

一种改进的基于 IOWA 算子的组合预测方法及应用

敖培, 牟龙华

(同济大学电子与信息工程学院, 上海 201804)

摘要: 按误差平方和准则建立的基于 IOWA 算子的组合预测模型并不能正确反映出各个时期观测点所引起误差对预测值的影响程度, 在实际预测时预测期数据是未知的, 无法直接利用该方法进行预测。针对以上缺陷, 提出以单项预测模型中精度较高者的预测值为标准, 计算其余模型的预测值与其偏差, 再按各个时期各单项偏差绝对值和的平均值大小赋予权系数, 建立按照加权误差平方和准则新的预测模型, 并利用遗传算法求解最优权系数。通过实例验证, 改进后的组合预测方法优于原来的基于 IOWA 算子的组合预测方法, 有效地提高了预测精度。

关键词: IOWA 算子; 组合预测; 遗传算法

An improved combination forecasting method based on IOWA operator and application

AO Pei, MU Long-hua

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

Abstract: The combination forecasting model based on induced ordered weighted averaging (IOWA) operators, which is built according to the criterion of error sum of squares, fails to reflect the influence of errors arising from observation points in various periods on the predictive values. Moreover, this method can not be used to predict directly because future data are unknown in actual forecasting. In order to overcome the above flaws, an improved method is proposed. First, individual forecasting model that has higher forecasting accuracy is chosen as a criterion. Then, the deviation of predictive values between other models and standard model is computed. The weights are given according to the mean value size of the absolute value sum of deviation in every individual forecasting model in every period. Finally, a new forecasting model is built in accordance with the weighted error sum of squares. And genetic algorithm is used to solve the optimal weights. Verified by an example, the improved combination forecasting method is better than the original combination forecasting method based on IOWA operator. Forecasting accuracy is improved effectively.

Key words: IOWA operator; combination forecasting; genetic algorithm

中图分类号: TM715 文献标识码: A 文章编号: 1674-3415(2011)01-0053-04

0 引言

基于诱导有序加权平均算子 (Induced Ordered Weighted Averaging operator, IOWA) [1-3] 的组合预测模型通过引进诱导有序加权平均 IOWA 算子, 对每个单项预测方法在样本区间上各个时点的拟合精度的高低按顺序赋权, 以误差平方和为准则建立组合预测模型, 可以提高预测精度。然而, 实际预测时, 由于预测期数据是未知的, 因而无法直接利用该方法进行预测。此外, 该方法是基于误差平方和准则建立的, 即所有各个时期观测点所引起的误差都是同等对待的, 未能正确反映出各个时期观测点对预测值的影响程度。

为了克服以上缺陷, 本文首先在单项预测模型

中, 选取精度较高模型的预测值为标准, 分别计算出其余模型的预测值与它的偏差, 并按各个时期各单项偏差绝对值和的平均值大小赋予权系数, 最后按照加权误差平方和准则建立新的预测模型, 并利用遗传算法求解最优权系数。实例验证表明改进后的模型预测精度不仅远远高于各单项预测模型, 而且也高于原有的基于 IOWA 算子的组合预测模型。

1 IOWA算子

定义 1

设 $(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots, \langle v_n, a_n \rangle)$ 为 n 个二维数组,

令

$$IOWA_W(\langle v_1, a_1 \rangle, \langle v_2, a_2 \rangle, \dots, \langle v_n, a_n \rangle) = \sum_{i=1}^n w_i a_{v\text{-index}(i)} \quad (1)$$

其中： $W=(w_1, w_2, \dots, w_n)^T$ 是与 $IOWA_W$ 有关的加权向量，满足 $\sum_{i=1}^n w_i = 1, w_i \geq 0, i=1,2,\dots,n$ ， $v\text{-index}(i)$ 是 v_1, v_2, \dots, v_n 中按从大到小的顺序排列的第 i 个大的数的下标，则称函数 $IOWA_W$ 是由 v_1, v_2, \dots, v_n 所产生的 n 维诱导有序加权算术平均算子，简称为 IOWA 算子， v_i 称为 a_i 的诱导值。

