文章编号:1673-0062(2018)01-0058-05

基于指数支撑度的 IGOWLA 算子的最优组合预测模型

孙　浩,杨桂元*

(安徽财经大学 统计与应用数学学院,安徽 蚌埠 233030)

摘　要:在广义诱导有序加权对数平均算子(IGOWLA 算子)的基础上,引入指数支撑度以及构建了一种基于支撑度的 IGOWLA 算子的最优组合预测模型,并对该模型的预测精度、优性及非劣性给出定义.实例分析表明,该组合预测模型优于传统的组合预测模型,能够充分利用各个单项预测方法的信息并能提高预测精度,是一种优性组合预测.

关键词:指数支撑度;组合预测;IGOWLA 算子

中图分类号:F224.0　　　**文献标志码**:A

The Combination Forecasting Model Based on Exponential Supporting Degree and Induced Geometric Ordered Weighted Logarithm Averaging(IGOWLA) Operator

SUN Hao,YANG Gui-yuan*

(School of Statistics and Applied Maths,Anhui University of Finance and Economics, Bengbu,Anhui 233030,China)

Abstract:Based on the induced geometric ordered weighted logarithm averaging operator (IGOWLA operator),the paper construct the optimal combination forecasting model of IGOWLA operator on the basis of an exponential supporting degree and power error are introduced.Then,the prediction accuracy,superiority and non-inferiority of this model are defined.Finally,an example was illustrated by using the model which can make full use of the information from the single forecasting method .The forecasting precision is superior to the traditional single forecasting model as well.It is concluded that the method is a superior combination forecasting method.

key words:average exponential supporting degree;combination forecast;IGOWLA operator

---

收稿日期:2017-12-27

基金项目:国家社会科学基金研究项目资助(12BTJ008);安徽财经大学研究生科研创新基金项目(ACYC2016113)

作者简介:孙　浩(1991-),男,硕士研究生,主要从事于预测理论与方法、金融计量分析方面的研究.E-mail:sunhao1535@163.com. *通讯作者:杨桂元,E-mail:yangguiyuan57@163.com

0 引　言

组合预测的概念是由 C.W.J.Granger 等[1]在 1969 年首次提出的的,即为了提高预测的精确度,将各种单项预测方法的有效信息进行组合.1988 年,美国学者 Yager 提出了一种存在于“or”和“and”之间的“orand”数据信息结集算子,即有序加权平均(OWA)算子[2-3].由于 OWA 算子及其拓展方法具有提高模型拟合精度和预测能力的优点,被广泛应用到人工神经网络、模糊系统控制与模糊建模、信息融合、决策分析、通信网络等诸多领域[4-10].2003 年,陈华友等在 OWA 算子的基础上通过增添诱导值,给出了一种新算子(IOWA 算子)[11],该算子中的系数只与该预测方法在各点处的预测精度有关.2004 年,R.R.Yager 等提出了 GOWA 算子[12],该算子将 OWA 算子扩展到 λ 次幂.在此基础上,J.M.Merigo 等结合诱导变量提出了诱导广义有序加权平均算子[13](IGOWA 算子)的概念,之后陈华友等将单项预测数值取对数后再进行 IGOWA 算子运算,得到了诱导广义有序加权对数平均算子(IGOWLA 算子)的定义[14].本文在上述研究的基础上,将指数支撑度[15]与 IGOWLA 算子结合,提出了一种基于指数支撑度的 IGOWLA 算子最优组合预测模型,并通过实证对该模型的合理性和有效性进行验证评价.

