

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## A New Method on Partially Linear Autoregression Forecasting

<sup>1</sup>Ang Wei, <sup>2</sup>F.U. Chunyao and <sup>3</sup>Wei Wei

<sup>1</sup>School of Natural Sciences, Nanjing University of Posts and Telecommunications, Nanjing 210046, China

<sup>2</sup>Key Laboratory for Organic Electronics and Information Displays and Institute of Advanced Materials,  
 Nanjing University of Posts and Telecommunications, Nanjing 210046, China

<sup>3</sup>XI'an University of Technology, School of Computer Science and Engineering, XI'an 710048, China

**Abstract:** In this study, a new method to acquire the decision rule of information system with continuous attributes was proposed. The modeling theory and the estimation method of partially linear auto-regression model are discussed and the lag variables, best model, the optimal bandwidth are determined. Partially linear auto-regression models of Shanghai stock index and Shenzhen component index based on partial residual estimate are established and the validity of the model are examined. Finally, forecasting with the model and judged the effect of the model based on the real data.

**Key words:** Partially linear autoregression, recursive entropy, discretizing methods

### INTRODUCTION

Partially linear auto-regression model with parametric variable and nonparametric variable has more adaptability and explanatory ability than the parametric model and the nonparametric model. The wavelet is a perfect tool to deal with instability signals. This thesis has established models of Shanghai stock index and Shenzhen component index based on wavelet and time series models.

Traditional methods discretize the attributes before apply linear auto-regression model to the information system. Fayyad and Irami (1993) use a recursive entropy minimization heuristic for attribute discretization. In (Wang, 2001), naïve and semi naïve scaler methods are proposed to find the appropriate cut point of the continuous attribute. In this study, we construct the indiscernibility set of  $x$  by defining the toleration relation on the information systems. Using the indiscernibility set of  $x$ , we partitioned  $U$  and acquired the decision rule. This decision rule is influence by the system permitted error. The relation between the system permitted error and the effectiveness of the acquired decision rule is discussed through a numerical experiment. And we compared our methods with the traditional ones under the optimum system permitted error. It shows that our method is better than the traditional ones in many cases.

### INFORMATION SYSTEM WITH CONT-INUOUS CONDITION ATTRIBUTES

First, let us have a look on the definition of decision information system.

**Definition 1:** A decision information system including, where:

- $U$  is a non-empty finite set of objects

$$C_\psi = \int_{-\infty}^{+\infty} |\omega^{-1}| |\hat{\psi}(\omega)|^2 d\omega < +\infty$$

- $A$  is a non-empty finite set of attributes

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R}; a \neq 0$$

- $D$  is a non-empty finite set of decisions

$$\hat{\psi}(\omega = 0) = 0$$

- $F$  is the relation set between  $U$  and  $A$

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0$$

- $f_j: U \rightarrow V_j$  ( $j \leq q$ ),  $V_j$  is the domain of  $a_j$
- $G$  is the relation set between  $U$  and  $D$

$$W_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

- $g_j: U \rightarrow V'_j$  ( $j \leq q$ ),  $V'_j$  is the domain of  $d_j$

In order to solve the problem of continuous condition attributes discretization, we define the relationship  $R_B^{\circ}$  and  $[x]_B^{\circ}$  in the following way.

**Definition 2:** A decision information systems  $X_t = (X_{t1}, \dots, X_{tp})^T$  we define the toleration relation as the following:

$$R_B^\delta = \{(x, y); |f_i(x) - f_i(y)| \leq \delta, a_i \in B\}$$

Here  $\delta \geq 0$  is called a system permitted error.

Under this relation, the indiscernibility set of  $x$  is  $W_\psi f(a, b)$ .

