

## NGHIÊN CỨU MÔ HÌNH HỆ LUẬT MỜ VÀ HỆ LAI CHO CÔNG TÁC PHÂN TÍCH DỰ BÁO

Vũ Thanh Nguyên\*

### 1. Áp dụng mô hình hệ luật mờ và thuật toán di truyền trong công tác phân tích dự báo

#### 1.1. Mô hình hệ luật mờ (Standard Additive Model - SAM)

Việc sử dụng mô hình mạng neuron [4] cho bài toán xấp xỉ hiện vẫn đang còn nhiều nhược điểm. Trước hết đó là khả năng học của mạng neuron. Hơn nữa, việc xác định cấu trúc mạng neuron phù hợp cho từng bộ số liệu vẫn là một công việc hết sức khó khăn. Chính vì các hạn chế nói trên của mạng neuron mà các chuyên gia đã xây dựng một cấu trúc khá đặc biệt để xây dựng các hệ thống xấp xỉ, đó là hệ luật mờ. Có thể nói hướng tiếp cận này thật sự mới mẻ và chưa được nhiều tác giả quan tâm.

**Mô hình hệ luật mờ cộng chuẩn :** hệ luật mờ là hệ thống m luật mờ dạng  $\square_j: IF x = A_j THEN y = B_j; j = \overline{1, m}$ , hoạt động theo cơ chế song song. Tuy nhiên, vì các đặc tính thuận lợi trong tính toán (tính tích phân để xác định trọng tâm), trong bài báo này chỉ đề cập đến hệ luật mờ hoạt động theo qui tắc kết hợp SUM-PRODUCT. Trong hệ mờ SAM, ứng với mỗi giá trị vào  $x = x_0$ , luật thứ j :  $\square_j$  được kích hoạt và cho kết quả là tập mờ  $B'_j$  xác định theo  $B_j$  và mức độ thỏa mãn về trái  $a_j(x_0)$  dựa trên qui tắc PRODUCT.

$$B'_j = a_j(x_0) \times B_j$$

m kết quả ra  $B'_j$  của các luật trong hệ luật được SAM kết hợp theo qui tắc SUM để cho kết quả chung của toàn hệ thống là tập mờ B. ta có :

$$B = \sum_{j=1}^m w_j \cdot B'_j = \sum_{j=1}^m w_j \cdot a_j(x_0) \cdot B_j \quad B = \sum_{j=1}^m w_j \cdot B'_j = \sum_{j=1}^m w_j \cdot a_j(x_0) \cdot B_j \quad (1)$$

Giá trị B sẽ được khử mờ để nhận được một giá trị rõ duy nhất.

---

\* TS, Sở Bưu chính Viễn thông Tp.HCM

**Ứng dụng SAM cho xấp xỉ hàm phi tuyến :** theo (1), hệ SAM hoạt động như một ánh xạ :  $F: \square_n \rightarrow \square_p$ . Chính nhờ đặc trưng này mà với bất kỳ hàm phi tuyến liên tục giới hạn  $f(x)$ ,  $f: U \subset \square_n \rightarrow \square_p$ , với  $U$  là tập compact, ta luôn có thể xây dựng một hệ mờ SAM :  $F: \square_n \rightarrow \square_p$  cho phép xấp xỉ  $f$  bởi  $F$ . Cũng như các mô hình xấp xỉ khác, mỗi mô hình xấp xỉ mờ hàm phi tuyến SAM luôn tương ứng với một giá trị sai số nhất định. Gọi  $e$  là giá trị sai số của mô hình xấp xỉ SAM, ta có

$$e = \max_{x \in X} \{|f(x) - F(x)|\}$$

Giả sử  $f$  là hàm số được xấp xỉ,  $\varepsilon > 0$  cho trước.

**Định nghĩa :**  $\varepsilon$ -close =  $\{F \text{ là hàm mờ xấp xỉ } f : |f(x) - F(x)| \leq \varepsilon, \forall x \in X\}$

Dưới góc độ khảo sát đồ thị, khả năng xấp xỉ của hệ mờ  $F$  đối với một hàm phi tuyến  $y = f(x)$  được thể hiện như sau :

- Mỗi luật mờ trong hệ mờ hoạt động theo cơ chế xấp xỉ cho phép tương ứng với mỗi tập mờ vào  $A_j$  là một tập mờ kết quả  $B_j$ .
- Thông qua việc kết hợp các khối mờ hình thành từ các luật mờ  $\square_j$  hoặc lấy trung bình giữa các khối mờ này nếu chúng chồng lấp lẫn nhau nhờ vào cơ chế khử mờ bằng phương pháp trọng tâm, hệ mờ SAM  $F$  có thể bao phủ đồ thị biểu diễn của hàm  $f(x)$  mà nó xấp xỉ.

**Cơ chế học trong SAM :** quá trình học của SAM thông thường bao gồm hai bước chính là học cấu trúc và học tham số. Tuy nhiên, để cho hiệu quả học của hệ được tốt hơn, nhóm nghiên cứu phối hợp thêm cơ chế học tối ưu hệ luật. Do đó, quá trình học của SAM ở đây bao gồm các giai đoạn sau :

- **Tự phát sinh cấu trúc luật :** thực hiện theo cơ chế tự học. Bằng cách thực hiện việc phân lớp mờ trên bộ dữ liệu học, hệ SAM sẽ tự phát hiện ra các luật mờ cần thiết cho việc xấp xỉ hàm phi tuyến cho bộ số liệu học đó.
- **Điều chỉnh các thông số :** điều chỉnh các thông số của hệ luật như : trọng số của từng luật, trọng tâm và kích thước của các tập mờ tham gia ở vế trái và vế phải của các luật.

