

# 基于 Shapley 值的农机总动力组合预测方法<sup>\*</sup>

张淑娟 赵 飞

**【摘要】** 应用合作博弈中的 Shapley 值方法,通过分配总误差来确定组合预测模型中各预测模型的权重,以此构建组合预测模型并对山西省农机总动力进行组合预测。结果表明,该组合预测模型的预测平均误差为 1.81,低于选定的一元线性回归模型、三次指数平滑模型和 BP 神经网络模型的 3.12、2.42 和 2.23;也低于基于方差倒数法以及基于离异系数法构建的组合预测模型的预测平均误差 1.88 和 1.90。

**关键词:** 农机总动力 Shapley 值 组合预测

**中图分类号:** S232.3; C931.1

**文献标识码:** A

## Combinatorial Forecast of Agricultural Machinery Total Power Based on Shapley Value

Zhang Shujuan Zhao Fei

(Shanxi Agricultural University, Taigu, 030801, China)

### Abstract

A combinatorial model was established to forecast the total power of agricultural machinery in Shanxi province. Applying the Shapley value method of cooperative game, the weight of each forecast model in the combinatorial forecast model, was determined by distributing the total combinatorial error. The results showed that the forecast average error of combinatorial forecast model is 1.81, which is lower than 3.12, 2.42 and 2.23 of one-variable liner regression model, cubic exponent smooth model and BP neural network model, and is also lower than 1.88 and 1.90 of the combinatorial forecast model based on the variance reciprocal method and the divergence coefficient method.

**Key words** Total power of agricultural machinery, Shapley value, Combinatorial forecast

### 引言

农业机械总动力是指用于农、林、牧、渔业生产和运输的所有机械动力之和,它反映了一个地区农业机械化发展的总体水平,是农业机械化发展规划的主要指标。农机总动力预测就是通过建立动力变量与时间变量之间的稳定关系,来推算动力变量未来发展趋势。目前,国内外已提出的预测方法多达几十种,如移动平均法、指数平滑法、趋势分析法、线性回归分析法、曲线拟合法、人工神经网络法及灰色预测法等。这些方法各有其优点和不足,所得的结

果也存在差异<sup>[1-4]</sup>。

组合预测的基本思想是把不同模型的计算结果综合起来,取长补短,达到提高预测精度和预测结果可靠性的效果。组合预测的关键问题是确定各预测模型的权重,合理的权重会提高预测精度。各预测模型分配的权重应反映这种预测方法对总预测结果贡献的大小:误差越大,预测效果越差,则在组合预测中的权重越小;预测误差越小,预测效果越好,则在组合预测中的权重越大<sup>[5-6]</sup>。

本文选用一元线性回归模型、三次指数平滑模型和 BP 神经网络模型 3 种预测模型,应用 Shapley

收稿日期: 2007-05-16

<sup>\*</sup> 山西省高校科技研究开发资助项目(项目编号: 20051220)

张淑娟 山西农业大学工程技术学院 教授 博士生导师, 030801 太谷县

赵 飞 山西农业大学工程技术学院 硕士生

值权重分配法<sup>[7-8]</sup>确定各预测模型的权重,从而构建组合预测模型并对山西省农机总动力进行组合预测,同时与基于方差倒数法构建的组合预测模型以及基于离异系数法构建的组合预测模型进行预测精度比较,以获得农业机械总动力预测的实用新方法。

### 1 预测模型

#### 1.1 一元线性回归模型

由表 1 可知,山西省农业机械总动力的历史统计数据随时间不断增加,因此采用时间序列一元线性回归分析法<sup>[1]</sup>进行预测。采用 SAS 统计软件建立预测模型

$$Y = 359.11966 + 61.24709T \quad (1)$$

式中  $Y$ ——农业机械总动力

$T$ ——时间变量,1979~2004 年的对应取值分别为 1~26

方差分析表明回归模型的显著性检验尾概率小于 0.000 1,决定系数  $R^2 = 0.9805$ ,模型极显著,并具有较高的拟合精度。

表 1 山西省农业机械总动力

Tab.1 Total power of agricultural machinery in Shanxi province

年份	Y/万 kW	年份	Y/万 kW	年份	Y/万 kW
1979	513.69	1988	922.44	1997	1463.22
1980	543.44	1989	985.90	1998	1506.67
1981	574.08	1990	1053.51	1999	1655.00
1982	603.01	1991	1089.00	2000	1701.30
1983	677.15	1992	1152.49	2001	1767.51
1984	774.94	1993	1211.84	2002	1869.39
1985	822.71	1994	1298.36	2003	1928.21
1986	852.92	1995	1359.60	2004	2186.48
1987	895.76	1996	1426.22		

