

Dự báo phụ tải điện theo mô hình tương quan dựa trên luật mờ

• Phan Thị Thanh Bình

• Lương Văn Mạnh

Trường Đại học Bách khoa, ĐHQG-HCM

(Bài nhận ngày 06 tháng 03 năm 2014, hoàn chỉnh sửa chữa ngày 28 tháng 04 năm 2014)

TÓM TẮT:

Các mô hình dự báo phụ tải điện theo phương pháp tương quan truyền thống thường có các dạng hàm hồi qui tường minh như $Y=f(x_1, x_2, \dots, x_n)$ hoặc $\log Y=f(\log x_1, \log x_2, \dots, \log x_n)$ trong đó f có dạng tuyến tính và x_i là các yếu tố tương quan: nhiệt độ, dân số, GDP, sản lượng công nghiệp. Tuy nhiên mô hình chỉ áp dụng được khi có tương quan tuyến tính giữa các đại lượng trên với phụ tải điện (thể hiện qua hệ số tương quan). Bài báo trình bày mô hình dự báo tương quan trên ý tưởng sử dụng các luật mờ dạng

Takagi-Sugeno theo giải thuật phân loại trừ nhóm cho trường hợp tổng quát, cả khi không có hàm dự báo kiểu tường minh. Khảo sát cho thấy mô hình cho kết quả khả quan khi hàm hồi qui có dạng hàm thường gặp (tuyến tính, tuyến tính theo log hóa), và cả khi không thể tìm được dạng hàm tường minh. Các dự báo điện năng tiêu thụ theo yếu tố nhiệt độ cho một trạm điện của thành phố Hồ Chí Minh được trình bày.

Từ khóa: Giải thuật trừ nhóm, Luật mờ Takagi-Sugeno, Tương quan, hồi qui.

1. GIỚI THIỆU

Các mô hình dự báo phụ tải điện theo phương pháp tương quan thường có các dạng hàm hồi qui tường minh: $Y=f(x_1, x_2, \dots, x_n)$ hoặc $\log Y=f(\log x_1, \log x_2, \dots, \log x_n)$ trong đó f có dạng tuyến tính và x_i là các yếu tố tương quan: nhiệt độ, dân số, GDP, sản lượng công nghiệp.... Mô hình tương quan truyền thống thường dựa trên các đánh giá tương quan giữa các đại lượng. Ví dụ như nếu hàm đề xuất có dạng tuyến tính thì cần phải tính hệ số tương quan r để đánh giá mức độ liên quan tuyến tính giữa phụ tải điện và các đại lượng liên quan [1].

Mối quan hệ giữa phụ tải điện với các yếu tố tương quan truyền thống như GDP và các yếu tố kinh tế, xã hội (mức tiêu thụ điện theo đầu

người, mức tiêu hao điện năng trên đơn vị sản phẩm, giá điện) bị ảnh hưởng nhiều theo yếu tố thời gian (công nghệ rẽ đi, mức độ điện khí hóa cao lên...). Tất cả điều này làm cho mối quan hệ giữa phụ tải điện với các yếu tố tương quan trở nên không tường minh. Điều này dẫn tới việc sử dụng công nghệ Neural-Fuzzy, Neural net để tìm mối tương quan bằng cách xấp xỉ các hàm phi tuyến. Một số tác giả lại tập trung vào kết hợp với kỹ thuật Wavelet như [2][4]. Cụ thể như trong [2], mô hình phức tạp được đề xuất với phân tích Wavelet kết hợp với lý thuyết tập mờ để xây dựng các đầu vào cho mạng Neural nhằm xấp xỉ mối tương quan giữa nhiệt độ và tải. Trong bài báo này, chúng tôi đề xuất mô hình dự báo tương quan với số luật mờ sẽ được xác định tự động dựa trên giải thuật trừ nhóm

(subtractive) của Chiu [3]. Tổ hợp các luật mờ sẽ cho ra mô hình xấp xỉ mối quan hệ giữa tải dự báo và các yếu tố tương quan.

