

DOI:10.3969/j.issn.1001-5337.2020.4.072

# 一种新的区间模糊数时间序列预测模型\*

谢小军<sup>①</sup>, 马虹<sup>②</sup>, 杨付贵<sup>①</sup>, 邱云兰<sup>①</sup>

(<sup>①</sup>广州工商学院, 510850; <sup>②</sup>广东金融学院, 510521, 广东省广州市)

**摘要:**通过对区间模糊数进行研究分析,发现已有处理方法是直接对两个界点建模和预测,这样做往往会导致不能很好的描述序列整体性的发展趋势以及模型所预测的结果容易发生错乱等,从而预测失效.首先基于等价和整体性考虑提出了模糊序列的面积序列和重心序列概念.然后对面积序列和重心序列分别建立了遗传优化 BP 神经网络模型进行回归和预测,并通过还原公式得到原区间模糊数序列的拟合值和预测值.最后通过实例验证了该方法有效可行,对比传统的 BP 神经网络模型和 ARIMA 模型,显著提高了预测精度.

**关键词:**遗传算法;BP 神经网络;区间模糊数;预测

**中图分类号:**O159 **文献标识码:**A **文章编号:**1001-5337(2020)04-0072-05

## 0 引言

1994 年, Song<sup>[1-2]</sup> 在原有的模糊集理论基础上, 研究并构建了模糊时间序列预测模型, 自此打开了人们对模糊时间序列应用研究的大门, 此后对于模糊时间序列的研究受到越来越多的专家和研究者的关注与重视, 并已经取得大量的研究成果. 例如袁潮清等<sup>[3]</sup> 提出了将区间灰数序列分别采用发展趋势序列和认知程度序列来进行描述, 此方法使得区间灰数预测过程中的灰数运算问题得到成功避免, 而且使得区间灰数序列本身所拥有的信息得到成功利用; 余文利等<sup>[4]</sup> 基于模糊 C 均值算法建立了模糊时间序列预测模型, 并利用该模型对杭州的温度进行了预测和分析; 还有很多学者结合一些新颖的机器学习的方法, 例如粒子群算法模型<sup>[5]</sup>、神经网络方法<sup>[6]</sup>、决策树<sup>[7]</sup>等, 它们不需要建立模糊关系矩阵, 而是利用各预测规则得到的值加权后求得预测值, 但是总体上这些方法的预测结果还是令人比较满意的, 而且在实践中其模糊规则也具有说服力, 可作为模糊时间序列预测模型未来的重要研究方向之一. 要瑞璞等<sup>[8]</sup> 对模糊数的 HFGB 算子进行扩展, 定义了区间值三角模糊数几何加权均值 (ITFGWM) 算子. 并对 Carlsson 定义的均值进行扩展, 定义了区间

值三角模糊数的均值, 且基于 ITFGWM 算子、均值及可能度给出了一种区间值三角模糊数预测方法. 高志方等<sup>[9]</sup> 提出了一种新的计算区间二型模糊数之间距离的测度, 然后结合该距离测度公式构造出一种新的基于距离测度来求权重的方法. 最近几年组合模型越来越受到研究者的青睐, 故一些学者提出了利用组合模型进行预测的方法<sup>[10-15]</sup>, 在一定程度上能够提高区间模糊数时间序列的预测精度, 但是已有的处理方法是直接对两个界点建模和预测, 这样做往往会导致不能很好的描述序列整体性的发展趋势以及模型所预测的结果容易发生错乱等缺陷, 甚至出现上界点小于下界点的情况, 从而预测失效.

文章基于等价和整体性考虑提出了模糊序列的面积序列和重心序列概念, 提出了一种基于遗传优化的 BP 神经网络<sup>[16,17]</sup> 区间预测模型. 首先将区间模糊数转换成精确数序列, 然后对精确数序列建立遗传 BP 神经网络区间模型, 最后利用训练出的模型进行预测, 并利用还原公式还原为区间模糊数, 从而实现了区间模糊数的预测.

