

Phối hợp đồ thị phụ tải và đồ thị tang góc cho bài toán phân loại đồ thị phụ tải của các khách hàng

• **Phan Thị Thanh Bình**

Khoa Điện-Điện tử - Trường Đại học Bách khoa, ĐHQG-HCM

(Bản nhận ngày 17 tháng 3 năm 2015, hoàn chỉnh sửa chữa ngày 25 tháng 5 năm 2015)

TÓM TẮT

Phân nhóm đồ thị phụ tải của các khách hàng thường dựa trên không gian đầu vào 24 chiều. Điều đó có nghĩa là mỗi đồ thị phụ tải được coi như một phần tử 24 đặc tính tương ứng với 24 giá trị tải trong ngày. Tuy nhiên trong một vài trường hợp nếu phân nhóm chỉ dựa trên đồ thị phụ tải thì có thể dẫn tới phân nhóm sai, khi mà hai đồ thị khác nhau hoàn toàn về hình dáng nhưng lại có cùng một khoảng cách Oclit tới đồ thị thứ ba. Để khắc phục điều này, bài báo này

quan tâm tới việc chọn không gian đầu vào. Từ một đồ thị phụ tải, một đồ thị tang của góc sẽ được xây. Việc phân nhóm bây giờ sẽ dựa không những trên đồ thị phụ tải mà còn dựa trên đồ thị tang góc này. Kỹ thuật phân nhóm dựa trên giải thuật Pulsa được bài báo cải biên cho phù hợp với không gian đầu vào. Khảo sát cho trường hợp thành phố Hồ Chí Minh cho thấy cách tiếp cận nêu trên cho kết quả tốt nhất.

Từ khóa: phân nhóm đồ thị phụ tải, khoảng cách Oclit

1. GIỚI THIỆU

Phân loại đồ thị phụ tải điện nhằm mục đích tìm ra các nhóm phụ tải có cùng hình dạng đồ thị dùng điện. Nó thường được dùng cho hoạch định giá điện và các chương trình quản lý nhu cầu dùng điện (DSM) của các công ty điện. Các bài báo tổng quan nhất về các kỹ thuật phân loại đồ thị phụ tải được trình bày trong [1] [2] [3].

Trong việc phân loại đồ thị phụ tải của một khách hàng, lượng đồ thị rất lớn và các đồ thị có thể được biểu diễn trong hệ đơn vị có tên. Điều

này hoàn toàn khác khi tiến hành phân loại đồ thị của các khách hàng. Số lượng các đồ thị điển hình cho mỗi loại khách hàng như công nghiệp bia, giấy, hóa chất... thường không lớn. Thay vì biểu diễn đồ thị trong hệ đơn vị có tên, và do công suất tiêu thụ của các khách hàng khác nhau chênh lệch nhau rất nhiều (từ vài MW tới vài chục MW), nên đồ thị sẽ được biểu diễn trong hệ đơn vị tương đối.

Sự phân loại được dựa trên khoảng cách d_{ij} giữa hai phần tử i và j . Dạng phổ biến nhất của d_{ij} là khoảng cách Oclit:

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (1)$$

Với x_{ik} - đặc tính thứ k của phân tử thứ i .

Khi phân loại đồ thị phụ tải, vấn đề chủ yếu là lựa chọn các đặc tính trong công thức (1). Hầu như các công trình đều coi đồ thị phụ tải như một phân tử với 24 đặc tính tương ứng với tải của 24 giờ trong ngày. Một số rất ít tác giả như trong [4] [5] lại sử dụng một vài chỉ số của đồ thị làm đặc tính là: $P_{\text{mean-day}}/P_{\text{max}}$; $P_{\text{min}}/P_{\text{max}}$; $P_{\text{min}}/P_{\text{mean-day}}$ với $P_{\text{mean-day}}$, P_{min} , P_{max} -trị tải trung bình, bé nhất và lớn nhất của đồ thị phụ tải ngày. Kết quả tính theo các chỉ số như vậy có độ tin cậy thấp.

