

# 基于主成分分析的模糊时间序列模型的平稳化算法

陈刚<sup>†</sup>, 丁慧玲

(大连海事大学 数学系, 辽宁 大连 116026)

**摘要:** 在模糊时间序列模型建立的过程中,对数据的预处理和模糊规则的优化往往是影响模型预测精确度的关键因素.针对上述问题,提出基于主成分分析(PCA)的平稳化算法.首先,对数据进行平稳化检验,并将非平稳的数据进行预处理使其平稳;其次,对论域进行划分并根据模糊关系构建广义的协方差矩阵,由此计算广义协方差矩阵的特征值和特征向量;再次,根据特征值的累计贡献率优化模糊规则,利用优化后模型进行预测;最后,通过实际案例验证新算法的可行性.

**关键词:** 模糊时间序列; 平稳化; 主成分分析; 广义协方差矩阵; 模糊规则优化

中图分类号: TP273

文献标志码: A

## Stabilization algorithm of fuzzy time series based on principal component analysis

CHEN Gang<sup>†</sup>, DING Hui-ling

(Department of Mathematics, Dalian Maritime University, Dalian 116026, China)

**Abstract:** In the process of establishing fuzzy time series model, the data pre-process and the optimization of the fuzzy rules are the two key factors which influence the precision of predication. Therefore this paper proposes a stabilization algorithm based on principal component analysis(PCA). Firstly, time series are changed into stationary time series by augmented dickey-fuller tests. Then, the fuzzy sets are set up, the generalized covariance matrix of fuzzy relationships is built, and its eigenvalues and eigenvectors are calculated. Furthermore, fuzzy rules are optimized according to the cumulative of the eigenvalues, and forecasting results for real time series are obtained. Finally, the experiment results show the feasibility of the proposed algorithm.

**Keywords:** fuzzy time series; stabilization; principal component analysis; generalized covariance matrix; fuzzy rule optimization

## 0 引言

所谓时间序列,就是一个有序的观测值序列,通常是按照时间观测的,其基本特点是每一个序列包含了产生该序列的历史行为的全部信息<sup>[1]</sup>.传统的时间序列模型在许多预测方面都有成功的应用,但对于解决数据是语言值或是不完整、不确定的预测问题还存在许多不足.1993年,基于模糊集理论, Song等<sup>[2-4]</sup>首次利用模糊时间序列模型进行预测,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)预测和去模糊化.在实际应用中,模糊规则的建立对预测结果有着非常重要的影响,因此,人们对步骤3)进行了

大量的研究.文献[5]提出了根据模糊逻辑关系组来建立模糊规则;文献[6-7]提出了利用粗糙集理论来处理模糊逻辑关系;文献[8]提出了权重趋势的模糊时间序列模型;文献[9]利用实际数据和区间中点的距离差来建立权重模型;文献[10]提出了建立权重模型时采用的主要模糊关系和次要模糊关系;文献[11]提出了一种基于规则归纳的模糊时间序列模型;文献[12]提出了一种基于模糊关系和相似性测度的模糊时间序列模型;文献[13]提出了一种基于证据理论的模糊时间序列预测模型;文献[14]提出了一种基于自相关函数的模糊时间序列优化算法.众所周知,时间序列预测方法的基本思想是:预测一个现象的未

收稿日期: 2017-05-17; 修回日期: 2017-06-17.

基金项目: 国家自然科学基金项目(11571056).

责任编委: 张庆灵.

作者简介: 陈刚(1964—),男,教授,博士后,从事时间序列分析、数据挖掘与信息提取等研究;丁慧玲(1992—),女,硕士生,从事数据处理与信息提取的研究.