1 基于指数支撑度的 IGOWLA 算子的最优组合预测模型

定义 1　假设 IGOWLA 是 n 元函数, $W=(w_1,w_2,\cdots,w_n)^{\mathrm T}$ 是和这些算子相对应的的加权向量,数据集($<v_1,a_1>$, $<v_2,a_2>$,$\cdots$, $<v_n,a_n>$)是包含 n 个数据的二维数组,且 W 满足 $\sum_{i=1}^{n} w_i=1$, $w_i\geqslant 0$, $i=1,2,\cdots,n$.令

$$IGOWLA(<v_1,a_1>,<v_2,a_2>,\cdots,<v_n,a_n>) = \exp\left\{\left(\sum_{i=1}^{n} w_i\,(\ln a_{v\text{-}index(i)})^{\lambda}\right)^{\frac{1}{\lambda}}\right\} \qquad (1)$$

在式(1)中, $v\text{-}index(i)$ 为 $v_1,v_2,\cdots,v_n$ 中第 i 大的数所对应的 a 的下标,其中 $v_1,v_2,\cdots,v_n$ 为诱导变量, $\lambda\in(-\infty,0)\cup(0,\infty)$,称式(1)为诱导广义有序加权对数平均算子(IGOWLA 算子).从式(1)中看出,IGOWLA 算子中的权系数 w_i 与诱导值 v_i 所在的位置有关,也就是说 $a_1,a_2,\cdots,a_n$ 是按照诱导值 $v_1,v_2,\cdots,v_n$ 从大到小的排序结果所进行的一种有序加权算术平均.

定义 2　设 $a,b,x,y\in R^+$,如果 $Sup(a,b)$ 满足下列条件:

(1) $Sup(a,b)\in[0,1]$,当且仅当 $a=b$ 时, $Sup(a,b)=1$

(2) $Sup(a,b)=Sup(b,a)$

(3)如果当 $|a-b|<|x-y|$,有 $Sup(a,b)\geqslant Sup(x,y)$

那么,则称 $Sup(a,b)$ 为 b 对 a 的支撑度.

特别地,在文献[15]中给出了一种支撑度,即

$$Sup\left(\frac{a}{a+b},\frac{b}{a+b}\right)=\exp\left(-\frac{|a-b|}{a+b}\right) \qquad (2)$$

称之是 b 对 a 归一化的平均指数支撑度,并简记为 $Sup(a,b)$.

若将定义 2 中的 a 当成某一经济现象的实际值, b 当成该经济现象的预测值,则 $Sup(a,b)$ 当可以看成是反映预测精度的一个指标.两个数据集 b 与 a 之间的差异越大时,二者之间的支撑度 $Sup(a,b)$ 越小;当且仅当 b 与 a 相同时,支撑度 $Sup(a,b)$ 达到最大值 1.

假设某一现象的实际值为 $\{x_t,t=1,2,\cdots,N\}$,并且使用 m 种可行的单项预测方法对其进行预测,使用 x_{it} 来表示第 i 种单项预测方法在第 t 时刻的拟合值, $i=1,2,\cdots,m$; $t=1,2,\cdots,N$.由于原始数据集 $\{x_t,t=1,2,\cdots,N\}$ 中的数据与使用预测方法所得到的数据之间一般都是量纲数据集,所以仍需要进行如下归一化处理:

$$y_t=\frac{x_t}{\max(x_1,x_{1t},x_{2t},\cdots,x_{mt})}$$

$$y_{it}=\frac{x_{it}}{\max(x_1,x_{1t},x_{2t},\cdots,x_{mt})}$$

$$(i=1,2,\cdots,m;t=1,2,\cdots,N)$$

定义 3　称 a_{it} 为第 i 种预测方法在 t 时刻的预测精度,其中:

$$a_{it}=\begin{cases}1-|(x_t-x_{it})/x_t| & ;\quad 当|(x_t-x_{it})/x_t|<1\\ 0 & ;\quad 当|(x_t-x_{it})/x_t|\geqslant 1\end{cases} \qquad (3)$$

显然 $a_{it}\in[0,1]$, $i=1,2,\cdots,m;t=1,2,\cdots,N$.