Ý tưởng trong [6] được áp dụng cho gần 30 khách hàng lại dựa trên đồ thị hệ số góc và không sử dụng khoảng cách Oclit. Đây là một thuật toán khó có tính khả thi vì thường cho ra số nhóm rất lớn. Ở đây quá trình phân nhóm tuân thủ theo sự tăng hoặc giảm tải theo thời gian một cách đồng bộ giữa các đồ thị và theo hệ số góc.

Bài báo này sẽ tập trung vào tìm kiếm các đặc tính của (1) và áp dụng thuật toán Pulsar [7] để phân nhóm.

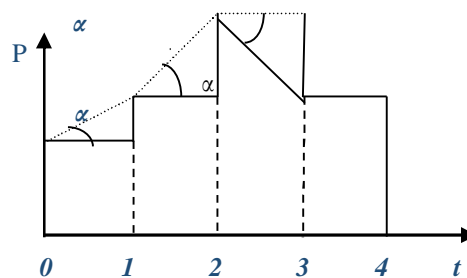
2. CÁC ĐẶC TÍNH CHO PHÂN NHÓM ĐỒ THỊ PHỤ TẢI

Như đã đề cập ở trên, các đặc tính được sử dụng trong [6] là các hệ số tang của các góc. Ở đây, từ một đồ thị phụ tải sẽ tính được 23 giá trị

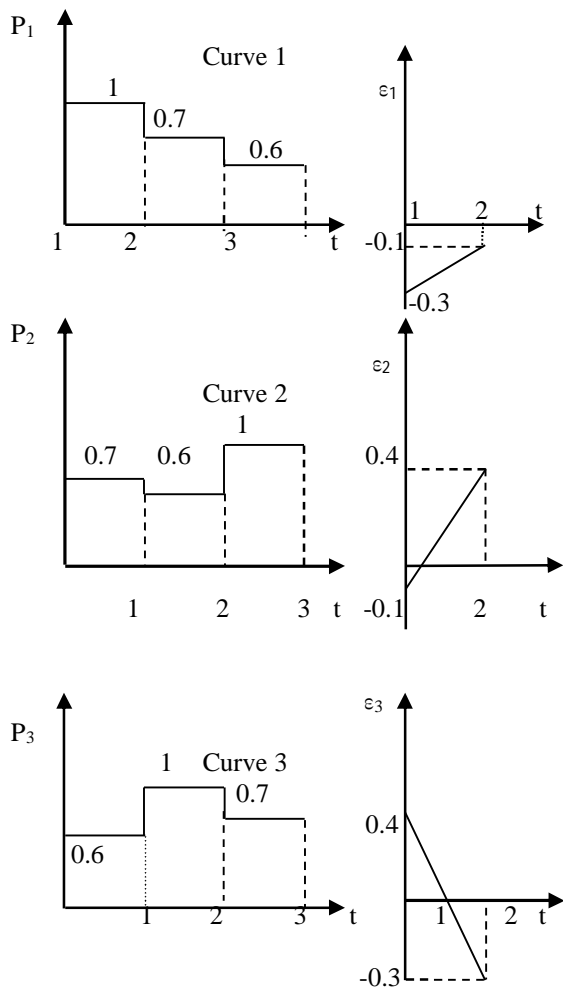
của hệ số góc ε . Các góc α của một đồ thị được trình bày trên Hình 1. Hệ số góc của α chính là tang góc ε .

Theo [6], hai phân tử i và j sẽ thuộc cùng một nhóm nếu trị tuyệt đối của $(\varepsilon_i(k)-\varepsilon_j(k))$ cho tất cả các phân đoạn thời gian k nhỏ hơn giá trị đủ nhỏ nào đó.

Với đồ thị phụ tải trong hệ đơn vị tương đối, khoảng cách Oclit của hai đồ thị sẽ thường là nhỏ và tình huống sau sẽ xảy ra: hai đồ thị hoàn toàn khác nhau về hình dạng song lại có cùng khoảng cách (1) tới một đồ thị khác (Hình 2). Trong Hình 2 hai đồ thị 2 và 3 có cùng khoảng cách tới đồ thị 1. Trong khi đó, xét đường cong hệ số góc ε thì đường cong 2 và 3 lại có hình dạng hoàn toàn khác nhau và do vậy, khoảng cách Oclit của hai đường cong hệ số góc 2 và 3 tới đường cong hệ số góc 1 sẽ hoàn toàn khác nhau



