

SỬ DỤNG THUẬT TOÁN SVM ĐỂ PHÂN LOẠI PHỤ TẢI ĐIỆN

Nguyễn Văn Bình⁽¹⁾

(1) Trường Đại học Thủ Dầu Một

Ngày nhận bài 24/6/2022; Ngày phản biện 30/6/2022; Chấp nhận đăng 30/7/2022

Liên hệ Email: binhnv@tdmu.edu.vn

<https://doi.org/10.37550/tdmu.VJS/2022.04.326>

Tóm tắt

Với sự phát triển của công nghiệp 4.0, hầu hết tất cả mọi lĩnh vực trong đời sống nói chung đều áp dụng các thuật toán thông minh, hệ thống IoT, trí tuệ nhân tạo với mục đích để tăng sự hiệu quả, năng suất trong công việc quản lý, giám sát, tăng hiệu quả kinh tế. Với ngành điện hiện tại cũng chú trọng việc sử dụng và truyền tải điện năng một cách thông minh, một trong những sự thông minh đó là tự phân loại được các loại tải sử dụng hàng ngày của các khu vực khác nhau như nhà hàng, khách sạn, khu vực vui chơi giải trí, ... bên cạnh việc phân loại tải sử dụng điện năng còn có thêm các thông tin khách hàng, thông tin người dùng, những nhà sản xuất và cung cấp điện năng có thể dựa vào sự phân loại tải những khu vực cụ thể để dự đoán được thói quen của người sử dụng dịch vụ, từ những dự đoán này nhà sản xuất có những phương án tối ưu hơn cho việc cân bằng pha phù hợp trong quá trình cung cấp điện năng cho từng khu vực. Trong bài báo này, tác giả trình bày kết quả thực hiện phân loại phụ tải dựa vào học sâu đối với các lưới điện, cụ thể là sử dụng phương pháp Support Vector Machine (SVM) để thực hiện việc phân loại phụ tải. Kết quả đạt được cao hơn các phương pháp phân loại truyền thống, tỷ lệ chính xác của phương pháp lớn hơn 0,93.

Từ khóa: học sâu, lưới điện thông minh, phân loại phụ tải điện năng, support vector machine (svm), tiêu thụ năng lượng

Abstract

USING SVM ALGORITHM TO CLASSIFY ELECTRICAL LOADS

With the development of Industry 4.0, almost all areas of life in general are applying smart techniques, IoT systems, and artificial intelligence with the goal of increasing efficiency and productivity at work. manage, monitor, increase economic efficiency, etc. With the electricity industry also focusing on the smart use and transmission of electricity, one of the smart ones is the self-categorization of loads used in daily life. In addition to load classification, there are information customers, information users, manufacturers, and suppliers. Power supply can rely on load classification of possible areas to predict service users. From these projects, manufacturers have more optimal solutions for suitable phase balance in the process. power supply for each area. In this paper, the author presents the results of performing deep learning dependent classification for electrical categories, specifically using the Support Vector Machine (SVM) method to perform subcategory

loading. The results obtained are higher than the method distribution system. The accuracy ratio of the method is greater than 0.93.

1. Giới thiệu

Phân loại là một trong những nhiệm vụ cơ bản trong lĩnh vực học sâu (deep learning) cũng như học máy (machine learning) (Nguyen và cs., 2019). Khi tiến hành nhiệm vụ phân loại chúng ta cần thu thập một lượng lớn dữ liệu nhất định. Công việc đi tìm một mô hình toán học thích hợp để mô tả sự quan hệ của bộ dữ liệu đã thu thập trước đó được gọi là phân loại.