## 1 区间模糊数时间序列的转换

大量的研究者对于二元区间模糊数序列的处理

\* 收稿日期: 2020-01-08

基金项目: 广州工商学院 2019 年院级科研课题立项项目 (KA201933).

作者简介: 谢小军, 男, 1990-, 硕士, 讲师; 研究方向: 智能算法; 决策分析与预测; E-mail: 454625759@qq.com.

通信作者: 马虹, 女, 1989-, 硕士; 研究方向: 智能算法、决策分析与预测; E-mail: 1037496299@qq.com.

是通过直接对两个界点进行建模和预测,这样做往往会存在的缺陷是:a.不能很好的描述序列整体性的发展趋势;b.建立的模型所预测的结果容易发生错乱,导致预测失效.因此出于整体性考虑,先将二元区间模糊数序列转换成等量信息的两个精确数序列,然后对转换后的两个精确数列构建组合模型并进行预测,最后再将预测结果利用还原公式转换为二元区间模糊时间数序列.

设二元区间模糊数  $X = (x(t_1), x(t_2), \dots, x(t_n))$ , 其中第  $i$  个二元模糊数为  $x(t_i) = (a_i, b_i)$ ,  $i = 1, 2, \dots, n$ , 二元模糊数的隶属函数为

$$u(x(t_i)) = \begin{cases} 0, & x < a_i, \\ 1, & a_i \leq x \leq b_i, \\ 0, & x > b_i. \end{cases} \quad (1)$$

隶属函数所覆盖的面积为

$$s(t_i) = \int_{a_i}^{b_i} 1 dx, \quad i = 1, 2, \dots, n. \quad (2)$$

比较和排序的指标以模糊数的重心代替,利用如下公式计算

$$\bar{X} = \frac{\int_s x u(x) d\sigma}{\int_s u(x) d\sigma}. \quad (3)$$

二元模糊数  $x(t_i) = (a_i, b_i)$  利用式(3)计算得重心为

$$\bar{X}_i = \frac{a_i + b_i}{2}. \quad (4)$$

**定义 1** 在二元模糊数序列中,将各个二元模糊数的隶属函数所覆盖的面积构成的实数序列定义为面积序列,记为  $S = (s(t_1), s(t_2), \dots, s(t_n))$ .

**定义 2** 在二元模糊数序列中,将各个二元模糊数计算所得的重心构成的实数序列定义为重心序列,记为  $\bar{X} = (\bar{X}(t_1), \bar{X}(t_2), \dots, \bar{X}(t_n))$ .

上述方法将二元模糊数序列转换成了两个精确数序列:面积序列和重心序列.由此两个精确数列也可以推导出原二元模糊序列的两个界点,由式(2)和式(4)可得

$$\begin{cases} a_i = \bar{X}_i - \frac{1}{2}s(t_i), \\ b_i = \bar{X}_i + \frac{1}{2}s(t_i). \end{cases} \quad (5)$$

转换后的两个精确序列同时受到二元模糊数的两个界点的约束,从而保持了模糊数的整体性,且避免了模糊数的界点可能产生跳跃,也让序列的光滑性更好,也避免了预测的两个界点相对位置出现错乱

的情况.此方法还可以推广到对三角模糊数序列和梯形模糊数序列的转换.

## 2 遗传 BP 神经网络模型的构建

### 2.1 BP 神经网络模型

反向传播(Back Propagation,简称 BP)神经网络的结构往往由输入层、输出层以及多个隐含层构成:以一个三层结构的为例,即含有一个输入层、一个输出层和一个隐含层,每层由  $n, l, m$  个神经元组成,其基本结构如下图 1 所示.