- **Tối ưu hóa hệ luật** : cho phép SAM có thể phát hiện và loại bỏ các luật mờ không cần thiết cho hoạt động xấp xỉ của nó, nhằm tăng đáng kể tốc độ xử lý cũng như giảm nhiễu.

**Học cấu trúc** : đây là giai đoạn bắt đầu đối với hệ SAM chưa có tri thức. Bài toán học cấu trúc có thể phát biểu như sau : Cho trước bộ dữ liệu học vào  $\{x_j\}$  và bộ kết quả mong muốn  $\{y_j\}$ ,  $j = \overline{1, ntd}$ ; với  $ntd$  : số bộ dữ liệu học. Cho trước dạng hàm phụ thuộc của các tập mờ. Hãy xây dựng một phân lớp mờ  $P(\{x_j | y_j\})$  trên các bộ số liệu học. Trên cơ sở đó, xác định các tập mờ và hàm phụ thuộc tương ứng để từ đó phát sinh các luật mờ của hệ mờ SAM có khả năng xấp xỉ một cách tốt nhất hàm phi tuyến  $y = f(x)$  đặc trưng của bộ dữ liệu học.

Việc giải quyết bài toán này được tiến hành theo hai bước sau :

- Xác định các tập mờ bằng thuật toán phân lớp dữ liệu mờ.

Gọi  $\square_n$  là không gian các vector có  $n$  thành phần thực. Đặt  $X = \{x_1, x_2, \dots, x_{ntd}\}$ ,  $x_j \in \square_n$  là tập hữu hạn bộ số liệu học, trong đó  $ntd$  là số bộ dữ liệu học.

Gọi  $V_{cn}$  là không gian vector các ma trận  $c \times n$ ,  $c \in \square^+$  cho trước,  $1 < c < n$ .

- Xác định một phân lớp mờ trên  $X$  biểu diễn bởi một bộ vector trọng tâm :

$V = \{v_1, v_2, \dots, v_c\}$ ,  $v_i \in \square_n$ . Cho tương ứng với 1 ma trận

$U = \{u_{ij}\} \in V_{cn}$ , với  $u_{ij}$  là giá trị thực trong đoạn  $[0,1]$  diễn tả mức độ phụ thuộc của bộ số liệu học  $x_j$  ứng với vector trọng tâm  $v_i$ , và thỏa hai điều kiện sau :

$$1. \forall x \in X, u_{ij} \in [0,1], \quad \sum_{k=1}^c u_{ik} = 1$$

$$2. \forall i, j = \overline{1, c} : \quad 0 < \sum_{j=1}^{ntd} u_{ij} < n$$

Nhiệm vụ đặt ra của bài toán phân lớp mờ là phải làm giảm thiểu giá trị hàm mục tiêu  $J$  xác định trên  $U$  và  $V$  có dạng như sau :

$$J(U, V) = \sum_{i=1}^c \sum_{x_j \in X} \sum_{k=1}^c g[w(x_i), u_{ij}] d(x_j, v_k)$$

trong đó :  $w(x_i)$  là trọng số khởi đầu của  $x_i$ ,  $d(x_j, v_k)$  là độ đo biểu diễn mức độ khác biệt giữa  $x_j$  và vector trọng tâm của phân lớp thứ  $k$  :  $v_k$ . Độ đo chọn phải thỏa hai tính chất sau :  $d(x_j, v_k) > 0$  và  $d(x_j, v_k) = d(v_k, x_j)$ .

Hiện nay có rất nhiều thuật toán đề cập đến vấn đề này như thuật toán phân lớp mờ trung bình (FCM) [1], thuật toán FCM cải tiến với phân lớp mờ dẫn đầu [1]... Đa số các thuật toán đều có mục tiêu chung là xác định  $V$ . Giá trị của  $U$  có thể được xác định một cách tuyệt đối hoặc tương đối thông qua một đại lượng khác nhằm mục đích hạn chế thao tác xử lý và tài nguyên sử dụng. Một phương pháp phân lớp theo hướng tiếp cận tựa FCM là phương pháp dùng vector lượng tử thích nghi [1]. Giống như các vector  $V$  của thuật toán FCM, các vector lượng tử được dùng như một công cụ để dò tìm các phân lớp mờ.

**Xây dựng các luật mờ :** Sau khi thực hiện quá trình phân lớp mờ, công việc tiếp theo là xây dựng các luật mờ từ các phân lớp đó. Dựa trên các thông tin về các phân lớp mờ : các trọng tâm của các vector lượng tử  $q_i$ , người ta tiến hành xây dựng các luật mờ. Trọng tâm của các tập mờ có thể dễ dàng xác định thông qua tọa độ các vector lượng tử. Nhưng để xác định dạng hàm thành viên đòi hỏi phải xác định được độ rộng của các tập mờ. Kosko [1] với đề nghị sử dụng các luật mờ dạng ellipse và thuật toán phân lớp mờ thông qua các vector lượng tử với cơ chế học cạnh tranh đã cung cấp một cơ chế giúp xác định chính xác độ rộng của các tập mờ thông qua tâm của các ellipse và độ nghiêng của chúng. Tuy nhiên phương pháp này có nhiều trở ngại do mức độ phức tạp của nó khi cài đặt.

