

ỨNG DỤNG PHƯƠNG PHÁP PHÂN CỤM k -MEANS ĐỂ XÁC ĐỊNH ĐIỂM TẮC NGHẼN TRONG SẢN XUẤT

AN APPLICATION OF k -MEANS CLUSTERING METHOD TO IDENTIFY BOTTLENECKS IN MANUFACTURING PROCESS

Nguyễn Thị Phương Quyên*, Lê Vy

Trường Đại học Bách khoa - Đại học Đà Nẵng, Việt Nam¹

*Tác giả liên hệ / Corresponding author: ntpquyen@dut.udn.vn

(Nhận bài / Received: 06/11/2024; Sửa bài / Revised: 12/02/2025; Chấp nhận đăng / Accepted: 18/02/2025)

DOI: 10.31130/ud-jst.2025.470

Tóm tắt - Những tiến bộ trong kỹ thuật học máy (ML) đã mang đến những cơ hội mới hỗ trợ các nhà quản lý sản xuất ra quyết định thông qua việc phân tích hệ thống sản xuất. Tìm ra các điểm nghẽn trên dây chuyền sản xuất là vấn đề quan trọng đối với các nhà quản lý vì tắc nghẽn sẽ làm giảm năng suất của hệ thống, tăng chi phí sản xuất và giảm hiệu suất tổng thể của cả chuyền sản xuất. Nghiên cứu này áp dụng phương pháp phân cụm dữ liệu k -means, một trong những phương pháp phân cụm phổ biến của ML, để phát hiện những điểm tắc nghẽn trong sản xuất. Phương pháp bắt đầu bằng cách xác định thời gian trạng thái hoạt động tại mỗi máy/công đoạn trên dây chuyền sản xuất. Phân cụm k -means được sử dụng để phân loại các công đoạn thành các nhóm khác nhau, trong đó mỗi nhóm có đặc tính máy tương tự nhau. Từ đó, tiến hành phân tích dữ liệu được phân nhóm dựa trên chuỗi thời gian đại diện cho mỗi nhóm để phát hiện các điểm nghẽn trong sản xuất.

Từ khóa - Điểm tắc nghẽn; học máy; phân cụm k -means; hệ thống sản xuất

1. Giới thiệu

Ngày nay, những tiến bộ của học máy (ML) đã được ứng dụng rộng rãi và mang lại sự phát triển đáng kể trong các lĩnh vực như tài chính, thương mại điện tử, năng lượng, địa chất, không gian và sinh học v.v [1, 2]. ML được sử dụng trong nhiều lĩnh vực khác nhau, từ việc tự động hóa các tác vụ thông thường đến cung cấp những thông tin chi tiết thông minh và hỗ trợ trong quá trình ra quyết định của con người. Do đó, việc ứng dụng những tiến bộ của ML trong việc tối ưu hóa hoạt động của dây chuyền sản xuất cũng trở nên phổ biến. Khi các dây chuyền sản xuất trở nên phức tạp hơn và khó vận hành thủ công hơn, những tiến bộ trong lĩnh vực ML có thể được khai thác để hỗ trợ tốt hơn cho việc ra quyết định của các chuyên gia trong lĩnh vực sản xuất [3-5].

Một trong những vấn đề quan trọng hiện nay của các nhà quản trị sản xuất là làm cách nào để tăng hiệu suất tổng thể của toàn hệ thống. Các nghiên cứu đã chỉ ra rằng, việc giảm thiểu tổn thất do dây chuyền sản xuất hoạt động kém hiệu quả gây ra là một phương pháp tiềm năng được áp dụng để tăng hiệu suất. Thông thường, có đến 20–30% tổn thất đầu ra là do tình trạng tắc nghẽn dây chuyền sản xuất gây ra [6]. Tắc nghẽn xảy ra khi một máy hoặc một nhóm máy cụ thể mất nhiều thời gian hơn để gia công các công việc so với các máy hoặc trạm khác trong nhà máy, từ đó dẫn đến hiệu suất đầu ra của cả dây chuyền sản xuất giảm

Abstract - Advances in machine learning (ML) techniques have provided new opportunities for manufacturing managers to support their decision-making through the analysis of manufacturing systems. Detecting bottlenecks on the production line is an important issue for manufacturing managers because bottlenecks reduce the productivity of the system. In addition, detecting bottlenecks in the manufacturing system also reduces costs and improves the overall efficiency of the entire production line. This study applies k -means clustering, one of the popular clustering methods of ML, to detect bottlenecks on production lines. The proposed method begins by identifying the operating state time at each machine/process on the production line. k -means clustering is used to classify the processes into different groups in which each group has similar machine characteristics. The analysis of clustered data based on its representative time series is conducted to detect bottlenecks.

Key words - Bottlenecks; machine learning; k -means clustering; manufacturing system.

xuống. Vì vậy việc xác định các điểm tắc nghẽn là hết sức cần thiết để nâng cao hiệu suất chuyền sản xuất, từ đó duy trì và nâng cao mức hiệu suất toàn hệ thống [7].

Một số phương pháp được đề xuất để xử lý điểm tắc nghẽn trong sản xuất dựa trên dữ liệu máy bao gồm các phương pháp dựa trên thống kê truyền thống và các phương pháp dựa trên ML [8, 9]. Các phương pháp tiếp cận dựa trên thống kê có thể được áp dụng tốt nhất trong các môi trường nơi có các bộ mô tả thống kê dữ liệu máy đã được biết trước. Bên cạnh đó, các phương pháp này còn dựa vào nhiều giả định cơ bản khác nhau về dữ liệu của máy để tính toán các mô tả thống kê như phân phối, độc lập thống kê và tính ổn định của dữ liệu. Do đó, vấn đề quan trọng nhất của việc sử dụng các phương pháp tiếp cận dựa trên thống kê là dữ liệu máy không được phép vi phạm các giả định thống kê. Trong khi đó, hệ thống sản xuất trên thực tế là một hệ thống động với dữ liệu máy thay đổi theo thời gian có thể làm thay đổi phân phối dữ liệu trong một khoảng thời gian nhất định. Trong trường hợp này, nhà quản lý cần xử lý thống kê đối với dữ liệu máy trước khi áp dụng phương pháp thống kê phù hợp để xác định điểm tắc nghẽn. Do đó, việc sử dụng các phương pháp thống kê trong các trường hợp này sẽ gặp khó khăn trong việc duy trì kết quả với độ chính xác và độ tin cậy cao.