Với mục đích quản lý hiệu quả cho việc cân bằng giữa người sử dụng và nhà cung cấp về năng lượng điện, việc phân loại tải sử dụng tại các khu vực là rất cần thiết. Để tăng hiệu quả quản lý thì công việc phân loại tải sử dụng và kế hoạch được lập ra một cách phù hợp từ việc phân loại là cần thiết và gắn kết với nhau (Nguyen và cs., 2018). Năng lượng có tầm quan trọng trong sự phát triển của tất cả lĩnh vực trên thế giới hiện tại, trong đó có năng lượng điện, một dạng năng lượng được sử dụng rất nhiều trong mọi lĩnh vực. Vì vậy, việc phân loại phụ tải điện trong hệ thống điện có ý nghĩa quan trọng trong việc phát triển kinh tế nói chung và sự sinh hoạt của mọi người dân nói riêng. Do đó, cần có một mô hình toán phân loại phụ tải điện thật chính xác để không ảnh hưởng đến việc quản lý chung cũng như hạn chế sự ảnh hưởng đến nền kinh tế của khu vực, của đất nước và lớn hơn là toàn thế giới.

Thực tế thì có rất nhiều nhu cầu cụ thể khác nhau nên phải có một mô hình phân loại thích hợp, lựa chọn kiểu phân loại phải hợp lý. Cụ thể như trường hợp cần chiến lược và kế hoạch phát triển lâu dài thì phải chọn kiểu phân loại dạng trung hạn hoặc dài hạn, ngược lại với việc khi sử dụng kiểu phân loại ngắn hạn thì phù hợp cho những mục đích vận hành. Một số kiểu dự báo thông thường được sử dụng đó là phân loại điều độ, ngắn hạn, trung hạn, dài hạn. Với kiểu phân loại điều độ thì phân loại tải theo giờ hoặc vài phút sử dụng, phân loại ngắn hạn thì thời gian phân loại sẽ dài hơn, thường sẽ phân loại theo ngày, tháng. Thời gian phân loại theo năm (khoảng 5-7 năm) là phân loại theo kiểu trung hạn, cuối cùng là phân loại theo thời gian dài nhất (từ 10-20 năm) là dạng phân loại dài hạn.

Thuật toán SVM sử dụng rất nhiều trong các lĩnh vực về điện, một số công trình nghiên cứu trước đây cũng đã sử dụng giải thuật SVM để nghiên cứu về mô hình dự báo phụ tải điện (Wei-Chiang Hong, 2009), phân loại hệ thống điện (Amine và cs., 2009), phân loại và mô hình hóa phụ tải điện trong hệ thống điện (Vignesh và cs., 2017), tối ưu hóa công suất phức (Yong và cs., 2017).

Như trình bày ở trên thì có nhiều dạng phân loại khác nhau, đề tài này với mục tiêu chính là nghiên cứu vấn đề phân loại phụ tải cho lưới điện bằng cách sử dụng thuật toán SVM (Cortes và cs., 1995; Nijhawan và cs., 2020). Hơn nữa, đề tài này cũng trình bày chi tiết hoạt động của giải thuật SVM và sử dụng ngôn ngữ Python để lập trình (Adrian Rosebrock, 2017), với dữ liệu thu được, dữ liệu được lấy từ lượng điện tiêu thụ của Hoa Kỳ năm 2004 được tác giả trích xuất ở địa chỉ. Kết quả phân loại trong bài báo này giúp người cung cấp và sử dụng chủ động trong việc vận hành thiết bị, trang bị thiết bị có công suất phù hợp.

2. Phương pháp phân loại và đề xuất

2.1. Ý tưởng

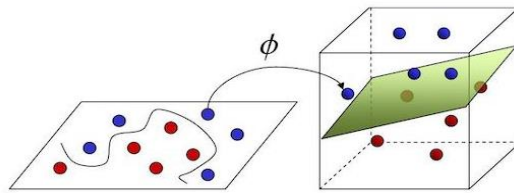
Trong mục này sẽ thảo luận về giải thuật phân loại nổi tiếng là Support Vector Machine (SVM) (Cortes và cs., 1995), nó là một phương pháp học có giám sát trong các mô hình nhận dạng mẫu. SVM không chỉ hoạt động tốt với các dữ liệu được phân tách tuyến tính mà còn tốt với cả dữ liệu phân tách phi tuyến.

Ý tưởng của thuật toán SVM khá đơn giản, nhưng để hiểu và cải thiện hiệu suất của nó chúng ta cần nắm rõ kiến thức về toán đại số và tối ưu. Đầu tiên, tác giả muốn trình bày về khái niệm siêu phẳng phân chia tuyến tính và biên độ.