**Điều chỉnh thông số :** khi các luật mờ đã được xác định, học điều chỉnh thông số giúp giảm sai số giữa kết quả của hệ và kết quả mong muốn. Bài toán được phát biểu như sau : Cho trước bộ dữ liệu học vào  $\{x_j\}$  và bộ kết quả mong muốn  $\{y_j\}$ ,  $j = \overline{1, ntd}$ ; với ntd : số bộ dữ liệu học. Cho hệ luật mờ SAM với các luật mờ và trọng số. Hãy điều chỉnh thông số của các tập mờ về trái, về phải và trọng số các luật mờ sao cho sai số giữa kết quả cho bởi hệ luật mờ và kết quả mong muốn là ổn định và nhỏ nhất. Quá trình học điều chỉnh thông số được tiến

hành dựa trên luật học sai số nhỏ nhất. Luật học sai số nhỏ nhất áp dụng cho tham số  $\xi$  trong SAM có dạng :

$$\xi(t+1) = \xi(t) - \mu_t \frac{\partial E}{\partial \xi}$$

Trong đó  $\mu_t$  là hệ số học biến đổi theo thời gian và có xu hướng giảm dần. Mục đích của giai đoạn học điều chỉnh thông số là tối thiểu bình phương sai số :

$$E(x) = \frac{1}{2} (f(x) - F(x))^2$$

Sai số trong xấp xỉ của hệ luật mờ phụ thuộc vào các tham số tham gia vào hệ, bao gồm : các tập mờ về trái  $A_{ji}$ , các tập mờ về phải  $B_{ji}$ , các trọng số  $w_j$ .

### **Tối ưu hệ luật bằng thuật toán di truyền (Genetic Algorithm - GA).**

Về mặt lí thuyết, một hệ SAM càng có nhiều luật thì độ chính xác trong hoạt động xấp xỉ của nó càng lớn. Tuy nhiên, nếu hệ có quá nhiều luật thì thời gian cho quá trình xử lí trong hệ luật sẽ là yếu tố đáng quan tâm. Một hệ luật tối ưu sẽ chỉ giữ lại một số (hoặc tất cả) các luật ban đầu trong hệ SAM. Mỗi luật được giữ lại cho tương ứng với giá trị 1, các luật bị loại cho tương ứng với giá trị 0. Bài toán tối ưu hệ luật được phát biểu như sau : Cho trước bộ dữ liệu học vào  $\{x_j\}$  và bộ kết quả mong muốn  $\{y_j\}$ ,  $j = \overline{1, ntd}$ ; với ntd : số bộ dữ liệu học. Cho hệ luật mờ SAM với các luật mờ và trọng số. Hãy tìm số vị trí 1 ít nhất sao cho sai số giữa kết quả cho bởi hệ SAM và kết quả mong muốn là ổn định và nhỏ nhất.

Một trong số các giải pháp cho bài toán trên là phương pháp sử dụng thuật toán GA ([1], [5], [7], [8]). Phương pháp này xem mỗi bộ kết hợp các luật là một nhiễm sắc thể, dùng bộ lọc Kalman với hai tiêu chuẩn tối ưu và đảm bảo chính xác để phát hiện các cá thể thích hợp. Từ đó chọn một cá thể tốt nhất làm kết quả của thuật toán. Quá trình thực hiện như sau :

- **Biểu diễn các nhiễm sắc thể** : mỗi nhiễm sắc thể là một chuỗi các giá trị nhị phân diễn tả trạng thái hoạt động của luật tương ứng trong hệ SAM.
- **Hàm thích nghi** : mối qua hệ giữa kích thước SAM và độ chính xác trong xấp xỉ của SAM được giải quyết bằng hàm thích nghi sau :

$$\text{Fit}(m) = \ln(\bar{\sigma}_\epsilon^2) + \frac{\log_n(m)}{n}, \quad \bar{\sigma}_\epsilon^2 = \frac{1}{n} \sum_{j=1}^n (y_j^d - F(x_j))^2 \quad (2)$$

ở đó m : Số luật (trạng thái 1) được sử dụng trong hệ SAM, n : Số bộ số liệu học.

– **Thuật toán :**

- b1. Khởi tạo 10 nhiễm sắc thể, có 1 nhiễm sắc thể biểu diễn đầy đủ các luật.
- b2. Tạo các nhiễm sắc thể mới bằng các phương pháp : Lai nhị phân (Tỉ lệ 0.5) và đột biến nhị phân (Tỉ lệ 0.01).
- b3. Dùng phương pháp bánh xe quay với hàm thích nghi (2) để giữ lại 10 nhiễm sắc thể tốt nhất (có hàm Fit(.) → min).
- b4. Nếu điều kiện lặp chưa kết thúc, quay lại b2.
- b5. Chọn nhiễm sắc thể tốt nhất trong 10 nhiễm sắc thể nhận được làm kết quả trả về. Chuỗi nhị phân tìm được sẽ được dùng làm cơ sở cho việc hủy bỏ các luật không cần thiết trong hệ SAM.