<sup>†</sup>通讯作者. E-mail: chengang@dmlu.edu.cn

来变化时,可利用该现象的过去行为来预测未来.即通过时间序列的历史数据揭示现象随时间变化的规律,将这种规律延伸到未来,从而对该现象的未来作出预测<sup>[15]</sup>.目前为止,经典时间序列所讨论的都是平稳过程,但生活中的许多应用,其时间序列过程是非平稳的,尤其是在自然灾害、工程控制、经济和金融等领域.而序列平稳性是时间序列建模的重要前提<sup>[16]</sup>,为了满足建模要求和确保预测的准确性,传统时间序列在进行建模前,要先对其进行平稳性检验及必要的变换,将非平稳信号平稳化,以达到更加精确的预测结果.

基于上述理由,本文针对如何优化模糊规则提出一种基于主成分分析的模糊时间序列模型的平稳化算法.在对模糊化的时间序列进行分析时,先对非平稳序列进行平稳化处理,再结合主成分分析法对模糊时间序列进行预测,在优化模糊规则的同时提高预测的准确性.最后本文进行实例研究,在表明主成分分析法对处理时间序列问题可行的同时,验证平稳化算法的有效性.

## 1 相关知识

### 1.1 模糊时间序列简介<sup>[2]</sup>

将论域 $U$ 划分为 $n$ 个子区间,即 $U = \{u_1 \cdots u_n\}$ ,则模糊集合为

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

其中: $f(A_i)(\cdot)$ 为模糊集合 $A_i$ 的隶属函数, $f(A_i)(\cdot) : U \rightarrow [0, 1]$ ;  $u_k$ 为模糊集合 $A_i$ 的一个因素, $f(A_i)(u_k)$ 对应模糊集合 $A_i$ 的隶属度, $f(A_i)(u_k) \in [0, 1]$ ,  $k = 1, 2, \cdots, n$ .

**定义1** 假设 $F(t)$ 仅由 $F(t-1)$ 确定, $F(t)$ 可表示为

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

称式(1)为 $F(t)$ 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.

**定义2** 设 $F(t-1) = A_i$ ,  $F(t) = A_j$ ,则一个模糊逻辑关系可以定义为 $A_i \rightarrow A_j$ ,其中 $A_i$ 为模糊关系的前件, $A_j$ 为模糊关系的后件.

### 1.2 主成分分析简介<sup>[1,15]</sup>

主成分分析方法是一种将相关的多个指标简化为少数几个不相关的综合指标的多元统计分析方法.主要步骤如下:

1) 观测样本矩阵的构建.设某评价指标体系有 $n$ 个评价指标,原始数据矩阵可以用 $n \times m$ 阶的矩阵 $X$ 表示,即

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \quad (2)$$

2) 样本协方差矩阵的构建.根据原始数据矩阵可定义样本协方差矩阵

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} = (c_{ij})_{n \times n}. \quad (3)$$

其中: $c_{ij} = \frac{1}{m-1} \sum_{k=1}^m (X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j)$ ,  $i, j = 1, 2, \cdots, n$ ,  $\bar{X}_i$ 为 $X_i$ 的均值, $\bar{X}_i = \frac{1}{m} \sum_{k=1}^m X_{ik}$ .

3) 对协方差阵 $C$ 进行特征分解,有

$$Q^T C Q = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}. \quad (4)$$

其中: $\lambda_i (i = 1, 2, \cdots, n)$ 为 $C$ 的特征值, $Q$ 为特征向量组成的正交阵.

4) 计算主成分的贡献率,确定主成分的个数.按照下式计算主成分的累积方差贡献率 $\rho$ ,确定主成分的个数:

$$\rho = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^n \lambda_i}. \quad (5)$$

一般取满足 $\rho \geq 85\%$ 的前 $m$ 个主成分.

