

Interval Combination Forecast Model Based on Arithmetic Average Approach Degree

Fengxiao Wang

College of Mathematics and Statistics, Kashigar University, Kashi 844000, China.

fxw-hz@126.com

Abstract. Using the arithmetic average minimum approach degree as a weighted approach degree, the new indicator of the relevance of the optimal interval combination forecast model is established. Some new concepts such as superior interval combination forecast model and non-inferior interval combination forecast model, which is on the basis of interval combination forecasting model based on arithmetic average degree are proposed. Some sufficient conditions of non-inferior and superior combination forecasting are given. Finally, Application example shows that this method can be effective for better forecast.

Keywords: Interval combination forecast; Arithmetic average approach degree; Superior interval combination forecasting

基于算术平均贴近度的区间组合预测模型

王丰效

(喀什大学 数学与统计学院, 新疆 喀什 844000)

摘要: 利用算术平均贴近度作为相关性指标, 建立了一种新的最优区间组合预测模型. 针对这类区间组合预测模型, 给出了非劣性区间组合预测、优性区间组合预测等基本概念, 并给出了这类区间组合预测方法的存在性的判定方法. 最后通过一个实例的分析, 结果表明该方法能够取得比较好的预测效果.

关键词: 区间组合预测; 算术平均贴近度; 优性区间组合预测

中图分类号: O 224 **文献标志码:** A

引言

组合预测方法一直是国内外研究的热点, 并取得了大量的研究成果[1-2]。传统的组合预测方法一般都是从改善某种拟合误差角度提出的, 文献[3]给出了以误差平方和达到最小的最优组合预测模型, 提出了非劣性组合预测, 优性组合预测的概念, 并利用组合预测绝对误差信息矩阵的性质判断简单平均方法是非劣性组合预测、优性组合预测的条件。文献[4-7]给出相关性指标的最优组合预测模型的有效性理论。然而实际的预测对象是较为复杂, 使得获得的信息往往具有不确定性。这时, 信息的处理用点值来表示就显得不合理, 将其看成一个区间数形式的模糊数更为合适, 关于区间组合预测的问题多是基于左右端点序列的相关性, 对左右序列分别建立模型。例如, 文献[8-9]针对单项预测值与实际值都为实数而组合预测值为区间数的组合预测问题, 提出了相应的确定组合预测区间值系数的方法。文献[10]针对预测值与实际值都以区间数形式给出的组合预测问题, 提出了基于 IOWA (Induced Ordered Weighted Averaging) 算子的区间组合预测方法。文献[11-12]在其基础上分别引进了向量夹角余弦、Theil 系数等相关性指标, 提出了不同准则的区间组合预测模型。区间数据通常可以用区间中点和区间半径来描述, 本文将模型中各数据按决策属性指标无量纲化处理, 以区间中点和半径算术平均贴近度的加权和作为新的贴近度, 根据加权贴近度建立最优组合预测模型, 并对该模型进行有效性理论研究, 提出了非劣性区间组合预测, 优性区

间组合预测存在的条件。最后给出实例，表明该方法的有效性和合理性。

1 基本概念

设 $\{x_t = [x_t^-, x_t^+], t = 1, \dots, n\}$ 为某社会经济现象指标的实际区间值序列， $x_{it} = [x_{it}^-, x_{it}^+]$ 为第 i 种单项预测方法在第 t 时刻的预测值 ($i = 1, 2, \dots, m$)，则 $\hat{x}_t = [\hat{x}_t^-, \hat{x}_t^+] = \sum_{i=1}^m l_i x_{it}$ 为实际区间值的组合预测值， $\hat{x}_t = [\hat{x}_t^-, \hat{x}_t^+] = \sum_{i=1}^m l_i x_{it}$ 为各种单项预测在组合预测中的非负加权系数，且满足 $l_1 + l_2 + \dots + l_m = 1$ 。原始区间数据序列和单项预测结果序列都可以利用区间中点和区间半径表示，记 $[x_t^-, x_t^+] = (c_t, r_t)$ ， $x_{it} = [x_{it}^-, x_{it}^+] = (c_{it}, r_{it})$ ， $t = 1, 2, \dots, n$ 。由于实际区间中点和半径以及单项预测区间中点和半径数据一般不满足归一化，因此就需要对指标数据进行无量纲化预处理。令

$$d_t = \frac{c_t - \min c_t}{\max c_t - \min c_t}, \quad d_{it} = \frac{c_{it} - \min c_{it}}{\max c_{it} - \min c_{it}}, \quad i = 1, 2, \dots, m, \quad t = 1, 2, \dots, n$$

$$\theta_t = \frac{r_t - \min r_t}{\max r_t - \min r_t}, \quad \theta_{it} = \frac{r_{it} - \min r_{it}}{\max r_{it} - \min r_{it}}, \quad i = 1, 2, \dots, m, \quad t = 1, 2, \dots, n$$

