文章编号:1008-1542(2010)06-0497-04

基于绝对误差的线性组合预测研究

韩　东[1],许葆华[1],马献果[2]

(1.军械工程学院导弹工程系,河北石家庄　050003;2.河北科技大学信息科学与工程学院,河北石家庄　050018)

摘　要:针对组合预测效果评价中存在着把拟合精度和预测精度相混淆的问题,阐述了区分样本区间和预测区间的组合预测精度评价方法,并在预测区间上对2种常用的基于绝对误差的线性组合预测方法的预测精度进行了评价,对评价结果进行实例和理论分析。在此基础上,提出一种新的基于绝对误差的线性组合预测方法——基于预测模型有效性的线性组合预测方法。

关键词:绝对误差;组合预测;有效性

中图分类号:O221　　**文献标志码**:A

Research on linear combination forecasting based on absolute error

HAN Dong[1], XU Bao-hua[1], MA Xian-guo[2]

(1. Department of Missile Engineering, Ordnance Engineering College, Shijiazhuang Hebei 050003, China; 2. College of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang Hebei 050018, China)

Abstract: In order to solve the problem of mixing fitting precision with forecasting precision in combination forecasting, the precision evaluation method for distinguishing forecasting interval from sample interval is expatiated. The forecasting precision of two common linear combination forecasting models based on absolute error are evaluated on forecasting interval, and the evaluating result is validated and analyzed. Based on above analysis, a new linear combination forecasting method based on the forecasting model's validity is proposed.

Key words: absolute error; combination forecasting; validity

目前,在采用组合预测方法对预测精度进行评价时,部分研究存在着这样的做法:首先,基于某种准则,利用前 N 个时刻的真实值 $\{x_t\}(t=1,2,\cdots,N)$ 和各个单项预测方法的预测值 $\{x_{it}\}(i=1,2,\cdots,m;t=1,2,\cdots,N)$ 求出加权系数 l_i;然后,将 l_i 代入前 N 个时刻的单项预测值 $\{x_{it}\}$,得出组合预测值 $\hat{x}_t=\sum l_i x_{it}$;最后,根据各种误差评价指标对预测效果进行评价[1-6]。笔者认为,这种做法混淆了拟合和预测的概念,用拟合精度代替了预测精度。对预测精度进行评价应该是将 l_i 代入 N 时刻以后的单项预测值 $\{x_{it}\}(t=N+1,N+2,\cdots,n)$,才能得到真正意义上的组合预测值,从而再根据 N 时刻以后的真实值 $\{x_t\}(t=N+1,N+2,\cdots,n)$ 对预测精度进行评价。

笔者利用上述预测精度评价方法对2种常用的基于绝对误差的线性组合预测方法(以绝对误差的2-范数达到最小、以绝对误差的1-范数达到最小的组合预测方法)的预测效果进行重新评价,对其评价结果进行实例和理论分析。在此基础上,提出一种新的基于绝对误差的线性组合预测方法。实例验证表明:该方法的预测精度优于前述常用的组合预测方法。

收稿日期:2010-06-18;责任编辑:张　军

基金项目:河北省自然科学基金资助项目(E2009001431)

作者简介:韩　东(1972-),男,河北安新人,副教授,博士研究生,主要从事状态监测、故障诊断与预测方面的研究。

1 问题描述

设对同一预测对象的某个指标序列$\{x_t\}(t=1,2,\cdots,N,N+1,\cdots,n)$有$m$种单项预测方法对其进行预测，第$i$种单项预测方法在第$t$时刻的预测值为$\{x_{it}\}(i=1,2,\cdots,m;t=1,2,\cdots,N,N+1,\cdots,n)$。

笔者将$t=1,2,\cdots,N$定义为样本区间，$t=N+1,N+2,\cdots,n$定义为预测区间。在样本区间上建立以绝对误差的2-范数达到最小和以绝对误差的1-范数达到最小2种常用的线性组合预测模型，在预测区间上对上述2种组合预测模型的预测精度进行评价。