**1.2. Mô hình hệ lai (kết hợp thuật toán di truyền GA và Mạng Neural Network)**

Các mô hình hệ lai sử dụng thuật toán di truyền GA kết hợp mạng Neural Network ([1], [5], [8]) là một cách tiếp cận tương đối phổ biến và có tính hiệu quả cao cho vấn đề nêu trên. Điều đó xuất phát từ khả năng phân lớp nhờ thuật toán GA và khả ghi nhớ, học của mô hình hệ lai này. Bên cạnh đó, tính ổn định của mạng Neuron Network cũng là một yếu tố quan trọng giúp nó được chọn vì đây là điều kiện quan trọng đặt ra cho bài toán mô hình.

***Phép toán chọn lọc.***

Lượng giá từng nhiễm sắc thể.

Tính độ thích nghi của từng nhiễm sắc thể.

Các độ thích nghi đều lớn hơn 0.

- Tính độ thích nghi cho mọi cá thể.

$$\text{eval}(v_i) \quad (i: 1 \rightarrow \text{GEMAX})$$

$$v_i = \frac{1}{\text{TestError}}$$

- Tính tổng giá trị thích nghi trên toàn quần thể.

$$F = \text{Sigma}(\text{eval}(v_i)) \quad (i:1 \rightarrow \text{GEMAX})$$

- Tính xác suất chọn  $p_i$  cho nhiễm sắc thể  $v_i$

$$p_i = \frac{\text{eval}(v_i)}{F}$$

$$q_i = \text{Sigma}(p_j) \quad (j:1 \rightarrow i)$$

- Chọn GEMAX lần. Với mỗi lần :
  - + Phát sinh một số ngẫu nhiên  $r$  ( $r:0 \rightarrow 1$ )
  - + Nếu  $r < q_i$  : chọn  $v_i$
  - Ngược lại chọn  $v_i : q_{(i-1)} < r < q_i$ .

**Khởi tạo mỗi nhiễm sắc thể là 1 mạng nơ ron (neural network).**

Các mạng nơ ron không sử dụng file lưu kết quả riêng. Toàn bộ các nhiễm sắc thể sử dụng chung một file kết quả.

***Phép toán lai.***

Xác định xác suất lai  $p_c = 0.25$ .

Đối với mỗi nhiễm sắc thể trong quần thể mới :

- + Phát sinh một số  $r$  trong khoảng  $[0,1]$
- + Nếu  $r < p_c$  . Chọn nhiễm sắc thể để lai

Ghép đôi các nhiễm sắc thể đã chọn một cách ngẫu nhiên

- + Phát sinh một số ngẫu nhiên  $pos$  ( $pos:1 \rightarrow n$ )
- $n$  : chiều dài nhiễm sắc thể

***Phép toán đột biến.***

Xác định xác suất đột biến  $p_n = 0.01$

Đối với mỗi nhiễm sắc thể trong quần thể hiện hành, đối với mỗi bit

- + Phát sinh một số  $r$  trong khoảng  $[0,1]$
- + Nếu  $r < p_n$  . Đột biến cá thể đã chọn (trọng số đột biến là 0.1).

## 2. Lựa chọn ngành để dự báo

Trong công tác phân tích dự báo, vấn đề quan trọng hàng đầu cần đặt ra là việc nắm bắt tối đa thông tin về lĩnh vực dự báo. Thông tin ở đây có thể hiểu một cách cụ thể nhất là bao gồm :

- Các số liệu quá khứ của lĩnh vực dự báo
- Diễn biến tình hình hiện trạng cũng như động thái phát triển của lĩnh vực dự báo
- Đánh giá một cách đầy đủ nhất các nhân tố ảnh hưởng cả về định lượng lẫn định tính.

Trong tất cả các lĩnh vực nghiên cứu, rất nhiều các lĩnh vực gặp khó khăn rất lớn về mặt thông tin, cụ thể như nếu phân chia theo ngành kinh tế, có nhiều ngành thiếu số liệu quá khứ, điều này do nhiều lí do, có thể nêu ra như : phân ngành kinh tế quốc dân chưa ổn định, hệ thống chỉ tiêu thông tin của ngành thống kê chưa ổn định, việc thu thập số liệu đối với các thành phần kinh tế ngoài quốc doanh gặp nhiều khó khăn ... Vì vậy, phần lớn các đề tài nghiên cứu hiện nay gặp rất nhiều khó khăn, nhất là những đề tài có tính chất dự báo [3], thường xuyên phải sử dụng các yếu tố định tính để phân tích dự báo, mà sử dụng các yếu tố định tính lại lệ thuộc rất lớn vào những nhận định chủ quan của người phân tích, do đó thường không đạt độ chính xác cao trong dự báo, không mang tính thuyết phục. Một khi tìm được phương pháp khả thi để dự báo là một tác nhân tích cực trong việc điều hành quản lí kinh tế.

## 3. Các chương trình dự báo

Nhóm nghiên cứu tiến hành cài đặt các chương trình máy tính dựa trên các mô hình hệ luật mờ và hệ lai đã được giới thiệu ở trên và những dữ liệu thử nghiệm cho các chương trình này, ứng dụng trực tiếp vào dự báo sự tăng trưởng của nền kinh tế quốc dân. **Các chương trình có các chức năng như huấn luyện dữ liệu, thử nghiệm dữ liệu, dự báo dữ liệu trong khoảng thời gian thực. Các chương trình dự báo mức độ tăng trưởng cao nhất, thấp nhất và tốc độ tăng trưởng trung bình của dữ liệu cần dự báo.** Ngoài ra để kiểm tra và so sánh khả năng dự báo của các mô hình này với các phương pháp đang được ứng dụng rộng



rãi trong kinh tế, các chương trình được cài đặt thêm phương pháp dự báo chuỗi ARIMA ([2],[6]) là mô hình phân tích dự báo kinh tế cổ điển đang được sử dụng rộng rãi trong các ngành dự báo của kinh tế ở Việt Nam và trên thế giới, nhằm đối chiếu và so sánh các phương pháp dự báo đã nghiên cứu trong bài báo với phương pháp dự báo chuỗi ARIMA.