Trong khi đó, ML đã có những bước tiến đáng kể trong việc cung cấp khả năng tự động hóa, thông tin chi tiết và

¹ The University of Danang - University of Science and Technology, Vietnam (Nguyen Thi Phuong Quyên, Le Vy)

hỗ trợ ra quyết định. Các phương pháp tiếp cận dựa trên ML để phát hiện điểm tắt nghẽn trong sản xuất trở nên hiệu quả hơn mà không cần dựa vào các giả định thống kê nghiêm ngặt. Ngoài ra, các phương pháp này còn có thể áp dụng hiệu quả cho dữ liệu máy thay đổi theo thời gian, mang lại kết quả nhanh chóng và đáng tin cậy hơn [10].

Một trong các phương pháp phổ biến nhất được sử dụng để phát hiện điểm tắt nghẽn là phân cụm dữ liệu. Phân cụm dữ liệu là quy trình phân chia bộ dữ liệu cho sẵn thành nhiều nhóm khác biệt dựa trên sự tương đồng giữa các điểm dữ liệu đã cho. Trong đó, các điểm dữ liệu trong cùng một cụm tương tự nhau theo một thang đo khoảng cách cụ thể [11]. Nghiên cứu này sử dụng phương pháp phân cụm dữ liệu để xác định các điểm tắt nghẽn và nâng cao năng suất của dây chuyền sản xuất. Đặc biệt, phương pháp phân cụm k -means [12], được biết đến như là kỹ thuật phân cụm được sử dụng rộng rãi nhất trong công nghiệp do tính đơn giản và dễ hiểu của nó, được áp dụng để phát hiện các khu vực tắt nghẽn trên chuyền sản xuất [13]. k -means sẽ phân loại các máy trong dây chuyền sản xuất thành nhiều nhóm dựa trên thời gian trạng thái hoạt động của chúng. Nhóm máy được phân cụm theo phương pháp k -means được phát hiện là các khu vực tắt nghẽn nếu chúng có thời gian trạng thái hoạt động dài nhất.

2. Cơ sở lý thuyết

2.1. Tổng quan về các phương pháp phát hiện tắt nghẽn

Nhiều phương pháp tiếp cận đã được đề xuất để xác định các điểm nghẽn trong hệ thống sản xuất, chẳng hạn như phương pháp chu trình vận hành của Roser và cộng sự [8], phương pháp điểm xoay của Li và các cộng sự [14], phương pháp biến đổi thời gian của Betterton và Silver [15] và phương pháp hiệu suất thiết bị tổng thể của Tang [16]. Các phương pháp này sử dụng các kỹ thuật thống kê truyền thống để phát hiện tắc nghẽn, bao gồm thống kê mô tả (như giá trị trung bình, độ lệch chuẩn, hệ số biến thiên) và các thống kê suy diễn (như kiểm định giả thuyết). Các phương pháp thống kê suy diễn rất quan trọng trong việc xử lý tính biến thiên của dữ liệu máy trong quá trình phát hiện điểm tắt nghẽn. Tuy nhiên, quá trình triển khai các mô hình thống kê suy diễn tương đối phức tạp với nhiều quy trình giám sát thủ công khác nhau như: 1) xác định các mô tả thống kê để phân tích cấu trúc dữ liệu máy, 2) xác định kiểm định thống kê phù hợp và 3) diễn giải kết quả. Trong số các bước này, việc xác định mô tả thống kê cho cấu trúc dữ liệu thu thập được trên các máy trong chuyền sản xuất là bước quan trọng nhất làm nền tảng cơ sở cho việc lựa chọn kiểm định thống kê phù hợp trong quá trình phân tích thống kê suy diễn. Việc lựa chọn phương pháp kiểm định thống kê không chính xác có thể tạo ra kết quả phát hiện điểm tắt nghẽn không hiệu quả. Trong môi trường sản xuất thực tế, sự thay đổi hành vi của mỗi máy trên dây chuyền sản xuất hoàn toàn khác nhau đã đặt ra những thách thức trong việc áp dụng các giả định tổng quát trên tất cả các máy. Li và các cộng sự [14] nhấn mạnh rằng, các giả định (đặc biệt là giả định về phân phối thống kê) khó có thể kiểm soát được do sự thay đổi của dữ liệu máy.

Roser và các cộng sự [8] đã đề xuất một phương pháp phát hiện điểm tắt nghẽn hiệu quả dựa trên giai đoạn hoạt động của trạng thái máy. Trạng thái máy biểu thị một nhiệm vụ mà máy

thực hiện. Phương pháp này chia trạng thái máy trong quá trình sản xuất thành hai loại riêng biệt: trạng thái hoạt động và trạng thái không hoạt động. Trạng thái hoạt động xảy ra khi một thao tác được thực hiện trên máy dẫn đến tình trạng tắc nghẽn hoặc thiếu hụt ở các máy khác. Ví dụ, trạng thái hoạt động có thể bao gồm: 1) trạng thái sản xuất được xác định khi máy móc đang sản xuất hàng hóa, 2) trạng thái tạm dừng máy khi máy bị hỏng và đang được sửa chữa, và 3) trạng thái thiết lập, trong đó các tác vụ thiết lập các thông số khởi tạo đang được thực hiện trên máy. Ngược lại, trạng thái không hoạt động gây ra bởi các tác vụ chờ đợi trong suốt quy trình sản xuất như chờ dỡ sản phẩm khỏi máy hoặc chờ các chi tiết được gia công từ các máy khác đến. Toàn bộ thời gian trạng thái hoạt động của từng máy trên dây chuyền sản xuất được tổng hợp và sau đó so sánh với các máy khác để xác định máy xảy ra hiện tượng tắt nghẽn là máy có thời gian trạng thái hoạt động dài nhất. Subramaniyan và các cộng sự [10] đã tích hợp phương pháp định hướng dữ liệu (data-driven method) với phương pháp giai đoạn hoạt động để xác định các điểm tắt nghẽn và cải thiện hiệu suất tổng thể của hệ thống sản xuất. Trong đó, phương pháp phân cụm thứ bậc được sử dụng để phân nhóm các máy và xác định máy tắt nghẽn dựa trên thời gian trạng thái hoạt động của máy cao nhất.