把预测精度 a_{it} 看成是预测值 x_{it} 归一化后的数据 y_{it} 的诱导值,从而 a_{it} 与 y_{it} 就构成 m 个二维数组: $<a_{1t},y_{1t}>$, $<a_{2t},y_{2t}>$,$\cdots$, $<a_{mt},y_{mt}>$, $i=1,2,\cdots,m$; $t=1,2,\cdots,N$.将 m 种单项预测方法第 t 时刻预测精度序列 $a_{1t},a_{2t},\cdots,a_{mt}$ 按由大到

小的次序进行排序,令 $a-index(it)$ 为第 i 大精度所对应的 y 的下标, $\hat{y}_t$ 为第 t 时刻由预测精度序列 $a_{1t},a_{2t},\cdots,a_{mt}$ 作为诱导变量所产生的 IGOWLA 算子的组合预测值,其中, $t=1,2,\cdots,N$. 根据定义 1 有:

$$\hat{y}_t = IGOWLA(<a_{1t},y_{1t}>,<a_{2t},y_{2t}>,\cdots,<a_{mt},y_{mt}>) = \exp\left\{\left(\sum_{i=1}^{m} w_i (\ln y_{a-index(it)})^{\lambda}\right)^{\frac{1}{\lambda}}\right\} \tag{4}$$

由式(4)可以看出,基于 IGOWLA 算子的组合预测的系数是根据各单项预测方法在不同时刻预测精度的相对高低排序结果来进行赋权的.所以,由式(4)可得

$$(\ln \hat{y}_t)^{\lambda} = \sum_{i=1}^{m} w_i (\ln y_{v-index(it)})^{\lambda} \tag{5}$$

令 $e_{it}=(\ln y_t)^{\lambda}-(\ln \hat{y}_{it})^{\lambda}$ 表示第 i 种单项预测方法在 t 时刻与相应归一化的实际值之间的对数 λ 次幂的预测误差, $i=1,2,\cdots,m;t=1,2,\cdots,N$.并令

$$e_{v-index(it)} = (\ln y_t)^{\lambda} - (\ln \hat{y}_{v-index(it)})^{\lambda} \tag{6}$$

由式(4)~(6)可以得到第 t 时刻 IGOWLA 组合预测值的无量纲化指标误差:

$$e_t = (\ln y_t)^{\lambda} - (\ln \hat{y}_t)^{\lambda} = (\ln y_t)^{\lambda} - \left(\sum_{i=1}^{m} w_i (\ln y_{v-index(it)})^{\lambda}\right) = \sum_{i=1}^{m} w_i[(\ln y_t)^{\lambda} - (\ln y_{v-index(it)})^{\lambda}] = \sum_{i=1}^{m} w_i e_{v-index(it)} \tag{7}$$

将式(2)与 IGOWLA 算子相结合,有如下定义:

定义 4 称 Sup_i 为第 i 种单项预测方法预测值序列与指标实际值序列的支撑度,Sup 为 IGOWLA 算子组合预测值序列与指标实际值序列的支撑度,其中:

$$Sup_i(y_t,y_{it}) = \exp\left(-\sum_{t=1}^{N}\left(\left|e_{it}\right| \Big/ \sum_{t=1}^{N}\left|(\ln y_t)^{\lambda} + (\ln y_{it})^{\lambda}\right|\right)\right) \tag{8}$$

$$Sup(y_t,\hat{y}_t) = \exp\left(-\sum_{t=1}^{N}\left|\sum_{i=1}^{m} w_i e_{v-index(it)}\right| \Big/ \sum_{t=1}^{N}\left|(\ln y_t)^{\lambda}\right| + \sum_{t=1}^{N}\left(\left|\sum_{i=1}^{m} w_i (\ln y_{v-index(it)})^{\lambda}\right|\right)\right) \tag{9}$$

由式(9)可以看出,IGOWLA 组合预测值序列与实际观测值序列的支撑度 Sup 为 $W=(w_1,w_2,\cdots,w_m)^{\mathrm{T}}$ 的函数,记为 $Sup(w_1,w_2,\cdots,w_m)$. $Sup(w_1,w_2,\cdots,w_m)$ 越大表示组合预测方法越有效.因此,基于指数支撑度的 IGOWLA 算子最优组合预测模型为:

$$\max Sup(w_1,w_2,\cdots,w_m) = \exp\left(-\sum_{t=1}^{N}\left|\sum_{i=1}^{m} w_i e_{v-index(it)}\right| \Big/ \sum_{t=1}^{N}\left|\left(\ln y_t\right)^{\lambda}\right| + \sum_{t=1}^{N}\left(\left|\sum_{i=1}^{m} w_i\left(\ln y_{v-index(it)}\right)^{\lambda}\right|\right)\right)$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{m} w_i = 1\\ w_i \geqslant 0, i=1,2,\cdots,m\end{cases}$$