而且易得

$$\hat{d}_t = \sum_{i=1}^m l_i d_{it}, \quad \hat{\theta}_t = \sum_{i=1}^m l_i \theta_{it}, \quad t = 1, 2, \dots, n \tag{1}$$

定义 1: 称 $e_{it} = d_t - d_{it}$ 为第 i 种预测方法区间预测值的中点在第 t 时刻预测的无量纲化指标误差，称 $e_t = d_t - \hat{d}_t$ 为组合预测区间中点在第 t 时刻预测的无量纲化指标误差。称 $f_{it} = \theta_t - \theta_{it}$ 为第 i 种预测方法区间预测值的半径在第 t 时刻预测的无量纲化指标误差，称 $f_t = \theta_t - \hat{\theta}_t$ 为组合预测区间半径在第 t 时刻预测的无量纲化指标误差。由 (1) 有

$$e_t = \sum_{i=1}^m l_i e_{it}, \quad f_t = \sum_{i=1}^m l_i f_{it} \tag{2}$$

定义 2: 称 $\Gamma_{1i}(d_t, d_{it})$ 为第 i 种预测方法预测区间中点值与指标实际区间中点值的算术平均最小贴程度，称 $\Gamma_{2i}(\theta_t, \theta_{it})$ 为第 i 种预测方法预测区间半径与指标实际区间半径的算术平均最小贴程度，称 $\Gamma_1(d_t, \hat{d}_t)$ 为组合预测中点值与指标实际中点值的算术平均最小贴程度，称 $\Gamma_2(\theta_t, \hat{\theta}_t)$ 为组合预测中点值与指标实际中点值的算术平均最小贴程度，其中

$$\Gamma_{1i}(d_t, d_{it}) = 1 - \frac{\sum_{t=1}^n |e_{it}|}{\sum_{t=1}^n (d_t + d_{it})}, \quad \Gamma_{2i}(\theta_t, \theta_{it}) = 1 - \frac{\sum_{t=1}^n |f_{it}|}{\sum_{t=1}^n (\theta_t + \theta_{it})} \tag{3}$$

$$\Gamma_1(d_t, \hat{d}_t) = 1 - \frac{\sum_{t=1}^n \left| \sum_{i=1}^m l_i e_{it} \right|}{\sum_{t=1}^n (d_t + \sum_{i=1}^m l_i d_{it})}, \quad \Gamma_2(\theta_t, \hat{\theta}_t) = 1 - \frac{\sum_{t=1}^n \left| \sum_{i=1}^m l_i f_{it} \right|}{\sum_{t=1}^n (\theta_t + \sum_{i=1}^m l_i \theta_{it})} \tag{4}$$

称 $\Gamma_i = \omega\Gamma_{1i}(d_t, d_{it}) + (1-\omega)\Gamma_{2i}(\theta_t, \theta_{it})$ 为第 i 种预测方法预测值与指标实际值的加权贴进度, 称 $\Gamma = \omega\Gamma_1(d_t, \hat{d}_t) + (1-\omega)\Gamma_2(\theta_t, \hat{\theta}_t)$ 为组合预测值与指标实际值的加权贴进度. ω 为加权系数. 显然, 贴进度 Γ_i 和 Γ 都是 $L = (l_1, l_2, \dots, l_m)^T$ 的函数, 分别记为 $\Gamma_i(l_1, l_2, \dots, l_m)$ 和 $\Gamma(l_1, l_2, \dots, l_m)$.