2.2. Lý thuyết tổng quan về phân cụm k -means

k -means là một trong những phương pháp phân cụm dữ liệu được sử dụng phổ biến nhất trong các ứng dụng công nghiệp. Ý tưởng chính của k -means là phân loại một tập dữ liệu cho trước thành k cụm riêng biệt, trong đó k là số được xác định trước, nhằm cực tiểu tổng bình phương khoảng cách giữa trung tâm cụm đến các điểm dữ liệu trong cụm đó [12]. Kỹ thuật phân cụm k -means bao gồm các bước sau:

- **Bước 1:** Chọn ngẫu nhiên k tâm cụm cho k nhóm. Mỗi nhóm được biểu diễn bằng tâm của nó.
- **Bước 2:** Tính khoảng cách giữa mỗi điểm dữ liệu đến k điểm tâm cụm.
- **Bước 3:** Gán mỗi đối tượng vào một cụm có điểm tâm cụm gần nhất với nó.
- **Bước 4:** Cập nhật tâm cụm.
- **Bước 5:** Dừng nếu không có thay đổi nào trong quá trình gán các đối tượng vào cụm hoặc tâm cụm giữ nguyên, nếu không thì lặp lại từ bước 2.

3. Áp dụng phương pháp phân cụm k -means để phát hiện tắc nghẽn

Phương pháp k -means dựa trên giai đoạn trạng thái hoạt động máy được đề xuất để xác định điểm tắt nghẽn trên dây chuyền sản xuất. Hình 1 thể hiện sơ đồ mô tả phương pháp đề xuất một cách trực quan. Phương pháp này gồm bốn bước chính được miêu tả cụ thể như sau:

Bước 1: Thu thập và xử lý dữ liệu.

Nghiên cứu này sử dụng dữ liệu nhật ký sự kiện máy thu thập từ các hệ thống sản xuất thực tế. Tất cả các sự kiện xảy ra trên máy đều được ghi lại bằng kỹ thuật số với các mốc thời gian tương ứng. Dựa trên các mốc thời gian của sự kiện, thời lượng của từng sự kiện được trích xuất. Một ví dụ về dữ liệu thời gian của sự kiện theo mốc thời gian trên dây chuyền sản xuất hàng may mặc, cụ thể là để may áo sơ mi, được thể hiện trong Bảng 1.

Bảng 1. Dữ liệu sự kiện máy theo mốc thời gian

Thứ tự.	Sự kiện	Mốc thời gian	Thời lượng (giây)
1	Cắt thân trước trái	15-02-2024 8:00:01	137
2	Cắt thân trước phải	15-02-2024 8:02:18	135
3	Cắt thân sau	15-02-2024 8:04:33	99
4	Cắt tay áo	15-02-2024 8:06:12	101
...
25	May viền cổ áo bằng dây chun	15-02-2024 8:21:46	145
26	May vai vào thân áo	15-02-2024 8:24:14	148
27	Gấp và may gấu áo	15-02-2024 8:27:39	205
...

Cần phải xem xét khối lượng dữ liệu quá khứ cần thu thập được sử dụng để phát hiện điểm tắc nghẽn trong bước này. Ví dụ, cần thu thập dữ liệu bao nhiêu lượt/ lô hàng sản xuất trước đây? Nếu quy mô dữ liệu sản xuất cần thu thập trong quá khứ quá nhỏ, bộ dữ liệu này không thể đại diện cho hệ thống sản xuất. Ngược lại, việc tính toán sẽ quá phức tạp nếu dữ liệu quá lớn. Do đó, việc lựa chọn kích cỡ bộ dữ liệu quá khứ phải được thực hiện cẩn thận bởi một chuyên gia có hiểu biết sâu sắc về dây chuyền sản xuất.

Ngoài ra, bước này cũng xử lý trước dữ liệu thu thập được trên từng máy trong dây chuyền sản xuất. Dữ liệu nhật ký sự kiện cần được làm sạch trước khi xử lý tiếp. Các bước làm sạch phổ biến bao gồm: xóa các sự kiện ngoài khoảng thời gian đã xác định trong các lần chạy sản xuất khác nhau; kiểm tra nhật ký sự kiện và loại bỏ các dữ liệu nhiễu. Trong các hệ thống sản xuất được minh họa ở mục 4, quá trình làm sạch dữ liệu được thực hiện cho từng máy bằng cách kiểm tra và loại bỏ các sự kiện dư thừa và những sự kiện không được các nhà quản trị quan tâm.

Việc tạo dữ liệu chuỗi thời gian sau khi dữ liệu được làm sạch từng máy cũng được thực hiện trước khi phân cụm bộ dữ liệu. Dựa vào các mốc thời gian của sự kiện máy để xác định thời lượng thực hiện từng công đoạn. Thời lượng này được tính là thời gian trôi qua giữa thời điểm bắt đầu và kết thúc của trạng thái hoạt động của công đoạn trong quá trình sản xuất. Kết quả thời lượng hoạt động được thể hiện trong ma trận T, trong đó mỗi hàng đại diện cho quá trình sản xuất và mỗi cột đại diện cho máy. Tất cả các giá trị trong một cột là chuỗi thời gian cho máy cụ thể đó.