又因为指数函数是单调函数,故上述模型等价于如下模型:

$$\min f(w_1,w_2,\cdots,w_m) = \sum_{t=1}^{N}\left|\sum_{i=1}^{m} w_i e_{v-index(it)}\right| \Big/ \sum_{t=1}^{N}\left|(\ln y_t)^{\lambda}\right| + \sum_{t=1}^{N}\left(\left|\sum_{i=1}^{m} w_i (\ln y_{v-index(it)})^{\lambda}\right|\right)$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{m} w_i = 1\\ w_i \geqslant 0, i=1,2,\cdots,m\end{cases} \tag{10}$$

定义 5 令

$$f_i = \sum_{t=1}^{N}\left|e_{it}\right| \Big/ \sum_{t=1}^{N}\left|(\ln y_t)^{\lambda}\right| + \sum_{t=1}^{N}\left(\left|(\ln y_{it})^{\lambda}\right|\right)$$

$$f = \sum_{t=1}^{N}\left|\sum_{i=1}^{m} w_i e_{v-index(it)}\right| \Big/ \sum_{t=1}^{N}\left|(\ln y_t)^{\lambda}\right| + \sum_{t=1}^{N}\left(\left|\sum_{i=1}^{m} w_i (\ln y_{v-index(it)})^{\lambda}\right|\right) \tag{11}$$

记 $f_{\min}=\min\{f_i,i=1,2,\cdots,m\}$,$f_{\max}=\max\{f_i,i=1,2,\cdots,m\}$;若 $f(w_1,w_2,\cdots,w_m)<f_{\min}$,则称由权系数 $w_1,w_2,\cdots,w_m$ 确定的组合预测模型为优性组合预测;若 $f_{\min}\leqslant f(w_1,w_2,\cdots,w_m)\leqslant f_{\max}$,则称之为非劣性组合预测;若 $f(w_1,w_2,\cdots,w_m)>f_{\max}$ 则称之为劣性组合预测.

2 实例分析

本文利用文献[16]中的数据,对该测模型的有效性进行验证和分析,并采用式(3)计算出各种单项预测方法的预测精度,具体数据见表 1.

根据表 1 中各种单项预测方法的预测值的精度大小对预测值进行排序,并进行归一化处理,处理后的结果见表 2.

表 1　实际值与各种预测方法下的预测值及其预测精度

Table 1　Actual value and forecasting value under various prediction methods and their prediction accuracy

年份	实际值	多元回归		残差自回归		GM(1,1)	
		预测值	精度	预测值	精度	预测值	精度
2005	14 679	12 209.50	0.832	—	—	—	—
2006	14 550	12 023.95	0.826	17 810.02	0.776	10 239.62	0.704
2007	13 463	16 968.76	0.740	17 653.50	0.689	12 668.55	0.941
2008	18 626	19 849.30	0.934	16 334.64	0.877	15 673.64	0.842
2009	19 947	22 710.61	0.862	22 598.91	0.867	19 391.57	0.972
2010	21 567	20 910.05	0.970	24 201.68	0.878	23 991.42	0.888
2011	28 708	29 040.71	0.988	26 167.22	0.912	29 682.41	0.966
2012	36 460	35 960.35	0.986	34 831.39	0.955	36 723.34	0.993
2013	49 295	46 999.27	0.953	44 236.88	0.897	45 434.44	0.923
2014	62 169	62 279.68	0.998	59 809.57	0.962	56 211.90	0.904
2015	71 853	72 364.82	0.993	75 429.59	0.950	69 545.86	0.968

