Article ID: 1000-8152(2001)03-0369-06

Research on the Technique of Nonlinear Combination Modeling and Forecasting Based on Fuzzy Inference System*

DONG Jingrong
(Department of Mathematics and Computer Science, Chongqing Normal College · Chongqing, 400047, P.R.China;
College of Bussiness Management, Chongqing University · Chongqing, 400044, P.R.China)

Abstract: Based on the property that fuzzy inference system can uniformly approximate any nonlinear multivariable continuous function arbitrarily well, a new nonlinear combination forecasting method is presented to overcome the difficulties and drawbacks in combined modeling non-stationary time series by using linear combination forecasting method. Furthermore, the optimization algorithm based on a hierarchical structure of learning automata is used to identify the membership functions in the antecedent part and the real numbers in consequent part of the inference rule. Theoretical analysis and forecasting results related to numerical examples all show that the new technique has reinforcement learning properties and universalized capabilities. With respect to combined modeling and forecasting of non-stationary time series in nonlinear systems, which has some uncertainties, the method has the excellent identification performance and forecasting accuracy superior to other existing linear combining forecasts for the same event.

Key words: nonlinear combination forecasting; fuzzy inference system; a hierarchical structure of learning automata
Document code: A

基于模糊推理系统的非线性组合建模与预测方法研究

董景荣
(重庆师范学院数学与计算机科学系·重庆,400047;重庆大学工商管理学院·重庆,400044)

摘要: 基于模糊推理系统在紧支集中能够逼近任意非线性连续函数的特性,提出了一种基于 Takagi-sugeno 模糊规则基的非线性组合建模与预测新方法,以克服线性组合预测方法在解决非平衡时间序列组合建模问题所遇到的困难和存在的不足,并给出了相应的基于学习自动机层次结构的优化算法确定模糊系统的参数和模糊子集的划分,理论分析和大量的经济预测实例表明:该方法具有很强的学习与泛化能力,在处理诸如经济时间序列这种具有一定程度不确定性的非线性系统组合建模与预测方法有很好的应用.

关键词: 非线性组合预测; 模糊推理系统; 学习自动机层次结构

1 Introduction

One of the most important functions of management at all levels in an organization is planning, and forecasting plays a crucial role in the planning process. Forecasting can asist a manager or planner to identify organization strategies to influence the future in a way that will fulfil organization's business objectives. Therefore, Accurate forecasts are essential for risk reduction and fulfilling business objectives. Combining forecasts is an accepted means of improving forecasts.

Since J.N. Bate and C.W.J. Granger originally proposed the idea and method of combining individual forecasts in a single overall assessment in 1969[1], much progress has been made in the field of forecast combination[2~6]. According to different ways of combining individual forecasts, the existing combining forecasts can be classified in two major types, namely linear combining forecasts and nonlinear combining forecasts. A quick glance at the combining literature, however, suggests that it should have focused almost exclusively on linear combination of forecasts with linear constraints or unrestricted, with little attention paid to how combining can be undertaken in the context of nonlinear combination of forecasting variable. Owing to nonlinear essence imply-

* Foundation item: supported by National Science Foundation (79770105).
Received date: 1999-11-30; Revised date: 2000-12-13.

ing among the data or individual forecasts, these linear combination forecasting methods suffers from several deficiencies and limitations, which make them severely inadequate for strategic organization planning[7]. It is therefore desirable to develop a new combination forecasting method that can overcome these deficiencies and limitations.

Recently, there have been some attempts to solve these problems by the application of nonlinear combining forecasts. If the actual values of some forecasting problem in a period are $y_t(t = 1,2,\cdots,n)$ and there are m kinds of feasible individual forecasting methods around it, whose forecasting values are $f_{tj}(t = 1,2,\cdots,n; j = 1,2,\cdots,m)$, respectively, then nonlinear combination forecasting model can be expressed by a general nonlinear function as follows:

$$f_t = g[f_{t1}, f_{t2}, \cdots, f_{tm}], \tag{1}$$

where g is a nonlinear function and f_t with higher accuracy or more excellent performance than these individual forecasts according to some measure. Unfortunately, nonlinear combining forecast hasn't been used widely as effective tools in various practical applications, because it is difficult for us to establish effecting nonlinear function g.

In recently years, the issue of fuzzy inference systems as universal approximators has drawn significant attention and progress has been made. References [8,9] have proved that fuzzy inference systems based on if-then rules are universal approximators of nonlinear mapping. Reference [10] also proves that Takagi-Sugeno fuzzy inference systems can uniformly approximate any multivariate continuous function arbitrarily well and they are suitable for the modeling of complex nonlinear systems[11]. So, the objective of this paper is to use the fuzzy inference system composed of Takagi-Sugeno rule base with constant output, a parametrized gaussian membership function, and simple multiplication operator as approximators of nonlinear mapping to simulate nonlinear combination function g of individual forecasts.