Bước 2: Áp dụng phương pháp k -means để phân cụm dữ liệu.

Để áp dụng phương pháp k -means, cần tính toán lựa chọn số cụm k . Việc chọn số lượng cụm k rất quan trọng trong quá trình phân cụm. Giá trị k quá lớn sẽ gây phức tạp trong quá trình phân cụm do có quá nhiều cụm (không hữu ích trong thực tế). Ngược lại, có quá ít cụm sẽ làm giảm độ phân giải khi cố gắng tìm các điểm gây ra tắc nghẽn. Phương pháp khuỷu tay (Elbow) là một phương pháp phổ biến được sử dụng để lựa chọn số lượng cụm hợp lý. Phương pháp này dựa trên biểu đồ trực quan hóa với trục tung biểu thị tổng biến thiên bình phương khoảng cách (WSS) và trục hoành biểu thị số lượng cụm cần phân chia trong thuật toán k -means. Mục tiêu của phân cụm k -means nhằm cực tiểu giá trị WSS. Do đó, số cụm k được chọn tại điểm khuỷu tay là điểm mà ở đó tốc độ suy giảm của WSS

sẽ thay đổi nhiều nhất. Sau vị trí khuỷu tay thì việc gia tăng thêm số lượng cụm cũng không làm WSS giảm đáng kể.

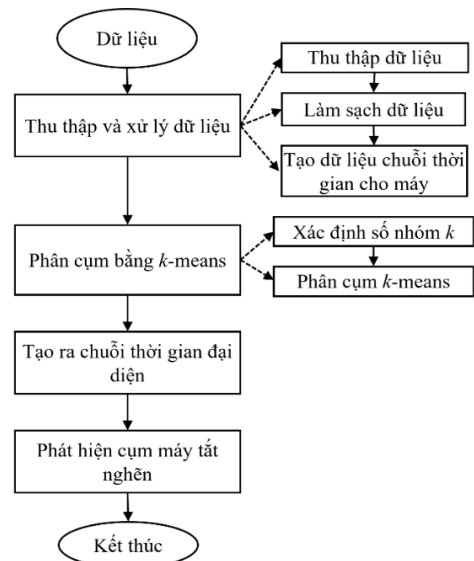
Sau khi chọn số lượng cụm k , thuật toán k -means để phân loại dữ liệu các máy đã được trích xuất và xử lý trước đó vào các cụm khác nhau. Mỗi cụm sẽ có những đặc trưng riêng về thời gian trạng thái hoạt động máy.

Bước 3: Tạo chuỗi thời gian đại diện

Thông tin về số lượng cụm và việc phân từng máy cho các cụm đó đã thực hiện ở bước trước. Ở bước này, các cụm cần được phân tích chi tiết để có thể phát hiện điểm tắc nghẽn xảy ra trên máy hoặc cụm máy nào. Với mục đích này, một chuỗi thời gian đại diện có thể được tạo cho mỗi cụm. Chuỗi thời gian đại diện là một bản tóm tắt của chuỗi thời gian riêng lẻ trong một cụm, được xây dựng bằng cách tính trung bình của mỗi điểm dữ liệu cho các chuỗi thời gian riêng lẻ khác nhau trong cụm đó được trích từ ma trận T.

Bước 4: Phát hiện điểm tắc nghẽn

Các điểm tắc nghẽn trên dây chuyền sản xuất được xác định bằng cách phân tích trực quan chuỗi thời gian đại diện của mỗi cụm được tạo ra bởi kỹ thuật phân cụm. Mục đích của việc kiểm tra trực quan là xác định xem chuỗi thời gian đại diện cho mỗi cụm có được tách biệt tốt với nhau về tổng thể hay không. Nếu chuỗi thời gian đại diện được phân tách tốt, các nhà quản lý có thể tiến hành phân tích chuỗi thời gian cho từng cụm và phát hiện cụm chứa điểm tắc nghẽn. Tuy nhiên, nếu chuỗi thời gian đại diện các cụm không được phân tách rõ ràng, cần phải đánh giá lại số lượng cụm và lặp lại các tính toán của các bước trước đó.

**Hình 1.** Sơ đồ mô tả phương pháp đề xuất

4. Phân tích kết quả

Ứng dụng phân cụm trên một hệ thống sản xuất trong thực tế để phát hiện điểm tắc nghẽn bằng cách sử dụng phương pháp phát hiện tắc nghẽn dựa trên thời gian trạng thái hoạt động máy. Trình tự thực hiện như các bước đã nêu ở mục 3. Bộ dữ liệu được sử dụng để đánh giá phương pháp đề xuất được thu thập trong một công ty may mặc tại Việt Nam với nhiều loại sản phẩm khác nhau. Trong đó, bộ dữ liệu kích cỡ nhỏ cho sản phẩm đơn giản dưới 20 công đoạn máy gia công, dữ liệu cỡ trung vừa cho sản phẩm tương đối

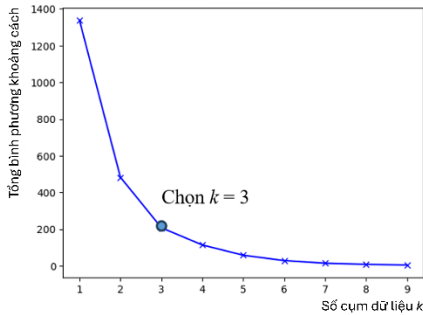
nhieu chi tiet tu 20 den 50 cong doan may va bo du lieu co lon len den 100 cong doan may. Cac bo du lieu da duoc xu ly lam sach theo nhu buoc 1 cua phuong phap de xuat.