Hình 1. Góc α

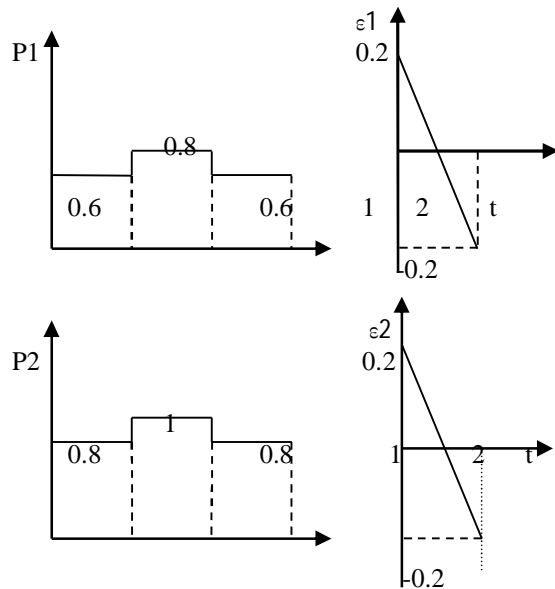


Hình 2. Đường cong hệ số góc

Tuy nhiên nếu đường cong hệ số góc (tang góc) được coi là một đặc thù duy nhất cho việc xét phân nhóm, thì chiều (hướng) và sự thay đổi của tải theo mỗi phân đoạn thời gian sẽ được xem xét, song lại không quan tâm tới giá trị của trục y tức là chính giá trị thực của tải. Hình 3 sẽ giải thích rõ hơn về điều này.

Trong hình này, hai đồ thị 1 và 2 có cùng đồ thị tang góc, song lại có sự tiêu thụ tải hoàn toàn khác nhau.

Như vậy, để phân nhóm đồ thị phụ tải trong hệ đơn vị tương đối, giải pháp tốt nhất là phối hợp cả đồ thị phụ tải và đường cong hệ số góc của nó.



Hình 3. Hai đường cong hệ số góc giống nhau nhưng lại có đồ thị phụ tải khác nhau.

3. THUẬT TOÁN PULSAR

3.1. Thuật toán Pulsar truyền thống

Một trong các kỹ thuật phân nhóm là giải thuật Pulsar [7]. Với quá trình phân nhóm khi số liệu đầu vào (số phần tử cần được phân nhóm) là ít thì giải thuật này tỏ ra đơn giản hơn và hiệu quả hơn. Thuật toán bao gồm nhiều giai đoạn. Trong mỗi giai đoạn, một nhóm sẽ được phát hiện và số phần tử trong nhóm đó sẽ bị loại trừ trong quá trình tiếp theo. Quá trình sẽ tiếp diễn cho tới khi tất cả dữ liệu được xem xét. Có nhiều bước lặp cho mỗi giai đoạn và trên mỗi bước lặp,

bán kính phân nhóm sẽ thay đổi tùy thuộc vào số phần tử rơi vào nhóm khi quét.

Chọn tâm ban đầu e^0 và tính các bán kính r_{max} , r_{min} (là các khoảng cách lớn nhất và nhỏ nhất giữa các phần tử).

Bán kính r là khoảng cách giữa hai phần tử x và y :

$$r = \|x - y\| = \sqrt{\sum_{j=1}^n (x_j - y_j)^2}$$

Ở đây n -số chiều (số đặc tính) của véc tơ đầu vào x và y (ví dụ nếu x và y là hai đồ thị phụ tải trong 24 giờ thì n bằng 24).

Giải thuật được trình bày như sau: Cho một giai đoạn:

1- Với $r_0 = \frac{r_{min} + r_{max}}{2}$ xác định nhóm

$$S^0 = \{x \in X : \|x - e^0\| \leq r_0 \}$$

Tính số phần tử n_0 rơi vào nhóm S^0 khi cho $\gamma_0 = 0$, với γ_0 là số lần dao động bán kính ở bước ban đầu.