表 2　指标实际值与按预测精度大小排列以及归一化后的预测值

Table 2　The actual value of the indicator is arranged according to the prediction accuracy and the normalized forecast value

归一化数据	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
y_t	0.816 96	0.762 62	0.938 37	0.878 31	0.891 14	0.988 54	0.992 83	1.000 00	0.998 22	0.952 58
$y_{a-index(1t)}$	0.675 12	0.717 62	1.000 00	0.853 86	0.863 99	1.000 00	1.000 00	0.953 43	1.000 00	0.959 37
$y_{a-index(2t)}$	1.000 00	0.961 21	0.822 93	0.995 08	0.991 31	1.022 10	0.979 22	0.921 68	0.960 34	0.922 00
$y_{a-index(3t)}$	0.574 94	1.000 00	0.789 63	1.000 00	1.000 00	0.901 05	0.948 48	0.897 39	0.902 57	1.000 00

将表 2 中的数据代入式(10)中,运用 MATLAB 软件中的最优化工具箱求出不同参数 $\lambda = 1$, $\lambda = 2$, $\lambda = 3$ 下的最优权系数均为 $w_1 = 1$, $w_2 = 0$, $w_3 = 0$.再将其代入下面的 IGOWLA 组合预测模型

$$\hat{x}_t = IGOWLA(<a_{1t}, x_{1t}>, <a_{2t}, x_{2t}>, \cdots, <a_{mt}, x_{mt}>) = \exp\left\{\left(\sum_{i=1}^{m} w_i \left(\ln x_{v\text{-}index(it)}\right)^{\lambda}\right)^{\frac{1}{\lambda}}\right\}$$

从而可以得到组合预测值(见表 3).

表 3　实际值与基于支撑度的 IGOWLA 算子组合预测值

Table 3　The actual value is combined with the supportive IGOWLA operator forecast

年份	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
实际值	14 550	13 463	18 626	19 947	21 567	28 708	36 460	49 295	62 169	71 853
组合预测值	12 023.95	12 668.55	19 849.3	19 391.57	20 910.05	29 040.71	36 723.34	46 999.27	62 279.68	72 364.82

为了验证本文提出的组合预测模型的有效性,使用下列 5 种误差指标进行衡量:

(1)平方和误差:$\delta_{SSE} = \sum_{t=1}^{N} (x_t - \hat{x}_t)^2$;

(2)均方误差:$\delta_{MSE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N} (x_t - \hat{x}_t)^2}$;

(3)平均绝对误差:$\delta_{MAE} = \frac{1}{N}\sum_{t=1}^{N} |x_t - \hat{x}_t|$;

(4)平均绝对百分比误差:$\delta_{MAPE} = \frac{1}{N}\sum_{t=1}^{N} \left|\frac{(x_t - \hat{x}_t)}{x_t}\right|$;

(5)均方百分比误差:$\delta_{MSPE} = \frac{1}{N}\sqrt{\sum_{t=1}^{N} \left(\frac{x_t - \hat{x}_t}{x_t}\right)^2}$

表 4 给出了各种预测方法的效果评价结果.由表 4 可以看出,与 3 种单项的预测方法以及误差平方和最小组合预测方法相比较,本文提出的模型的上述五种误差的数值均比单一预测方法的误差要小得多.这说明本文提出的模型优于 3 种单一的预测方法和误差平方和最小组合预测方法,并且能够得到预测精度更高的结果.

另外,使用式(12)能够测算出 3 种单项的预测方法以及在不同权系数下的组合预测方法的预测值与实际值之间的指数支撑度,结果见表 5.由表 5 可以看出,文中提出的模型的支撑度均小于

单项预测方法的支撑度,即二者满足定义 5 中的 $f < \min(f_1, f_2, f_3)$,由此说明本文提出的组合预测模型是优性的和有效的.