4.1. Kết quả thực nghiệm trên bộ dữ liệu cỡ nhỏ

Bộ dữ liệu này được thu thập từ một hệ thống may mặc nối tiếp gồm 14 công đoạn được thực hiện trên 14 máy khác nhau để tạo ra sản phẩm. Thời gian trạng thái hoạt động ở từng máy được thu thập. Sau khi đã xử lý dữ liệu gồm thời gian sản xuất trong 10 ngày, các số liệu có được trong ma trận T_{10x14} như sau:

$$T_{10 \times 14} = \begin{bmatrix} 8,99 & 1,22 & \dots & 6,59 & 17,01 & 14,71 \\ 8,87 & 1,11 & \dots & 7,50 & 18,49 & 15,73 \\ 8,56 & 1,13 & \dots & 8,11 & 17,90 & 14,08 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 8,69 & 1,54 & \dots & 8,12 & 18,15 & 15,50 \\ 8,67 & 1,62 & \dots & 7,70 & 18,03 & 15,23 \end{bmatrix}$$

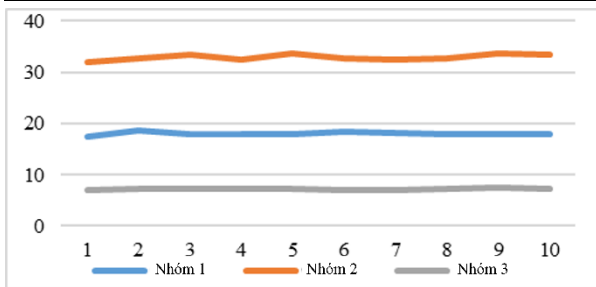
Tiếp theo, tạo đồ thị Elbow để tìm ra số lượng cụm thích hợp để. Hình 2 thể hiện biểu đồ Elbow cho bộ dữ liệu nhỏ. Kết quả cho thấy rằng tại số đồ thị có sự thay đổi độ dốc đáng kể trước cụm thứ 3. Vậy số cụm k= 3 cụm được lựa chọn để đưa vào thuật toán phân cụm k-means. Kết quả phân cụm được thể hiện trong như Bảng 2.



Hình 2. Minh họa phương pháp Elbow với k=3

Bảng 2. Các công đoạn theo từng cụm với dữ liệu nhỏ

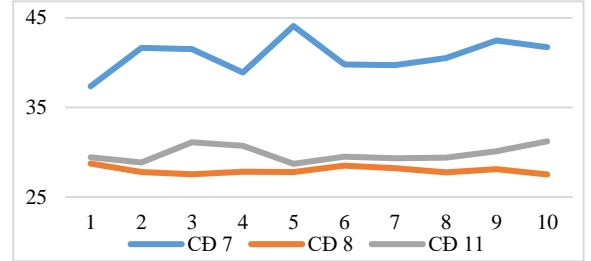
Cụm	Công đoạn
1	3, 4, 5, 9, 10, 13, 14
2	7, 8, 11
3	1, 2, 6, 12



Hình 3. Chuỗi thời gian đại diện cho các cụm với bộ dữ liệu nhỏ

Sau khi nhóm các công đoạn vào các cụm, tiến hành xây dựng chuỗi thời gian đại diện cho các cụm và phân tích biểu đồ thể hiện chuỗi thời gian đó. Hình 3 cho thấy, chuỗi thời gian đại diện các cụm tách biệt nhau. Do đó, có thể sử dụng kết quả phân cụm này để phân tích điểm tắt nghẽn. Cụm 2 có thời gian cao nhất trong tất cả các ngày quan sát, nên cụm 2 có thể là cụm chứa điểm tắt nghẽn của hệ thống. Hình 4 phân tích thời gian các công đoạn 7, 8 và 11 trong cụm 2. Có thể thấy, công đoạn 7 có thời gian cao nhất trong tất cả các

ngày quan sát, nên công đoạn 7 chính là điểm tắt nghẽn của hệ thống này. Mặt khác, từ Hình 3 có thể thấy, cụm 2 có thời gian rất cao so với các cụm 1, 3 nên là các công đoạn 8,11 cũng cần được xem xét nhiều hơn.



Hình 4. Thời gian các máy trong cụm 2 chứa điểm tắt nghẽn

4.2. Kết quả thực nghiệm trên bộ dữ liệu cỡ vừa

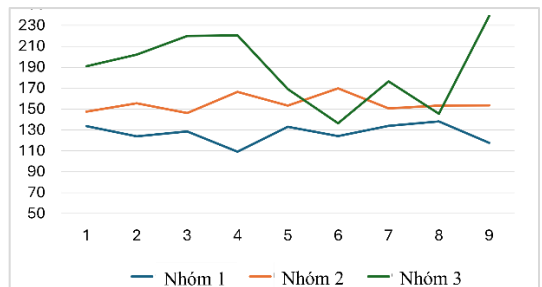
Dữ liệu này cũng được thu thập tại công ty may mặc gồm thời gian chạy 50 máy trong 9 ngày, sau khi xử lý các số liệu được ma trận T_{9x50} như sau:

$$T_{9 \times 50} = \begin{bmatrix} 166 & 191 & \dots & 259 & 73 & 198 \\ 27 & 169 & \dots & 198 & 159 & 49 \\ 259 & 219 & \dots & 166 & 169 & 52 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 166 & 182 & \dots & 222 & 251 & 210 \\ 73 & 184 & \dots & 219 & 86 & 198 \end{bmatrix}$$

Quy trình được thực hiện tương tự như trên bộ dữ liệu cỡ nhỏ. Phương pháp khuỷu tay được áp dụng với số lượng cụm được xác định k=3. Dữ liệu này được sử dụng vào phương pháp phân cụm k-means. Kết quả phân cụm được trình bày trong Bảng 3.