1.1 基于样本区间的线性组合预测方法

设$l_1,l_2,\cdots,l_m$分别为m种单项预测方法的加权系数，且满足$\sum_{i=1}^{m}l_i=1,l_i\geqslant 0,\hat{x}_t=\sum_{i=1}^{m}l_ix_{it},t=1,2,\cdots,N$为$\{x_t\}$的组合预测值，称$e_t=(x_t-\hat{x}_t)$为组合预测在第$t$时刻的绝对误差。

设J_1表示组合预测绝对误差的2-范数，则以绝对误差的2-范数达到最小的线性组合预测模型为

$$\min J_1=\sum_{t=1}^{N}e_t^2,\quad \text{s.t.}\quad \begin{cases}\sum_{i=1}^{m}l_i=1,\\ l_i\geqslant 0,i=1,2,\cdots,m。\end{cases} \tag{1}$$

设J_2表示组合预测绝对误差的1-范数，则以绝对误差的1-范数达到最小的线性组合预测模型为

$$\min J_2=\sum_{t=1}^{N}|e_t|,\quad \text{s.t.}\quad \begin{cases}\sum_{i=1}^{m}l_i=1,\\ l_i\geqslant 0,i=1,2,\cdots,m。\end{cases} \tag{2}$$

1.2 基于预测区间的预测精度评价指标

笔者采用以下几种形式的误差指标对组合预测的精度进行评价[7]。

1）平方和误差（SSE）

$$\mathrm{SSE}=\sum_{t=N+1}^{n}(\hat{x}_t-x_t)^2。 \tag{3}$$

2）均方根误差（RMSE）

$$\mathrm{RMSE}=\sqrt{\frac{1}{n-N}\sum_{t=N+1}^{n}(\hat{x}_t-x_t)^2}。 \tag{4}$$

3）平均绝对误差（MAE）

$$\mathrm{MAE}=\frac{1}{n-N}\sum_{t=N+1}^{n}|\hat{x}_t-x_t|。 \tag{5}$$

4）平均相对误差（ARE）

$$\mathrm{ARE}=\frac{1}{n-N}\sum_{t=N+1}^{n}\left|\frac{\hat{x}_t-x_t}{x_t}\right|。 \tag{6}$$

5）均方根相对误差（RMSRE）

$$\mathrm{RMSRE}=\sqrt{\frac{1}{n-N}\sum_{t=N+1}^{n}\left(\frac{\hat{x}_t-x_t}{x_t}\right)^2}。 \tag{7}$$

2 常用线性组合预测精度评价

2.1 实例分析

笔者应用文献[7]的数据进行实例分析。指标实际值和各单项预测方法预测值数据如表1所示。

将例1的前7组数据、例2的前8组数据、例3的前9组数据作为样本，建立上述2个组合预测模型，并对这3个例子的后3组数据进行预测，计算其预测误差指标值，结果如表2所示。在表2中，方法1表示以绝对误差的2-范数达到最小的组合预测方法，方法2表示以绝对误差的1-范数达到最小的组合预测方法。可以看出，组合方法1预测效果明显优于组合方法2。

表 1　指标实际值和各单项预测方法预测值

Tab. 1　Actual value and forecasting value of single prognostics method

所测点	例 1			例 2			例 3		
	x_t	x_{1t}	x_{2t}	x_t	x_{1t}	x_{2t}	x_t	x_{1t}	x_{2t}
1	14.90	10.00	12.00	57.00	54.52	64.68	11.49	18.47	10.03
2	18.60	14.90	15.48	65.40	62.89	64.74	13.06	14.54	11.23
3	22.20	23.30	18.95	75.40	72.54	68.72	15.34	12.84	15.24
4	17.60	26.10	22.43	82.50	83.67	76.61	20.58	13.38	18.67
5	19.60	17.50	25.90	92.80	96.51	88.42	23.28	16.15	27.78
6	24.00	20.20	29.38	102.70	111.32	104.15	26.46	21.16	26.36
7	31.60	26.40	32.85	119.50	128.41	123.79	27.33	28.40	29.67
8	43.70	36.80	36.33	143.80	148.11	147.35	34.22	37.87	27.40
9	37.00	52.50	39.80	169.70	170.84	174.82	40.19	49.58	42.73
10	47.20	38.50	43.28	201.00	197.06	206.21	53.37	63.53	47.36
11				251.20	227.31	241.51	77.79	79.00	71.00
12							100.63	98.12	109.32