2- Giả sử trên bước thứ m , với tâm e^m và bán kính r_m :

$$\text{Xác định } S^m = \{x \in X : \|x - e^m\| \leq r_m \} \quad (2)$$

Tính số phần tử n_m trong nhóm S_m và xác định $\gamma = \gamma_m$

$$\text{Tính : } e^{m+1} = \frac{1}{n_m} \cdot \sum_{x_i \in S^m} x_i \quad (3)$$

$$r_{m+1} = \begin{cases} \min(r_m + \lambda\delta, r_{max}), & n_m \leq n_{min} \\ \max(r_m - \lambda\delta, r_{min}), & n_m > n_{max} \\ \text{và } \gamma_m < \gamma_{cp} \text{ hoặc } e^{m+1} \neq e^m \\ r_m - \text{Cho trường hợp còn lại} \end{cases} \quad (4)$$

$$\text{Với : } \lambda = \frac{1}{1 + \gamma_m}$$

n_{max} và n_{min} là số phần tử lớn nhất và nhỏ nhất trong nhóm. δ là ngưỡng nào đó để điều chỉnh bán kính; γ_{cp} là số lần cho phép dao động bán kính, thường được gán bằng 2.

Cho $\gamma_1 = \gamma_0 = 0$ và với $m \geq 1$:

$$\gamma_{m+1} = \begin{cases} \gamma_m & \text{if } \Delta r_m \cdot \Delta r_{m+1} \geq 0 \\ \gamma_m + 1 & \text{if } \Delta r_m \cdot \Delta r_{m+1} < 0 \end{cases}$$

Từ (4) cho thấy là bán kính quét sẽ thay đổi khi số phần tử nhóm vượt quá n_{max} hoặc nhỏ hơn n_{min} .

3-Nếu $e^{m+1} = e^m$, $r_{m+1} = r_m$ thì dừng lại, nếu không quay về bước 2.

Như vậy một nhóm sẽ được hình thành sau giai đoạn đầu tiên. Giai đoạn thứ hai sẽ được lặp lại cho các phần tử còn lại và quá trình cứ tiếp diễn như thế.

3.2. Thuật toán Pulsar cải biên

Như đã đề cập ở mục 2, mỗi khách hàng sẽ được xem xét phân nhóm theo hai đặc thù: đồ thị phụ tải và đồ thị tang góc. Do đó giải thuật Pulsar cần thiết phải được cải biên lại. Trong bài báo này, hai thuật toán cải biên đã được xây dựng. Trong giải thuật thứ nhất, giải thuật Pulsar sẽ được dùng hai lần một cách tuần tự cho mỗi giai đoạn. Trước hết, giải thuật Pulsar được áp dụng cho tập đồ thị phụ tải. Sau khi xác định được một nhóm, các đồ thị tang góc của các phần tử trong nhóm này sẽ được phân nhóm bằng giải thuật Pulsar một lần nữa. Và như vậy sẽ tạo được một nhóm đầu tiên. Các phần tử còn lại lại được phân nhóm theo qui trình như trên để xác định nhóm thứ hai. Rồi lại tiếp tục cho nhóm thứ ba và cứ tiếp tục như thế. Ký hiệu giải thuật này là Pulsar cải biên 1.

Giải thuật thứ hai (được ký hiệu là Pulsar cải biên 2) được trình bày như sau: Giả sử tại bước lặp m nào đó, với mỗi phần tử (khách hàng), đầu tiên tính khoảng cách từ đồ thị phụ tải của nó tới tâm 1 và kiểm tra điều kiện (2). Nếu (2) được thỏa mãn, tính tiếp khoảng cách từ đồ thị tang góc tới tâm 2. Nếu lại tuân thủ điều kiện (2), phần tử này sẽ thành phần tử của nhóm đầu tiên. Điều này có nghĩa là có hai tâm và hai bán kính cần hiệu chỉnh ở các bước 2,3. Tâm thứ nhất là tâm tính theo đồ thị phụ tải và được gọi là tâm 1. Tâm 2 chính là tâm theo đồ thị tang góc.