2 基于IOWA算子的组合预测模型^[4]

设 $\{x_t, t=1, 2, \dots, N\}$ 为观测序列，有 m 种可行的单项预测方法对其进行预测， x_{it} 为第 i 种预测方法第 t 时刻的预测值（或称拟合值）， $i=1, 2, \dots, m, t=1, 2, \dots, N$ 。设 $L=(l_1, l_2, \dots, l_m)^T$ 为 m 种单项预测在组合预测中的加权系数，它满足 $\sum_{i=1}^m l_i = 1, l_i \geq 0, i=1, 2, \dots, m$ 。

定义 2

$$\hat{x}_t = \sum_{i=1}^m l_i x_{it}, t=1,2,\dots,N \quad (2)$$

则称 \hat{x}_t 为第 t 时刻传统的加权算术平均的组合预测值。令

$$a_{it} = \begin{cases} 1 - \left| \frac{x_t - x_{it}}{x_t} \right|, & \text{当 } \left| \frac{x_t - x_{it}}{x_t} \right| < 1 \\ 0, & \text{当 } \left| \frac{x_t - x_{it}}{x_t} \right| \geq 1 \end{cases}$$

其中： $i=1, 2, \dots, m; t=1, 2, \dots, N$ 。

则 a_{it} 表示第 i 种预测方法第 t 时刻预测精度，显然 $a_{it} \in [0,1]$ 。

定义 3

$$IOWA_L(\langle a_{1t}, x_{1t} \rangle, \langle a_{2t}, x_{2t} \rangle, \dots, \langle a_{mt}, x_{mt} \rangle) = \sum_{i=1}^m l_i x_{a\text{-index}(it)} \quad (3)$$

则式 (3) 称为第 t 时刻由预测精度序列 $\langle a_{1t}, a_{2t}, \dots, a_{mt} \rangle$ 所产生的 IOWA 算子组合预测值。式中 $a\text{-index}(it)$ 是第 i 个大的预测精度的下标。

令 $e_{a\text{-index}(it)} = x_t - x_{a\text{-index}(it)}$ ，于是 N 期总的组合预测误差平方和 S 为

$$S = \sum_{t=1}^N \left(x_t - \sum_{i=1}^m l_i x_{a\text{-index}(it)} \right)^2 =$$

$$\sum_{i=1}^m \sum_{j=1}^m l_i l_j \left(\sum_{t=1}^N e_{a\text{-index}(it)} e_{a\text{-index}(jt)} \right) \quad (4)$$

以误差平方和为准则的基于 IOWA 算子的组合预测模型可表示成如式 (5) 最优化模型。

$$\min S(L) = \sum_{i=1}^m \sum_{j=1}^m l_i l_j \left(\sum_{t=1}^N e_{a\text{-index}(it)} e_{a\text{-index}(jt)} \right) \quad \text{s.t.} \begin{cases} \sum_{i=1}^m l_i = 1 \\ l_i \geq 0, i=1,2,\dots,m \end{cases} \quad (5)$$

3 遗传算法

遗传算法^[5] (Genetic Algorithm, GA) 是以自然选择和遗传理论为基础，将生物进化过程中适者生存规则与群体内部染色体的随机信息交换机制相结合的高效全局寻优搜索算法，具有广泛的可行解表示、采用内在启发式群体随机搜索、无需适应度函数数值以外的辅助信息、不易陷入局部最优解等特点。目前，遗传算法作为一种全新的优化方法，已经在约束规划^[6-8]、非线性规划^[9-11]、控制系统优化^[12-13]等方面得到广泛的应用。遗传算法的主要实现步骤如下：