Bảng 3. Các công đoạn theo từng cụm với dữ liệu vừa

Cụm	Công đoạn
1	3, 5, 8, 9, 15, 18, 20, 21, 22, 24, 25, 26, 28, 29, 32, 39, 42, 43, 44, 45, 46, 48, 49
2	1, 2, 4, 6, 7, 10, 11, 12, 13, 14, 19, 27, 30, 31, 33, 34, 35, 36, 37, 40, 41, 50
3	16, 17, 23, 38, 47

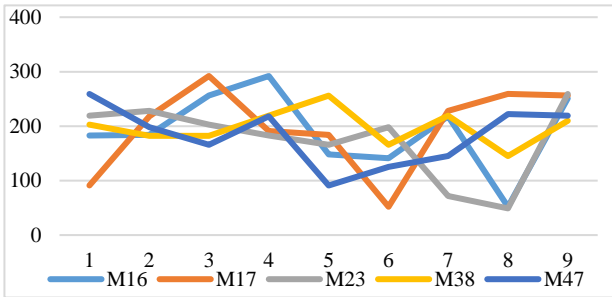


Hình 5. Chuỗi thời gian đại diện cho các cụm với bộ dữ liệu vừa

Hình 5 cho thấy, chuỗi thời gian đại diện cho từng cụm tương đối khác biệt rõ ràng với nhau, chỉ duy nhất ngày thứ 6 và ngày thứ 8, chuỗi thời gian đại diện của cụm 3 tương đối thấp hơn cụm 2. Tuy nhiên, sự giao thoa dữ liệu giữa 3 cụm được phân nhóm này không đáng kể chứng tỏ kết quả phân cụm của k-means đáng tin cậy. Bên cạnh đó, thời gian trạng thái hoạt động của cụm thứ 3 cao hơn các cụm khác trong hầu hết các ngày sản xuất. Dựa vào phương pháp thời gian trạng thái hoạt động, thời gian hoạt động cao nhất là điểm tắt nghẽn. Do đó, các máy trong cụm 3: M16, M17, M23, M38, M47 có thể là các điểm tắt nghẽn chính trong dây chuyền sản xuất. Các máy trong cụm 3 thể hiện sự hạn

chế hiệu suất trong hầu hết các hoạt động sản xuất. Hình 5 cũng cho thấy, cụm 2 có thời lượng hoạt động cao nhất đối với các ngày 6 và 8. Nói cách khác, đối với các hoạt động sản xuất này, điểm tắc nghẽn chính thay đổi giữa các máy trong cụm 3 và các máy cụm 2. Cần phải phân tích sâu hơn, dựa trên các điều kiện, dữ liệu khác trong các lần chạy sản xuất để xác định nguyên nhân dẫn đến tắc nghẽn dịch chuyển. Ngoài ra, từ Hình 5 có thể thấy, cụm 1 có thời gian hoạt động thấp hơn cụm 2 và 3 trong tất cả các lần chạy sản xuất cho nên được đánh giá là không có điểm tắc nghẽn.

Ở trên đã chỉ ra rằng, cụm 3 là điểm tắc nghẽn chính của dây chuyền. Tiếp theo sẽ phân tích về các máy trong cụm 3 để đánh giá sâu hơn. Từ ma trận T trích ra thời gian hoạt động trong 9 ngày của các máy trong cụm 3: M16, M17, M23, M38, M47. Hình 6 có thể thấy, các máy trong cụm 3 thay phiên nhau có thời gian hoạt động cao nhất trong các ngày chạy sản xuất hay là các điểm tắc nghẽn rất dễ bị dịch chuyển giữa các máy trong cụm 3. Do đó, có thể kết luận rằng M16, M17, M23, M38, M47 tạo thành một nhóm các tắc nghẽn chính tiềm ẩn trong hệ thống sản xuất. Những máy này có thể là trọng tâm của các hành động cải tiến nhằm tăng năng xuất đầu ra.



Hình 6. Thời gian các máy trong cụm 3 chứa điểm tắc nghẽn

4.3. Kết quả thực nghiệm trên bộ dữ liệu cỡ lớn

Bộ dữ liệu kích cỡ lớn cho sản phẩm may mặc phức tạp nhiều công đoạn trên 100 máy. Dữ liệu thu thập trong 15 ngày sản xuất. Sau khi xử lý các số liệu được ma trận $T_{15 \times 100}$ như sau:

$$T_{15 \times 100} = \begin{bmatrix} 58 & 251 & 79 & \dots & 386 & 295 \\ 129 & 282 & 79 & \dots & 377 & 300 \\ 85 & 221 & 49 & \dots & 411 & 359 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 112 & 233 & 58 & \dots & 365 & 353 \\ 51 & 256 & 122 & \dots & 415 & 310 \end{bmatrix}$$

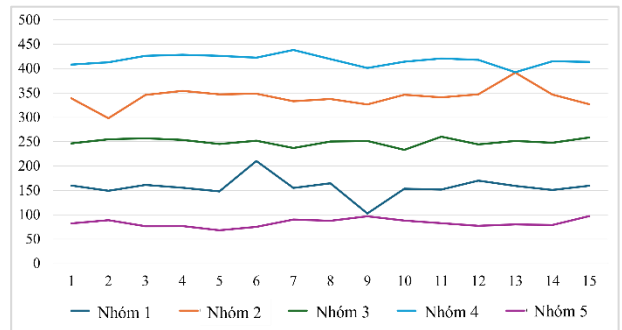
Phương pháp khuỷu tay được sử dụng để xác định số cụm $k=5$. Do đó, số lượng 100 máy sản xuất sẽ được phân thành 5 nhóm bằng kỹ thuật phân cụm k -means. Kết quả phân cụm được trình bày trong Bảng 4.