2.2. Những khái niệm cơ bản

2.2.1. Siêu phẳng phân cách

Siêu phẳng phân cách là một hàm số tương tự như phương trình đường thẳng: $y = ax + b$. Trong giải thuật SVM, siêu phẳng lúc này chính là một đường thẳng.



Hình 1. Siêu phẳng phân chia tập dữ liệu trong không gian n chiều

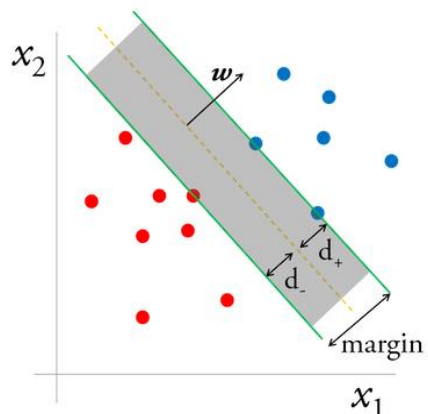
Về ý tưởng thì SVM sử dụng phương pháp để ánh xạ tập dữ liệu đầu vào sang không gian nhiều chiều hơn. Khi đó, SVM sẽ xem xét và quyết định chọn ra siêu phẳng phù hợp nhất để phân lớp tập dữ liệu đầu vào như hình 1.

2.2.2. Biên độ (Margin)

Như được hiển thị trong hình 2, biên độ là khoảng cách giữa hai siêu phẳng đến hai điểm dữ liệu gần nhất tương ứng với hai phân lớp.

Trong thực tế, ta có thể tìm được vô số các siêu phẳng phân chia trên cùng một tập dữ liệu đầu vào và mong muốn tìm đường thẳng phân chia sao cho tối ưu nhất. Nói một cách đơn giản hơn, điều này có nghĩa là có sai số phân loại phải nhỏ nhất trên bộ dữ liệu đầu vào. Do đó, mục tiêu của phương pháp phân lớp SVM là tìm một siêu phẳng phân cách giữa hai lớp sao cho khoảng cách lề (margin) giữa hai lớp đạt cực đại. Theo nghiên cứu trước đây, biên độ gồm có hai loại chủ yếu là:

- Biên độ mềm (soft margin)
- Biên độ cứng (hard margin)



Hình 2. Minh họa về biên độ (margin)

Để tránh vấn đề quá khớp (overfitting), chúng ta cần có margin cao và phải chấp nhận một vài điểm dữ liệu không được phân chia chính xác. Các điểm dữ liệu này được gọi là nhiễu. Khi đó, biên độ trong trường hợp này gọi là **biên độ mềm**. Ngược lại, **biên độ cứng** có nghĩa là giải thuật SVM tạo ra biên độ mà hầu như không có bất kì điểm dữ liệu nhiễu nào hết, mà tất cả chúng phải đều thỏa mãn sự phân chia. Với các bài toán thực tế, việc tìm được biên độ cứng là một nhiệm vụ vô cùng khó khăn. Vì thế, việc chấp nhận sai lệch ở một mức độ chấp nhận được là một cách hiệu quả để đạt được kết quả phân loại tối ưu.

Một điều vô cùng quan trọng khi lựa chọn tham số cho giải thuật SVM, đặc biệt là tham số C, cái mà được giới thiệu với quy tắc sau (Cortes và cs., 1995):

- $C = \infty$ có nghĩa là không cho phép sai lệch, đồng nghĩa với biên độ cứng.
- C lớn cho phép sai lệch nhỏ, thu được biên độ nhỏ.
- C nhỏ cho phép sai lệch lớn, thu được biên độ lớn.