2. MÔ HÌNH TÌM KIẾM LUẬT MỜ

Bài báo đề xuất tìm kiếm xấp xỉ mối quan hệ giữa đại lượng dự báo và các yếu tố tương quan bằng cách tìm kiếm các luật mờ. Ý tưởng tìm kiếm luật mờ được thực hiện qua giải thuật leo núi bởi Yager và Filev [3]. Tuy nhiên giải thuật này khi áp dụng cho số lượng lớn các số liệu đầu vào lại không hữu hiệu. Để cải tiến thuật toán này, Chiu năm 1994 đề xuất giải thuật trừ nhóm.

Xem phụ tải điện và các yếu tố tương quan như là vector x gồm 2 phần: phần input (đầu vào) chứa các đại lượng tương quan và phần output là phụ tải điện. Các vector này sẽ được đưa vào để phân loại và sẽ cho ra số nhóm nhất định. Mỗi tâm nhóm tìm được có thể xem như là một nguyên mẫu đặc tính hành vi của hệ thống. Do đó mỗi tâm nhóm có thể được sử dụng như là một luật mờ (fuzzy rule) dùng để mô tả hành vi của hệ thống. Giả sử tìm được một tập hợp c

tâm nhóm $\{x_1^*, x_2^*, \dots, x_c^*\}$ trong không gian M

chiều. Trong đó, mỗi vector x_i^* có $M-1$ chiều đầu tiên chứa biến ngõ vào (các yếu tố tương quan tới phụ tải tiêu thụ) và chiều còn lại chứa biến ngõ ra chính là phụ tải. Phân chia mỗi

vector x_i^* thành hai thành phần y_i^* và z_i^* , trong

đó y_i^* chứa $M-1$ phần tử đầu vào của x_i^* (tọa

độ tâm nhóm trong không gian ngõ vào) và z_i^*

chứa phần tử còn lại của x_i^* (tọa độ tâm nhóm trong không gian ngõ ra). Xem mỗi tâm nhóm

x_i^* như là một luật mờ mô tả hệ thống. Với mỗi

vector ngõ vào y , độ thỏa mãn của luật mờ thứ i được xác định theo công thức :

$$\mu_i = e^{-\alpha \|y - y_i^*\|^2} \quad (1)$$

$$\text{Trong đó: } \alpha = \frac{4}{r_a^2} \quad (2)$$

với r_a là bán kính hiệu quả. Ngõ ra z được tính như sau:

$$z = \frac{\sum_{i=1}^c \mu_i z_i^*}{\sum_{i=1}^c \mu_i} \quad (3)$$

Có thể xem mô hình tính toán trên là một mô hình Fuzzy với các luật IF-THEN. Nếu giả thiết z trong phương trình (3) là một hàm tuyến tính của biến ngõ vào thì z_i^* của nhóm i được viết lại như sau [3]:

$$z_i^* = G_i y + h_i \quad (4)$$

với G_i là một ma trận hằng số $(1 \times (M-1))$ chiều và h_i là một vector cột hằng số với một phần tử. Luật IF – THEN lúc này trở thành luật Takagi-Sugeno (Takagi and Sugeno, 1985), trong đó mỗi hậu thức là một phương trình tuyến tính của các biến đầu vào.

Gán:

$$\rho_i = \frac{\mu_i}{\sum_{j=1}^c \mu_j} \quad (5)$$

Phương trình (3) được viết lại như sau :

$$z = \sum_{i=1}^c \rho_i z_i^* = \sum_{i=1}^c \rho_i (G_i y + h_i) \quad (6)$$

Hay:

$$z^T = \begin{bmatrix} \rho_1 y^T & \rho_1 \dots \rho_c y^T & \rho_c \end{bmatrix} \begin{bmatrix} G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T \end{bmatrix} \quad (7)$$

Trong đó z^T và y^T là các vector cột. Cho một tập hợp n điểm ngõ vào $\{y_1, y_2, \dots, y_n\}$ thì kết quả tập hợp đầu ra $[Z]$ sẽ là:

$$\begin{bmatrix} z_1^T \\ \vdots \\ z_n^T \end{bmatrix} = \begin{bmatrix} \rho_{1,1} y_1^T & \rho_{1,1} & \dots & \rho_{c,1} y_1^T & \rho_{c,1} \\ & \vdots & & \vdots & \\ \rho_{1,n} y_n^T & \rho_{1,n} & \dots & \rho_{c,n} y_n^T & \rho_{c,n} \end{bmatrix} \begin{bmatrix} G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T \end{bmatrix} \quad (8)$$

Để ý rằng ma trận đầu tiên trong vế phải biểu thức trên là một ma trận hằng số, trong khi ma trận thứ hai chứa tất cả các tham số của mô hình được tối ưu. Việc ước lượng bình phương cực tiểu (8) cho phép tìm ra G và h .