Bảng 4. Kết quả phân cụm với dữ liệu lớn

Cụm	Công đoạn
1	5, 6, 13, 14, 16, 17, 23, 25, 26, 37, 43, 47, 49, 64, 71, 78, 91, 94, 97
2	7, 18, 19, 28, 34, 42, 48, 53, 56, 59, 65, 68, 80, 81, 86, 100
3	2, 4, 10, 15, 20, 22, 27, 31, 32, 35, 46, 52, 54, 58, 62, 70, 73, 79, 85, 87, 90
4	11, 21, 29, 36, 39, 40, 41, 50, 51, 55, 63, 66, 72, 74, 76, 77, 82, 83, 89, 93, 95, 99
5	1, 3, 8, 9, 12, 24, 30, 33, 33, 38, 44, 45, 57, 60, 61, 67, 69, 75, 84, 88, 92, 96, 98

Từ ma trận $T_{15 \times 100}$, dữ liệu chuỗi thời gian đại diện cho từng cụm được xác định. Hình 7 thể hiện dữ liệu đại diện một cách trực quan hóa tương ứng với 5 cụm máy được xác định bằng phương pháp k -means. Rất rõ ràng để nhận thấy rằng các chuỗi thời gian đại diện này phân tách riêng biệt, không chồng chéo lên nhau nên kết quả này có thể dùng cho các phân tích tiếp theo. Cụm 4 có chuỗi thời gian đại diện cao nhất trong tất cả các ngày. Do đó, các máy trong cụm 4 bao gồm M11, M21, M29, M36, M39, M40, M41, M50, M51, M55, M63, M66, M72, M74, M76, M77, M82, M83, M89, M93, M95 và M99 được nghi ngờ gây ra tắc nghẽn trên dây chuyền sản xuất và giảm hiệu suất của toàn hệ thống.

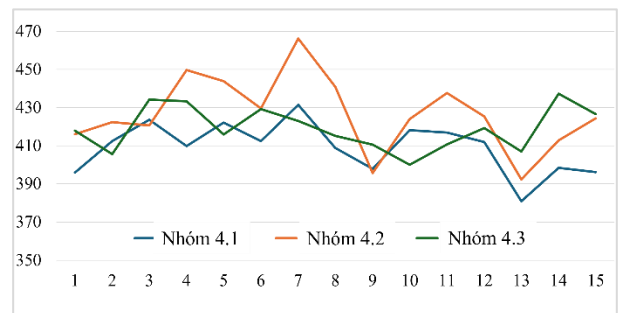
Phân tích sâu hơn các máy trong cụm 4 nhằm xác định chính xác địa điểm gây ra tắc nghẽn hệ thống. Tuy nhiên, do số lượng máy trong cụm 4 quá nhiều (22 máy), kỹ thuật k -means một lần nữa được áp dụng cho riêng cụm 4 để chia nhỏ và phân tích sâu từng máy trong cụm này. Tương tự, phương pháp khuỷu tay được sử dụng và $k=3$ được xác định để chia cụm 4 ra thành 3 cụm nhỏ hơn nữa. Kết quả phân cụm bằng k -means của cụm 4 được thể hiện trong Bảng 5.



Hình 7. Chuỗi thời gian đại diện cho các cụm với bộ dữ liệu lớn

Bảng 5. Kết quả phân cụm với cho dữ liệu của cụm 4

Cụm	Công đoạn
4.1	11, 21, 29, 39, 63, 72, 74, 76, 99
4.2	41, 50, 51, 66, 77, 95
4.3	36, 40, 55, 82, 83, 89, 93

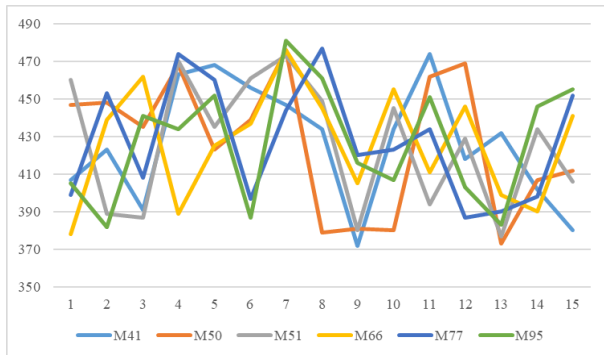


Hình 8. Biểu đồ chuỗi thời gian các cụm nhỏ trong cụm 4

Tương tự, chuỗi dữ liệu thời gian đại diện cho từng cụm con 4.1, 4.2 và 4.3 cũng được xác định và thể hiện trong Hình 8. Trong đó, thời gian trạng thái hoạt động của cụm 4.2 cao hơn hẳn các cụm con 4.1 và 4.3 trong tất cả các ngày sản xuất. Do đó, các máy trong cụm 4.2 được xác định là nguyên nhân gây ra tắc nghẽn trên hệ thống. Cũng có thể thấy từ Hình 8, đối với các ngày 3, 9, 13, 14, cụm 4.3 có thời lượng hoạt động cao nhất. Nói cách khác, đối với các hoạt động sản xuất này, điểm tắc nghẽn chính thay đổi giữa

các máy trong cụm 4.2 và các máy cụm 4.3. Cần phải phân tích sâu hơn, dựa trên các điều kiện, dữ liệu khác trong các lần chạy sản xuất để xác định nguyên nhân dẫn đến tắc nghẽn dịch chuyển.

Hình 9 cho thấy, các máy trong cụm 4.2 thay phiên nhau có thời gian hoạt động cao nhất trong các ngày chạy sản xuất hay là các điểm tắc nghẽn rất dễ bị dịch chuyển giữa các máy trong cụm 4.2. Do đó, có thể kết luận rằng, M41, M50, M51, M66, M77, M95 tạo thành một nhóm các tắc nghẽn chính tiềm ẩn trong hệ thống sản xuất. Những máy này có thể là trọng tâm của các hành động cải tiến nhằm tăng năng suất đầu ra.