Hence,  $\forall X \subseteq U$ , we can get its approximation under the relation:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad t \in Z$$

$$W_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \Delta t \sum_{k=1}^N f(k\Delta t) \bar{\psi}\left(\frac{k\Delta t - b}{a}\right);$$

$$W_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \bar{\psi}\left(\frac{t - b}{a}\right) dt$$

Let:

$$\begin{cases} \int t^p \psi(t) dt = 0 & p = 1 \sim n \\ \bar{\psi}(\omega) = \omega^{n+1} \bar{\psi}_0(\omega) & \bar{\psi}_0(\omega = 0) \neq 0 \end{cases}$$

then  $R_D$  is equivalent relation in  $X$ , its equivalent class is denoted as:

$$U/R_D = \{D_1, \dots, D_i\}$$

**DECISION RULE ACQUISITION**

According to Zhang *et al.* (2001, 2003a, b), Let:

$$\psi(t) = \begin{cases} 1 & 0 \leq t \leq 1/2 \\ -1 & 1/2 \leq t \leq 1 \\ 0 & \text{else} \end{cases} \mu_B^\delta(x) = \left( D\left(\frac{D_j}{[x]_B^\delta}\right), \dots, D\left(\frac{D_j}{[x]_B^\delta}\right) \right)$$

$$\tau_B^\delta(x) = \{D_j : [x]_B^\delta \cap D_j \neq \emptyset\},$$

$$\gamma_B^\delta(x) = \left\{ D_j : D\left(\frac{D_j}{[x]_B^\delta}\right) = \max_{j \in \tau} D\left(\frac{D_j}{[x]_B^\delta}\right) \right\}$$

$$m_B^\delta(x_i) = \max_{j \in \tau} D\left(\frac{D_j}{[x_i]_B^\delta}\right) = D\left(\frac{D_{j_i}}{[x_i]_B^\delta}\right)$$

**Definition 1:**

$$Y_i = X_i^T \beta + g(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

let:

$$Y_i = X_i^T \beta + g(t_i) + \varepsilon_i, \quad i = 1, 2, \dots, n$$

**Theorem 1:** Suppose that  $(U, A, F, D, G)$  is an information system with continuous condition attributes, then all the non-empty:  $0 \leq \alpha \leq 1$  ( $h_i(t) = E(X_i|T = t)$ ) form a partition of  $U$ .

**Proof:** (1) first, to prove that with different  $Y_t = \beta^T X_t + g(Z_t) + \varepsilon_t$ ,  $X_n (I = 1, \dots, p) \cap X_t = (X_{t1}, \dots, X_{tp})^T = \phi$  ( $i \neq j$ ), reduction to absurdity is used, supposed that  $Y_t = \beta Y_{t-1} + g(Y_{t-2} + \varepsilon_t (t \geq 3))$ , then  $\Psi(t) \in L^2(R)$  lead to  $\hat{\psi}(\omega = 0) = 0$ :

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0$$

so:

$$\gamma_B^\delta(x_0) = D^i, \gamma_B^\delta(x_0) = D^j$$

we can get the result as  $D^i = D^j$ , which is obviously a confliction.

So:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in R; a \neq 0$$

(2) Then, let us prove that to all:

$$W_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt$$

According to the definition of  $C_B^\delta(D^j)$ , we know that  $C_B^\delta(D^j) \subseteq U$  and then  $\cup C_B^\delta(D^j) \subseteq U$ , Whereas,  $\forall x \in U$ , suppose that:

$$\gamma_B^\delta(x) = D^b$$

and then:

$$W_\psi f(a, b) = \frac{1}{\sqrt{|a|}} \Delta t \sum_{k=1}^N f(k\Delta t) \bar{\psi}\left(\frac{k\Delta t - b}{a}\right)$$

therefore:

$$U \subseteq C_B^\delta(D^j)$$

From (1) and (2) we can draw a conclusion that:

$$\cup C_B^\delta(D^j) = U$$

According to theorem 1, we can partition  $U$  as  $U = \{C_1, C_2, C_3, \dots, C_i\}$ , in which:

$$t = \left\{ (D^j \subseteq U/R_D, C_B^\delta(D^j) \neq \emptyset) \mid C_j = C_B^\delta(D^j) \right\}$$

And hence gain the decision rule:

$$\begin{cases} \int t^p \psi(t) dt = 0 & p = 1 \sim n \\ \hat{\psi}(\omega) = \omega^{n+1} \hat{\psi}_0(\omega) & \hat{\psi}_0(\omega = 0) \neq 0 \end{cases}$$

Now, we call:

$$q_B^{\delta}(D^j) = \min_{x \in C_B^{\delta}(D^j)} \{m_B^{\delta}(x)\}$$

the reliability of  $C_B^{\delta}(D^j) \Rightarrow D^j$ ,  $q_B^{\delta}$  is called the precision of  $C_B^{\delta}$ . Then we obtain the rule with precision:

$$\psi(t) = \begin{cases} 1 & 0 \leq t \leq 1/2 \\ -1 & 1/2 \leq t \leq 1 \\ 0 & \text{other} \end{cases}$$