### 3.1. Dữ liệu thực nghiệm

Dữ liệu thực nghiệm được chọn trong lĩnh vực thị trường giá cả vì lĩnh vực này có ảnh hưởng và tác động rất lớn đến nền kinh tế thành phố nói riêng và Việt Nam nói chung. Vì vậy có thể lựa chọn phương án thử nghiệm bằng cách sử dụng dữ liệu của giá cả một số mặt hàng quan trọng, cụ thể như : **vàng, đô la, gạo, cà phê, xi măng**. Đây là các mặt hàng quan trọng trong nền kinh tế VN, sự biến động giá của chúng có tác động rất lớn đến các hoạt động kinh tế khác [1], [2]. Dữ liệu về giá cả các mặt hàng này được cập nhật hàng ngày, và có thể sử dụng dữ liệu từ nhiều năm trước, từ năm 1994 đến 2004, tức bao gồm khoảng hơn 3000 số liệu cho mỗi bộ dữ liệu của mỗi loại mặt hàng thử nghiệm. Ngoài ra, nhóm nghiên cứu nhận thấy việc nghiên cứu giá cả các mặt hàng này là vấn đề cần thiết và có thể nói là rất quan trọng do :

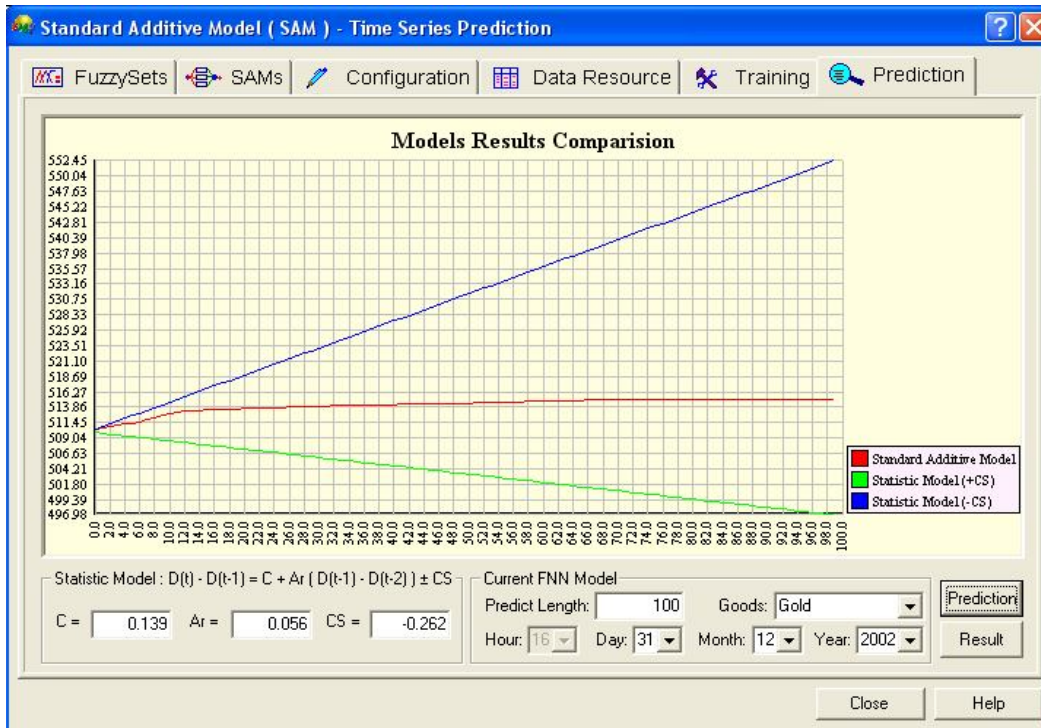
- Giá cả vàng và đô la thể hiện khá rõ nét tình trạng sức khỏe của nền kinh tế. Khi kinh tế đi xuống, người dân sẽ có hành vi tích trữ vàng, đô la ..., do đó sẽ đẩy giá lên. Khi kinh tế phát triển, người dân không giữ vàng và USD nữa mà sẵn sàng bỏ vốn ra kinh doanh, làm ăn, do đó vàng sẽ ổn định giá và có xu hướng hạ giá.
- Gạo, cà phê và xi măng là ba trong số các mặt hàng chiến lược của nước ta phục vụ cho xuất khẩu và ổn định kinh tế trong nước. Sự ổn định giá cả các mặt hàng này sẽ là thước đo có ý nghĩa cho sự ổn định và phát triển kinh tế. Vàng, đô la, gạo, cà phê và xi măng một khi có biến động giá sẽ tác động rất mạnh lên thị trường tất cả các loại hàng hoá khác.

### 3.2. Kết quả thử nghiệm

#### 3.2.1. Ứng dụng mô hình hệ luật mờ

- Kết quả thử nghiệm (có đối chiếu với mô hình ARIMA)

Thử nghiệm trên dữ liệu vàng.