3. Thực nghiệm

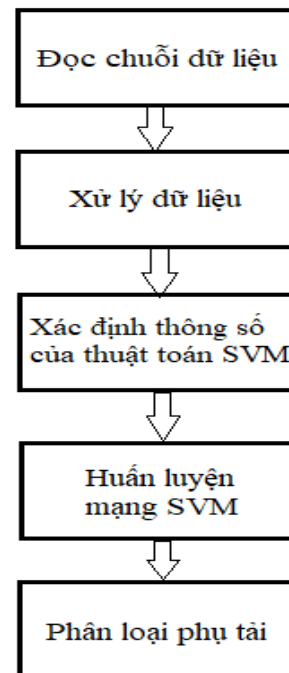
3.1. Giới thiệu bộ dữ liệu

Bộ dữ liệu lấy từ lượng điện tiêu thụ của Hoa Kỳ năm 2004 được tác giả trích xuất ở địa chỉ. Bộ dữ liệu này giúp ích rất nhiều cho việc mô hình giải thuật SVM mà đề tài này đã đề xuất. Nói cách khác, đây là bộ dữ liệu quý giá mà SVM có thể phân loại hiệu quả phụ tải điện cho lưới điện thông minh ở Hoa Kỳ.

Sơ đồ khối

Các phần chính trong giải thuật (hình 3):

- Đọc chuỗi dữ liệu.
- Xử lý dữ liệu ngõ ra và tạo nhãn.
- Xác định thông số của thuật toán SVM
- Huấn luyện mạng SVM.
- Phân loại phụ tải



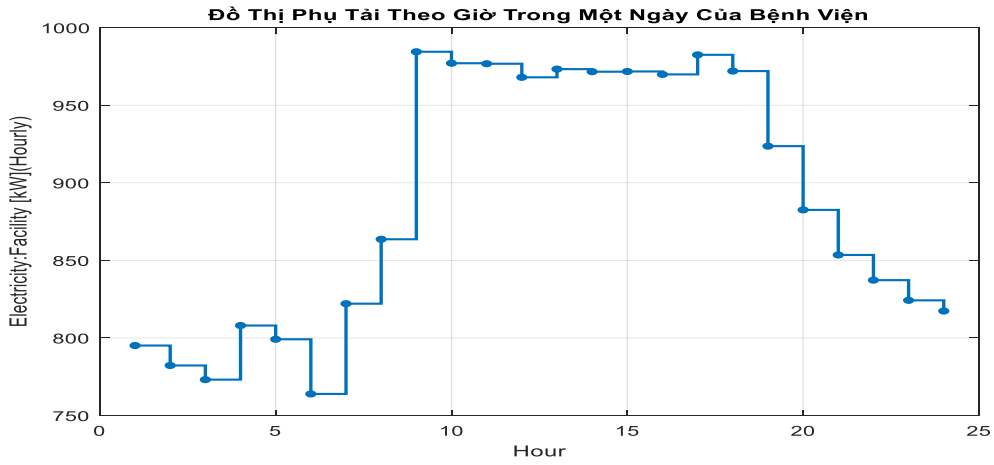
Hình 3. Sơ đồ khối chung

3.2. Kết quả thực nghiệm từng bước

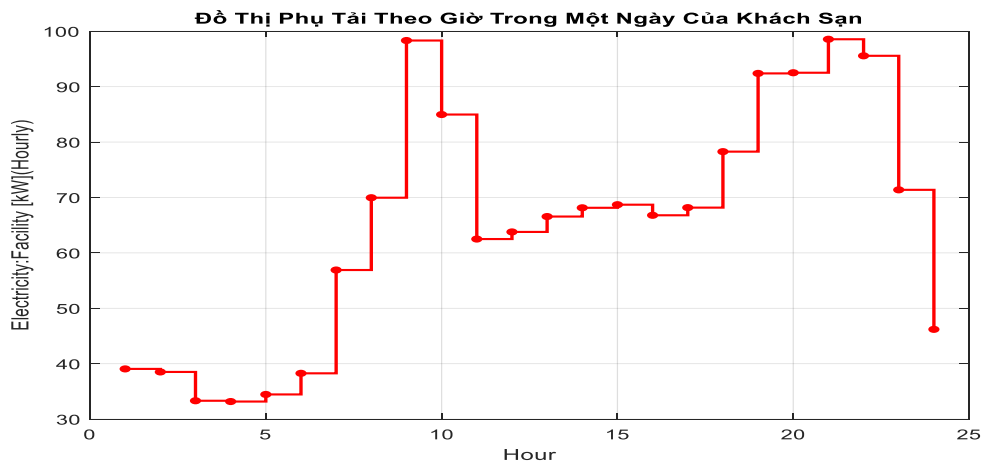
3.2.1. Đọc chuỗi dữ liệu

Nhập chuỗi dữ liệu là lượng điện năng tiêu thụ phân bố theo giờ trong năm 2004 của Hoa Kỳ, trong đó các bước thời gian là đơn vị giờ và các giá trị tương ứng với tổng năng lượng điện tiêu thụ (Hình 4, 5, 6, 7, 8).