Để tìm được tâm các nhóm, bài báo dựa trên giải thuật leo núi, được đề xuất bởi Yager Yager và Filev (1992) và được Chiu (1994) cải tiến như sau: cho một tập hợp n dữ liệu $\{x_1, x_2, \dots, x_n\}$ trong không gian M chiều, tiến hành chuẩn hóa trong mỗi chiều sao cho chúng nằm trong một đường cong đơn vị (trong mỗi chiều). Giả thuyết rằng mỗi điểm dữ liệu là một thể năng của tâm nhóm. Định nghĩa thể năng tâm nhóm như sau:

$$P_i = \sum_{k=1}^n e^{-\alpha \|x_k - x_i\|^2} \quad (9)$$

Kí hiệu $\|.\|$ biểu thị khoảng cách toán học và r_a là một hằng số dương (thường là 0.5). Việc tính toán thể năng của một điểm dữ liệu là một hàm khoảng cách từ điểm đó đến các điểm dữ liệu khác. Một điểm dữ liệu với nhiều điểm lân cận sẽ có thể năng cao. Hằng số r_a là bán kính hiệu quả định nghĩa một lân cận; những điểm

nằm ngoài bán kính ít ảnh hưởng đến thể năng nhóm. Sau khi thể năng của tất cả các điểm dữ liệu được tính toán, chọn điểm dữ liệu có thể năng cao nhất làm tâm nhóm thứ nhất. Gọi x_i^* là tọa độ và P_1^* là thể năng tâm nhóm thứ nhất. Tính lại thể năng của mỗi điểm dữ liệu x_i theo công thức sau:

$$P_i \Leftarrow P_i - P_1^* e^{-\beta \|x_i - x_1^*\|^2} \quad (10)$$

$$\text{Với } \beta = \frac{4}{r_b^2} \quad (11)$$

Ở đây r_b là một hằng số dương. Điểm dữ liệu càng gần tâm nhóm thứ nhất thì thể năng của nó giảm đi càng nhiều, và vì thế nó sẽ càng không được chọn làm tâm nhóm tiếp theo. Hằng số r_b là bán kính hiệu quả xác định lân cận giảm thể năng. Để tránh sự quá gần nhau giữa các tâm nhóm, thường chọn r_b lớn hơn r_a , giá trị tốt nhất là $r_b = 1.5 r_a$ [3].

Khi tất cả thể năng của các điểm dữ liệu được tính lại theo phương trình (10), chọn điểm dữ liệu với thể năng cao nhất làm tâm nhóm thứ hai. Sau đó tiếp tục giảm thể năng của các điểm dữ liệu dựa trên khoảng cách giữa nó đến tâm nhóm thứ hai. Tổng quát, sau khi tìm được tâm thứ k , tiến hành tính lại thể năng của mỗi điểm dữ liệu theo phương trình:

$$P_i \Leftarrow P_i - P_k^* e^{-\beta \|x_i - x_k^*\|^2} \quad (12)$$

Trong đó x_k^* và P_k^* lần lượt là tâm và giá trị thể năng tâm nhóm thứ k . Quá trình trên sẽ tiếp tục cho đến khi thể năng tâm nhóm giảm đến một ngưỡng nào đó phụ thuộc thể năng tâm nhóm đầu tiên:

$$P_k^* < \varepsilon P_1^* \quad \text{trong đó } \varepsilon \text{ là một số đủ nhỏ.}$$

Như vậy khi cho biết các đại lượng tương quan (vector đầu vào y_{n+i}), có thể sử dụng (7) để dự báo phụ tải

3. KHẢO SÁT CHO MỘT SỐ HÀM ĐIỆN HÌNH TRONG DỰ BÁO THEO PHƯƠNG PHÁP TƯƠNG QUAN

Như trên đã đề cập, các mô hình tương quan dự báo phụ tải điện thường có dạng hàm: $y=ax+b$; $y=ax_1+bx_2+\dots$; hay $\log y=a\log x+b$; $\log y=a\log x_1+b\log x_2+\dots$