In this paper, A new nonlinear combining forecasting method based on fuzzy inference system is presented. Furthermore, a hierarchical structure of learning automata is used to automatically identify the parameter of the fuzzy model and partitions of fuzzy subsets. Theoretical analysis and forecasting examples all show that the new technique is feasible and effective.

2 Fuzzy inference system

The fuzzy inference system is a popular framework based on the concepts of fuzzy set theory, fuzzy IF-THEN rules and fuzzy reasoning. With respect to fuzzy control and modeling application, the existing fuzzy inference systems can be classified in two major types, namely Mamdani fuzzy inference systems and Takagi-Sugeno fuzzy inference systems. Takagi-Sugeno rules differ from Mamdani rules in that their "outputs" are not defined by membership function but by non-fuzzy analytical functions (frequently, constant or affine functions). This feature should permit to exppress complicated knowledge with small number of rules. From mathematics point of view, Takagi-Sugeno fuzzy rules base is just functions mapping their input to output. For Takagi-Sugeno fuzzy rules base used as models, a fuzzy system can always be established which is capable of approximating any continuous and nonlinear physical system arbitarily well.

In this paper, the Takagi-Sugeno rule type with constant output are employed for the modeling of n-input and single-output combination forecasting system. Each rule R_i can be written as follows:

R_i: IF x_1 is $A_{i1}, \cdots x_m$ is A_{im},

THEN y is $w_i = f_i(x_1, x_2, \cdots x_m)$, $i = 1, \cdots, n$. (2)

where R_i is a label of the i-th rule, $x^T = (x_1, \cdots, x_m)$ is the input vector, A_{ij} are the membership functions for jth input of ith rule, y the output variable, W_i the real value of output for ith rule, n the total number of rules, and m the total number of input variables.

The type of A_{ij} in the antecedent part in Eq.(2) is parametrized gaussian membership function expressed as follows:

$$A_{ij}(x_j) = \exp\left[-\left(\frac{x_i - a_{ij}}{b_{ij}}\right)^2\right]. \tag{3}$$

a_{ij} and b_{ij} in Eq.(3) are the center point and the width of the membership function, respectively.

The membership value for ith rule, μ_i, is defined through simple multiplication operations such as

$$\mu_i(x) = \prod_{j=1}^{m} A_{ij}(x_j). \tag{4}$$

The output of the rule base can be obtained from the center of gravity method such as

$$y = \left(\sum_{i=1}^{n} u_i w_i\right) / \left(\sum_{i=1}^{n} u_i\right). \tag{5}$$

The learning of the rule base is realized from a set of learning samples. Let us define the quadratic error E:

$$E = \frac{1}{2}\sum_{p=1}^{D}(y_p - y_p^d)^2, \tag{6}$$

where D is the total number of training data, y_p is the output calculated from procedures of Eqs. (2) ~ (5) for pth training data $x_p = (x_1, \cdots, x_m)_p$ and y_p^d is the desired or reference output value.

The purpose of learning is to minimize the quadratic error E. Minimization of E is performed by tuning the a_{ij}, b_{ij} and w_1 parameters. This is a supervised learning scheme, therefore, a hierarchical structure of learning automata could be used for the simultaneous identification of these parameters.

3 Learning algorithm

3.1 The hierarhical structure of automata

The automaton is composed of a performance valuation unit and a hierarchical structure of stochastic automata with variable structures. It may be connected in a feedback loop to the random environment.

A learning automaton is a sextuple $\{W, A, B, P, R, G\}$. $W(t)$ is the response of the performance evaluation system. $A = \{a_1, \cdots, a_s\}^T$ is the set of internal states. $B = \{b_1, \cdots, b_N\}^T$ with $N \leqslant s$ is the output or the set actions. $P(t) = \{p_1(t), \cdots, p_N(t)\}^T$ is the state probability distribution at iteration t. R is the learning algorithm (reinforcement scheme or updating scheme) that changes the probability vector from $p(t)$ to $p(t+1)$. G is the mapping of the set A onto the set B ($G: A \rightarrow B$).

The hierarchical system of automata is at different levels that are composed of single automaton with a limited number of actions (N)[12].

The first level consists of a single automaton with N internal states, the second one consists of N single automata (of N actions each), and the kth level is formed by (N^{k-1}) automata (Fig.1).