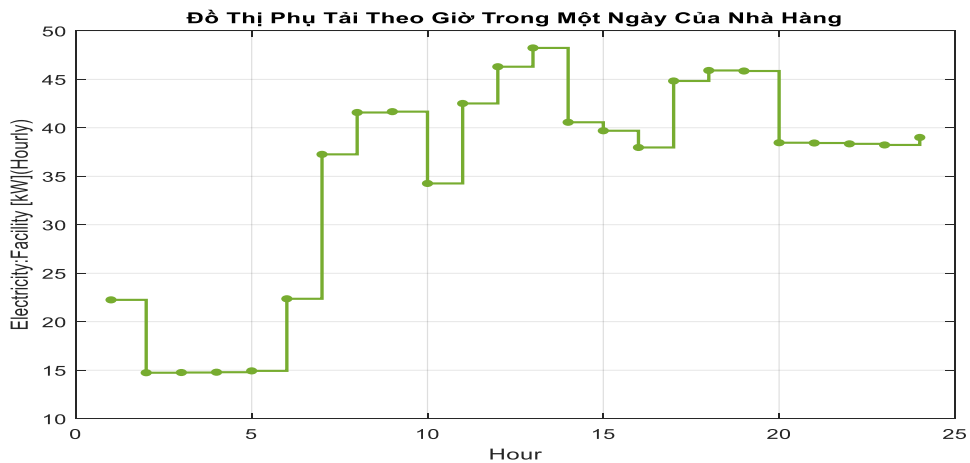
Phân chia dữ liệu để huấn luyện và kiểm tra là dùng 80% chuỗi dữ liệu để huấn luyện và 20% để kiểm tra.



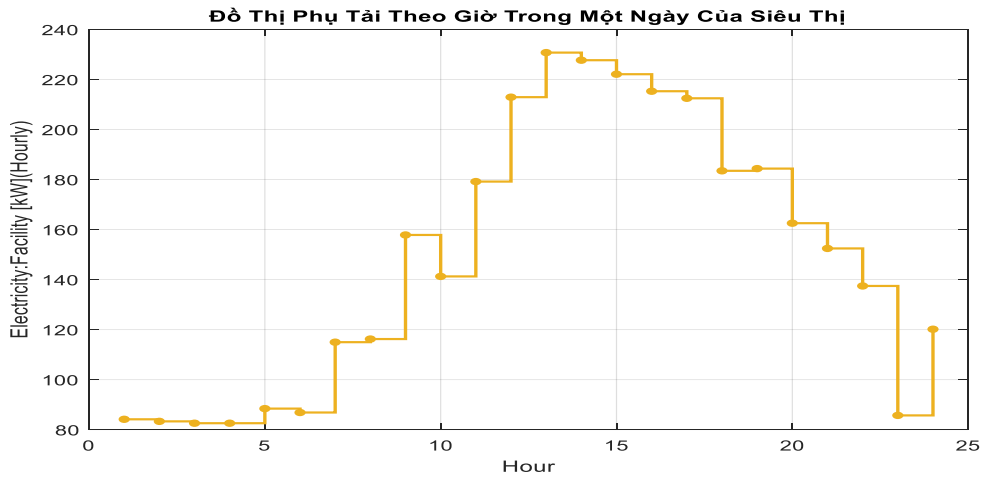
Hình 4. Đồ thị năng lượng điện sử dụng của phụ tải ở bệnh viện trong một ngày