### 1.3 非平稳时间序列的平稳化

迄今为止,所讨论的时间序列过程都是平稳过程,但是许多应用时间序列过程是非平稳的,这些非平稳时间序列可能有随时间变化的均值 $\mu_t$ 、二阶距(如方差 $\sigma_t^2$ ),或者二者皆有<sup>[16]</sup>.对于均值非平稳序列,通常采用的方法是差分法.一般做法是通过差分算子 $\Delta = (1 - B)^d$ (通常 $d = 1$ 或者 $2$ ),一个均值非平稳序列可以转化为一个平稳序列.对于方差非平稳序列,既可以由均值平稳的过程导致其方差及自协方差非平稳,也可以由均值非平稳的过程导致其方差和自协方差的非平稳.为了克服这个问题,需要适当地进行方差稳定变换.一般地,可以使用由Box和Cox<sup>[17]</sup>引入对数变换 $T(Z_t) = \ln(Z_t)$ .通常,如果需要做方差稳定变换,则必须在任何其他分析之前进行.

## 2 模糊关系的广义协方差矩阵构造方法

作为主成分分析的关键步骤,协方差矩阵的构建对主成分的提取有着非常重要的影响,而主成分的提取对后续的预测效果又有着至关重要的影响.本文从贴进度的角度提出模糊关系的广义协方差矩阵构造方法<sup>[18]</sup>.

设一阶模糊时间序列模型中,模糊规则库包含以下单输入单输出的模糊规则:  $A_p \rightarrow A_q, A_k \rightarrow A_j, A_p \rightarrow A_j, \dots, A_m \rightarrow A_n$  按下式计算得到由模糊规则形成的模糊关系矩阵  $R_i (i = 1, 2, \dots, n)$ :

$$R = (r_{ij}) = A_k^T \times A_q. \quad (6)$$

其中:  $r_{ij} = \min(A_{ki}, A_{qj}), A_{ki}$  为  $A_k$  的第  $i$  行,  $A_{qj}$  为  $A_q$  的第  $j$  列. 计算得到模糊关系矩阵如下:

$$R_i = \begin{bmatrix} r_{11}^i & r_{12}^i & \dots & r_{1n}^i \\ r_{21}^i & r_{22}^i & \dots & r_{2n}^i \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^i & r_{n2}^i & \dots & r_{nn}^i \end{bmatrix}.$$

接下来计算模糊关系矩阵的贴进度. 通常情况下用距离来描述两个矩阵的相关程度,称该相关程度为矩阵的贴进度.贴进度的计算公式如下:

$$(R_i, R_j) = \begin{cases} 1, & i = j; \\ 1 - d(R_i, R_j), & i \neq j. \end{cases} \quad (7)$$

其中  $d(R_i, R_j)$  为矩阵  $R_i$  和  $R_j$  的距离. 计算距离的方法有许多种,本文采用常用的曼哈顿距离法.

$$d(X, Y) = \frac{\sum |x_{ij} - y_{ij}|}{n^2}. \quad (8)$$

其中:  $x_{ij}$  和  $y_{ij}$  为两个矩阵的对应元素,  $n^2$  为矩阵元素的个数.

通过上述计算,可以得到模糊关系矩阵  $R_1, R_2, \dots, R_n$  的协方差矩阵为

$$C = \begin{bmatrix} 1 & \dots & 1 - d(R_1, R_n) \\ 1 - d(R_2, R_1) & \dots & 1 - d(R_2, R_n) \\ \vdots & \ddots & \vdots \\ 1 - d(R_n, R_1) & \dots & 1 \end{bmatrix}. \quad (9)$$

## 3 新算法描述

本文提出了一种基于主成分分析法的模糊时间序列模型的平稳化算法,步骤如下.

**Step 1:** 时间序列的平稳化. 若时间序列过程平稳,则直接进行 Step 2; 若时间序列过程非平稳,则按照第 1.3 节的方法将时间序列平稳化.

**Step 2:** 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. 基于平均划分法将论域划分为等长度的区间  $u_1, u_2, \dots, u_m$ .

**Step 3:** 数据模糊化.

定义模糊集合为

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n},$$

通过区间对数据进行模糊化. 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  $u_k, 1 \leq k \leq n$ , 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为  $A_k$ . 因而,样本数据按上述模糊化后得到相应的模糊时间序列.

