。预测精度通过式(11)定义的均方根误差(RMSE)和式(12)平均预测误差率(average forecasting error rate, AFER)进行测量, AFER定义如下:

$$\text{AFER} = \frac{1}{N} \sum_{i=1}^N \frac{|\text{预测值} - \text{真实值}|}{\text{真实值}} \times 100\%. \quad (12)$$

5.1 训练阶段

本文中模型阶数为3, 模糊区间划分个数 $n = 7$, IWPA算法中相关参数设置如下: 人工狼的数量 $N = 100$, 算法循环次数 $M = 100$, 探狼比例因子 $\alpha = 4$, 游走方向个数 $h = 10$, 最大游走次数 $T_{\text{max}} = 20$, 步长因子 $S = 1000$, 群体更新比例因子 $\beta = 6$, 趋向行为的最大移动步数 $N_{\text{max}} = 4$, 游走行为中死亡概率 $P_e = 0.5$ 。

根据实际数据确定论域 $U = [13000, 20000]$, 利

用本文所提IWPA算法对论域进行划分, 得到的最优划分结果为

$$\mathbf{X}_{\text{best}} = [14509 \ 15296 \ 15634 \ 16695 \ 17251 \ 18498],$$

即

$$u_1 = [13000 \ 14509], u_2 = (14509 \ 15296],$$

$$u_3 = (15296 \ 15634], u_4 = (15634 \ 16695],$$

$$u_5 = (16695 \ 17251], u_6 = (17251 \ 18498],$$

$$u_7 = (18498 \ 20000].$$

相对应的模糊化结果、3阶FLR以及训练阶段的预测值都如表1所示。图1为训练阶段的实际入学人数与预测入学人数的对比。从图1中可以清楚地看出, 运用本文的模型计算出的预测值与实际值非常接近。

表1 实际数据和训练阶段结果

Table 1 Actual data and training phase results

年份	实际值/人	模糊集	3阶FLR	预测值/人
1971	13055	A_1	—	—
1972	13563	A_1	—	—
1973	13867	A_1	—	—
1974	14696	A_2	$A_1, A_1, A_1 \rightarrow A_2$	14771
1975	15460	A_3	$A_1, A_1, A_2 \rightarrow A_3$	15465
1976	15311	A_3	$A_1, A_2, A_3 \rightarrow A_3$	15409
1977	15603	A_3	$A_2, A_3, A_3 \rightarrow A_3$	15521
1978	15861	A_4	$A_3, A_3, A_3 \rightarrow A_4$	15988
1979	16807	A_5	$A_3, A_3, A_4 \rightarrow A_5$	16880
1980	16919	A_5	$A_3, A_4, A_5 \rightarrow A_5$	16973
1981	16388	A_4	$A_4, A_5, A_5 \rightarrow A_4$	16341
1982	15433	A_3	$A_5, A_5, A_4 \rightarrow A_3$	15465
1983	15497	A_3	$A_5, A_4, A_3 \rightarrow A_3$	15465
1984	15145	A_2	$A_4, A_3, A_3 \rightarrow A_2$	15034
1985	15163	A_2	$A_3, A_3, A_2 \rightarrow A_2$	15034
1986	15984	A_4	$A_3, A_2, A_2 \rightarrow A_4$	15988
1987	16859	A_5	$A_2, A_2, A_4 \rightarrow A_5$	16880
1988	18150	A_6	$A_2, A_4, A_5 \rightarrow A_6$	18082
1989	18970	A_7	$A_4, A_5, A_6 \rightarrow A_7$	18999
1990	19328	A_7	$A_5, A_6, A_7 \rightarrow A_7$	19249
1991	19337	A_7	$A_6, A_7, A_7 \rightarrow A_7$	19249
1992	18876	A_7	$A_7, A_7, A_7 \rightarrow A_7$	18999

为了直观地说明IWAP的优越性, 本文对混合人工鱼群算法(hybrid artificial fish swarm algorithm, HAF-SA)^[26]、WPA和IWPA3种算法划分论域进行了比较, 这里区间划分个数为7, 模型阶数为2, 仿真结果取20次运行结果的平均值, 它们的收敛曲线如图2所示, 优化结果及耗时情况如表2所示。由图2可知, IWPA具有更快的收敛速度和更高的优化精度。在表2中, 根据最好的最优值、最差的最优值、平均最优值3项指标可以看出, IWPA获得的结果最好, 其次为WPA和HAFSA。另外, 根据平均耗时情况来看, 虽然WPA的耗时比

HAFSA多,但是WPA的优化精度要比HAFSA高. IW-PA耗时最少获得的精度最高. 说明本文在WPA算法的游走行为中引入的趋向行为和死亡概率能有效改善WPA算法的收敛速度和优化精度.