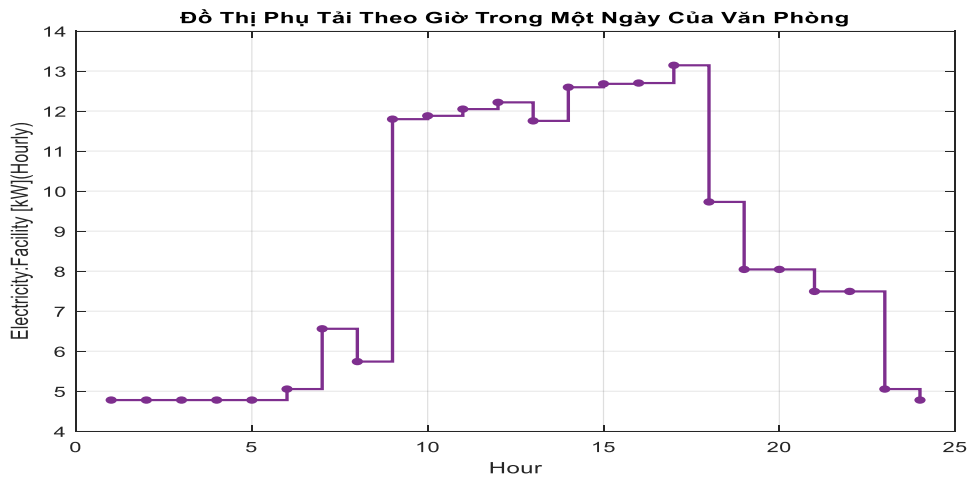
Hình 5. Đồ thị năng lượng điện sử dụng của phụ tải ở khách sạn trong một ngày



Hình 6. Đồ thị năng lượng điện tiêu thụ của phụ tải ở nhà hàng trong một ngày



Hình 7. Đồ thị năng lượng điện tiêu thụ của phụ tải ở siêu thị trong một ngày



Hình 8. Đồ thị năng lượng điện tiêu thụ của phụ tải ở văn phòng trong một ngày

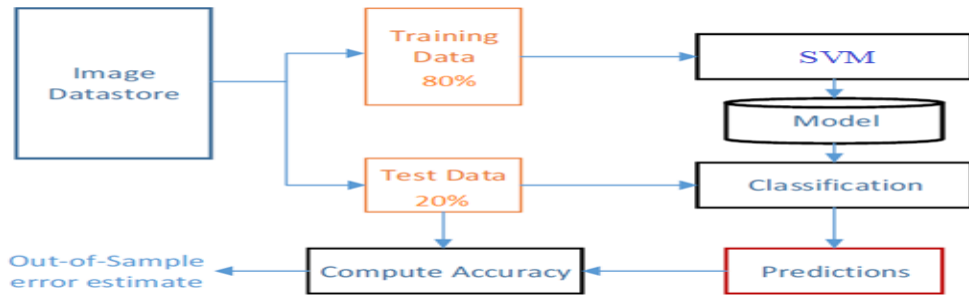
3.2.2. Xác định cấu trúc và huấn luyện mạng

Ngõ vào của hệ thống sẽ là tập hợp 5 chuỗi gồm các mảng có 2 phần tử (bảng 1). Phần tử đầu tiên sẽ là thời điểm (giờ) mà cơ sở tiêu thụ điện năng. Phần tử thứ hai sẽ là lượng điện năng tiêu thụ (kW) tương ứng với thời điểm đó