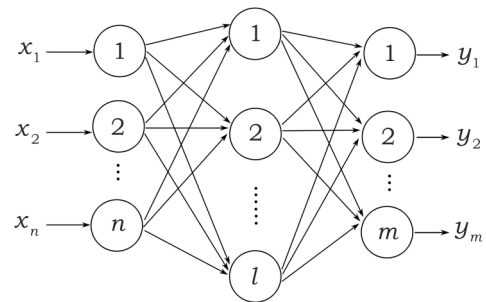


图 1 BP 神经网络基本结构

### 2.2 遗传 BP 神经网络模型

遗传算法(Genetic Algorithm)是一种并行随机搜索最优化算法,一通过模拟自然界遗传机制和生物进化论而成.由于 BP 神经网络易陷入局部最优的缺点,因此通过如何优化 BP 神经网络模型初始权阈值寻找最优解一直是研究的热点问题<sup>[17]</sup>.遗传算法优化的 BP 神经网络算法,是通过结合遗传算法和 BP 神经网络模型,遗传算法的输入信息是以 BP 神经网络各隐含层节点的阈值和权值,并将它们编码生成染色体,利用遗传算法的选择算子、交叉算子、变异算子产生新的后代作为 BP 算法的初始值,然后接着训练网络 BP 神经网络,重复进行此过程,直到满足问要求的误差精度.遗传 BP 神经网络模型建立的基本流程如下页图 2 所示,主要流程包括三部分:BP 神经网络连接结构的确定、遗传算法优化 BP 神经网络的权值和阈值以及 BP 神经网络预测.

本文利用遗传算法优化 BP 神经网络模型主要包括如下步骤<sup>[17]</sup>:

**Step 1** 种群初始化.个体编码方法为实数编码,每个个体均为一个实数串,由输入层与中间层连接权值、中间层阈值、中间层与输出层连接权值及输出层阈值 4 部分组成,个体包含了神经网络全部权值和阈值,在已知网络结构为  $n-l-m$  时,则个体

编码的长度共计为:  $L = n \times l + n \times m + l \times m$ . 其中  $n, l, m$  表示输入层、中值和阈值, 在已知网络结构为  $n-l-m$  时, 则个体编码的长度共计为:  $L = n \times l + n \times m + l \times m$ . 其中  $n, l, m$  表示输入层、中间层和输出层神经元的个数.

**Step 2** 适应度函数的建立. 根据个体初始化得到网络的初始权值和阈值, 利用训练数据训练 BP 神

经网络并得到预测值, 鉴于适应度值越大越好, 而预测值输出与期望输出之间的误差平方和越小说明该个体越优, 因此采用误差平方和的倒数作为适应度值

$$F = 1 / \sum_{i=1}^n (y_i - o_i)^2,$$

式中  $n$  为输入样本数,  $y_i$  表示第  $i$  个样本的期望输出值,  $o_i$  为第  $i$  个样本的预测输出值.