(1) 编码：GA 在进行搜索之前，将解空间的数据表示成遗传空间的基因型串结构数据，编码方法主要包括二进制编码、符号编码和浮点数编码等；

(2) 初始群体的生成：随机产生 N 个初始串结构数据，每个串结构数据称为一个个体， N 个个体构成一个群体；

(3) 适应性值评估检测：适应性函数表明个体或解的优劣性；

(4) 选择：从当前群体中选择适应性强的优良个体，使其作为父代以大概率繁殖后代；

(5) 交叉：组合其父辈个体特性，得到新一代个体；

(6) 变异：在群体中随机选择一个个体，以一定概率随机地改变串结构数据中某个串的值。

4 改进的基于IOWA算子的组合预测模型

实际预测时，预测期数据是未知的，即 x_{it} 未知，因此无法直接利用第 2 部分所述方法进行预测。此外，在式 (4) 中，所有各个时期观测点所引起的误差都是同等对待的，未能正确反映出各个时期观测点对预测值的影响程度。基于以上两点，本文针对第 2 部分方法做以下改进。

在单项预测模型中，选取精度较高模型的预测值 x_{jt} 代替第 2 部分中的 x_{it} 作为标准，再分别计算出

其余模型的预测值与它的偏差,然后赋予一个合适的权系数 α_t ,最后按照加权误差平方和准则建立新的预测模型,并利用遗传算法求解最优权系数 l_i 。

这里,权系数 α_t 按照各个时期各单项偏差绝对值和的平均值大小进行取值。

按上述加权误差平方和准则改进的基于 IOWA 算子的组合预测模型如为:

$$IOWA_L(\langle a_{1t}, x_{1t} \rangle, \langle a_{2t}, x_{2t} \rangle, \dots, \langle a_{(m-1)t}, x_{(m-1)t} \rangle) =$$

$$\sum_{i \neq j, i=1}^{m-1} l_i x_{a-\text{index}(it)} \quad (6)$$

$$\min S(L) = \sum_{t=1}^N \alpha_t \left(x_{jt} - \sum_{i \neq j, i=1}^{m-1} l_i x_{a-\text{index}(it)} \right)^2$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{m-1} l_i = 1 \\ l_i \geq 0, \quad i=1, 2, \dots, m-1 \end{cases} \quad (7)$$

式中, $\alpha_t = \alpha \frac{1}{m-1} \sum_{i=1}^{m-1} |x_{jt} - x_{it}|$, $0 < \alpha \leq 1$, $t=1, 2, \dots, N$ 。

5 实例分析

表1为某地区1990年~2008年电力消费量、工业总产值、农业总产值和国内生产总值等相关数据,本文根据1990年~2004年的相关数据,预测2005年~2008年该地区的电力消费量,以验证本文所提方法的有效性。

表1 某地区 1990 年~2008 年相关数据

Tab.1 Related data in a certain area from 1990 to 2008

年份	电力消费量	工业总产值	农业总产值	国内生产总值
1990	469.27	1 348.17	273.8	1 062.7
1991	497.43	1 544.39	295.9	1 200.1
1992	547.17	1 876.95	330.1	1 473
1993	605.17	2 609.99	425.7	2 010.8
1994	594.28	3 117.82	546.8	2 461.8
1995	622.81	3 055.54	691.8	2 793.4
1996	661.30	3 354.61	804.7	3 157.7
1997	711.56	3 644.88	834.7	3 582.5
1998	674.56	3 147.86	969.8	3 881.7
1999	756.11	3 390.27	977.1	4 171.7
2000	748.89	4 249.46	967.4	4 669.1
2001	764.77	4 480.32	1 045.7	5 033.1
2002	809.45	4 888.02	1 132.5	5 458.2
2003	907.91	6 112.96	1 215.0	6 002.5
2004	1 019.78	8 603.90	1 510.5	6 872.7
2005	1 110.56	10 814.51	1 671.60	8 005
2006	1 228.27	14 167.95	1 738.10	9 257.05
2007	1 359.51	18 249.53	2 128.00	11 021.7
2008	1 412.00	24 769.09	2 476.9	13 461.6

注:工业总产值、农业总产值和国内生产总值的单位为:亿元;