#### 1.2 三次指数平滑模型

为了使模型更能精确地反映农机总动力的发展趋势,选用了三次指数平滑预测法<sup>[1]</sup>,计算公式为

$$y_{t+T'} = a_t + b_t T' + c_t T'^2 \quad (2)$$

式中  $y_{t+T'}$ ——农机总动力预测值

$T', t$ ——预测的时间周期数和预测起始年

$a_t, b_t, c_t$ ——三次指数平滑的平滑系数

$$\begin{cases} a_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)} \\ b_t = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_t^{(1)} - 2(5-4\alpha)S_t^{(2)} + (4-3\alpha)S_t^{(3)}] \\ c_t = \frac{\alpha^2}{2(1-\alpha)^2} (S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}) \end{cases} \quad (3)$$

式中  $\alpha$ ——指数平滑加权系数,一般在 0.01~0.5 之间

$S_t^{(1)}, S_t^{(2)}, S_t^{(3)}$ ——起始年  $t$  的一次、二次、三次指数平滑值

为了使农机装备近期较快的发展趋势在预测模型中得到充分反映, $\alpha$  的取值应大一些,经反复测算比较后,选取  $\alpha = 0.4$ 。指数平滑值计算公式为

$$\begin{cases} S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} \\ S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)S_{t-1}^{(2)} \\ S_t^{(3)} = \alpha S_t^{(2)} + (1-\alpha)S_{t-1}^{(3)} \end{cases} \quad (4)$$

式中  $y_t$ ——起始年  $t$  的对应原始数据

初始平滑值  $S_0^{(1)}, S_0^{(2)}, S_0^{(3)}$  计算公式为

$$\begin{cases} S_0^{(1)} = a - \frac{b(1-\alpha)}{\alpha} + \frac{c(1-\alpha)(2-\alpha)}{\alpha^2} \\ S_0^{(2)} = a - \frac{2b(1-\alpha)}{\alpha} + \frac{2c(1-\alpha)(3-2\alpha)}{\alpha^2} \\ S_0^{(3)} = a - \frac{3b(1-\alpha)}{\alpha} + \frac{3c(1-\alpha)(4-3\alpha)}{\alpha^2} \end{cases} \quad (5)$$

式中,参数估计量  $a, b, c$  用待定系数法估算,将  $t_1 = 1, t_2 = 2, t_3 = 3$  代入方程组

$$\begin{cases} y_{t_1} = a + bt_1 + ct_1^2 \\ y_{t_2} = a + bt_2 + ct_2^2 \\ y_{t_3} = a + bt_3 + ct_3^2 \end{cases} \quad (6)$$

采用逐步改进法,依据已有数据计算。

根据表 1 中 1979~2004 年的农业机械总动力数据,利用式(2)~(6)进行计算,且令  $t = 2004$ ,则山西省农业机械总动力的三次指数平滑预测公式为

$$y_{2004+T'} = 2149.17 + 147.17T' + 6.28T'^2 \quad (7)$$

预测结果见表 2。

#### 1.3 BP 神经网络模型

BP 神经网络在预测领域中的应用较为广泛<sup>[3-4]</sup>,本文采用 3 层 BP 网络建模进行预测。根据山西省 1979~2004 年农业机械化总动力的历史数据,以前 4 年的农机总动力数据作为输入,即输入层选定 4 个节点。用第 5 年的农机总动力作为输出数据,即输出层节点数为 1,构造输入输出样本对:网络的输入为  $X_{t'-4}, X_{t'-3}, X_{t'-2}, X_{t'-1}$ ,输出为  $X_{t'}$ ,其中  $t' = 5, 6, 7, 8 \dots$ 。

网络隐含层节点数的确定参考隐含层单元数计算公式<sup>[4]</sup>

$$n_1 = \sqrt{n+m} + r \quad (8)$$

式中  $n, m$ ——输入、输出神经元个数

$r$ ——1~10 之间的常数

因此,隐含层单元数为

$$\sqrt{4+1}+1 \leq n_1 \leq \sqrt{4+1}+10$$

$$\text{即 } 3.24 \leq n_1 \leq 12.24$$

根据  $n_1$  的计算值,由小到大改变节点数,训练并检验其精度。当节点数增加到误差不再减小时,其临界值即为应采用的数值。最后,选定网络隐含层节点数为 4,此时网络能较快地收敛至所要求的

精度。因此,该选定网络的拓扑结构为 4-4-1。

取 1994~2004 年山西省农机总动力历史数据作为网络校验数据,1979~1993 年的数据用于网络训练。当误差给定为  $E = 0.000\ 01$ ,学习步长为 0.5,动量项为 0.5,经 50 000 次训练,网络精度达到要求。用检验样本集进行预测检验,结果见表 2。