Bảng 1. Giá trị của 5 loại phụ tải tiêu biểu trong 1 ngày

Data	Fans:Electricity [kW](Hourly)																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
BV	79 5.1	78 2.2	773 .1	80 8	799 .1	763 .9	822 .1	863 .6	984 .4	977	976 .7	968	973 .2	971 .6	971 .7	969 .8	982 .5	972	923 .6	882 .5	853 .5	837 .2	824 .2	817 .3
KS	39	38. 5	33. 3	33 .1	34. 4	38. 2	56. 9	69. 9	98. 3	84. 9	62. 4	63. 7	66. 5	68. 1	68. 6	66. 7	68. 1	78. 2	92. 3	92. 5	98. 5	95. 5	71. 3	46. 1
NH	22. 2	14. 7	14. 7	14 .7	14. 9	22. 3	37. 2	41. 5	41. 6	34. 2	42. 5	46. 2	48. 2	40. 5	39. 6	37. 9	44. 8	45. 9	45. 8	38. 4	38. 4	38. 3	38. 2	39
ST	84. 1	83. 3	82. 5	82 .5	88. 4	86. 8	114 .9	116 .2	157 .8	141 .2	179 .1	212 .9	230 .7	227 .7	222 .1	215 .3	212 .4	183 .4	184 .3	162 .5	152 .4	137 .4	85. 7	120 .1
VP	4.2 2	4.2 2	4.2 2	4. 22	4.2 2	4.2 2	3.8	1.9 2	1.9 2	1.9 2	1.9 2	1.9 2	1.9 2	1.9 2	1.9 2	1.9 2	2.6	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2	4.2 2

Sơ đồ huấn luyện được thể hiện ở hình 9. Thời gian huấn luyện SVM mất 10 phút với số lượng ngõ vào là 43800 (Bảng 2). Độ chính xác đạt được là 93,6%.



Hình 9. Sơ đồ huấn luyện cho mô hình phân loại phụ tải

Bảng 2. Thời gian huấn luyện và độ chính xác của mô hình

Số lượng ngõ vào	43800
Thời gian huấn luyện (s)	632
Độ chính xác	0.93678

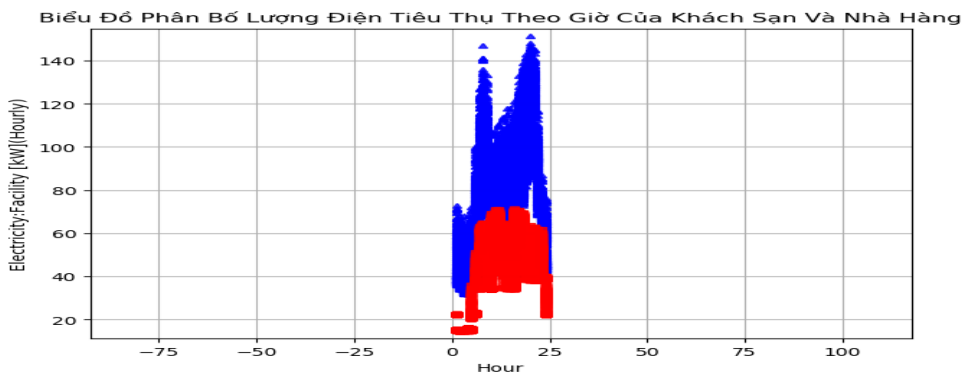
3.2.3. Xác định thông số thuật toán SVM

Xác định loại kernel sẽ được sử dụng trong thuật toán. Nó phải là một trong ‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’, ‘precomputed’. Ở đây tác giả thử nghiệm là chọn kernel là linear và cho kết quả tốt nhất.

C ở đây là hệ số cân bằng giữa nhiều và không nhiều. Nếu C càng lớn thì các nhiều càng nhiều điểm được coi là nhiều hơn tức là nhiều được coi trọng hơn. Trong đề tài này tác giả chọn $C = 1e5$.

3.2.4. Kết quả phân loại phụ tải

Mô hình phân loại phụ tải khá chính xác ở 3 địa điểm là bệnh viện, siêu thị và văn phòng với tỷ lệ độ chính xác tương ứng là 100%, 97%, 88% tương ứng. Còn 2 địa điểm còn lại là khách sạn và nhà hàng có độ chính xác phân loại không được cao, chỉ chiếm khoảng 63% và 58% (bảng 3). Nguyên nhân do lượng tiêu thụ điện năng của 2 địa điểm này tiêu thụ giống nhau ở khoảng 35-70 kWh ở khoảng thời gian 5-23h (Hình 10).