Prediction result data on: by SAM

	Day	by SAM	+CS Model	-CS Model
1	31/12/2002	510.43	509.88	510.40
2	01/01/2003	510.65	509.75	510.82
3	02/01/2003	510.84	509.62	511.25
4	03/01/2003	511.00	509.49	511.67
5	04/01/2003	511.18	509.36	512.10
6	05/01/2003	511.37	509.23	512.52
7	06/01/2003	511.59	509.10	512.95
8	07/01/2003	511.89	508.97	513.37
9	08/01/2003	512.24	508.84	513.80
10	09/01/2003	512.61	508.70	514.22

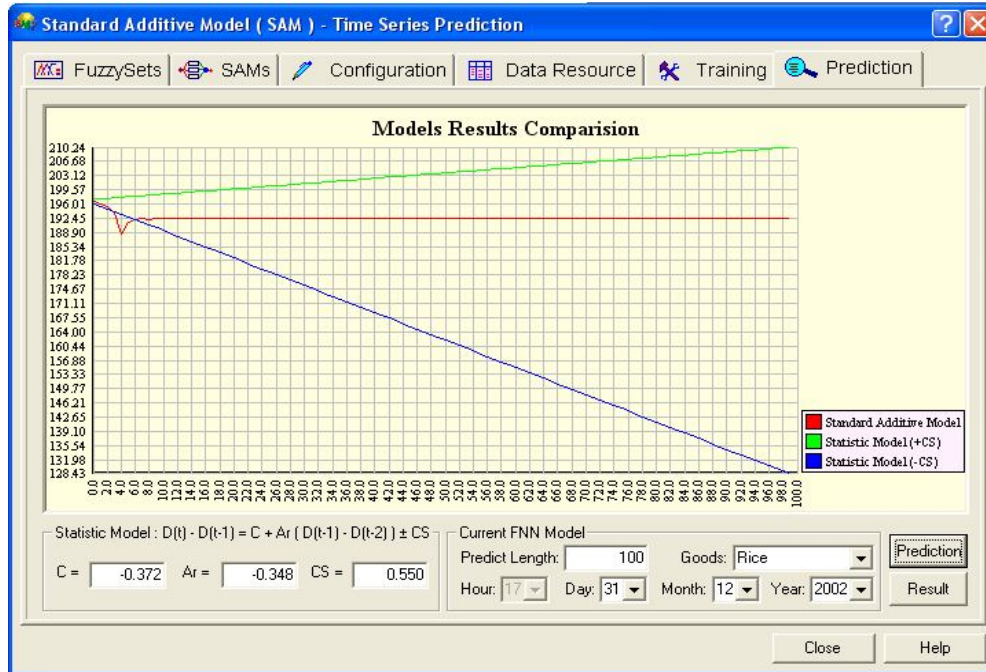
  

	by SAM	+CS Model	-CS Model
- Maximal Value	510.57	509.88	510.40
- Minimal Value	506.06	505.88	506.40
- Average Value	271.18	271.00	271.28
- Mean Average Error	0.34	0.34	0.26
- Mean Average Percentage Error	0.00	0.00	0.00
- Mean Square Error	0.28	0.34	0.14
- Maximal ascillation value	1.37	1.00	1.00

Length of Post Period: 15    Refresh    Save Predict Result    Save Statistic Info    Close

(Lưu ý: các đường màu xanh biểu diễn mô hình dữ liệu dự báo theo phương pháp ARIMA, đường màu đỏ theo phương pháp dự báo của mô hình hệ luật mờ).

Thử nghiệm trên dữ liệu gạo.



Prediction result data on: by SAM

	Day	by SAM	+CS Model	-CS Model
1	31/12/2002	196.55	197.18	196.08
2	01/01/2003	196.15	197.29	195.48
3	02/01/2003	195.47	197.43	194.76
4	03/01/2003	193.66	197.56	194.09
5	04/01/2003	188.58	197.69	193.40
6	05/01/2003	191.49	197.83	192.72
7	06/01/2003	192.06	197.96	192.04
8	07/01/2003	192.40	198.09	191.35
9	08/01/2003	192.18	198.22	190.67
10	09/01/2003	192.55	198.35	189.98