表 4 单种预测与组合预测方法预测效果评价

Table 4 Prediction effect evaluation of single prediction and combined prediction

误差种类	预测方法 1	预测方法 2	预测方法 3	误差平方和最小组合预测	IGOWLA 组合预测
δ_{SSE}	34 142 000	100 460 000	90 846 000	41 300 000	14 973 000
δ_{MSE}	584.31	1002.3	953.13	642.68	386.95
δ_{MAE}	1 442.6	3 019.2	2 440	1 613.6	927.05
δ_{MAPE}	0.074 947	0.123 7	0.090 143	0.063 907	0.043 089
δ_{MSPE}	0.035 346	0.046 807	0.038 384	0.025 76	0.020 506

表 5 单项预测与组合预测在不同参数下的指数支撑度

Table 5 Exponential degree of support for different predictors and combined predictors under different parameters

参数值	预测方法 1	预测方法 2	预测方法 3	组合预测
$\lambda = 1$	0.454 90	0.574 52	0.527 68	0.345 70
$\lambda = 2$	0.589 46	0.581 92	0.723 52	0.534 60
$\lambda = 3$	0.703 10	0.568 80	0.858 70	0.553 60

3 结 论

本文从广义诱导有序加权对数平均算子(IGOWLA 算子)与指数支撑度入手,将二者有效的结合在一起,从而构建了基于一种指数支撑度的 IGOWLA 算子的最优组合预测模型.通过实例进行验证,结果表明该组合预测模型能够充分有效的利用各单一方法所呈现出的信息,能够有效的提高模型的预测精度,即该模型是合理的和有效的.另外,对于该模型的理论问题如对冗余预测方法的判定的问题,仍需要继续探究.

参考文献:

[1] BATES J M,GRANGER C W J.Combination of forecasts [J].Operation research quarterly,1969,20(4):451-468.

[2] YAGER R R.On ordered weighted averaging aggregation operators in multicriteria decision making[J].IEEE Transactions on systems,man,and cybernetics,1988,18(1):183-190.

[3] FILEV D,YAGER R R.On the issue of obtaining OWA operator weights[J].Fuzzy sets and systems,1998,94 (2) :157-169.

[4] 马永开,杨桂元,唐小我.非负权重组合预测的冗余定理[J].系统工程理论方法应用,1995,4(4):33-39.

[5] 姜晨,徐宗昌,肖国军.用神经网络组合预测法估算反舰导弹研制费用[J].系统工程与电子技术,2004,26 (3):348-372.

[6] 滕云龙,师奕兵,康容雷.软件可靠性组合预测模型研究[J].计算机应用,2008,28(12):3092-3094.

[7] 唐小我,马永开,曾勇,等.现代组合预测和组合投资决策方法及应用研究[M].北京:科学出版社,2003.

[8] 王守相,张娜.基于灰色神经网络组合模型的光伏短期出力预测[J].电力系统自动化,2012,36(19):37-41.

[9] 谭华,谢赤,孙柏,等.证券市场灰色神经网络组合预测模型应用研究[J].湖南大学学报(自然科学版),2007,34(9):86-89.

[10] 陈华友.组合预测方法有效性理论及其应用[M].北京:科学出版社,2008.

[11] 陈华友,刘春林.基于 IOWA 算子组合预测方法[J].预测,2003,22(6):61-65.

[12] YAGER R R.Generalized OWA aggregation operators [J].Fuzzy optimization and decision making,2004,3 (1):93-107.

[13] MERIGO J M,GIL-LAFUENTE A M.The induced generalized OWA operator[J].Information sciences,2009,179 (6):729-741.

[14] 江立辉,陈华友,丁芳清,等.基于 IGOWLA 算子的最优组合预测模型及其应用[J].统计与决策,2015 (4):82-85.

[15] 袁宏俊,陈华友,胡凌云.基于指数支撑度的最优组合预测模型及其性质研究[J].应用概率统计,2012,28 (2):150-160.

[16] 储震,杨桂元.基于灰关联度的 IGOWLA 算子中国楼市库存的预测分析[J].佳木斯大学学报,2016,34 (4):599-605.

(责任编辑:龙威)