4. ỨNG DỤNG

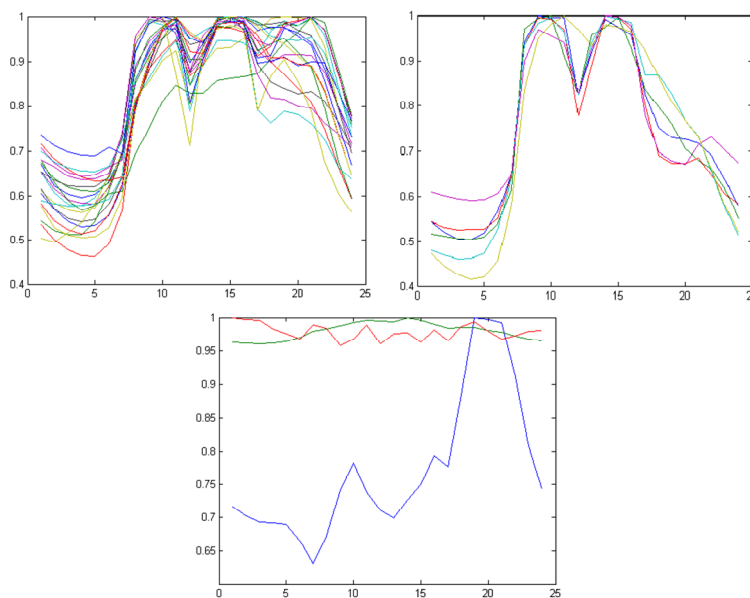
Do không có thông tin về đồ thị phụ tải từng

khách hàng và như để minh họa về áp dụng ý tưởng của bài báo, các số liệu đồ thị phụ tải của thành phố Hồ Chí Minh sẽ được sử dụng. Các đồ thị này là của hai năm: 2011 và 2012. Với năm 2011 sẽ có 29 trạm, còn năm 2012 số trạm được tăng lên 41.

Khảo sát được tiến hành cho 4 giải thuật: 1- sử dụng giải thuật Pulsar truyền thống với đồ thị phụ tải; 2-theo ý tưởng của [6] dựa trên đồ thị tang góc; giải thuật 3 và 4 là Pulsar cải biên 1 và 2.

A. Năm 2011

4.1. Giải thuật Pulsar truyền thống



Hình 4. Phân nhóm theo giải thuật Pulsar dựa trên đồ thị phụ tải-Năm 2011

Khi không gian đầu vào là đồ thị phụ tải, có ba nhóm được tạo thành như trên Hình 4. Nhóm đầu tiên có ba đỉnh với đỉnh tối tương đối cao.

Trong khi đó ở nhóm thứ hai thì tải vùng thấp điểm là nhỏ hơn cả trong 3 nhóm. Nhóm thứ ba bao gồm các đồ thị tương đối phẳng.

4.2. Dựa trên đồ thị tang góc

Số lượng các nhóm là rất lớn, gần bằng số trạm. Điều này có nghĩa là khả năng tương tự (giống nhau) theo đồ thị tang góc là rất nhỏ.