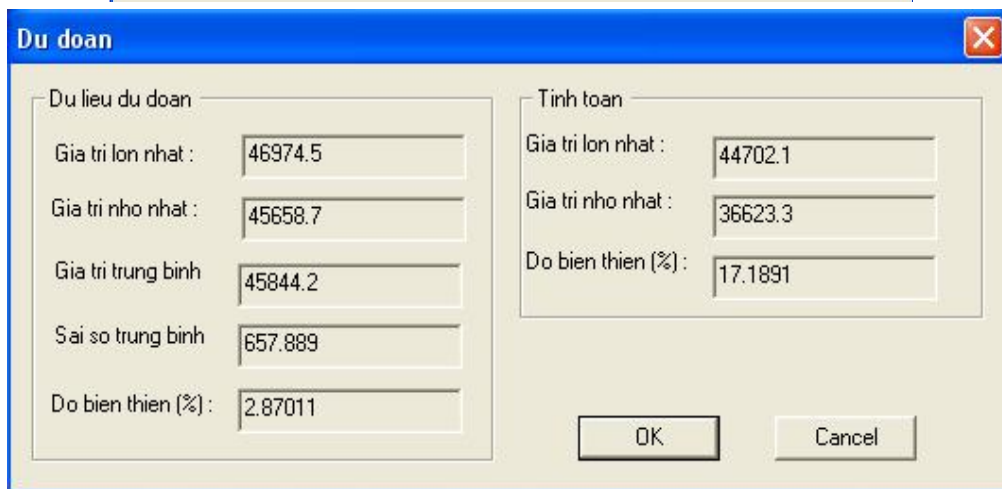
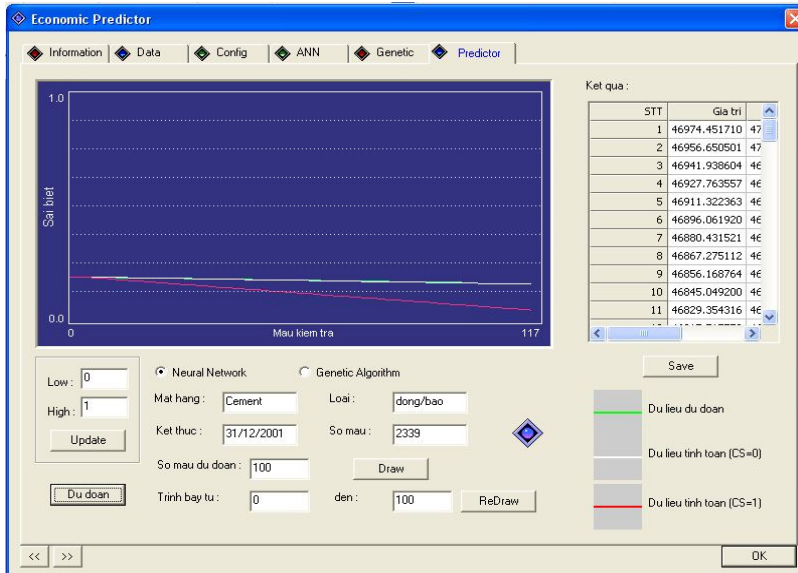
  

	by SAM	+CS Model	-CS Model
- Maximal Value	206.33	205.18	204.08
- Minimal Value	192.64	192.18	191.08
- Average Value	104.77	104.72	104.13
- Mean Average Error	1.58	1.56	1.62
- Mean Average Percentage Error	0.01	0.01	0.01
- Mean Square Error	13.99	14.00	14.97
- Maximal ascillation value	13.70	13.00	13.00

Length of Post Period: 15 Refresh Save Predict Result Save Statistic Info Close

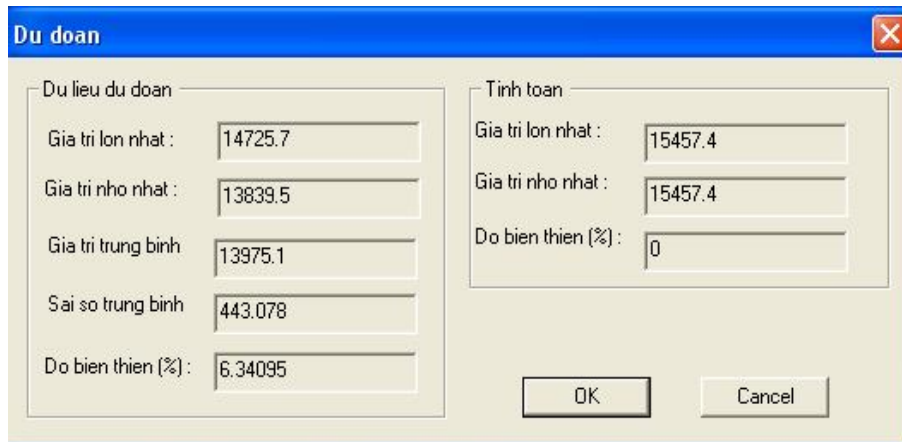
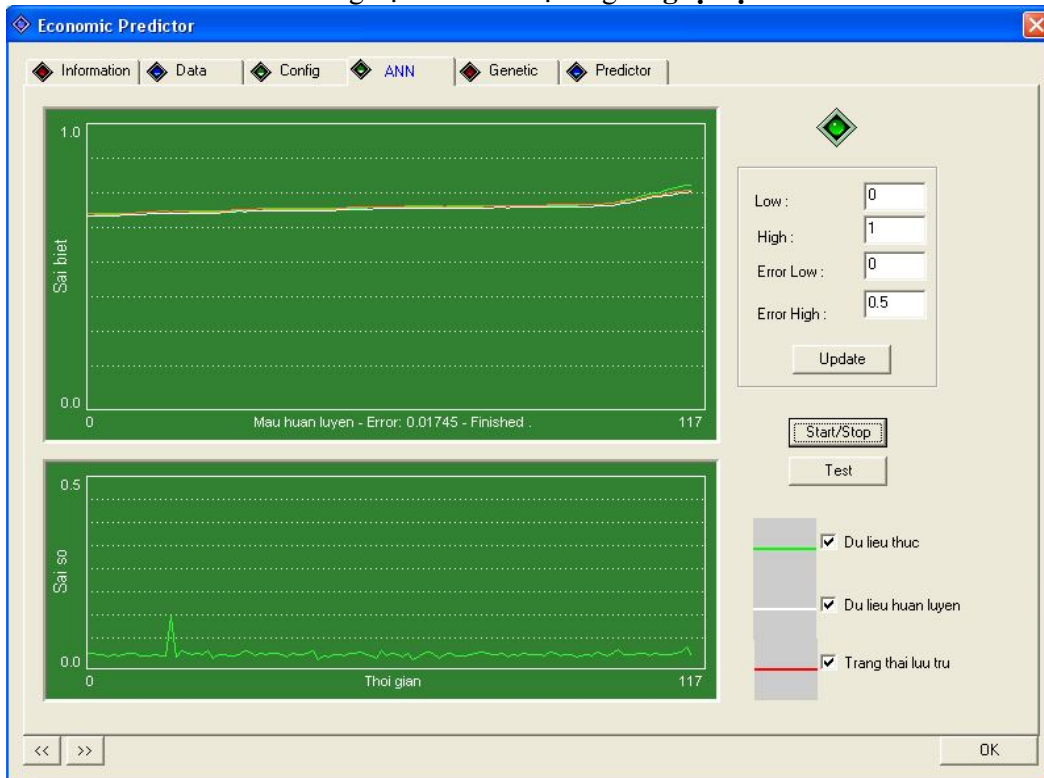
### 3.2.2. Ứng dụng mô hình kết hợp thuật toán di truyền và mạng Neural Network