**Step 4:** 构建模糊规则. 设第  $i$  个数据的模糊化结果为  $A_i$ , 第  $i + 1$  个数据为  $A_j$ , 模糊规则可以表示为  $A_i \rightarrow A_j$ .

**Step 5:** 构建模糊关系矩阵. 按照式(7)构建模糊关系矩阵.

**Step 6:** 构建模糊关系的广义协方差矩阵. 按照第 2 节介绍的方法构建广义协方差矩阵

$$C = \begin{bmatrix} 1 & c_{12} & \dots & c_{1n} \\ c_{21} & 1 & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & 1 \end{bmatrix}. \quad (10)$$

**Step 7:** 优化模糊规则. 根据主成分分析法对协方差矩阵进行主成分分析,得到  $m$  个主成分  $K_1, K_2, \dots, K_m$ , 原来的多个模糊规则可被优化为一个综合的模糊规则  $R = \bigcup_{i=1}^m K_i$ .

**Step 8:** 预测和去模糊化. 利用 Step 7 优化的模糊规则进行预测,得到预测值属于每个模糊集的隶属度

$$\begin{cases} F_1 = \frac{f_{11}}{u_1} + \frac{f_{12}}{u_2} + \dots + \frac{f_{1k}}{u_k}, \\ \vdots \\ F_t = \frac{f_{t1}}{u_1} + \frac{f_{t2}}{u_2} + \dots + \frac{f_{tk}}{u_k}. \end{cases}$$

根据如下公式去模糊化:

$$F(t) = \sum_{i=1}^k c_i \left( \frac{f_{ti}}{\sum_{i=1}^k f_{ti}} \right). \quad (11)$$

其中:  $F(t)$  为预测值,  $f_{ti}$  为隶属度,  $c_i$  为区间中心值.

## 4 算例分析

### 4.1 Alabama大学的入学人数

为了验证新算法的有效性,本文选取 1971~1992 年 Alabama 大学的入学人数作为样本数据进行比较研究.

**Step 1:** 时间序列的平稳化检验.

经 ADF 平稳化检验, Alabama 入学人数的时间序列非平稳, 因此先对时间序列进行平稳化处理。

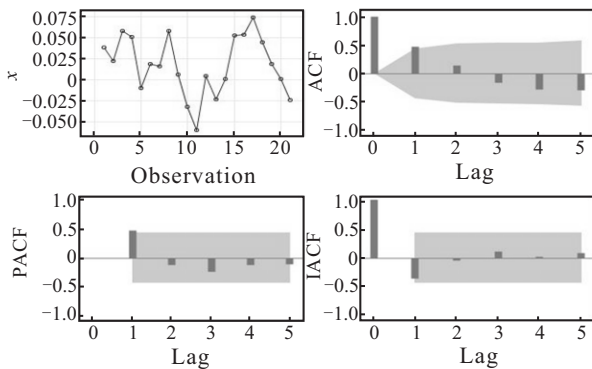
对该序列实施对数变换及一阶差分变换, 再对新序列进行 ADF 平稳化检验。

如图 1 所示,  $p = 0.0275 < 0.05$ , 自相关系数和偏相关系数除一阶之外都在 2 倍的标准差范围内波动, 得出该序列在对数和差分变换后为平稳序列。

|                        |           |
|------------------------|-----------|
| Name of variable = x   |           |
| Mean of working series | 0.017 558 |
| Standard deviation     | 0.034 263 |
| Number of observations | 21        |