3.1. Nếu giữa phụ tải điện và đại lượng tương quan có mối quan hệ tuyến tính $y= ax+b$

Khảo sát cho một chuỗi phụ tải có dạng gần tuyến tính theo x . Không làm mất tính tổng quát lấy hàm minh họa là chuỗi phụ tải xấp xỉ theo hàm $y=2x+5$, gồm 120 mẫu. Lấy 80 mẫu đầu tiên của chuỗi đưa vào mô hình để dự báo cho 40 mẫu liên tiếp của chuỗi. Kết quả sai số của dự báo theo mô hình cho 15 giá trị cuối được cho trong bảng 1. Sai số trung bình của dự báo 40 lần là 2.57 %.

Bảng 1. Bảng kết quả sai số cho dự báo 15 giá trị cuối phỏng theo hàm $y=2x+5$

Thứ tự	106	107	108	109	110
Sai số	0.065466	0.038604	0.013167	0.01558	0.009296
Thứ tự	111	112	113	114	115
Sai số	0.012731	0.032911	0.020549	0.017393	0.046104
Thứ tự	116	117	118	119	120
Sai số	0.036072	0.017327	0.006334	0.017075	0.031857

Bảng 2. Sai số cho 15 lần dự báo cuối phỏng theo hàm $y=2x_1+2x_2+2x_3+5$

Thứ tự	106	107	108	109	110
Sai số	0.0042	0.0174	0.03	0.0076	0.0093
Thứ tự	111	112	113	114	115
Sai số	0.0115	0.0186	0.01103	0.0235	0.00802
Thứ tự	116	117	118	119	120
Sai số	0.0276	0.0016	0.0131	0.0027	0.0143

3.2. Dự báo cho chuỗi dữ liệu có hàm dạng $\log y=a\log x+b$

Trong dự báo phụ tải có nhiều mô hình dạng $\log(y)=a\log(x)+b$ (ví dụ : y -điện năng, x -nhiệt độ, giá điện, GDP...). Ví dụ minh họa là phỏng theo hàm $y=2\log x+5$. Kết quả dự báo cho 10 lần liên tiếp có sai số trung bình là 2.43 %.

3.3. Khảo sát chuỗi dữ liệu có hàm dạng $y=ax_1+bx_2+cx_3+d$

Khảo sát cho chuỗi xấp xỉ theo hàm $y=2x_1+2x_2+2x_3+5$. Kết quả về sai số dự báo cho 15 giá trị cuối được cho trong bảng 2. Sai số trung bình cho 40 lần dự báo là 1.52%.

3.4. Dự báo cho hàm dạng $\log y=a\log x_1+b\log x_2+c\log x_3+d$

Xấp xỉ theo hàm $\log y=2\log x_1+2\log x_2+2\log x_3+5$. Kết quả dự báo cho 10 thời điểm có sai số trung bình là 1.93 %.

4. KHẢO SÁT KHI KHÔNG CÓ MỐI TƯƠNG QUAN TUYẾN TÍNH

Trong số các trạm tại thành phố Hồ Chí Minh thì trạm Gò Vấp 1 có đồ thị phụ tải ngày với đỉnh xảy ra vào buổi tối. Có nghĩa là phụ tải chủ yếu của trạm thuộc về tải dân dụng và thương mại. Do đó trong chừng mực nào đó sẽ có một mối tương quan giữa tải và nhiệt độ. Tuy nhiên khảo sát cho thấy không có mối quan hệ tuyến tính $y=ax+b$ hay mô hình dạng $\log y=a\log x+b$ (y -điện năng tiêu thụ ngày hoặc điện năng tiêu thụ vào các giờ đỉnh; x -nhiệt độ trung bình ngày hoặc nhiệt độ lớn nhất trong ngày). Điều này thể hiện qua hệ số tương quan tuyến tính rất thấp (r xấp xỉ 0.5). Việc tìm kiếm một mối quan hệ tường minh (hàm hồi qui) giữa phụ tải và nhiệt độ là rất khó khăn. Việc áp dụng mô hình đề xuất sẽ giúp ta tìm được kết quả dự báo.