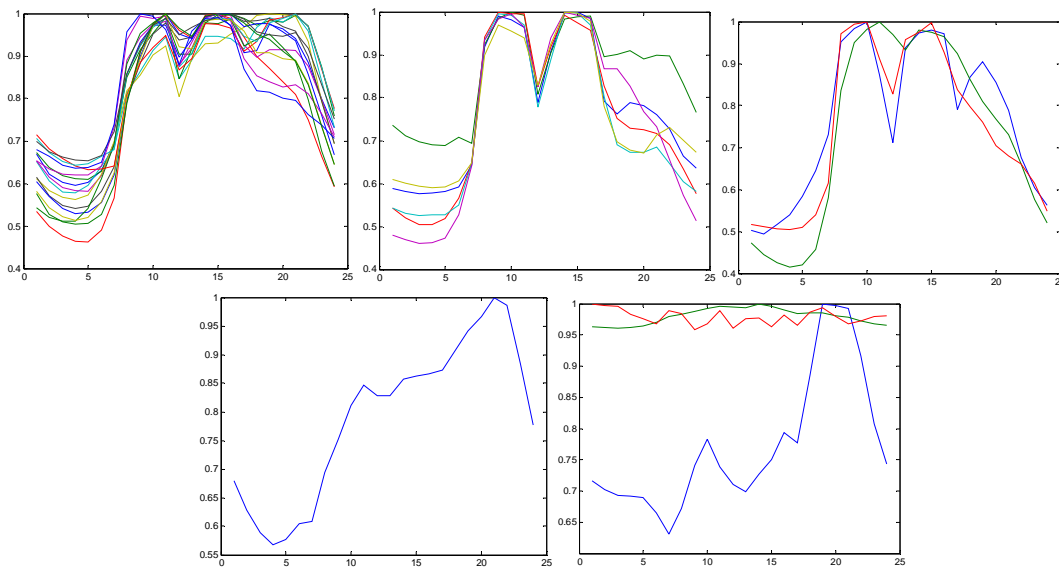
4.3. Giải thuật Pulsar cải biên 1

Số nhóm thu được là 5 và các nhóm được biểu diễn trên Hình 5. Ba nhóm đầu tiên có tải đỉnh sáng và đầu giờ chiều. Nhóm đầu tiên có tải thời điểm tối tương đối cao (lớn hơn 0.8). Nhóm thứ ba đặc trưng bởi tải ban đêm nhỏ nhất. Nhóm thứ tư có một đỉnh và đó là đỉnh tối. Nhóm cuối gồm

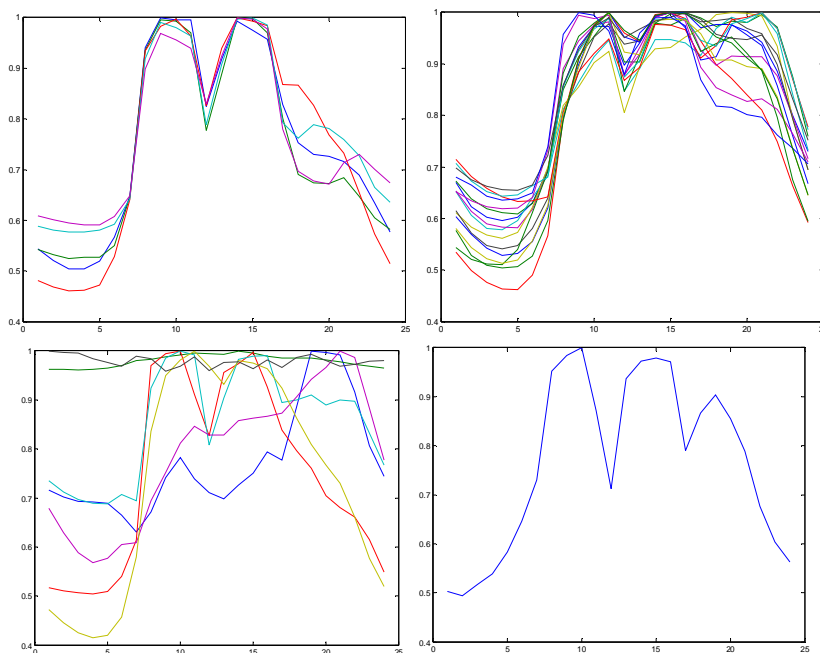
các đồ thị tương đối phẳng. Như vậy so với giải thuật Pulsar truyền thống, các nhóm bây giờ có các đặc thù rõ nét hơn.

4.4. Giải thuật Pulsar cải biên 2

Các kết quả được trình bày trên Hình 6. Trong 4 nhóm thu được thì nhóm thứ ba có đặc thù rất lộn xộn về hình thù đồ thị tải. Điều này có nghĩa là giải thuật này có kết quả không thuyết phục.



Hình 5. Các nhóm của 29 đồ thị phụ tải theo giải thuật Pulsar cải biên 1



Hình 6. Các nhóm của 29 đồ thị phụ tải theo giải thuật Pulsar cải biên 2.

B. Năm 2012

Các đồ thị của 41 trạm được đưa vào phân nhóm. Với giải thuật Pulsar truyền thống sẽ có 2 nhóm được hình thành và không rút được bất kỳ đặc thù nào từ mỗi nhóm. Với giải thuật dựa trên đường cong tang góc, một kết quả không tốt thu được. Lý do là số nhóm tìm được quá lớn (hơn 30 nhóm).