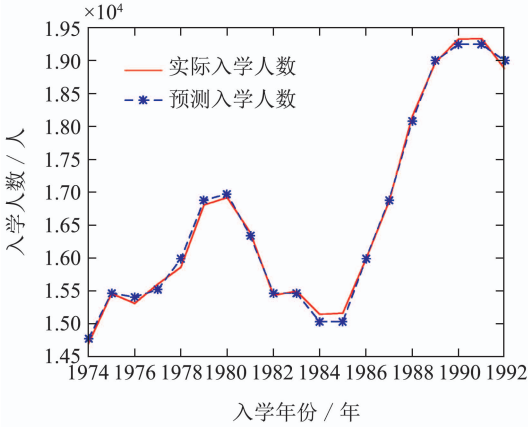


图 1 实际值与预测值对比结果

Fig. 1 Actual vs forecasted

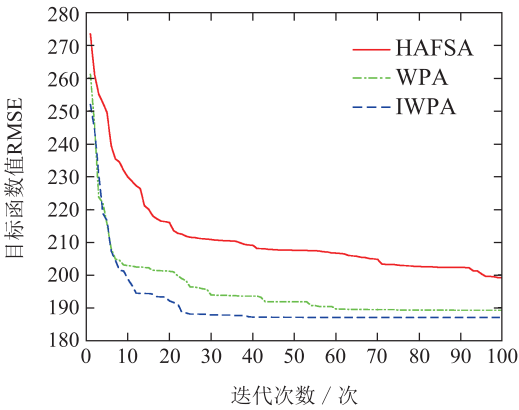


图 2 3种算法仿真结果比较

Fig. 2 Comparison results of three different algorithms

表 2 优化结果比较

Table 2 The comparison of optimization results

算法	最好的 最优值	最差的最优值	平均 最优值	平均耗时/s
HAFSA	187.95	217.83	199.08	138.40
WPA	185.66	207.03	189.20	204.24
IWPA	184.57	192.11	186.97	104.94

为了验证IWPA-FTS模型预测精度的提高,本文选取C96^[6], Q11^[29], W13^[30], PSO^[25], IFTS^[31]以及HAFSA^[26]多种预测模型进行比较,所有模型的区间划分数都取7. 比较结果如图3所示,从图3可以明显看出本文所提模型在训练阶段的预测精度最高.

为了验证本文所提模型在不同阶数情况下预测精度的有效性,本文与4个现有的预测模型: TVFTS^[32], GA^[33], PSO^[11], HAFSA^[26]分别在2阶、3阶和4阶的

情况下做了一个对比. 所有预测模型的区间划分个数为7, 对比结果如表3所示. 可以看出本文的模型在不同阶情况下均得到了最小的RMSE值以及最小的平均RMSE值.

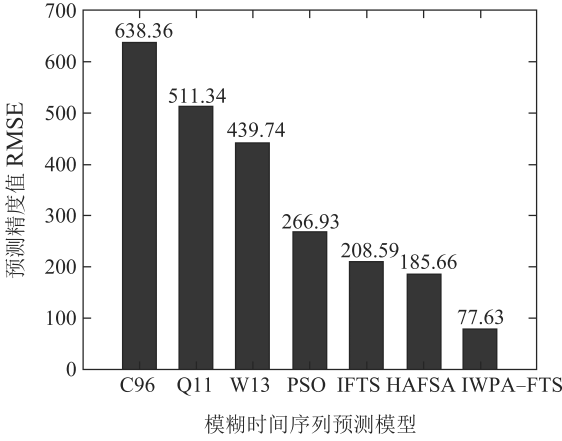


图 3 不同模型预测精度比较

Fig. 3 Comparison results of various forecast models

表 3 不同阶数情况下的对比结果

Table 3 Comparison results of different orders

阶数	TVFTS	GA	PSO	HAFSA	新模型
2	577.21	260.45	259.08	185.66	184.57
3	547.39	176.42	177.89	88.05	77.63
4	561.68	178.91	152.54	62.87	59.28
平均RMSE	562.09	205.26	196.50	112.19	107.16