表 2 各预测模型的预测结果与误差

Tab.2 Forecast result and error of single forecast method

万 kW

年份	实际值	一元线性回归预测结果	三次指数平滑预测结果	BP 神经网络预测结果	预测误差		
					一元线性回归模型	三次指数平滑模型	BP 神经网络模型
1994	1 298.36	1 339.07	1 269.68	1 263.25	40.71	-28.68	-35.11
1995	1 359.60	1 400.32	1 364.61	1 317.18	40.72	5.01	-42.42
1996	1 426.22	1 461.57	1 434.41	1 394.06	35.35	8.19	-32.16
1997	1 463.22	1 522.81	1 501.29	1 459.82	59.59	38.07	-3.40
1998	1 506.67	1 584.06	1 531.38	1 540.72	77.39	24.71	34.05
1999	1 655.00	1 645.31	1 561.69	1 597.50	-9.69	-93.31	-57.50
2000	1 701.30	1 706.56	1 721.77	1 666.19	5.26	20.47	-35.11
2001	1 767.51	1 767.80	1 788.57	1 800.01	0.30	21.06	32.50
2002	1 869.39	1 829.05	1 849.30	1 880.79	-40.34	-20.09	11.40
2003	1 928.21	1 890.30	1 952.38	2 039.27	-37.91	24.17	111.06
2004	2 186.48	1 951.54	2 013.77	2 182.41	-234.94	-172.71	-4.07
平均值					52.93	41.50	36.25

## 2 组合预测模型

### 2.1 模型的建立

设对于同一预测问题,用  $N$  种不同的预测模型分别进行预测,则由这  $N$  个模型构成的组合预测模型<sup>[5]</sup>为

$$f_t = \sum K_i f_{it} \quad (9)$$

式中  $f_t$ ——组合预测模型的预测值

$f_{it}$ ——第  $i$  种预测模型的预测值,  $i = 1, 2, \dots, N$

$K_i$ ——第  $i$  种模型的权重,  $i = 1, 2, \dots, N$ ,

且  $\sum K_i = 1$

### 2.2 最优权重的计算方法

Shapley 值法是一种用于解决多人合作对策问题的数学方法<sup>[7~8]</sup>。它主要应用于合作收益在各合作方之间的分配问题,Shapley 值实现的是每个合作成员对该合作联盟的贡献大小,突出反映了各个成员在合作中的重要性。Shapley 值法的最大优点在于其原理和结果易于被各合作方视为公平,结果易于被各方接受。因此,本文应用 Shapley 值权重分配法来确定各模型预测的权重。

组合预测中各预测方法之间共同作用产生了组

合预测的总误差。可以认为各预测方法之间为了共同目的而形成“合作关系”。把组合预测所产生的总误差看作合作的“总收益”。把这个总收益向各预测方法进行分配,利用分配的结果来确定各预测方法在组合预测中的权重。

有  $N$  种预测方法来进行组合预测,则记为  $I = \{1, 2, \dots, N\}$ ,对于  $I$  的任何子集  $u, v$  (表示  $N$  种方法中的任一组合),  $E(u), E(v)$  表示各自组合的误差。定义如下:

(1) 对于  $I$  的任一子集  $u, v$ , 都有  $E(u) + E(v) \geq E(u \cup v)$ 。

(2)  $N$  种预测方法参与组合预测产生的总误差  $E$ , 将在  $N$  种预测方法之间进行完全分配, 即  $E = \sum_{i \in N} E'_i$ , 其中  $E'_i$  为模型  $i$  分得的误差, 即 Shapley 值,  $i = 1, 2, \dots, N$ 。

设第  $i$  种预测方法预测误差的绝对值平均值为  $E_i, E'_i \leq E_i$ , 并有

$$E = \frac{1}{N} \sum_{i=1}^N E_i \quad (10)$$

Shapley 值法的权重分配公式为

$$E'_i = \sum_{u_i \in u} \omega(|u_i|) [E(u_i) - E(u_i - \{i\})] \quad (11)$$

其中  $w(|u_i|) = \frac{(N - |u_i|)! (|u_i| - 1)!}{N!}$   
 式中  $w(|u_i|)$ ——加权因子,表示组合中  $i$  应承担的组合边际贡献  
 $u_i - \{i\}$ ——组合中去除模型  $i$   
 $u$ ——包含  $i$  的所有子集  
 $|u_i|$ ——组合中预测模型的个数