| Augmented dickey-fuller unit root tests |      |           |         |       |         |      |
|---|------|-----------|---------|-------|---------|------|
| Type                                    | Lags | Rho       | Pr<Rho  | Tau   | Pr<Tau  | F    |
| Zero mean                               | 0    | -8.038 1  | 0.037 2 | -2.24 | 0.027 5 |      |
|   | 1    | -8.533 1  | 0.030 4 | -1.98 | 0.047 6 |      |
| Single mean                             | 0    | -9.683 1  | 0.096 3 | -2.30 | 0.180 2 | 2.75 |
|   | 1    | -11.956 1 | 0.042 4 | -2.12 | 0.240 0 | 2.31 |
| Trend                                   | 0    | -9.725 6  | 0.355 2 | -2.26 | 0.436 3 | 2.68 |
|   | 1    | -11.886 2 | 0.202 7 | -2.05 | 0.538 0 | 2.20 |

(a) The ARIMA procedure



(b) Trend and correlation analysis for x

图 1 变换后时间序列的 ADF 检验结果

Step 2: 定义论域和划分区间. 令  $D_{\min}$  和  $D_{\max}$  分别表示得到的平稳过程的最小值和最大值, 定义论域

$$U = [D_{\min} - D_1, D_{\max} + D_2],$$

其中  $D_1$  和  $D_2$  为适当的正数. 由 Alabama 入学人数可知

$$D_{\min} = -0.060 0, D_{\max} = 0.073 8.$$

令  $D_1 = 0.005, D_2 = 0.001 2$ , 因此, 得到论域  $U = [-0.065, 0.075]$ . 然后将论域等分成 7 个区间。

Step 3: 数据模糊化。

根据样本数据的大小, 将模糊集分别定义为  $A_1, A_2, \dots, A_7$ , 它们为入学人数模糊化后对应的模糊集, 有

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7},$$

⋮

$$A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7}.$$

之后建立一阶模糊时间序列模型, 获得的模糊规

则见表 1.

表 1 数据模糊化结果及对应的模糊规则

| 年份   | 模糊集   | 模糊规则                  | 年份   | 模糊集   | 模糊规则                  |
|------|-------|-----------------------|------|-------|-----------------------|
| 1971 | $A_6$ |                       | 1982 | $A_4$ | $A_1 \rightarrow A_4$ |
| 1972 | $A_6$ | $A_6 \rightarrow A_5$ | 1983 | $A_3$ | $A_4 \rightarrow A_3$ |
| 1973 | $A_7$ | $A_5 \rightarrow A_7$ | 1984 | $A_4$ | $A_3 \rightarrow A_4$ |
| 1974 | $A_6$ | $A_7 \rightarrow A_6$ | 1985 | $A_6$ | $A_4 \rightarrow A_6$ |
| 1975 | $A_3$ | $A_6 \rightarrow A_3$ | 1986 | $A_6$ | $A_6 \rightarrow A_6$ |
| 1976 | $A_5$ | $A_3 \rightarrow A_5$ | 1987 | $A_7$ | $A_6 \rightarrow A_7$ |
| 1977 | $A_5$ | $A_5 \rightarrow A_5$ | 1988 | $A_6$ | $A_7 \rightarrow A_6$ |
| 1978 | $A_7$ | $A_5 \rightarrow A_7$ | 1989 | $A_5$ | $A_6 \rightarrow A_5$ |
| 1979 | $A_4$ | $A_7 \rightarrow A_4$ | 1990 | $A_4$ | $A_5 \rightarrow A_4$ |
| 1980 | $A_2$ | $A_4 \rightarrow A_2$ | 1991 | $A_3$ | $A_4 \rightarrow A_3$ |
| 1981 | $A_1$ | $A_2 \rightarrow A_1$ |      |       |                       |

Step 4: 建立模糊规则。

由表 1 可得模糊规则 (重复的只记一次)

$$A_6 \rightarrow A_5, A_5 \rightarrow A_7, A_7 \rightarrow A_6, A_6 \rightarrow A_3, \\ A_3 \rightarrow A_5, A_5 \rightarrow A_5, A_7 \rightarrow A_4, A_4 \rightarrow A_2, \\ A_2 \rightarrow A_1, A_1 \rightarrow A_4, A_4 \rightarrow A_3, A_3 \rightarrow A_4, \\ A_4 \rightarrow A_6, A_6 \rightarrow A_6, A_6 \rightarrow A_7, A_5 \rightarrow A_4.$$