表 2　预测效果评价指标

Tab. 2　Forecasting effect evaluation

项目	组合预测方法	加权系数	SSE	RMSE	MAE	ARE	RMSRE
例 1	方法 1	$l_1 = 0.4253, l_2 = 0.5747$	154.12	7.17	7.11	0.1706	0.1751
	方法 2	$l_1 = 0.5861, l_2 = 0.4139$	200.43	8.17	8.02	0.1939	0.2027
例 2	方法 1	$l_1 = 0.4726, l_2 = 0.5274$	280.25	9.67	6.84	0.0296	0.0394
	方法 2	$l_1 = 0.5414, l_2 = 0.4586$	310.85	10.18	6.87	0.0293	0.0412
例 3	方法 1	$l_1 = 0.2677, l_2 = 0.7323$	56.83	4.35	4.01	0.0493	0.0509
	方法 2	$l_1 = 0.1730, l_2 = 0.8270$	85.15	5.33	5.12	0.0656	0.0657

组合预测模型的精度信息包含在其绝对误差序列$\{e_t\}(t=1,2,\cdots,N,N+1,\cdots,n)$中。其预测精度可以根据拟合绝对误差序列$\{e_t\}(t=1,2,\cdots,N)$作出估计。由于无法对预测绝对误差的正负号作出估计，因此在对预测精度进行估计时往往不考虑绝对误差的符号，而只考虑绝对误差的绝对值，即根据序列$\{|e_t|\}$进行估计。依据统计学观点，该序列的特征可以用其均值和标准差来刻画，均值反映了模型的准确性，标准差反映了模型的稳定性。

下面在建立组合预测模型的样本区间上进行分析。各单项预测方法和组合预测方法在样本区间上的均值 M 和标准差 σ 如表 3 所示。

从表 3 中可以看出，2 种组合方法的均值 M 明显小于各单项预测方法，虽然组合方法 1 的均值 M 略大于组合方法 2，但其标准差明显小于组合方法 2。这说明组合方法 2 在建立模型时只注重模型的准确性，而组合方法 1 在建立模型时则兼顾了模型的准确性和稳定性两方面的因素，因此，其预测效果要优于组合方法 2。

以上只是通过实例数据从直观上解释了组合方法 1 优于组合方法 2 的原因，下面从理论上作进一步的分析。

表 3　样本区间上 $|e_t|$ 的均值与标准差

Tab. 3　Average and standard deviation of $|e_t|$ in swatch section

预测方法	例 1		例 2		例 3	
	M	σ	M	σ	M	σ
单项方法 1	4.19	2.22	4.32	2.71	4.97	2.76
单项方法 2	3.86	1.60	4.32	2.28	2.40	2.00
组合方法 1	2.94	1.66	3.45	1.82	2.12	1.34
组合方法 2	2.73	2.20	3.33	2.09	2.10	1.51

2.2 理论分析

组合预测方法在样本区间上的绝对误差为 $|e_t|$，其绝对值的标准差 σ 可以表示为

$$\sigma = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(|e_t| - \frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2}, \tag{8}$$

那么，标准差的平方也就是方差为

$$\sigma^2 = \frac{1}{N}\sum_{t=1}^{N}\left(|e_t| - \frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2, \tag{9}$$