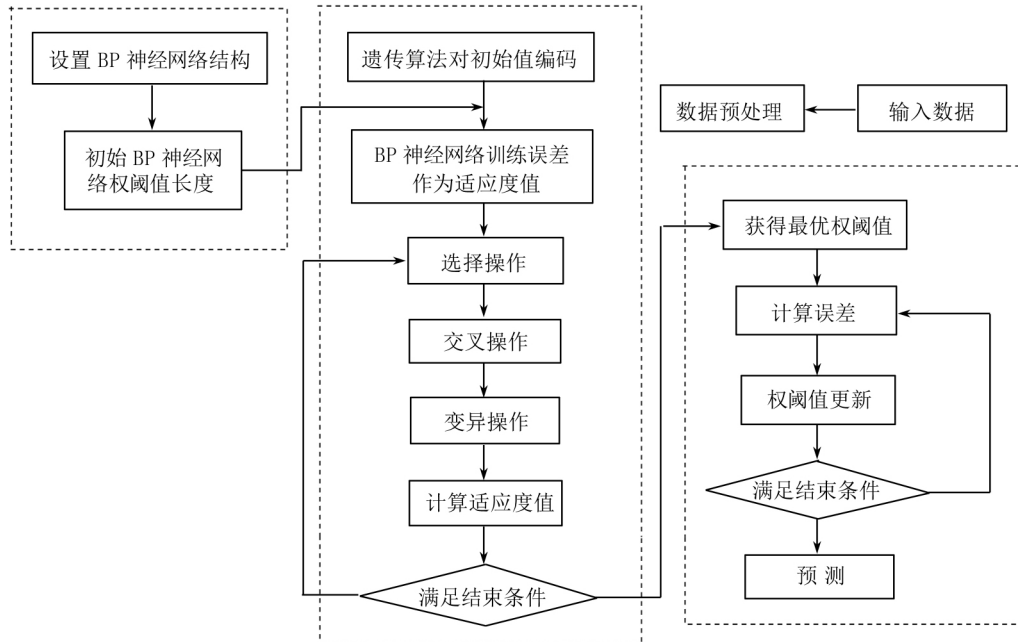


图2 遗传BP神经网络模型流程图

**Step 3** 选择操作. 本文采用轮盘赌法进行选择操作, 即以适应度比例的选择方法, 第  $i$  个个体的选择概率  $p_i$  采用如下公式计算

$$p_i = F_i / \sum_{j=1}^N F_j,$$

式中  $F_i$  为第  $i$  个个体的适应度值,  $N$  为种群个体数目.

**Step 4** 交叉操作. 以实数交叉法进行交叉, 染色体  $a_k$  与染色体  $a_s$  在  $j$  位按如下方法进行交叉操作

$$\begin{cases} a_{kj} = a_{kj}(1-b) + a_{sj}b, \\ a_{sj} = a_{sj}(1-b) + a_{kj}b, \end{cases}$$

式中  $b$  为  $[0, 1]$  之间的随机数.

**Step 5** 变异操作. 将第  $i$  个个体与第  $j$  个基因进行变异, 变异方法如下:

$$a_{ij} = \begin{cases} a_{ij} + (a_{\max} - a_{ij}) \cdot f(g), & r > 0.5, \\ a_{ij} - (a_{ij} - a_{\min}) \cdot f(g), & r \leq 0.5, \end{cases}$$

式中  $f(g) = r'(1 - g/G_{\max})^2$ ,  $a_{\max}$ ,  $a_{\min}$  分别表示基因的上、下界,  $g$  表示当前迭代次数,  $G_{\max}$  表示最大进化次数,  $r, r'$  为  $[0, 1]$  之间的随机数.

**Step 6** 将遗传算法获得的最优权阈值代入 BP 神经网络模型进行训练, 并用此模型进行预测.

### 3 数值实验

以 2019 年 1 月 10 日至 2019 年 3 月 8 日上证指数每日的最小值和最大值作为二元区间模糊数的两个界点, 数据来源于东方财富网 (<http://quote.eastmoney.com/zs000001.html>), 总共样本数据 37 组. 以 2019 年 1 月 10 日至 2019 年 3 月 1 日 32 天原始数据作为训练样本, 2019 年 3 月 4 日至 2019 年 3 月 8 日 5 天的值作为测试样本. 首先, 对 2019 年 1 月 2 日至 2019 年 3 月 2 日二元区间模糊数序列的两个界点  $a_i, b_i$  利用式(2)和(4)转换为两个精确数  $s(t_i)$ ,  $\bar{X}_i$ , 然后对精确序列  $S, \bar{X}$  分别建立遗传 BP 神经网络模型. 对精确序列  $S, \bar{X}$  具体建模过程如下:

**Step 1** 训练数据预处理. 利用 MATLAB 软件 premnmx 函数归一化后的数据变换到  $[-1, 1]$  的范围之间, 训练结束再利用 postmnmx 函数反归一

化为正常的。

**Step 2** BP 神经网络的构建.总的训练样本  $N=32$ .隐含层的激励函数采用 Log-sigmoid 型函数,输出层激励函数为纯线性函数,网络训练函数为共轭梯度法,设置 BP 神经网络参数:网络结构为 3—7—1,预设精度要求 0.000001,最大训练次数为 100,步长为 0.05.利用 MATLAB 软件编程,网络的初始权值、学习率和阈值由网络自动选取。