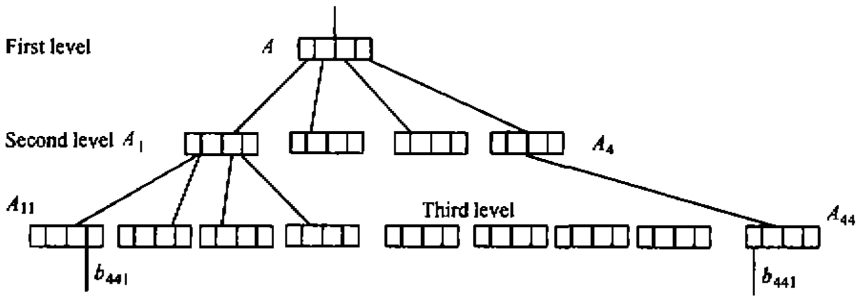

Fig. 1 Hierarchical structure of learning automata

Let us consider a hierarchical system of automata with N levels. the variation domain of the control variable is discretized in N^N intervals. The actions chosen at the various levels are denoted by $(b_{j1}, b_{j1j2}, \cdots, b_{j1j2\cdots jN})$, and so the corresponding automata are A, $A_{j1}, A_{j1j2}, \cdots, A_{j1j2\cdots(jN-1)}$.

If the action selected at iteration t by the first level automaton is b_{j1} the automaton concerned is A and the probability distribution will be adapted according to the following relations[12].

If $w(t) = 0$ (reward):

$$\begin{cases} p_{j1}(t+1) = p_j(t) + \eta\{p_{j1}(t)\}[1 - p_{j1}(t)], \\ p_{m1}(t+1) = p_{m1}(t) - [\eta\{p_j(t)\}]. \end{cases} \tag{7}$$

If $w(t) = 1$ (penalty):

$$\begin{cases} p_{j1}(t+1) = p_{j1}(t), \\ p_{m1}(t+1) = p_{m1}(t). \end{cases} \tag{8}$$

with $m1 \neq j1$ (actions not selected at time t).

By assumption, $\mu\{p(t)\}$ is positive for all $p(t)(0 < \eta\{p(t)\} < 1)$. To carry out the adaptation mechanism for the Nth level, let us introduce the following notations:

$j_1 j_2 \cdots jN$ is denoted by JN,

$m_1 m_2 \cdots mN$ is denoted by MN.

In the Nth level, if the selected action at time t is b_{JN} the automaton concerned is denoted by A_{JN-1}. The probability distribution is adjusted by the following algorithm.

If $w(t) = 0$ (reward):

$$\begin{cases} p_{JN}(t+1) = p_{JN}(t) + \mu_{JN-1}\{p(t)\}[1 - p_{JN}(t)], \\ p_{JN-1mi}(t+1) = p_{JN-1mi}(t)[1 - \mu_{JN-1}\{p(t)\}]. \end{cases} \tag{9}$$

If $w(t) = 1$ (penalty):

$$\begin{cases} p_{jN}(t+1) = p_{jN}(t), \\ p_{JN-1mi}(t+1) = p_{JN-1mi}(t), \end{cases} \tag{10}$$

with $mi \neq JN$.

For all the other actions

$$PMN(t+1) = PMN(t), \text{ for } MN \neq JN, JN-1mi. \tag{11}$$

A $\mu\{\cdot\}$ function is associated with every automaton in the hierarchy. The following algorithm[12] which ensures ϵ- optimality is chosen to adapt the $\mu\{\cdot\}$ functions at every level:

$$\mu_{JN}\{t\} = \mu_{JN-1}\{t\}/p_{JN}(t+1). \tag{12}$$

In summary, based on the probability distribution, the first level automaton randomly selects an action b_{j1}. This, in turn, activates the automaton A_{ji} at the second level, which chooses an action b_{j1j2} from its action probability distribution. Consequently, the automaton A_{j1j2} is activated, etc. The final action selected at the last level generates an action that corresponds to a value of the rule parameter with which the multilevel automaton is associated. Each parameter of the fuzzy inference rules is associated with such a multilevel automaton. At each iteration, and complete set of parameters is generated and a new performance index criterion $E(t)$ is calculated and used to update the probability distribution of each level of all multilevel automata. This procedure is repeated at each iteration until the desired convergence is reached. The probability vector $P(t)$ at each level depends only on the corresponding level and on the action selected at the previous level.

3.2 Implementation aspects

Let us assume that each of the parameters of the fuzzy inference system varies in a predetermined domain. This domain is discretized into N^N. A number of hierarchical structures of automata equal to the parameter number need to be used for the identification task. Each set of actions of these different pyramidal structures of automata is associated with one of the parameters to be identified.