- Kết quả thử nghiệm (có đối chiếu với mô hình ARIMA)  
Thử nghiệm trên dữ liệu **cerment**.



(Lưu ý: các đường màu xanh biểu diễn mô hình dữ liệu dự báo theo phương pháp ARIMA, đường màu đỏ theo phương pháp dự báo của mô hình hệ lai kết hợp thuật toán di truyền và mạng Neural Network).

Thử nghiệm trên dữ liệu tỉ giá ngoại tệ USD.



**4. Kết luận**

Kết quả thử nghiệm các loại dữ liệu trên so sánh với các phương pháp dự báo đã được sử dụng trước đây ([1], [2], [4], [6]) (như phương pháp ARIMA, các phương pháp định tính và định lượng khác) có thể chấp nhận được. Tuy nhiên, để có thể kiểm tra thêm phương pháp dự báo này cho công tác phân tích dự báo đối

với trung hạn và dài hạn cần phải có thêm nhiều dữ liệu hơn nữa mới có thể đánh giá hết được tính ổn định và độ chính xác của các mô hình hệ luật mờ và mô hình hệ lai đang được thử nghiệm.

### TÀI LIỆU THAM KHẢO

- [1]. Vũ Thanh Nguyên (2003), *Giải quyết một số vấn đề phân tích dự báo kinh tế ứng dụng trong ngành công nghiệp tại thành phố Hồ Chí Minh*; Đề tài nghiên cứu khoa học của Sở Khoa Học và Công Nghệ thành phố Hồ Chí Minh - Nghiệm thu 10/2003.
- [2]. Nguyễn Quốc Tòng (2000), *Các phương pháp định tính và định lượng được ứng dụng trong các công tác phân tích dự báo của Viện Kinh Tế thành phố*; Đề tài nghiên cứu khoa học của Viện Kinh Tế Thành phố Hồ Chí Minh - Nghiệm thu 2000.
- [3]. Nguyễn Thông (1999), *Phân tích dữ liệu và áp dụng vào dự báo*; Nhà xuất bản thanh niên.
- [4]. Duc Truong Pham, Liu Xing (1998), *Neural Networks for Identification, Prediction and Control*. Springer – Verlag London Limited.
- [5]. T.T. Chow, Z.Lin and C.L. Song (2001). *Applying Neural Network and Genetic Algorithm In System Optimization*. 7<sup>th</sup> International IBPSA Conference.
- [6]. ARIMA Models to Predict Next-Day Electricity Price, IEEE 2003.
- [7]. Vũ Thanh Nguyên, Nguyễn Thanh Phong (2000), *Sử dụng thuật toán di truyền trong vấn đề thiết kế mạng*. Hội nghị khoa học lần thứ II – ĐHKHTN, 05/2000.
- [8]. Vũ Thanh Nguyên, *Fuzzy Measure, Fuzzy Integral and Using them with genetic algorithm in Hybrid System*. IT@EDU2000.

### Tóm tắt :

#### **Nghiên cứu mô hình hệ luật mờ và hệ lai cho công tác phân tích dự báo**

Hiện nay, công tác phân tích dự báo sử dụng các mô hình ứng dụng các lí thuyết về logic mờ, lí thuyết mạng Neural Network, thuật toán di truyền đem lại các kết quả tương đối khả quan. Bài báo nghiên cứu mô hình hệ luật mờ, cải tiến từ mô hình mạng Neural kết hợp cùng lí thuyết logic mờ

và thuật toán di truyền nhằm cải tiến khả năng học và tối ưu hoá bộ luật. Ngoài ra bài báo còn sử dụng mô hình hệ lai kết hợp giữa thuật toán di truyền và mạng Neural Network ứng dụng cho công tác phân tích dự báo.

**Abstract :**

**Using models of fuzzy rule system and hybrid system for forecast analysis**

At present, forecast analysis which applied models on theories of fuzzy logic, neural network, genetic algorithm, gains some relatively positive achievements. This paper is about model of fuzzy rule system, improved from the model of Neural network with theory of fuzzy logic and genetic algorithm aiming at accelerating ability of learning and optimizing the rules. In addition to this, the article is about the use of hybrid model with the genetic algorithm and the Neural network for forecast analysis.