**Step 3** 利用遗传算法优化 BP 神经网络的权值和阈值.在训练遗传神经网络时,参数选取与之前的 BP 神经网络一致,设置具体的遗传参数为:种

群规模为 10,最大迭代次数为 50,交叉概率为 0.4,变异概率为 0.2.在对序列  $S, \bar{X}$  在寻优的过程中,通过不断优化初始权值和阈值,分别经过 77 次和 96 次训练达到精度要求。

**Step 4** 利用 Step 3 优化后得到初始权阈值重新对 BP 神经网络进行训练和回测,通过还原公式将序列  $S, \bar{X}$  的预测值还原为上证指数预测值  $\tilde{a}_i, \tilde{b}_i$ 。

将本文构建的预测模型的预测结果与传统 BP 神经网络模型和 ARIMA 时间序列模型分别进行对比,见表 1。

表 1 实际值与预测值

原始序列	ARIMA 模型	BP 神经网络模型	本文预测模型
[3006.94,3090.80]	[3021.12,3065.04]	[2884.89,3010.12]	[22954.96,3033.18]
[3009.45,3054.96]	[3054.84,3125.21]	[2999.58,2925.42]	[3004.93,3052.60]
[3050.11,3103.83]	[3047.09,3125.53]	[2941.26,2967.34]	[3036.38,3106.50]
[3074.97,3129.94]	[3041.49,3114.83]	[2945.49,2963.91]	[3061.34,3141.55]
[2969.58,3075.05]	[3052.23,3143.68]	[2865.28,3059.92]	[2953.84,3052.58]

为了验证本文提出的基于遗传优化的 BP 神经网络区间预测模型的有效性,下面分别计算传统 BP 神经网络模型和 ARIMA 时间序列模型以及本文预测模型的各项误差指标.结果见表 2.由 4 个误差指标可以看出,本文提出的预测方法均小于传统 BP 神经网络模型和 ARIMA 时间序列模型,故本文提出的基于模糊序列的面积序列和重心序列所构建遗传优化的 BP 神经网络区间预测模型可有效的提高区间模糊时间序列的预测精度。

表 2 3 种方法预测效果评价指标

预测模型	MAEP	MAEL	MAEI	MRIE
ARIMA 模型	35.744	40.29	76.034	0.0057569
BP 神经网络模型	94.91	105.57	200.48	0.015567
本文预测模型	19.92	19.346	39.266	0.0032362

## 4 结论及建议

文章研究了区间模糊序列的预测方法,首先基于等价和整体性考虑提出了模糊时间序列的面积序列和重心序列,然后对面积序列和重心序列分别建立了遗传优化的 BP 神经网络模型进行回归和预测,最后区间模糊数序列的拟合值和预测值可利用还原公式得到.数值实验表明,所提出的方法相对于

4 个误差指标计算公式:

$$\text{MAEP} = \frac{1}{N} \sum_{i=1}^N |a_i - \tilde{a}_i|,$$

$$\text{MAEL} = \frac{1}{N} \sum_{i=1}^N |b_i - \tilde{b}_i|,$$

$$\text{MAEI} = \text{MAEP} + \text{MAEL},$$

$$\text{MRIE} = \frac{1}{N} \sum_{i=1}^N \frac{|a_i - \tilde{a}_i|}{b_i + \tilde{b}_i}.$$