The performance evaluation unit contains the following rules:

$$\begin{aligned} &\text{IF } E(t) < E(t-1), \\ &\text{THEN } w(t) = 0, \text{ else } w(t) = 1, \end{aligned} \tag{13}$$

where $E(t)$ is the identification criterion (6) and $w(t)$ is the response of the performance evaluation system.

At each iteration t, a uniformly distributed random ζ is generated, and an action b_i is randomly selected, based on the probability distribution associated with the selected automata in the corresponding level. The index i is the last value of the index, verifying the following constraint:

$$\sum_{j=1}^{i} p_j(t) \geqslant \zeta, \ \zeta \in [0,1]. \tag{14}$$

This procedure is repeated for the N automata activated at the different N levels and for the set of different hierarchical structure of learning automata associated with the different parameters of the fuzzy inferene system. The last actions selected by different hierarchical structures of the learning automata correspond to the estimation of the inference system parameters.

The cost function is calculated and introduced in the performance evaluation unit, which generates an output $w(t)$ according to rules (13). The probability vectors are then updated according to this response, this procedure is repeated each time.

Nonlinear combination forecasting principle based on fuzzy inference system is as follows: let Y_t states reality observation data of some forecasting in a period ($t = 1, 2, \cdots, n$) and f_{tj} states forecasting values produced by m different individual forecasting models around it ($t = 1, 2, \cdots, n; j = 1, 2, \cdots, m$), respectively. We take $f = (f_{t1}, f_{t2}, \cdots, f_{tm}) \in \mathbb{R}^m$ and $Y_t \in \mathbb{R}$ as the input variables and the output variable of the fuzzy inference system, respectively, then tune the fuzzy inference system to make fitting precision of f_t approaching y_t according to some measure the less the better. After tuning, we have constructed the nonlinear mapping relation of the data about individual forecasts and an actual value and apply the system to combine these forecasts or information from individual forecasting model or different sources into a single forecast f_t with higher accuracy or excellent performance.

4 Evaluation of forecasting effects

In order to evaluate and compare nonlinear combining forecasting method based fuzzy system, a feasible evaluation index system must be adopted. According to the evaluation conventions and principle of forecasting effects, the following evaluation indexes and criteria have been employed here:

1) Sum of squares error, $SSE = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2$, where y_t and $\hat{y}_t$ represent the real values and the forecasting values, respectively;

2) Mean absolute error, $MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$;

3) Mean square error, $MSE = \frac{1}{n}\sqrt{SSE}$;

4) Mean absolute percent error,

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|(y_t - \hat{y}_t)/y_t|;$$

5) Mean square percent error,

$$MSPE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}[(y_t - \hat{y}_t)/y_t]^2}.$$

5 Application

In order to illustrate the efficiency of nonlinear combining forecasting based on fuzzy inference system (NCFS) and compare NCFS with other combination ways, several application examples are given in this section. The original data in Examples 1,2, respectively, are taken from references [13 ~ 15], as shown in Table 1. The evaluations of forecasting effects for NCFS are described in Table 2 in detail. For convinces in comparative analysis, we also give all the evaluations of forecasting effects about all methods both individual and combining in Table 2.

Table 1 The original data of forecasting examples

T		1	2	3	4	5	6	7	8	9	10	11
Example 1	y_t	14.9	18.6	22.2	17.6	19.6	24.0	31.6	43.7	37.0	47.2	
	f_{1t}	10	14.9	23.3	26.1	17.5	20.2	26.4	36.8	52.5	38.5	
	f_{2t}	12	15.48	18.95	22.43	25.9	29.38	32.85	36.33	39.80	43.82	
Example 2	y_t	57.0	65.4	75.4	82.5	92.8	102.7	119.5	143.8	169.7	201.0	251.2
	f_{1t}	54.52	62.89	72.54	83.67	96.51	111.32	128.41	148.11	170.84	197.06	227.31
	f_{2t}	64.68	66.74	68.72	76.61	88.42	104.15	123.79	147.35	174.82	206.21	241.51