从区间组合预测模型的角度, 考虑自然希望加权算术平均贴进度 $\Gamma(l_1, l_2, \dots, l_m)$ 越大越好, $\Gamma(l_1, l_2, \dots, l_m)$ 越大表明组合预测方法越有效, 因此, 可以利用区间中点的算术平均贴进度和区间半径的算术平均贴近进行加权, 就得到了基于加权算术平均贴进度的最优组合预测模型为(以下称该模型为加权算术平均贴进度区间组合预测模型)

$$\max \Gamma(l_1, l_2, \dots, l_m) = \omega\Gamma_1(d_t, \hat{d}_t) + (1-\omega)\Gamma_2(\theta_t, \hat{\theta}_t) \tag{5}$$

$$s. \quad t. \quad \sum_{i=1}^m l_i = 1, l_i \geq 0, i = 1, 2, \dots, m \quad \text{记} \quad \Gamma_{\min} = \min\{\Gamma_i, i = 1, 2, \dots, m\},$$

$$\Gamma_{\max} = \max\{\Gamma_i, i = 1, 2, \dots, m\}.$$

定义 3: 如果 $\Gamma(l_1, l_2, \dots, l_m) > \Gamma_{\max}$, 则称由权系数 l_1, l_2, \dots, l_m 所确定的加权贴进度区间组合预测模型为优性区间组合预测模型. 若 $\Gamma_{\min} < \Gamma(l_1, l_2, \dots, l_m) < \Gamma_{\max}$, 则称由权系数 l_1, l_2, \dots, l_m 所确定的加权贴进度区间组合预测模型为非劣性区间组合预测模型, 若 $\Gamma(l_1, l_2, \dots, l_m) < \Gamma_{\min}$ 则称之为劣性区间组合预测模型.

2 非劣性区间组合预测模型

定理 1 加权算术平均贴进度区间组合预测模型(2)的任一可行解对应的区间组合预测模型至少是非劣性区间组合预测模型.

证明 假定 $L = (l_1, l_2, \dots, l_m)^T$ 为加权算术平均贴进度组合预测模型的一个可行解, 则有 $\sum_{i=1}^m l_i = 1$, $l_i \geq 0, i = 1, 2, \dots, m$, 对应于该可行解的区间组合预测模型的加权算术平均贴进度为 $\Gamma(l_1, l_2, \dots, l_m) = \omega\Gamma_1(d_t, \hat{d}_t) + (1-\omega)\Gamma_2(\theta_t, \hat{\theta}_t)$, 为了证明 $\Gamma > \Gamma_{\min}$, 仅需证明 $\Gamma_1(d_t, \hat{d}_t) > \Gamma_{\min}$ 和 $\Gamma_2(\theta_t, \hat{\theta}_t) > \Gamma_{\min}$ 成立. 由于

$$\begin{aligned} 1 - \Gamma_1(d_t, \hat{d}_t) &= \frac{\sum_{t=1}^n \left| \sum_{i=1}^m e_{it} \right|}{\sum_{t=1}^n (d_t + \sum_{i=1}^m l_i d_{it})} \leq \frac{\sum_{i=1}^m l_i \sum_{t=1}^n |e_{it}|}{\sum_{i=1}^m l_i \sum_{t=1}^n (d_t + d_{it})} \\ &= \frac{\sum_{i=1}^m l_i \sum_{t=1}^n (1 - \Gamma_{1i}(d_t, d_{it}))(d_t + d_{it})}{\sum_{i=1}^m l_i \sum_{t=1}^n (d_t + d_{it})} \leq 1 - \Gamma_{\min} \end{aligned} \tag{6}$$

因此 $\Gamma_2 > \Gamma_{\min}$. 累似地有

$$1 - \Gamma_2(\theta_t, \hat{\theta}_t) = \frac{\sum_{t=1}^n \left| \sum_{i=1}^m f_{it} \right|}{\sum_{t=1}^n (\theta_t + \sum_{i=1}^m l_i \theta_{it})} \leq \frac{\sum_{i=1}^m l_i \sum_{t=1}^n |f_{it}|}{\sum_{i=1}^m l_i \sum_{t=1}^n (\theta_t + \theta_{it})}$$

$$= \frac{\sum_{i=1}^m l_i \sum_{t=1}^n (1 - \Gamma_{2i}(\theta_t, \theta_{it})) (\theta_t + \theta_{it})}{\sum_{i=1}^m l_i \sum_{t=1}^n (\theta_t + \theta_{it})} \leq 1 - \Gamma_{\min}$$

故 $\Gamma = \varpi \Gamma_1 + (1 - \varpi) \Gamma_2 > \Gamma_{\min}$ (7)