传统的 BP 神经网络模型和 ARIMA 模型具有更好的拟合效果和预测精度.此方法可以进一步推广实现对三角模糊数以及梯形模糊数建模。

## 参考文献:

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

- (1):1-9.
- [2] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series (Part II)[J]. Fuzzy Sets and Systems, 1994, 62(1):1-8.
- [3] 袁潮清,等.基于发展趋势和认知程度的区间灰数预测[J].控制与决策,2011,26(2):313-316.
- [4] 余文利,方建文,廖建平.一种新的基于模糊C均值算法的模糊时间序列确定性预测模型[J].计算机工程与科学,2010,32(7):112-116.
- [5] Kuo I H, Horng S J, Chen Y H, et al. Forecasting TAI-FEX based on fuzzy time series and particle swarm optimization[J]. Expert Systems with Applications, 2010, 37(2):1494-1502.
- [6] Egrioglu E, Aladag C H, Yolcu U, et al. A new approach based on artificial neural networks for high order multivariate fuzzy time series[J]. Expert Systems with Applications, 2009, 36(7):10589-10594.
- [7] Qiu W, Liu X, Wang L. Forecasting shanghai composite index based on fuzzy time series and improved C-fuzzy decision trees[J]. Expert Systems with Applications, 2012, 39(9):7680-7689.
- [8] 要瑞璞,沈惠璋.基于区间值三角模糊数的多属性群决策方法[J].数学的实践与认识,2015,45(20):197-203.
- [9] 高志方,盛冠帅,彭定洪.基于区间二型模糊数的多属性群体灰色关联决策方法[J].数学的实践与认识,2016,46(24):221-229.
- [10] 杨雷,等.属性值和权重均是区间三角模糊数的群体决策方法[J].模糊系统与数学,2017,31(01):183-190.
- [11] 朱家明,等.基于ICOFWA算子的连续区间模糊组合预测模型及其应用[J].模糊系统与数学,2016,30(3):172-184.
- [12] 丁珍妮,陈华友,朱家明.基于面积型中心误差准则的模糊组合预测模型及应用[J].模糊系统与数学,2018,32(3):165-175.
- [13] 杜康,袁宏俊,郑亚男.基于三角模糊数及GOWA算子的区间型组合预测模型[J].统计与决策,2019,35(16):22-28.
- [14] 袁宏俊,杜康,胡凌云.基于三角模糊数相似度的区间型组合预测模型[J].统计与决策,2019,35(6):24-29.
- [15] 谢小军,邱云兰,时凌.基于ARIMA和BP神经网络组合模型的能源消费预测[J].数学的实践与认识,2019,45(10):292-298.
- [16] 吴慧静,赫晓慧.基于GA-BP神经网络的空气质量指数预测研究[J].安徽师范大学学报(自然科学版),2019,42(4):360-365.
- [17] 李松,刘力军,解永乐.遗传算法优化BP神经网络的短时交通流混沌预测[J].控制与决策,2011,26(10):1581-1585.

## The BP neural network based on GA optimization in interval fuzzy number time series prediction

XIE Xiaojun<sup>①</sup>, MA Hong<sup>②</sup>, YANG Fugui<sup>①</sup>, QIU Yunlan<sup>①</sup>

(<sup>①</sup> Guangzhou College of Technology and Business, 510850; <sup>②</sup> Guangdong University of Finance, 510521, Guangzhou, Guangdong, PRC)

**Abstract:** Through the research and analysis of interval fuzzy number, it is found that the existing processing method is to directly model and predict the two boundary points, which often leads to the defects of not well describing the development trend of integrity and the results predicted by the model are prone to be disordered, so as to predict the failure. Firstly, the concepts of area sequence and center of gravity sequence of fuzzy sequences are proposed based on equivalence and integrity. Then, the BP neural network model of genetic optimization is established for regression and prediction of area sequence and center of gravity sequence respectively, and the fitting value and prediction value of original interval fuzzy number sequence are obtained by reduction formula. Finally, example is given to verify the effectiveness and feasibility of this method. Compared with the traditional BP neural network model and ARIMA model, the prediction accuracy is significantly improved.

**Key words:** genetic algorithms; BP neural network; interval fuzzy numbers; prediction