电力消费量的单位为:亿kWh。

首先,分别采用弹性系数法、修正指数增长曲线法、指数平滑法、多元线性回归法等单项方法和基于IOWA算子的组合预测模型进行预测,预测结果如表2所示。选择预测误差平方和(SSE)、均方误差(MSE)、平均绝对误差(MAE)、平均绝对百分比误差(MAPE)、均方百分比误差(MSPE)等指标评价预测效果(如表3)。

表2 某地区 2005 年~2008 年电力消费量预测值

Tab.2 Power consumption predictive value of a certain area from 2005 to 2008

	2005	2006	2007	2008
弹性系数法	1 166.91	1 254.42	1 348.51	1 449.64
修正指数增长曲线法	1 067.11	1 116.61	1 168.53	1 222.99
指数平滑法	1 112.01	1 192.55	1 278.82	1 370.83
多元线性回归法	1 156.45	1 259.14	1 376.41	1 510.43
基于 IOWA 组合预测	1 153.87	1 249.93	1 357.76	1 478.93
本文方法	1 146.44	1 229.94	1 319.55	1 415.59

表3 预测效果评价

Tab.3 Evaluation of predictive results

	SSE	MSE	MAE	MAPE	MSPE
弹性系数法	5 396.914 6	18.365 9	32.785	0.026 7	0.015 4
修正指数增长曲线法	86 554.110 0	73.550 2	133.775 3	0.101 1	0.054 5
指数平滑法	9 483.632 2	24.334 6	39.757 8	0.029 7	0.018 1
多元线性回归法	13 033.303 0	28.540 9	48.021 9	0.037 1	0.021 4
基于 IOWA 组合预测	4 648.386 9	17.044 8	28.924 7	0.022 6	0.013 1
本文方法	2 900.454 2	13.464	20.275 9	0.016 4	0.010 9

表3中: $SSE = \sum_{t=1}^N (x_t - \hat{x}_t)^2$, $MSE = \frac{1}{N} \sqrt{\sum_{t=1}^N (x_t - \hat{x}_t)^2}$,

$MAE = \frac{1}{N} \sum_{t=1}^N |x_t - \hat{x}_t|$, $MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{x_t - \hat{x}_t}{x_t} \right|$,

$MSPE = \frac{1}{N} \sqrt{\sum_{t=1}^N \left[\frac{(x_t - \hat{x}_t)}{x_t} \right]^2}$ 。

综合考虑各单项预测方法的模型精度和预测精度,选取多元线性回归模型的预测值作为标准,按照本文提出的改进的基于IOWA算子的组合预测方法进行预测。取权系数 $\alpha=0.8$,应用Matlab遗传算法工具箱求解最优权系数 l_i ,遗传算法部分参数设置如下:

Population type='Double Vector',

Population size=20,

Creation function='Uniform';

Selection function='Stochastic uniform';

Crossover function='Scattered';

Crossover probability=0.8;

Mutation function='Gaussian',
Scale=Shrink=1.0;

经过 51 代进化后, 得 $l_1=0.673$, $l_2=0.271$, $l_3=0.056$ 。预测值和评价结果分别如表 2 和表 3 所示。

从表 3 中可以看出, 本文提出的改进的基于 IOWA 算子的组合预测模型的各种误差指标不仅远远低于各单项预测模型, 而且也低于改进前的基于 IOWA 算子的组合预测模型, 这表明本文所提出的组合预测方法优于改进前的基于 IOWA 算子的组合预测方法, 能够有效地提高预测精度。

6 结论

在实际预测时, 预测期数据是未知的, 因而无法直接利用基于 IOWA 的组合预测方法进行预测。而且该方法是按误差平方和准则建立的, 未能正确反映出各个时期观测点所引起误差对预测值的影响程度。本文以单项预测模型中精度较高者的预测值为标准, 计算其余模型的预测值与它的偏差, 再按各个时期各单项偏差绝对值的平均值大小赋予权重系数, 最后按照加权误差平方和准则建立新的预测模型, 并利用遗传算法求解最优权重系数。通过实例分析, 改进后的预测模型各种误差指标不仅远远低于各单项预测模型, 而且也低于改进前的基于 IOWA 算子的组合预测模型, 证明本文所提出的组合预测方法能够有效地提高预测精度。

参考文献

[1] Yager R R. Induced ordered weighted averaging operators[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 1999, 29 (2): 141-150.