定理 1 表明在一定条件下加权算术平均贴近度区间组合预测模型(2)的任一个归一化非负权系数所对应的区间组合预测均不会比最差的单项预测方法差。

定理 2 假定 $(l_1^*, l_2^*, \dots, l_m^*)^T$ 是 m 个单项预测模型构成的加权算术平均贴近度组合预测模型(2)的最优解, $(\bar{l}_1, \bar{l}_2, \dots, \bar{l}_m, \bar{l}_{m+1})^T$ 是再增加 1 个单项模型构成的加权算术平均贴近度组合预测模型(2)的最优解. 则有

$$\Gamma(l_1^*, l_2^*, \dots, l_m^*) \leq \bar{\Gamma}(\bar{l}_1, \bar{l}_2, \dots, \bar{l}_m, \bar{l}_{m+1})$$
 (8)

也就是说, 组合预测模型(5)的最大组合预测加权贴近度一定是 m 的非减函数.

证明 注意到 $(l_1^*, l_2^*, \dots, l_m^*, 0)^T$ 是由 $m+1$ 个单项预测方法组成的区间组合预测模型(2)的可行解, 而且 $\bar{\Gamma}(l_1^*, l_2^*, \dots, l_m^*, 0) = \Gamma(l_1^*, l_2^*, \dots, l_m^*)$, 因此定理结论是显然的。

3 实例分析

为了说明本文基于加权贴近度组合预测方法的有效性和可行性, 利用文献[10]的数据进行比较分析, 相关数据见表 1。

表 1 实际区间值及单项预测区间值和区间组合预测值

实际区间值	单项方法 1	单项方法 2	单项方法 3	区间组合预测
[3,4]	[2,4,5]	[3,6,5,4]	[3,3,6]	[3.03,4.23]
[5,5.6]	[2,2,6]	[5, 6]	[4,5,2]	[3.91,5.51]
[4,6]	[3,8]	[5,2,7]	[4,3,5,1]	[4.28,6]
[6,10]	[4,6,11]	[6,8,11,6]	[6,1,7,3]	[6,8.86]
[6.6,8.8]	[6,12,4]	[8,9,6]	[7,8]	[7.05,9.09]
[9,11]	[7,15]	[9,6,12]	[9,1,9,9]	[8.86,11.21]

对表 1 中的数据进行无量纲化处理, 并选取加权系数 $\varpi = 0.5$, 利用 Matlab 软件编程求解可得组合模型的最优权系数为 $l_1 = 0.1683, l_2 = 0.2171, l_3 = 0.6146$, 从而可计算出区间组合预测模型的预测值(表 1 最后一行)。

为了评价组合预测方法的效果, 这里采用平均区间位置误差平方和(Mean Squared Error of Interval Position, MSEPI), 平均区间长度误差平方和(Mean Squared Error of Interval Length, MSEL)以及平均区间相对误差(Mean Relative Interval Error, 即 MRIE)。几种单项预测方法和本文所建立的区间组合预测方法的相关评价指标值见表 2。

表 2 几种预测方法预测效果评价

预测方法	MSEP	MSEL	MRIE
单项预测方法 1	0.3833	0.8367	0.2311
单项预测方法 2	0.0833	0.9233	0.4598
单项预测方法 3	0.5283	0.4333	0.3707
本文区间组合预测方法	0.1312	0.1576	0.1587

从表 2 可以看出, 基于区间中点和半径的算术平均加权贴近度的最优区间组合预测模型的各项误差指标均远低于三种单项预测模型预测误差指标值, 表明本文所采用的加权贴近度的区间组合预测方法能够有效提高预测精度。同时, 本文组合预测模型的各项误差指标与文献 [11] 采用的基于误差平方和最小为准则的区间组合预测模型的各项误差指标也很接近, 表明基于加权贴近度区间组合预测方法是有效的。

4 结论

区间组合预测模型都是基于改善某种拟合误差建立起来的, 文献中从不同的角度提出了基于各类相关性的区间组合预测新方法。本文将算术平均贴近度的概念用于区间组合预测, 提出了利用单项预测区间值的半径和中点的算术平均贴近度的加权作为优化指标, 得到了一类加权贴近度的线性组合预测模型, 并针对该模型提出了非劣性区间组合预测, 优性区间组合预测等概念, 讨论了它们的存在性。从理论和实际应用上说明基于加权贴近度区间组合预测模型的可行性和有效性。