由式(11)可得出组合预测中各预测方法的权重计算公式

$$w_i = \frac{E - E'_i}{E(N-1)} \quad (i = 1, 2, \dots, N) \quad (12)$$

选取山西省农机总动力的历史数据序列,取1994~2004年的数据为样本,用  $y_1$ 、 $y_2$ 、 $y_3$  分别代表一元线性回归、三次指数平滑、BP神经网络3种不同的预测模型对农机总动力进行预测,并将3种预测模型的结果与实际值进行比较,预测误差见表2。

通过比较上述3种模型预测误差绝对值的平均值可以看出,一元线性回归模型误差最大,三次指数平滑模型次之,神经网络模型的误差最小。按表2的计算结果,根据式(10)可得组合预测的总误差为:  
 $E = (52.93 + 41.50 + 36.25) / 3 = 43.56$ 。

根据 Shapley 值的概念,参与组合预测模型总误差分配的“合作关系”成员为: $I = \{1, 2, 3\}$ ,它所有子集的组合误差分别为  $E(\{1\})$ 、 $E(\{2\})$ 、 $E(\{3\})$ 、 $E(\{1, 2\})$ 、 $E(\{1, 3\})$ 、 $E(\{2, 3\})$ 、 $E(\{1, 2, 3\})$ ,其数值的大小为该子集所包括的向量误差平均值。

则根据式(11)计算各成员的 Shapley 值

$$E'_1 = \frac{2! \times 0!}{3!} [E(\{1\}) - E(\{1\} - \{1\})] + \frac{1! \times 1!}{3!} [E(\{1, 2\}) - E(\{1, 2\} - \{1\})] + \frac{1! \times 1!}{3!} [E(\{1, 3\}) - E(\{1, 3\} - \{1\})] + \frac{0! \times 2!}{3!} [E(\{1, 2, 3\}) - E(\{1, 2, 3\} - \{1\})] = \frac{1}{3}(52.93 - 0) + \frac{1}{6}(47.215 - 41.50) + \frac{1}{6}(44.59 - 36.25) + \frac{1}{3}(43.56 - 38.875) = 21.5475$$

同样可得预测模型  $y_2$ 、 $y_3$  应当分摊的误差量为: $E'_2 = 12.975$ ,  $E'_3 = 9.0375$ ,而  $E'_1 + E'_2 + E'_3 = 43.56$ ,说明3种预测方法分配的误差之和等于总误差  $E$ 。各个分配值的大小反映了各预测模型的精度大小。

各预测方法在组合模型中的最终权重依据式(12)计算为: $w_1 = 0.25$ ,  $w_2 = 0.35$ ,  $w_3 = 0.40$ 。

组合预测模型 I 为: $Y = 0.25y_1 + 0.35y_2 + 0.40y_3$ 。利用组合预测模型 I 对1994~2004年的农机总动力进行预测,并计算一元线性回归、三次指数平滑、BP神经网络3种预测模型和组合预测模型的相对误差,见表3。

### 2.3 3种权重确定方法比较

为了进一步证实基于 Shapley 值法的组合预测模型 I 的预测效果,基于方差倒数法构建组合预测模型 II,计算其权重系数为: $w_1 = 0.17$ ,  $w_2 = 0.29$ ,

表3 组合预测模型的预测结果及各预测模型的相对误差

Tab.3 Forecast result of the combinatorial forecast model and relative error of each combinatorial forecast model

年份	实际值 /万 kW	组合预测 模型 I 结果 /万 kW	相对误差/%					
			组合预测 模型 I	组合预测 模型 II	组合预测 模型 III	一元线性 回归模型	三次指数 平滑模型	BP 神经 网络模型
1994	1 298.36	1 284.46	1.08	1.59	0.84	3.04	2.26	2.78
1995	1 359.60	1 354.57	0.37	1.08	0.23	2.91	0.37	3.22
1996	1 426.22	1 425.06	0.08	0.63	0.40	2.42	0.57	2.31
1997	1 463.22	1 490.08	1.80	1.30	2.25	3.91	2.54	0.23
1998	1 506.67	1 548.29	2.69	2.50	2.70	4.89	1.61	2.21
1999	1 655.00	1 596.92	3.64	3.75	3.66	0.59	5.97	3.60
2000	1 701.30	1 695.74	0.33	0.72	0.14	0.31	1.19	2.11
2001	1 767.51	1 787.95	1.14	1.32	1.01	0.02	1.18	1.81
2002	1 869.39	1 856.83	0.68	0.35	0.97	2.21	1.09	0.61
2003	1 928.21	1 971.62	2.20	3.04	1.46	2.01	1.24	5.45
2004	2 186.48	2 065.67	5.85	4.40	7.26	12.04	8.58	0.19
平均值			1.81	1.88	1.90	3.12	2.42	2.23