[2] Yager R R. Using fuzzy methods to model nearest neighbors[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2002, 32 (4): 512-525.

[3] Fonooni B, Moghadam S. Applying induced aggregation operator in designing intelligent monitoring system for financial market[C]. //Proceedings of Computational Intelligence for Financial Engineering. IEEE/IAFE Conference on Computational Intelligence for Financial Engineering. 2009: 80-84.

[4] 陈华友, 刘春林. 基于 IOWA 算子的组合预测新方法[J]. 预测, 2003, 22 (6): 61-65.
CHEN Hua-you, LIU Chun-lin. A kind of combination forecasting method based on induced ordered weighted averaging (IOWA) operators[J]. Forecasting, 2003, 22 (6): 61-65.

[5] 王小平, 曹立明. 遗传算法——理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002.
WANG Xiao-ping, CAO Li-ming. Genetic algorithm—theory, application and software realization[M]. Xi'an: Xi'an Jiaotong University Press, 2002.

[6] 蔺红, 晁勤, 吐尔逊, 等. 基于相似性自适应学习的遗传算法在无功优化中的应用[J]. 继电器, 2007, 35(16): 46-49, 54.
LIN Hong, CHAO Qin, TUERXUN, et al. A genetic algorithm based on similitude frame of evolutionary computation applied in reactive power optimization[J]. Relay, 2007, 35(16): 46-49, 54.

[7] 张敬平, 梁志瑞, 苏海峰, 等. 基于改进拥挤小生境遗传算法配网无功优化研究[J]. 继电器, 2007, 35(10): 19-22.
ZHANG Jing-ping, LIANG Zhi-rui, SU Hai-feng, et al. Research on reactive power optimization of distribution network based on the improved crowding niche genetic algorithm[J]. Relay, 2007, 35(10): 19-22.

[8] 李晶, 王素华, 谷彩连. 基于遗传算法的含分布式发电的配电网无功优化控制研究 [J]. 电力系统保护与控制, 2010, 38(6): 60-63.
LI Jing, WANG Su-hua, GU Cai-lian. Research of reactive power optimization control in distribution networks based on genetic algorithm [J]. Power System Protection and Control, 2010, 38(6): 60-63.

[9] HE Da-kuo, WANG Fu-li, JIA Ming-xing. An improved genetic algorithm for a type of nonlinear programming problems[C]. //IEEE International Conference on Automation and Logistics, 2008: 2582-2585.

[10] DAI Chao-hua, CHEN Wei-rong, ZHU Yun-fang, et al. Seeker optimization algorithm for optimal reactive power dispatch[J]. IEEE Transactions on Power Systems, 2009, 24 (3): 1218-1231.

[11] 王洪涛, 刘玉田. 基于 NSGA2 II 的多目标输电网架最优重构[J]. 电力系统自动化, 2009, 33 (23): 14-18.
WANG Hong-tao, LIU Yu-tian. Multi-objective optimization of power system reconstruction based on NSGA2 II [J]. Automation of Electric Power Systems, 2009, 33 (23): 14-18.

[12] Hansen J V. Genetic search methods in air traffic control[J]. Computers and Operations Research, 2004, 31 (3): 445-459.

[13] 江全元, 邹振宇, 曹一家. 基于极大极小值原理的控制器设计[J]. 电力系统自动化, 2009, 29 (14): 23-28.
JIANG Quan-yuan, ZOU Zhen-yu, CAO Yi-jia. Design of UPFC damping controller based on minimax principle[J]. Automation of Electric Power Systems, 2009, 29 (14): 23-28.

收稿日期: 2010-01-14

作者简介:

教培 (1979-), 女, 博士研究生, 助理研究员, 研究方向为智能测试理论与装置; E-mail: aopei16.student@sina.com

牟龙华 (1963-), 男, 教授, 博士生导师, 研究方向为电力系统微机保护与智能电器. E-mail: lhmu@vip.163.com