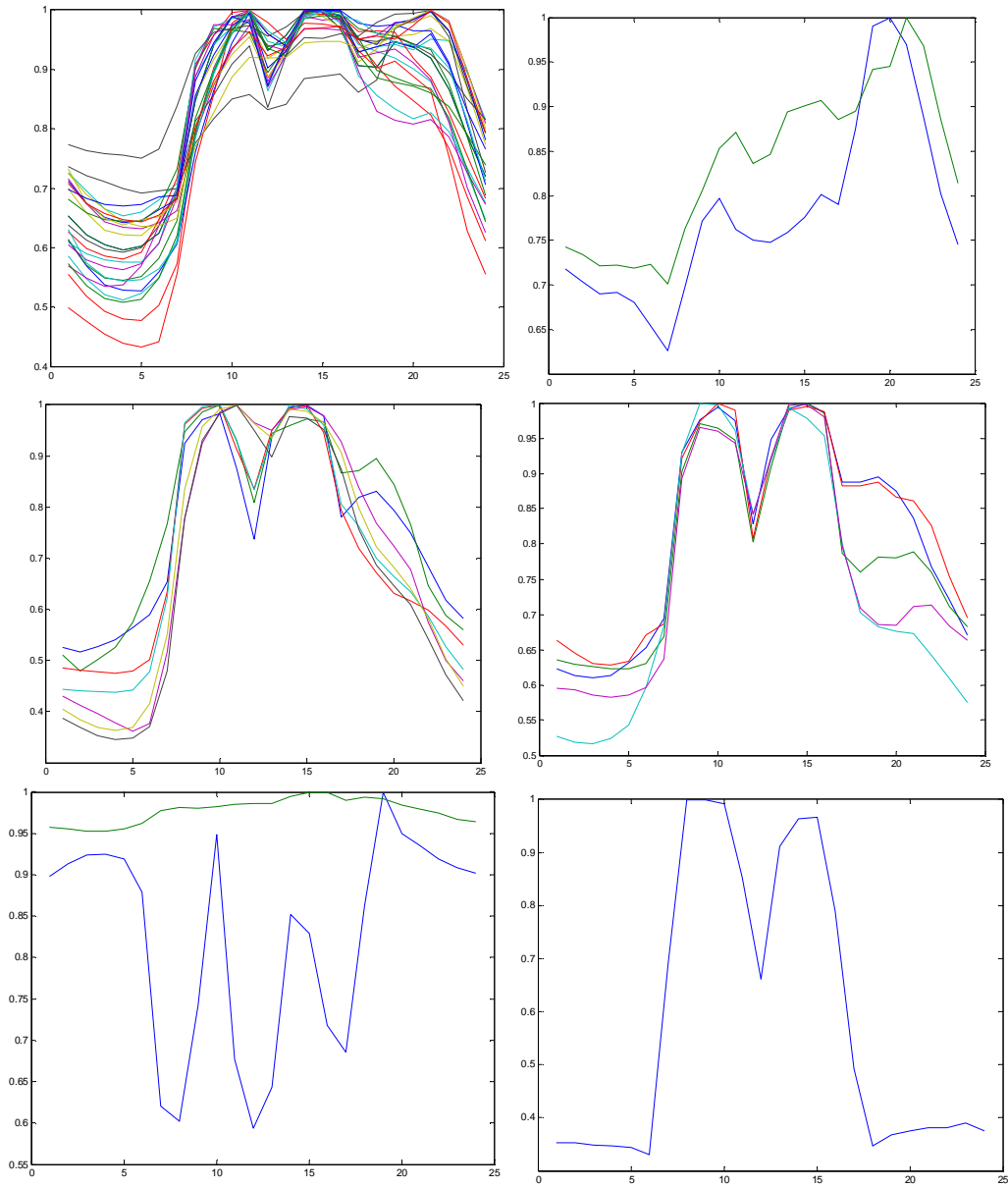
4.5. Giải thuật Pulsar cải biên 1

Các kết quả được trình bày trên Hình 7 và Bảng 1.

4.6. Giải thuật Pulsar cải biên 2

Kết quả thu được cũng giống như của giải thuật Pulsar truyền thống.

Qua phân tích đồ thị của hai năm, kết luận là: đa số các đồ thị rơi vào một nhóm với tải tối tương đối cao, ngoài ra còn có đỉnh sáng và đỉnh đầu giờ chiều.