展开上式，得

$$\begin{aligned}\sigma^2 = &\frac{1}{N}\sum_{t=1}^{N}\left(|e_t|^2 - 2|e_t|\frac{1}{N}\sum_{t=1}^{N}|e_t| + \left(\frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2\right) = \\ &\frac{1}{N}\sum_{t=1}^{N}|e_t|^2 - 2\left(\frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2 + \left(\frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2 = \\ &\frac{1}{N}\sum_{t=1}^{N}|e_t|^2 - \left(\frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2 = \\ &\frac{1}{N}J_1 - \left(\frac{1}{N}J_2\right)^2,\end{aligned} \tag{10}$$

可以得到：

$$J_1 = N\left(\left(\frac{1}{N}J_2\right)^2 + \sigma^2\right)。 \tag{11}$$

从式(11)可以看出，组合方法 1 比组合方法 2 多考虑了一项反映模型稳定性的指标 —— 绝对误差的绝对值标准差 σ，因此，其预测效果优于组合方法 2，结论与实例分析一致。

3 基于预测模型有效性的线性组合预测方法

预测模型的有效性是指模型的准确性和稳定性，可以用绝对误差的绝对值均值和标准差来刻画。虽然组合方法 1 同时考虑了模型的准确性和稳定性，但是它对绝对误差的绝对值均值和标准差都进行了平方运算，这样会对 2 个指标产生“放大”或“缩小”效应[7-8]。笔者将其平方运算去掉，提出一种新的基于绝对误差的线性组合预测方法 —— 基于预测模型有效性的线性组合预测方法。

设 J_3 表示组合预测绝对误差的绝对值均值和标准差之和，则基于预测模型有效性的线性组合预测模型为

$$\min J_3 = \frac{1}{N}\sum_{t=1}^{N}|e_t| + \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(|e_t| - \frac{1}{N}\sum_{t=1}^{N}|e_t|\right)^2}, \quad \text{s.t.} \quad \begin{cases}\sum_{i=1}^{m} l_i = 1, \\ l_i \geqslant 0, i = 1,2,\cdots,m。\end{cases} \tag{12}$$

4 实例验证

运用上述新方法对前述实例进行分析，结果如表 4 所示。

表 4 新方法预测效果评价指标

Tab.4 Forecasting effect evaluation of new method

项目	加权系数	SSE	RMSE	MAE	ARE	RMSRE
例 1	$l_1 = 0.3675, l_2 = 0.6325$	139.78	6.83	6.78	0.1623	0.1656
例 2	$l_1 = 0.3733, l_2 = 0.6267$	241.17	8.97	6.81	0.0300	0.0370
例 3	$l_1 = 0.2704, l_2 = 0.7296$	56.14	4.33	3.98	0.0488	0.0505

从表 4 中可以看出，新方法的各项误差指标均小于组合方法 1，说明笔者提出的新方法是有效的，并且优于上述 2 种常用的方法。

（下转第 522 页）

或噪声强度不仅能提高系统输出幅度增益,而且在某个位置输出幅度增益会达到最大值,即出现随机共振现象,这样会使夹杂在噪声中的被测信号突现出来,从而实现信号的检测[10-11]。当然,笔者虽然只对简单乘性高斯白噪声成功消噪处理,但对于一类乘性噪声[12],比如常常碰到的乘性色噪声,所研究方法同样适用。另外,对于复杂噪声背景下的欠阻尼二阶线性系统是否也存在参数调节随机共振现象,是否也能有效消噪,还有待于进一步的研究。

参考文献:

[1] 郭 锋,周玉荣,蒋世奇,等.具有乘性噪声的线性系统的随机共振[J].电路与系统学报(Circuits and Systems Technology),2008,13(2):6-8.

[2] 郭 锋,周玉荣,蒋世奇,等.具有乘性噪声的线性振荡器的随机共振[J].电子科技大学学报(Journal of University of Electronic Science and Technology),2008,37(1):77-80.

[3] 周登荣,郭 锋,鲁加国,等.乘性与信号调制噪声在线性模型中的随机共振[J].四川师范大学学报(自然科学版)(Sichuan Normal University (Natural Science)),2010,33(2):270-272.