Hình 9. Biểu đồ chuỗi thời gian các máy trong cụm 4.2

4.4. Thảo luận

Một trong những thách thức mà các chuyên gia quản trị sản xuất gặp phải là xác định tập hợp các cụm máy tắc nghẽn năng suất trên dây chuyền sản xuất. Nghiên cứu này đề xuất phương pháp phân cụm dựa trên k -means để phát hiện máy hoặc cụm máy gây ra tắc nghẽn năng suất hệ thống và đã áp dụng thành công trên ba bộ dữ liệu sản xuất thực tế với các kích cỡ dữ liệu đặc trưng cho quy mô hệ thống sản xuất lớn, trung bình và nhỏ. Phương pháp phát hiện tắc nghẽn dựa trên k -means giúp các chuyên gia có cái nhìn toàn cục về hệ thống sản xuất đặc biệt là mối tương quan giữa các máy được phân loại vào trong cùng một cụm. Phương pháp đề xuất đóng góp tiềm năng trong xu hướng tiếp cận các kỹ thuật ML không giám sát trong thực tế sản xuất công nghiệp. Bên cạnh đó, Subramaniyan và các cộng sự [10] đã chỉ ra rằng, việc áp dụng các phương pháp ML không giám sát hỗ trợ xác định điểm tắc nghẽn năng suất hệ thống nhanh hơn và hiệu quả hơn so với phương pháp truyền thống thông qua quan sát tại chuyền sản xuất.

5. Kết luận

Phát hiện tắc nghẽn là cần thiết để cải thiện hiệu suất dây chuyền sản xuất và tăng năng suất. Khi các dây chuyền sản xuất ngày càng trở nên phức tạp hơn và khi dữ liệu máy quy mô lớn trở nên phổ biến thì việc sử dụng các công cụ hiệu quả để phát hiện tắc nghẽn có tầm quan trọng đặc biệt. Nghiên cứu này trình bày một cách tiếp cận dựa trên kỹ thuật phân cụm dữ liệu bằng k -means kết hợp với phương pháp thời gian trạng thái hoạt động máy để phát hiện tắc nghẽn. Nghiên cứu các đặc điểm của từng cụm giúp xác định các máy bị tắc nghẽn có thể xảy ra trong dây chuyền sản xuất. Các kết quả thực nghiệm trên các bộ dữ liệu kích cỡ đa dạng lên đến 100 công đoạn máy cho thấy, sự hiệu

quả cho phương pháp đề xuất xác định được các điểm tắc nghẽn gây giảm hiệu suất toàn hệ thống sản xuất.

Nghiên cứu này có thể được mở rộng trong tương lai bằng việc tối ưu hóa số cụm để quá trình phân tích cụm máy chính xác hơn. Bên cạnh đó, phương pháp phân cụm k -means còn tồn tại một số hạn chế như kết quả phân cụm bị ảnh hưởng bởi giá trị khởi tạo cụm ban đầu và tương đối nhạy cho bộ dữ liệu chứa các điểm nhiễu và ngoại lai. Các nghiên cứu trong tương lai có thể áp dụng các biện pháp khắc phục các khuyết điểm này của k -means trong ứng dụng thực tế. Hơn nữa, các kỹ thuật phân cụm khác phức tạp hơn cũng có thể được áp dụng để nâng độ chính xác của việc phân cụm nhóm máy.

TÀI LIỆU THAM KHẢO

- [1] S. M. Miller, "AI: Augmentation, more so than automation", *Asian Management Insight*, vol. 5, no. 1, pp. 1-20, 2018.
- [2] H. J. Wilson and P. R. Daugherty, "Collaborative intelligence: Humans and AI are joining forces", *Harvard Business Review*, vol. 96, no. 4, pp. 114-123, 2018.
- [3] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine learning in manufacturing: advantages, challenges, and applications", *Production & Manufacturing Research*, vol. 4, no. 1, pp. 23-45, 2016.
- [4] M. Sharp, R. Ak, and T. Hedberg Jr, "A survey of the advancing use and development of machine learning in smart manufacturing", *Journal of manufacturing systems*, vol. 48, pp. 170-179, 2018.
- [5] C. Cimini, F. Pirola, R. Pinto, and S. Cavalieri, "A human-in-the-loop manufacturing control architecture for the next generation of production systems", *Journal of manufacturing systems*, vol. 54, pp. 258-271, 2020.
- [6] P. Alavian, Y. Eun, S. M. Meerkov, and L. Zhang, "Smart production systems: automating decision-making in manufacturing environment", *International Journal of Production Research*, vol. 58, no. 3, pp. 828-845, 2020.
- [7] K. Wu, Y. Zhou, and N. Zhao, "Variability and the fundamental properties of production lines", *Computers & Industrial Engineering*, vol. 99, pp. 364-371, 2016.
- [8] C. Roser, M. Nakano, and M. Tanaka, "A practical bottleneck detection method", in *Proceeding of the 2001 winter simulation conference (Cat. No. 01CH37304)*, 2001, pp. 949-953.
- [9] C. Yu and A. Matta, "Data-driven bottleneck detection in manufacturing systems: A statistical approach", in *2014 IEEE international conference on automation science and engineering (CASE)*, 2014, pp. 710-715.
- [10] M. Subramaniyan, A. Skoogh, A. S. Muhammad, J. Bokrantz, B. Johansson, and C. Roser, "A generic hierarchical clustering approach for detecting bottlenecks in manufacturing", *Journal of Manufacturing Systems*, vol. 55, pp. 143-158, 2020.
- [11] A. K. Jain, "Data clustering: 50 years beyond K-means", *Pattern recognition letters*, vol. 31, no. 8, pp. 651-666, 2010.
- [12] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm", *IEEE access*, vol. 8, pp. 80716-80727, 2020.
- [13] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing", *Expert Systems with Applications*, vol. 166, p. 114060, 2021.
- [14] L. Li, Q. Chang, and J. Ni, "Data driven bottleneck detection of manufacturing systems", *International Journal of production research*, vol. 47, no. 18, pp. 5019-5036, 2009.
- [15] C. E. Betterton and S. Silver, "Detecting bottlenecks in serial production lines—a focus on interdeparture time variance", *International Journal of Production Research*, vol. 50, no. 15, pp. 4158-4174, 2012.
- [16] H. Tang, "A new method of bottleneck analysis for manufacturing systems", *Manufacturing letters*, vol. 19, pp. 21-24, 2019.