Hình 7. Các nhóm của 41 đô thị theo giải thuật cải biên Pulsar 1

Bảng 1. Các đặc thù chính của các nhóm theo giải thuật Pulsar cải biên 1-Năm 2012

Nhóm	Đặc thù chính
1	2 đỉnh áp đảo: sáng và đầu giờ chiều, tải chiều tương đối cao (>0.8) và là cao nhất trong các nhóm có đỉnh sáng và đầu giờ chiều.
2	Một đỉnh và là đỉnh tối.
3	Tải giờ thấp điểm là thấp nhất; hai đỉnh: sáng và đầu giờ chiều.
4	Tải đêm tương đối cao; hai đỉnh: sáng và đầu giờ chiều.
5	Tính phẳng của đồ thị tương đối cao.
6	Tải tối là rất nhỏ

5. KẾT LUẬN

Phân loại đồ thị phụ tải dựa trên khoảng cách Oclit trong không gian 24 chiều là chưa đủ tin cậy. Một cách tiếp cận hợp lý hơn sẽ được dựa trên đồ thị phụ tải và đồ thị tang góc. Với số lượng đồ thị phân loại ít thì giải thuật Pulsar là hữu hiệu, tuy nhiên giải thuật này cần phải được cải biên để phù hợp với không gian đầu vào bao gồm cả hai đặc thù trên. Nghiên cứu cho trường hợp thành phố Hồ Chí Minh cho thấy giải thuật này cho kết quả tốt hơn cả. Nhóm có nhiều đồ thị rơi vào nhất là nhóm có hai đỉnh tải: buổi sáng và đầu giờ chiều cùng với tải buổi tối khá cao.

Combination of load curves and tangent curve for customer's load curve clustering

- Phan Thi Thanh Binh

Department of Electrical and Electronics Engineering, Ho Chi Minh city University of Technology, VNU-HCM

ABSTRACT

The load curve clustering for electrical customers traditionally is based on the 24-dimension input space. It means that every load curve is considered as an element with 24 attributes corresponding to 24 load values per day. But in some cases, the

load curve itself can not lead to the right cluster when the two curves have different forms but have the same distance to the third one. To overcome this limitation, the present paper pays attention to the selection of the input space. From each load curve,

the tangent curve will be received. Now the clustering will be based not only on the load curve but on the tangent curve. The clustering techniques used the Pulsar algorithm with some modifications to fit the

Từ khóa: The load curve clustering, the Euclidean distance.

input space. Examining for the electrical load curve of Ho Chi Minh city costumers, the propose approach has achieved the best results.

REFERENCES

- [1]. Chico G. et al, Comparison among Clustering teacniques for Electricity Classification, *IEEE Trans, Power Syst*, Vol.21, N.2, 2006
- [2]. S. Valero, M. Ortiz, C. Senabre, C. Alvarez, F.J.G. Franco and A. Gabaldon, Methods for customer and demand response policies selection in new electricity markets, *IET Gener. Transm. Distrib.*, Vol. 1, No. 1, January 2007
- [3]. G. Chicco et al., Customer characterization for improve the tariff offer. *IEEE Trans, Power Syst.*, (2003), vol. 18, pp. 381- 387
- [4]. Gianfranco Chicco, Roberto Napoli and Federico Piglione, Application of Clustering Algorithms and Self – Organising Maps to Classify Electricity Customers, *IEEE Proc. Bologna Power Tech. Conference*, 2003
- [5]. Y.-H.Pao and D.J.Sobajic, Combined use of unsupervised and distribute the load diagrams among the groups formed. An supervised learning for dynamic security assessment, *IEEE Trans. Overall evaluation of the algorithms leads to consider the on Power Systems* 7, 2 (May 1992) 878-884.
- [6]. Lei Wen, The Application of Temporal Pattern Clustering Algorithms in DSM, *Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)* Volume 1, 2006
- [7]. Aivazyan S.A., *Classification and reduction of dimensionality*, Financial and statistics, Moscow, 1989, 606p.
- [8]. Địa chỉ liên hệ: Phan Thị Thanh Bình
- [9]. Khoa Điện-Điện tử, ĐHBK, ĐHQG-HCM
- [10]. Bộ môn Cung cấp điện, ĐHBK Tp.HCM
- [11]. Email: thanhbinh055@yahoo.com.