[4] 周登荣,周玉荣.乘性噪声作用下线性模型中的随机共振[J].重庆师范大学学报(自然科学版)(Journal of Chongqing Normal University (Natural Science)),2008,25(3):70-73.

[5] GUO Feng,ZHOU Yu-rong,JIANG Shi-qi,et al. Stochastic resonance in a bias linear system with multiplicative and additive noise[J]. Chinese Physics,2006,15(5):947-952.

[6] ROUSSEAU D, ANAND G V,CHAPEAU-BLONDEAU F. Noise-enhanced nonlinear detector to improve signal detection in non-Gaussian noise[J]. Signal Processing,2006,86(11):3 456-3 465.

[7] GITTERMAN M. Underdamped oscillator with fluctuating damping[J]. Physica A,2004,37(22):5 729-5 736.

[8] 周玉荣.耦合白噪声作用下线性系统的随机共振[J].四川文理学院学报(自然科学版)(Journal of Sichuan University of Arts and Science (Natural Science)),2007,17(2):23-26.

[9] 周登荣,周玉荣.信号调制白噪声下线性系统的随机共振[J].振动与冲击(Vibration and Shock),2008,27(7):138-140.

[10] 蒋世奇,古天祥.随机振幅周期信号驱动的一阶线性系统的随机共振[J].电子测量与仪器学报(Journal of Electronic Measurement and Instrument),2008,22(1):104-108.

[11] 孙水发,郑 胜,姚陆锋,等.随机共振用于非周期信号处理的仿真[J].海军工程大学学报(Journal of Naval University of Engineering),2007,19(5):6-9.

[12] 周玉荣,郭 锋,蒋世奇,等.色噪声作用下线性系统的随机共振[J].电子科技大学学报(Journal of University of Electronic Science and Technology),2008,37(2):232-234.

(上接第 500 页)

5 结 语

在预测区间上对 2 种常用的基于绝对误差的线性组合预测方法重新进行了预测效果评价,得出了以绝对误差的 2- 范数达到最小的方法优于以绝对误差的 1- 范数达到最小的方法的结论,通过实例和理论分析找出了原因,并认为该原因才是以绝对误差的 2- 范数达到最小的线性组合预测方法目前在各领域实际预测问题中应用最为广泛的主要原因。受此启发,提出了一种新的基于预测模型有效性的线性组合预测方法,通过实例验证表明了该方法优于前述的 2 种常用方法。

参考文献:

[1] 岳艳春,黄廷祝.误差倒数变权组合预测方法[J].电子科技大学学报(Journal of University of Electronic Science and Technology of China),2007,36(2):349-351.

[2] 谢庆华,梁 剑,张 琦.基于统计粗集的航空发动机维修成本组合预测模型[J].兵工学报(Acta Armamentarii),2006,27(5):857-861.

[3] 王秋萍,刘素兵,王晓峰,等.图书出版量的优化组合预测模型及其应用[J].计算机工程与应用(Computer Engineering and Applications),2008,44(12):246-248.

[4] 胡 彦,李秀美,陈华友.基于 IOWA 算子的税收组合预测模型[J].统计与决策(Statistics and Decision),2009(10):33-35.

[5] 高 尚,张绍彪,梅 亮.基于相对误差的线性组合预测研究[J].系统工程与电子技术(Systems Engineering and Electronics),2008,30(3):481-484.

[6] 杨廷方,刘 沛,李 浙,等.应用新型多方法组合预测模型估计变压器中溶解气体浓度[J].中国电机工程学报(Proceedings of the Chinese Society for Electrical Engineering),2008,28(31):108-113.

[7] 陈华友.组合预测方法有效性理论与应用[M].北京:科学出版社,2008.

[8] 张 强,赵 艳,高曼莉.可测干扰过程的预测控制性能监控[J].河北科技大学学报(Journal of Hebei University of Science and Technology),2010,31(4):352-354.