To be more practical, we normally transform the rule to attribute expression. We denote:

$$\int t^p \psi(\omega) dt = 0, p = 0 \sim N \quad Y_B^{\delta}(D_j) = \{\wedge(a_i^j, b_i^j), a_i \in B\}$$

where:

$$\psi(t) = \sum_k g_k \phi(2t - k)$$

$$q_B^{\delta}(D_j) = \min\{q_B^{\delta}(D^j); C_B^{\delta}(D^j) \cap D_j \neq \emptyset\}$$

This formula of rule can be expressed as:

$$\hat{\psi}(t) = \sqrt{2\pi} e^{-(\omega - \omega_0)^2 / 2}$$

To illustrate this decision rule acquisition process clearly, let us see some examples.

**Example 1:**

U	a <sub>i</sub>	a <sub>j</sub>	d
x <sub>1</sub>	0	0.5	1
x <sub>2</sub>	0.1	0.7	1
x <sub>3</sub>	0.3	1	2
x <sub>4</sub>	0.4	1	2
x <sub>5</sub>	0.5	1	3
x <sub>6</sub>	1	0.6	3

Let  $\delta = 0.3$ .

We can accomplish the decision rule acquisition:

$$\psi(t) = e^{i\omega t} e^{-t^2/2} \quad \mu_A^{\delta}(x_2) = \{\frac{1}{2}, \frac{1}{2}, 0\}$$

$$\hat{\psi}(t) = \sqrt{2\pi} \omega^2 e^{-\frac{\omega^2}{2}} \quad \mu_A^{\delta}(x_4) = \{0, \frac{2}{3}, \frac{1}{3}\}$$

$$T = \left[ \frac{4\pi}{\omega_0 + \sqrt{2 + \omega_0^2}} \right] \times a$$

$$\begin{aligned} \gamma_A^{\delta}(x_1) &= \{D_1\} & \gamma_A^{\delta}(x_2) &= \{D_1, D_2\} & \gamma_A^{\delta}(x_3) &= \{D_2\} & \gamma_A^{\delta}(x_4) &= \{D_2\} \\ \gamma_A^{\delta}(x_5) &= \{D_2\} & \gamma_A^{\delta}(x_6) &= \{D_3\} \end{aligned}$$

Hence, we have:

$$\begin{aligned} D^1 &= \{D_1\} & C_A^{\delta}(D^1) &= \{x_1\} \\ q_A^{\delta}(D^1) &= \min\{1\} = 1 \end{aligned}$$

$$\begin{aligned} D^2 &= \{D_2\} & C_A^{\delta}(D^2) &= \{x_3, x_4, x_5\} \\ V_k &\subset V_{k+1}, & k &\in Z \end{aligned}$$

$$\bigcap_{m \in Z} V_m = \{0\}, \quad \bigcup_{m \in Z} \bar{V}_m = L^2(\mathbb{R})$$

$$\begin{aligned} D^4 &= \{D_1, D_2\} & C_A^{\delta}(D^4) &= \{x_2\} \\ \phi(t) \in V_0 &\Leftrightarrow \phi(t - k) \in V_0 \end{aligned}$$

$$\forall D^j \subseteq \bigcup_{R_D} / R_D, j > 4 \quad C_A^{\delta}(D^j) = \emptyset$$

$$Y_A^{\delta}(D_1) = ((0, 0.2), (0.3, 0.8)) \quad q_A^{\delta}(D_1) = \frac{1}{2}$$

$$f(t) = f_1(t) + g_1(t) \quad q_A^{\delta}(D_2) = \frac{1}{2}$$

$$f(t) = f_2(t) + g_2(t) + g_1(t) \quad q_A^{\delta}(D_3) = 1$$

**NUMERICAL EXPERIMENT**

We apply the method of rule acquisition to the famous Iris data.