4.1. Dự báo điện năng tiêu thụ ngày

Mô hình dự báo được xây trên số liệu điện năng tiêu thụ ngày và nhiệt độ trung bình ngày từ ngày 1/2/2012 tới ngày 9/7/2012. Để kiểm tra mô hình, sẽ dự báo từ ngày 10/7 tới 24/7 để xem sai số trung bình (MAPE). Bài báo cũng tiến hành tìm kiếm một mô hình tường minh với rất nhiều phép thử và mô hình tốt nhất được chọn là: $y = 35.648271 x^{1.03919}$ (ký hiệu là mô hình 2 trong bảng 3) với sai số trung bình là 2.655%.

Trong khi đó mô hình 1 là mô hình của bài báo đề xuất có MAPE là 2.59%.

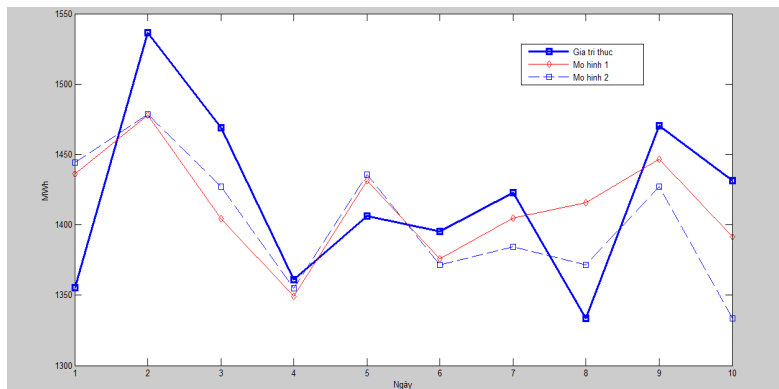
4.2. Dự báo tải đỉnh

Dữ liệu là tổng phụ tải tiêu thụ trạm Gò Vấp vào các giờ tải đỉnh của hệ thống và nhiệt độ trung bình ngày của các ngày như ở mục trên. Khảo sát của bài báo cho thấy nhiệt độ lớn nhất trong ngày có ảnh hưởng yếu hơn đến tải đỉnh. Theo phương pháp đề xuất (mô hình 1 trong bảng 4) thì sai số trung bình là 2.86%. Trong khi đó, sau khi thử các hàm hồi qui khác nhau thì dạng hàm tường minh tốt nhất tìm được (ký hiệu là mô hình 2 trong bảng 4) là $y = -525.132 - 0.542x^2 + 40.9131x$ với MAPE là 2.954%.

Lưu ý là hai dạng hàm hồi qui tường minh nêu trên hoàn toàn không phải là dạng hàm hồi qui truyền thống trong dự báo phụ tải điện. Việc tìm chúng thu được sau rất nhiều lần thử ngẫu nhiên dựa trên sai số nhỏ nhất thu được và tốn nhiều thời gian. Các thông số của mô hình được ước lượng theo phương pháp bình phương cực tiểu.

Bảng 3. Mười giá trị cuối của dự báo điện năng tiêu thụ ngày của trạm Gò Vấp1

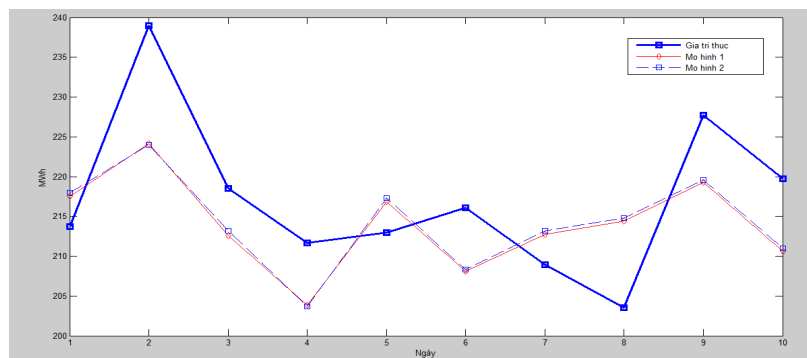
Ngày	15/7	16/7	17/7	18/7	19/7	20/7	21/7	22/7	23/7	24/7
Giá trị thực (MWh)	1355.6	1536.5	1468.9	1361.2	1406	1395.1	1423	1333.6	1470.6	1431.4
Mô hình 1	1436.4	1478	1404.5	1349.3	1431.4	1375.8	1405	1415.6	1446.8	1391.5
Sai số -1	0.05961	0.0381	0.0438	0.0087	0.0181	0.0139	0.0127	0.0615	0.0162	0.0279
Mô hình 2	1444.2	1478.3	1427.1	1354.7	1435.6	1371.7	1384.5	1371.7	1427.1	1333.5
Sai số -2	0.065	0.0378	0.0284	0.0047	0.0211	0.0167	0.027	0.0286	0.0295	0.0683