Step 5: 构建模糊关系矩阵。

按式 (6) 计算模糊关系矩阵

$$R_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 1 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0 \end{bmatrix},$$

⋮

$$R_{16} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 1 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Step 6: 构建模糊关系的广义协方差矩阵。

根据曼哈顿距离法 (8) 构建模糊关系矩阵  $R_1$ ,

$R_2, \dots, R_{16}$  的广义协方差矩阵

$$C = \begin{bmatrix} 1 & 0.8673 & \cdots & 0.8776 \\ 0.8673 & 1 & \cdots & 0.8265 \\ \vdots & \vdots & \ddots & \vdots \\ 0.8776 & 0.8265 & \cdots & 1 \end{bmatrix}_{16 \times 16}.$$

Step 7: 优化模糊规则.

对得到的协方差矩阵  $C$  进行主成分分析, 计算广义协方差矩阵  $C$  的特征值和特征向量, 特征值为 0.031 7, 0.034 2,  $\dots$ , 13.727 0.

计算特征值的累计贡献率

$$\sum_{i=1}^3 \lambda_i / \sum_{i=1}^{16} \lambda_i = 85.48\% > 85\%.$$

特征值对应的特征向量为

$$\begin{aligned} \alpha_1 &= (0.252\ 1, 0.254\ 3, \dots, 0.251\ 3), \\ \alpha_2 &= (0.295\ 2, 0.211\ 9, \dots, -0.120\ 4), \\ \alpha_3 &= (-0.234\ 9, 0.212\ 2, \dots, -0.244\ 7). \end{aligned}$$

主成分可以表示为

$$\begin{aligned} K_1 &= 0.252\ 1R_1 + 0.254\ 3R_2 + \dots + 0.251\ 3R_{16}, \\ K_2 &= 0.295\ 2R_1 + 0.211\ 9R_2 + \dots + \\ &\quad (-0.120\ 4)R_{16}, \\ K_3 &= (-0.234\ 9)R_1 + 0.212\ 2R_2 + \dots + \\ &\quad (-0.244\ 7)R_{16}. \end{aligned}$$

则优化的模糊规则可以表示为

$$R = \bigcup_{i=1}^3 K_i = \begin{bmatrix} 0.123\ 9 & 0.123\ 9 & \dots & 0 \\ 0.247\ 9 & 0.123\ 9 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0.122\ 8 & \dots & 0.498\ 6 \end{bmatrix}_{7 \times 7}.$$

Step 8: 预测和去模糊化.

利用优化后的模糊规则进行预测, 具体方法是按照式 (11) 去模糊化, 然后采用逆差分变换  $Z_t^{*'} = Z_{t-1}' + X_t$  及逆对数变换  $Z_t^* = e^{Z_t^{*'}}$  得到序列  $Z_t^*$

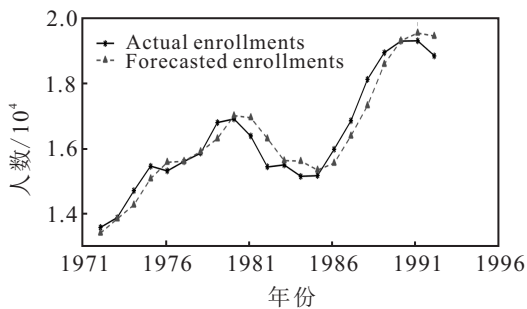


图2 本文方法下预测人数与实际入学人数

表2 不同方法下预测精度的比较

| 方法                    | MSE     | RMSE |
|-----------------------|---------|------|
| Song <sup>[2-4]</sup> | 423 027 | 650  |
| Huang <sup>[6]</sup>  | 226 611 | 476  |
| Chen <sup>[7]</sup>   | 261 142 | 511  |
| Cheng <sup>[8]</sup>  | 228 918 | 478  |
| Lee <sup>[9]</sup>    | 226 576 | 476  |
| 本文                    | 175 710 | 419  |