$\omega_3 = 0.54$ ; 基于离异系数法构建组合预测模型 III, 计算其权重系数为:  $\omega_1 = 0.28$ ,  $\omega_2 = 0.47$ ,  $\omega_3 = 0.25$ 。分别计算这两种预测模型的预测值(篇幅所限, 没列出), 并计算其相对误差, 见表 3。

由表 3 可知, 组合预测模型 I 的预测平均误差为 1.81, 而神经网络模型的预测平均误差为 2.23。因此, 该组合预测模型 I 的预测精度高于 3 个单一模型的预测精度。计算组合预测模型 II 的预测平均误差为 1.88, 组合预测模型 III 的预测平均误差为 1.90, 因此, 组合预测模型 I 的预测精度也高于组合预测模型 II 和 III, 具有更高的预测精度。

采用组合预测模型 I 对 2010、2015 年的山西省农机总动力进行预测, 得到其农机总动力分别为 2 487.64、3 073.19 万 kW。

### 3 结论

(1) 对山西省农机总动力进行预测, 采用 3 种预测模型。各预测模型的预测精度从高到低依次为: BP 神经网络模型、三次指数平滑模型、一元线性回归模型。

(2) 应用 Shapley 值权重分配法来确定各预测方法的权重, 构建组合预测模型 I, 对山西省农机总动力进行组合预测。该方法具有较好的预测精度, 其预测精度高于选定的各预测模型的预测结果。

(3) 基于 Shapley 值的组合预测方法精度高于基于方差倒数法以及基于离异系数法构建的组合预测方法, 因此, 为农业机械总动力预测提供了一种实用新方法。

### 参 考 文 献

- 1 陈玉祥, 张汉亚. 预测技术与应用[M]. 北京: 机械工业出版社, 1985.  
Chen Yuxiang, Zhang Hanya. Forecasting technology and application[M]. Beijing: Mechanical Industry Publishing House, 1985. (in Chinese)
- 2 白丽, 李行, 马成林. 2005~2015 年吉林省农机化作业水平定量预测[J]. 农业机械学报, 2005, 36(9): 64~67.  
Bai Li, Li Hang, Ma Chenglin. Quantificational forecast of farm mechanization level in Jilin province from year 2005 to 2015 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(9): 64~67. (in Chinese)
- 3 陈丽能, 谢永良. 基于 BP 神经网络的农机拥有量预测技术[J]. 农业机械学报, 2001, 32(1): 118~121.  
Chen Lineng, Xie Yongliang. Agricultural machinery possessive quantity prediction technique based on BP neural networks [J]. Transactions of the Chinese Society for Agricultural Machinery, 2001, 32(1): 118~121. (in Chinese)
- 4 张淑娟, 何勇, 方慧. 人工神经网络在作物产量与土壤空间分布信息关系分析中的应用[J]. 系统工程理论与实践, 2003, 23(12): 121~127.  
Zhang Shujuan, He Yong, Fang Hui. Application of artificial neural network on relationship analysis of crop yield and soil space distributing information[J]. Systems Engineering—Theory & Practice, 2003, 23(12): 121~127. (in Chinese)
- 5 唐小我, 马永开, 曾勇, 等. 现代组合预测和组合投资决策方法[M]. 北京: 科学出版社, 2003.  
Tang Xiaowo, Ma Yongkai, Zeng Yong, et al. Modern combination forecasting and investing decision-making methods[M]. Beijing: Science Press, 2003. (in Chinese)
- 6 项静恬, 史久恩. 非线性系统中数据处理的统计方法[M]. 北京: 科学出版社, 1997.  
Xiang Jingtian, Shi Jiuen. Statistical methods of data processing in nonlinear system[M]. Beijing: Science Press, 1997. (in Chinese)
- 7 陈华友. 组合预测权系数确定的一种合作对策方法[J]. 预测, 2003, 22(1): 75~77, 32.  
Chen Huayou. A kind of cooperative games method determining weights of combination forecasting[J]. Forecasting, 2003, 22(1): 75~77, 32. (in Chinese)
- 8 卢奇, 顾陪亮, 邱世明. 组合预测模型在我国能源消费系统中的构建及应用[J]. 系统工程理论与实践, 2003, 23(3): 24~30.  
Lu Qi, Gu Peiliang, Qiu Shiming. The construction and application of combination forecasting model in Chinese energy consumption system[J]. Systems Engineering—Theory & Practice, 2003, 23(3): 24~30. (in Chinese)