Hình 1. Giá trị thực và dự báo theo hai mô hình của điện năng ngày trạm Gò Vấp 1

Bảng 4. Dự báo tải đỉnh cho trạm Gò Vấp 1 (10 giá trị cuối)

Ngày	15/7	16/7	7/7	18/7	19/7	20/7	21/7	22/7	23/7	24/7
Trị thực (MWh)	213.7	239	218.5	211.7	213	216.1	208.9	203.5	227.7	219.7
Dự báo mô hình 1	217.56	224.18	212.62	203.84	216.88	208.1	212.74	214.38	219.31	210.61
Sai số -1	0.01806	0.062	0.0269	0.0371	0.0182	0.037	0.0184	0.0535	0.0369	0.0414
Mô hình -2	218.0	224.0	213.2	203.7	217.3	208.3	213.2	214.8	219.6	211.
Sai số-2	0.0206	0.063	0.024	0.0378	0.0202	0.0357	0.0206	0.0559	0.0354	0.0394



Hình 2. Giá trị thực và dự báo theo hai mô hình của điện năng các giờ tải đỉnh trạm Gò Vấp 1

5. KẾT LUẬN

Bài báo trình bày cách tiếp cận sử dụng thuật toán của Chiu tìm kiếm luật mờ cho bài toán dự báo phụ tải điện theo mô hình tương quan. Mô hình dự báo ở đây không cần biết dạng hàm hồi qui, cũng như không cần đánh giá mức độ tương

quan giữa các đại lượng. Khảo sát cho thấy mô hình cho kết quả khả quan khi hàm hồi qui có dạng hàm thường gấp (tuyến tính, tuyến tính theo log hóa), và cả khi không thể tìm được dạng hàm tường minh.

Load forecasting by regression model based on fuzzy rules

• Phan Thi Thanh Binh

• Luong Van Manh

University of Technology-VNU-HCM

ABSTRACT:

The forecasting models by traditional regression function have the crisp functions such as $Y=f(x_1, x_2, \dots, x_n)$ or $\log Y=f(\log x_1, \log x_2, \dots, \log x_n)$. Here f has the linear form and x_i are the factors such as GDP, temperature, industrial output, population... But these models are able to be used only when the linear correlation existed (expressed by the correlation coefficient). This paper introduced the regression model based on the fuzzy Takagi-Sugeno rules. These rules are built by using the

subtractive clustering. The model is used for the general case, even when there are no the crisp function f . Examining shows that the good results are obtained in the case of traditional correlation such as linear or linear by logarithm. The results are also satisfactory for the case of unknown correlation. The electricity consumption forecasting due to the temperature factor for one substation of HochiMinh city was carried out.

Key words: Subtractive clustering algorithm, Takagi-Sugeno Fuzzy rules, Correlation, Regression.

TÀI LIỆU THAM KHẢO

- [1]. Đặng Ngọc Dinh, Hệ thống điện, NXB Khoa học Kỹ thuật Hà nội, (1986).
- [2]. Bhavesh Kumar Chauhan¹, Madasu Hanmandlu, Load forecasting using wavelet fuzzy neural network, International Journal of Knowledge-Based and Intelligent Engineering Systems, IOS Press, Volume 14, 57-71, (2010).
- [3]. Chiu S., Fuzzy Model Identification Based on Cluster Estimation, Journal of Intelligent & Fuzzy Systems, Vol. 2, 267-278, (1994).
- [4]. Y Chen, P.B. Luh, Short-term Load forecasting: Similar Day-Based Wavelet Neural Networks, IEEE Trans, Power Syst Vol.25, N.1 322-327, (2010)