(133 89, 138 27,  $\dots$ , 194 84). 预报值与真值的比较如图2所示. 本文将它与目前已有方法进行对比, 结果如表2所示, 可知本文算法的预测精度优于其他方法.

### 4.2 TAIFEX数据

使用本文算法, 选取 TAIFEX 在 1998-08-03~1998-09-30 期间的 47 个观测值<sup>[19]</sup> 进行比较研究. 对数据进行 ADF 平稳性检验, 由于检验结果非平稳, 因此先对该时间序列进行平稳化处理, 得到平稳的时间序列; 然后, 将所得序列使用主成分分析法 (PCA) 进行规则优化; 最后, 进行预测和去模糊化, 通过逆差分变换和逆对数变换得到预测数据. 预报值与真值的比较如图3所示.

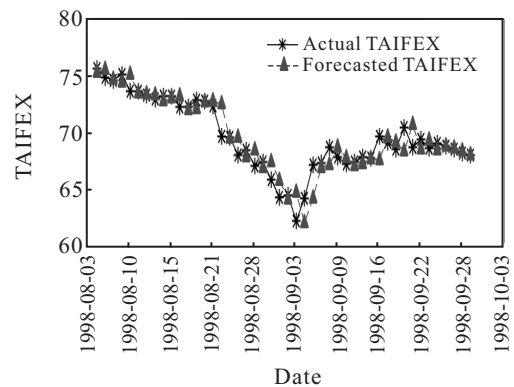


图3 本文方法下的预测值及 TAIFEX 实际值

由表3可以看出, 在不同方法下的 MSE 和 RMSE 的比较来看, 本文方法依然有效.

表3 不同方法下预测精度的比较

| 方法                    | MSE    | RMSE |
|-----------------------|--------|------|
| Chen <sup>[20]</sup>  | 10 652 | 103  |
| Huang <sup>[21]</sup> | 8 505  | 92   |
| Huang <sup>[22]</sup> | 6 397  | 80   |
| Lee <sup>[23]</sup>   | 1 399  | 37   |
| Lee <sup>[24]</sup>   | 105    | 10   |
| 本文                    | 7 397  | 816  |

## 5 结论

本文对以往的算法进行了改进, 取得了预期效果, 但同时具有局限性. 它的局限性在于主成分分析仅适用于协方差矩阵为正定的情况, 而实际应用中, 因素间的协方差矩阵也存在非正定的情况. 今后将从协方差矩阵的构造方法或协方差矩阵的正定性等问题上继续深入研究.

### 参考文献 (References)

[1] 卢世坤, 李夕海, 牛超, 等. 时间序列的非线性非平稳特性研究综述[J]. 中国地理科学, 2012, 9(6): 304-314. (Lu Shi-kun, Li XI-hai, Niu chao, et al. A review on the research of the nonstationarity of time series[J]. Chinese