5.2 测试阶段

为了验证本文所提模型在测试阶段的效果,即预测未来的能力. 本文将实际数据的最后3个作为测试数据集,根据剩余的19个数据建立相应的FLR和FLRG,对最后3个数据进行预测. 将本文的预测结果同模型C96^[6], HPSO^[11], AFPSO^[12]的预测结果进行对比,所有模型都采用MV方法进行预测, $W_h = 15$, $\lambda = 3$, 区间划分个数 $n = 7$, 模型阶数为3, 比较结果如表4所示. 从表4可以看出, 本文所提模型得到的预测结果具有最小的RMSE值和AFER值.

表 4 测试阶段预测结果比较

Table 4 Comparison of the predicted results for the testing phase

年份	实际值/人	C96/人	HPSO/人	AFPSO/人	新模型/人
1990	19328	18500	18988	18975	19034
1991	19337	19500	19167	19156	19168
1992	18876	19500	19265	19214	19249
RMSE	—	605.95	327.54	300.90	291.05
AFER	—	2.81%	1.57%	1.52%	1.46%

6 结论

本文提出了一种基于改进狼群算法的模糊时间序列预测模型, 并用它来预测Alabama大学的入学人数. 本文的主要创新点和贡献如下: 首先, 提出了改进的狼群算法(IWPA), 有效地提高了算法在游走阶段的收敛速度和跳出局部极值的能力. 其次, 本文首次将模糊时间序列与狼群算法相结合提出了IWPA-FTS预测模型, 拓宽了狼群算法的应用范围. 最后, 通过与各种预测模型相比, 本文所提模型在训练阶段和测试阶段均能获得较好的预测性能, 证实了该模型的有效性.

参考文献:

- [1] ZADEH L A. Fuzzy set. *Information and Control*, 1965, 8(3): 338 – 353.
- [2] SONG Q, CHISSOM B S. Forecasting enrollments with fuzzy time series part I. *Fuzzy Sets and Systems*, 1993, 54(1): 1 – 9.
- [3] SONG Q, CHISSOM B S. Fuzzy time series and its models. *Fuzzy Sets and Systems*, 1993, 54(3): 269 – 277.
- [4] SONG Q, CHISSOM B S. Forecasting enrollments with fuzzy time series part II. *Fuzzy Sets and Systems*, 1994, 62(1): 1 – 8.
- [5] HUARNG K. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets and Systems*, 2001, 123(3): 387 – 394.
- [6] CHEN S M. Forecasting enrollments based on fuzzy time-series. *Fuzzy Sets and Systems*, 1996, 81(3): 311 – 319.
- [7] LEE M H, EFENDI R, ISMAIL Z. Modified weighted for enrollment forecasting based on fuzzy time series. *Matematika*, 2009, 25(1): 67 – 78.
- [8] CHEN S M, TANUWIJAYA K. Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. *Expert Systems with Applications*, 2011, 38(12): 15425 – 15437.
- [9] SINGH P, BORAH B. An efficient time series forecasting model based on fuzzy time series. *Engineering Applications of Artificial Intelligence*, 2013, 26(10): 2443 – 2457.
- [10] LU W, CHEN X, PEDRYCZ W, et al. Using interval information granules to improve forecasting in fuzzy time series. *International Journal of Approximate Reasoning*, 2015, 57: 1 – 18.
- [11] 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. *Expert Systems with Applications*, 2009, 36(3): 6108 – 6117.
- [12] HUANG Y L, HORNG S J, HE M, et al. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 2011, 38(7): 8014 – 8023.
- [13] USLU V R, BSA E, YOLCU U, et al. A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations. *Swarm and Evolutionary Computation*, 2014, 15: 19 – 26.
- [14] XIAN Sidong, ZHANG Jianfeng. Fuzzy time series model based on improved artificial fish swarm algorithm. *Statistics & Decision*, 2018, 34(8): 68 – 72.
(鲜思东, 张建锋. 基于改进人工鱼群算法的模糊时间序列模型. 统计与决策, 2018, 34(8): 68 – 72.)
- [15] SINGH P, BORAH B. High-order fuzzy-neuro expert system for time series forecasting. *Knowledge-Based Systems*, 2013, 46: 12 – 21.
- [16] CHEN S M, TANUWIJAYA K. Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques. *Expert Systems with Applications*, 2011, 38(8): 10594 – 10605.
- [17] CHEN M Y, CHEN B T. A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Information Sciences*, 2015, 294(2): 227 – 241.
- [18] CAI Q, ZHANG D, ZHENG W, et al. A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowledge-Based Systems*, 2015, 74(1): 61 – 68.
- [19] PAL S S, KAR S. Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory. *Mathematics and Computers in Simulation*, 2019, 162: 18 – 30.
- [20] GARG B, GARG R. Enhanced accuracy of fuzzy time series model using ordered weighted aggregation. *Applied Soft Computing*, 2016, 48: 265 – 280.
- [21] ALADAG C H, EGRIOLGU E, YOLCU U, et al. A high order seasonal fuzzy time series model and application to international tourism demand of turkey. *Journal of Intelligent & Fuzzy Systems Applications in Engineering & Technology*, 2014, 26(1): 295 – 302.
- [22] CHENG C H, CHANG J R, YEH C A. Entropy-based and trapezoid fuzzificationbased fuzzy time series approaches for forecasting IT project cost. *Technological Forecasting & Social Change*, 2006, 73(5): 524 – 542.
- [23] IZAKIAN H, PEDRYCZ W, JAMAL I. Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, 2015, 39: 235 – 244.
- [24] BOSE M, MALI K. High order time series forecasting using fuzzy discretization. *International Journal of Fuzzy System Applications*, 2016, 5(4): 147 – 164.
- [25] QIU W, ZHANG C, PING Z. Generalized fuzzy time series forecasting model enhanced with particle swarm optimization. *International Journal of u- and e-Service, Science and Technology*, 2015, 8(7): 129 – 140.
- [26] XIAN S D, ZHANG J F, XIAO Y, et al. A novel fuzzy time series forecasting method based on the improved artificial fish swarm optimization algorithm. *Soft Computing*, 2018, 22(12): 3907 – 3917.
- [27] WU Husheng, ZHANG Fengming, WU Lushan. New swarm intelligence algorithm-wolf pack algorithm. *Systems Engineering & Electronics*, 2013, 35(11): 2430 – 2438.
(吴虎胜, 张凤鸣, 吴庐山. 一种新的群体智能算法—狼群算法. 系统工程与电子技术, 2013, 35(11): 2430 – 2438.)
- [28] KUO I H, HORNG S J, CHEN Y H, et al. Forecasting TAIEX based on fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 2010, 37(2): 1494 – 1502.
- [29] QIU W R, LIU X D, LI H L. A generalized method for forecasting based on fuzzy time series. *Expert Systems with Applications*, 2011, 38(8): 10446 – 10453.
- [30] WANG J, CHEN K. Frequency-weighted fuzzy time series model based on time variations. *Classics in Applied Mathematics (CIAM)*, 2013, 11(2): 76 – 90.
- [31] LEE C L, KUO S C, LIN C J. An efficient forecasting model based on an improved fuzzy time series and a modified group search optimizer. *Applied Intelligence*, 2017, 46(3): 641 – 651.
- [32] HWANG J R, CHEN S M, LEE C H. Handling forecasting problems using fuzzy time series. *Fuzzy Sets & Systems*, 1998, 100(1/3): 217 – 228.
- [33] CHEN S M, CHUANG N Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *International Journal of Intelligent Systems*, 2006, 21(5): 485 – 501.

作者简介:

鲜思东 硕士, 教授, 目前研究方向为不确定决策与优化、统计分析, E-mail: sidx@163.com;

李堂金 硕士研究生, 目前研究方向为模糊决策与控制, E-mail: tangjinli.cqupt@qq.com.