Table 2 Evaluation results of forecasting effects

Indices of forecasting effects				SSE	MAE	MSE	MAPE	MSPE
Example 1	Individual Forecasting		Method(I)	520.60	6.04	2.28	0.2251	0.0825
			Method(II)	199.76	4.11	1.41	0.1696	0.0599
	CF	WAM	$W_1 = 0.1158$, $W_2 = 0.8842$	194.16	4.05	1.39	0.1649	0.579
		WGM	$W_1 = 0.2159$, $W_2 = 0.7841$	191.35	3.97	1.38	0.159	0.0561
		NCFS		183.72	3.60	1.31	0.149	0.0516
Example 2	Individual Forecasting		Method(I)	795.59	5.78	2.56	0.0440	0.0156
			Method(II)	338.25	4.96	1.67	0.0474	0.0179
	CF	WAM	$W_1 = 0.1259$, $W_2 = 0.8741$	328.56	4.80	1.65	0.0443	0.0159
		WGM	$W_1 = 0.5652$, $W_2 = 0.4348$	449.04	4.28	1.92	0.0334	0.0159
		NCFS		283.44	3.98	1.53	0.0306	0.0144

CF: combining forecasting; WAN: weighted arithmetic mean combining forecasts; WGM: weighted geometric mean combining forecasts; NCFS: nonlinear combining based on fuzzy inference system.

The simulation evaluation results, relative to two examples, show that nonlinear combining forecasting method based on fuzzy system can get the best forecasting effects of all the methods in fitting precision, stability, smoothing and trend analysis to forecasting target. Besides, other big numbers of examples also show the similar results. Therefore, the method (NCFS) is always recommended such that the best combining forecasting effects can be obtained.

6 Conclusion

The analysis reported in this paper has demonstrated that the nonlinear combining forecasting method based on fuzzy inference system forms a fruitful extension to the existing methodology, and contributes new insights into practical ways to model problems involving large number of complex and possibly interrelated alternatives. In other words, the nonlinear combining forecasting method incorporating fuzzy inference system possesses adequate generalization capability, learning properties and universalized capabilities. Further empirical studies of combined modeling and forecasting of non-stationary time series in nonlinear systems with this method are therefore under way.

References

[1] Bates J M and Granger C W J. Combination of forecasts [J]. Operations Research Quarterly, 1969, 20(4): 451 - 468

[2] Armstrong J S. Research on forecasting: a quarter-century review [J]. Interfaces, 1986, 16(1): 89 - 109

[3] Bunn D W. Combining forecasts [J]. European Journal of Operational Research, 1988, 33(3): 223 - 229

[4] Bunn D W. Forecasting with more than one model [J]. Journal of Forecasting, 1989, 8(3): 161 - 166

[5] Granger C W J. Combining forecasts-twenty years later [J]. Journal of Forecasting, 1989, 8(3): 167 - 173

[6] Clemen R T. Combining forecasts: a review and annotated bibliography [J]. International Journal of Forecasting, 1989, 5(4): 559 - 583

[7] Zhang Guoping. Analysis on the theory of B-G combining forecasts [J]. Journal of Forecasting, 1988, 7(5): 54 - 59

[8] Kosko B. Fuzzy system as universal approximators [A]. Proc. FUZZ-IEEE'92 [C], San Diego, California, 1992, 1153 - 1162

[9] Wang C X. Fuzzy inference systems are universal approximators [A]. Proc. FUZZ-IEEE'92 [C], San Diego, California, 1992, 1163 - 1170

[10] Hao Ying. Sufficient conditions on uniform approximation of multivariate functions by general Takagi-Sugeno fuzzy inference systems with linear rule consequent [J]. IEEE Transactions on System, Man, And Cybernetics, Part A: Systems and Humans, 1998, 28(4): 515 - 520

[11] Dong Jingrong and Yang Xuiai. Predicting of the economic development for the circumjacent reservoir region of the Longtan hydropower project using fuzzy neural network [J]. Journal of Forecasting, 1998, 17(4): 54 - 58

[12] Thathachar M A L and Ramakrishnal K R. A hierarchical system of learning automata [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1981, 11(3): 236 - 241

[13] Yu Tao and LI Fangzhong. Principeles of Social Economic Statistics [M]. Wuhan: Wuhan University Publishing House, 1992, 138 - 143

[14] Zhou Chuanshi and Lou Guoming. A weight geometric-mean combining forecasting model and application [J]. Application of Statistics and Management, 1992, 20(2): 17 - 19

[16] Yang Guiyuan and Tang Xiaowo. Research on the optimal method of combining forecasting based on geometric mean [J]. Statistical Research, 1996, 28(2): 55 - 58

本文作者简介

董景荣 1966年生.1996年毕业于重庆大学系统工程及应用数学专业,获硕士学位.1999年毕业于重庆大学管理学院,获博士学位.现为重庆师范学院数学与计算机科学系副教授、系副主任、硕士研究生导师.在重庆大学工商管理博士后流动站做博士后研究.在国内外发表学术论文50余篇,学术专著2部.目前主要研究方向是经济系统的演化建模,非线性组合预测及其控制.