Hình 10. Lượng điện năng tiêu thụ phân bố theo giờ của khách sạn và nhà hàng

Bảng 3. Độ chính xác của mô hình phân loại phụ tải cho từng địa điểm khảo sát

Địa điểm	Độ chính xác
Khu vực bệnh viện	100%
Khu vực siêu thị	97%
Khu vực khách sạn	63%
Khu vực nhà hàng	58%
Khu vực văn phòng	88%

4. Kết luận

4.1. Kết luận

Phương pháp support vector machine (SVM) đã giải quyết các nhược điểm mà mạng truyền thống chưa làm được. Trong đề tài này tác giả đã phân loại phụ tải cho lưới điện, hệ thống điện dựa trên tập dữ liệu thực cụ thể là dữ liệu được lấy từ lượng điện tiêu thụ của Hoa Kỳ năm 2004 được tác giả trích xuất ở địa chỉ.

Trong quá trình nghiên cứu tác giả phân tích lý thuyết và kiểm chứng trên phần mềm. Qua đó thấy được phương pháp đề tác giả đề xuất là dùng giải thuật SVM để phân loại phụ tải có độ chính xác cao hơn so với các phương pháp truyền thống trước đây. Ngoài ra, trong nghiên cứu vẫn còn có một số trường hợp, mô hình phân loại vẫn chưa chính xác, nguyên nhân là do dữ liệu phụ tải quá đa dạng và luôn thay đổi liên tục. Vì thế phương pháp sử dụng SVM vẫn chưa là phương pháp phân loại tối ưu nhất cho bài toán phân loại phụ tải điện. Tuy nhiên, đây là cơ sở quản lý và vận hành lưới điện thông minh.

4.2. Hướng phát triển

Mặc dù đã đạt được một kết quả khả quan nhưng trong tương lai gần, tác giả sẽ tiếp cận một số nghiên cứu sau đây:

Tìm kiếm dữ liệu về lượng điện năng tiêu thụ của công nghiệp, dân dụng, thương mại ở Việt Nam để dự báo phụ tải điện năng để góp phần phát triển hệ thống lưới điện thông minh của Quốc gia.

Chạy trên nhiều tập dữ liệu khác nhau, số lượng dữ liệu nhiều hơn để cho kết quả chính xác hơn.

So sánh với nhiều phương pháp cạnh tranh.

TÀI LIỆU THAM KHẢO

- [1] Adrian Rosebrock (2017). *Deep Learning for Computer Vision with Python*.
- [2] Commercial and residential hourly load profiles for all TMY3 locations in the united states (2016). <https://openei.org/datasets/dataset/hourly-load-profiles-for-all-tmy3/locations-in-the-united-states>.
- [3] Cortes, C., Vapnik, V (1995). Support-vector networks. *Mach Learn*, 20, 273-297. <https://doi.org/10.1007/BF00994018>.

- [4] David De Yong, Sudipto Bhowmik, Fernando Magnago (2017). Optimized Complex Power Quality Classifier Using One vs. Rest Support Vector Machines. *Energy and Power Engineering*.
- [5] Ladjici Admed Amine, Dr. Boudour Mohamed, Rachedi Nadjib (2009). *Power System Applications of Support Vector Machine in Classification and Regression*. 3rd International Conference on Electrical Engineering.
- [6] Parag Nijhawan, Vinod Kumar Bhalla, Manish Kumar Singla, Jyoti Gupta (2020). Electrical Load Forecasting using SVM Algorithm. *International Journal of Recent Technology and Engineering*, 8(6).
- [7] T. D. Nguyen, T. D. Do, M. H. Le, N. T. Le and W. Benjapolakul (2019). Appliance Classification Method Based On K-Nearest Neighbors for Home Energy Management System. *2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 53-56, DOI: 10.1109/ICA-SYMP.2019.8646278.
- [8] T. D. Nguyen, V. K. Tran, T. D. Nguyen, N. T. Le and M. H. Le (2018). IoT-Based Smart Plug-In Device for Home Energy Management System. *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, 734-738, DOI: 10.1109/GTSD.2018.8595615.
- [9] Vignesh V, S.Chakrabarti, S.C. Srivastava (2015). *Classification and modelling of loads in power systems using SVM and optimization approach*. IEEE power and energy society general meeting.
- [10] Wei-Chiang Hong (2009). Electric load forecasting by support vector model. *Applied Mathematical Modelling*, 33.