- Geophysical Society, 2012, 9(6): 304-314.)
- [2] Song Q, Chissom B S. Fuzzy time series and its models[J]. Fuzzy Sets and Systems, 1993, 54(1): 269-277.
- [3] Song Q, Chissom B S. Forecasting enrollments with fuzzy time series: Part I[J]. Fuzzy Sets and Systems, 1993, 54(1): 1-10.
- [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] 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.
- [6] Chen S M, Hwang J R. Temperature prediction using fuzzy time series[J]. IEEE Trans on Systems, 2000, 30(2): 263-275.
- [7] Chen S M. Forecasting enrollments based on high-order fuzzy time series[J]. Cybernetics and Systems, 2002, 33(1): 1-16.
- [8] 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, 4232(III): 469-477.
- [9] Lee M I I, Efendi R, Ismail Z. Modified weighted for enrollment forecasting based on fuzzy time series[J]. Matematika, 2009, 25(1): 67-78.
- [10] Qiu W R, Liu X D, Wang L D. Forecasting in time series based on generalized fuzzy logical relationship[J]. Int J of Research and Surveys, 2010, 4(5): 1431-1437.
- [11] Chen Ching-hsue, Yang Jun-he. Rough-set rule induction to build fuzzy time series model in forecasting stock price[C]. The 2nd Int Conf on Fuzzy Systems and Knowledge Discovery(FSKD). Zhangjiajie, 2015: 456-463.
- [12] Cheng Shou-hsiung, Chenc Shyi-ming, Jian Wen-shan. A novel fuzzy time series forecasting method based on fuzzy logical relationships and similarity measures[C]. 2015 IEEE Int Conf on Systems, Man, and Cybernetics. Hong Kong: IEEE Press, 2015: 1175-1184.
- [13] 邱望仁, 刘晓东. 基于证据理论的模糊时间序列模型[J]. 控制与决策, 2012, 27(1): 99-103.  
(Qiu W R, Liu X D. Fuzzy time series model for forecasting based on Dempster-Shafer theory[J]. Control and Decision, 2012, 27(1): 99-103.)
- [14] 陈刚, 王鹏飞, 李金玲. 基于自相关函数的模糊时间序列优化算法[J]. 控制与决策, 2014, 29(2): 251-256.  
(Chen G, Wang P F, Li J L. Optimization algorithm for fuzzy time series model based on autocorrelation function[J]. Control and Decision, 2014, 29(2): 251-256.)
- [15] 王丽娜, 肖冬荣. 基于 ARMA 模型的经济非平稳时间序列的预测分析[J]. 武汉理工大学学报, 2004, 2(1): 133-135.  
(Wang L N, Xiao D R. Analysis of non-steady time-series forecast for economy based on ARMA model[J]. J of Wuhan University of Technology, 2004, 2(1): 133-135.)
- [16] William W S Wei. Time Series analysis-univariate and multivariate methods[M]. Beijing: China Renmin University Press, 2009: 65-81.
- [17] Box G E P, Cox D R. An analysis of transformations[J]. J of the Royal Statistical Society, Series B, 1964, 26(2): 211-243.
- [18] Zhang Xianda. Matrix analysis and applications[M]. Beijing: Tsinghua University Press, 2004: 186-210.
- [19] 赵柯柯. 模糊时间序列的一类混合预测模型研究[D]. 大连: 大连理工大学信息与通信工程学院, 2014: 43-45.  
(Zhao K K. Reserch on one class of hybrid forecasting models for fuzzy time series[D]. Dalian: School of Information and Communication Engineering, University of Technology, 2014: 43-45.)
- [20] Chen S M. Forecasting enrollments based on fuzzy time series[J]. Fuzzy Sets and Systems, 1996, 81(3): 311-319.
- [21] Huarng K H. Effective lengths of intervals to improve forecasting in fuzzy time series[J]. Fuzzy Sets and Systems, 2001, 123(3): 375-378.
- [22] Huarng K H. Heuristic models of fuzzy time series for forecasting[J]. Fuzzy Sets and Systems, 2001, 123(3): 369-386.
- [23] Lee L W, Wang L H, Chen S M. Temperature prediction and TAIEX forecasting based on fuzzy logical relationships and genetic algorithms[J]. Expert Systems with Applications, 2007, 33(6): 539-550.
- [24] Lee L W, Wang L H, Chen S M. Temperature prediction and TAIEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques[J]. Expert Systems with Applications, 2008, 34(2): 328-336.

(责任编辑: 孙艺红)