We conduct the experiment in the following way:

- Step 1:** Given  $\delta = 0$
- Step 2:** Randomly chose these samples from the whole data set as training samples 15, 30, 45, 60, 75, 90, 105, 120, 135, 150
- Step 3:** Then acquire the rules by applying our methods on the training samples
- Step 4:** We gain the rules of  $f(t) = f_m(t) + g_{m-1}(t) + \dots + g_1(t)$
- Step 5:** Judge the whole data set by the rules we acquired and compare our judgments with the original decision of the data sets. We can gain the percentage of the data that are being judged correctly
- Step 6:** Repeat step (2-4) for 500 times. Then, compute the average percentage of correctness
- Step 7:**  $\delta = \delta + 2.0$  if  $\delta < 51$  go to step 2

or else end

We convert the result into chart. With the value of  $\delta$  as abscissa and the average percentage of correctness as y-axis, we plot ten curves according to different capability of training data.

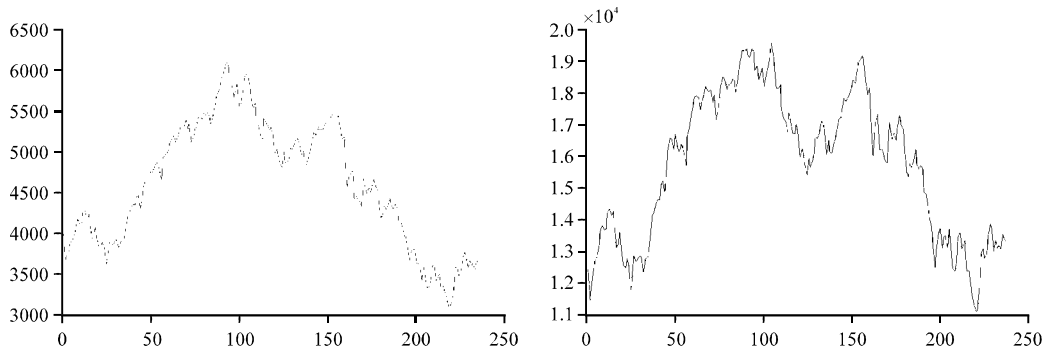


Fig. 1: The initial data of (a) stock index and (b) component index

Table 1: Comparison of all the methods

Method	Correctness (%)
Our method	86.67
Naïve scale	70.67
Semi Naïve scaler	78.00
Minimum entropy	84.00

From Fig. 1, we can see that as the capability of the training data becomes larger, the optimum  $\delta$  slides away from the point  $\delta = 0$ . When the capability of the training data exceed 75, most the optimum  $\delta$  stays at the point  $\delta = 8.0$ .

After the optimum system permitted error was gained, we compare our methods to the traditional ones. 90 samples from the whole data set were chosen randomly and the decision rules under the optimum system permitted error acquire were acquired. Then, we apply the acquired decision rule to the whole data set and gain the correctness of the rule. The comparison between our methods with traditional methods of discretizing the data is showed as follows.

**Comparison:** Rules derived from 90 samples applying to the whole data set.

In fact, our methods greatly improved the percentage of correctness in many cases. The Table 1 is only an example.

### CONCLUSION

The main work of this study can be summarized as: First, a method of decision rule acquisition on decision information system with continuous condition attributes is given. This method avoids the step of data discretization and hence decreases the information which

is lost in pretreatment. Second, the relation between system permitted error and the effectiveness of the decision acquisition is discussed through a numerical experiment. We also demonstrate that our method gets better results than traditional discretizing methods by comparison.

### ACKNOWLEDGMENTS

This work is supported by the financial support from the Natural Science Research Project of Jiangsu Ordinary University (09KJB430008), the Opening Project of State Key Laboratory of High Performance Ceramics and Superfine Microstructure (SKL201111SIC) and Education Reform Project of NJUPT (JG00711JX39).

### REFERENCES

Fayyad, U. and K. Irani, 1993. Multi-interval discretization of continuous-valued attributes for classification learning. *Int. Joint Conf. Artificial Intel.*, 13: 1022-1029.

Wang, G., 2001. *Theory and Knowledge Acquisition of Rough Sets*. 1st Edn., Xi'an Jiaotong University Press, Xi'an, China, Pages: 146.

Zhang, W., W. Wu and J. Liang, 2001. *Theory and Methods in Rough Sets*. Science Press, Beijing, China.

Zhang, W., J. Mi and W. Wu, 2003a. *Information System and Knowledge Discovery*. Science Press, Beijing, China.

Zhang, W., J. Mi and W. Wu, 2003b. Knowledge reductions in inconsistent information systems. *Chin. J. Comput.*, 26: 12-18.