5 致谢

本文为国家社科基金项目《基于灰色系统理论的统计组合预测方法及其应用研究》(11XTJ001) 的阶段性成果之一。

Acknowledgement

This work is supported by the National Social Science Fund Project (11XTJ001)

参考文献:

- [1] 杨桂元, 唐小我, 马永开. 关于非负权重组合预测若干问题的探讨[J]. 电子科技大学学报, 1996, 25(2): 210-215.
- [2] 王应明. 基于相关性的组合预测方法研究[J]. 预测, 2002, 21(2): 58-62.
- [3] 陈华友, 赵佳宝, 刘春林. 基于灰色关联度的组合预测模型的性质[J]. 东南大学学报, 2004, 34(1): 130-134.
- [4] 陈华友. 基于相关系数的优性组合预测模型研究[J]. 系统工程学报, 2006, 21(4): 353-360.
- [5] 程玲华, 陈华友. 基于对数灰关联度的加权几何平均组合预测模型的有效性[J]. 运筹与管理, 2007, 16(6): 69-73.
- [6] 王丰效. 不同优化准则统计组合预测权系数的优化[J]. 数学的实践与认识, 2013, 43(13): 135-139.
- [7] 沈家骅, 严振祥. 基于区间分析的组合预测系数确定方法[J]. 武汉理工大学学报: 交通科学与工程版, 2006, 30(6): 107-109.
- [8] 江立辉, 陈华友, 丁芳清. 基于 IOWC-GOWA 算子的区间组合预测模型[J]. 计算机工程与应用, 2015, 51(3): 50-54.
- [9] 王晓, 刘兮, 陈华友. 基于 IOWA 算子的区间组合预测方法[J]. 武汉理工大学学报: 信息与管理工程版, 2010, 32(2): 221-225.
- [10] 陶志富, 张进, 陈华友. 基于向量夹角余弦的区间组合预测多目标规划方法[J]. 西华大学学报, 2010, 29(3): 35-37.

References:

- [1] Yang Gui-yuan, Tang Xiao-wo, Ma Yong-kai: “Research on several problems for combined prediction of non-negative weight” [J]. Journal of University of Electronic Science and Technology of China, Vol.25(1996), No.2, pp. 210-215 (in Chinese)
- [2] Wang Ying-ming: “Research on the methods of combining forecasts based on correlativity” [J]. Forecasting, Vol.21 (2002), No.2, pp. 58-62(in Chinese)
- [3] Chen Hua-you, Zhao Jia-bao and C. L. Liu Chun-lin: “Properties of combination forecasting model based on degree of grey incidence” [J]. Journal of Southeast University, Vol.34 (2004), No.1,pp. 130-134(in Chinese)
- [4] Chen Hua-you: “Research on properties of superior combined forecasting models based on correlation coefficients” [J]. Journal of Systems Engineering, Vol. 21(2006), No.4, pp. 353-360 (in Chinese)
- [5] Cheng Ling-hua, Chen Hua-you: “Efficiency of weighted geometric means combination forecasting model based on degree of logarithm grey incidence” [J]. Research Operations Research and Management Science, vol. 16(2007), No.6, pp. 69-73 (in Chinese)
- [6] Wang Feng-xiao: “Weight coefficient optimization of combination forecast by different criteria” [J]. Mathematics in Practice and Theory, vol. 43(2013), No.13, pp. 135-139 (in Chinese)
- [7] Shen Jia-hua, Z. X. Yan Zhen-xiang: “A method of finding weight in combination forecasting model based on interval analysis” [J]. Journal of Wuhan University of Technology, vol. 30(2006),No.6, pp. 1077-1080 (in Chinese)
- [8] Jiang Li-hui, Chen Hua-you, Ding Fang-qing: “Combination forecasting model based on Induced Ordered Weighted Continuous Generalized Ordered Weighted Averaging operator” [J]. Computer Engineering and Applications, Vol.51 (2015), No.3, pp.50-54 (in Chinese)
- [9] Wang Xiao, Liu Xi, Chen Hua-you: “An Interval Combination Forecasting Method Based on IOWA Operators”[J]. Journal of WUT (Information and management engineering), Vol.32(2010), No.2, pp.221-225(in Chinese)
- [10] Tao Zhi-fu, Zhang Jin, Chen Hua-you: “Multi-objective programming method of interval combination forecasting based on vectorial angle cosine” [J]. Journal of Xihua University, Vol. 29(2010), No.3, pp. 35-37(in Chinese)