

PHÂN TÍCH VÀ ỨNG DỤNG PHƯƠNG PHÁP HỌC MÁY ĐỂ CHUẨN ĐOÁN LỖI CHO CHU TRÌNH SẢN XUẤT CÔNG NGHIỆP

ANALYZE AND APPLICATION MACHINE LEARNING TECHNIQUE FOR DIAGNOSIS INDUSTRIAL PRODUCTION PROCESS

Trần Ngọc Hoàng

Trường Đại học Sư phạm Kỹ thuật – Đại học Đà Nẵng, Việt Nam

Ngày toà soạn nhận bài 1/6/2020, ngày phản biện đánh giá 10/6/2020, ngày chấp nhận đăng 15/6/2020

TÓM TẮT

Bài viết này đề xuất một giao thức ứng dụng của quá trình chuẩn đoán trong một máy phản ứng của một quy trình phức tạp. Bằng cách sử dụng Kỹ thuật học tập Bayes, giao thức này được đào tạo bằng cách học cơ sở dữ liệu sản xuất lịch sử để chuẩn đoán nguyên nhân thất bại của lò phản ứng này trong quá trình sản xuất. Ứng dụng trong lĩnh vực tự động hóa, mô hình đề xuất được cấu trúc tự động từ trích xuất dữ liệu được thu thập trực tiếp từ các cảm biến. Dựa trên thuật toán Tối đa hóa kỳ vọng trong học máy, chúng tôi cho thấy kết quả của mô hình này là phân loại và cũng để xác định nguyên nhân gốc rễ của vấn đề trôi dạt trong một mô phỏng kịch bản cụ thể. Do đó, đóng góp chính của chúng tôi để hỗ trợ máy công cụ bảo trì để tăng kỹ thuật vòng đời.

Từ khóa: Điều khiển hệ thống sản xuất; Học mạng Bayes; Hệ thống sản xuất phức tạp; Chu trình chuẩn đoán; Bảo trì sửa chữa.

ABSTRACT

This paper proposes an application protocol of diagnosis process in a reactor machine of a complex process. By using Bayes Learning Technique, this protocol is trained by learning historical production database in order to diagnosis the failure cause of this reactor in production process. Application in automation field, the model propose is structured automatically from collected data extract directly from sensors. Based on Expectation Maximization algorithm in machine learning, we show that the result of this model is to classify and also to identify the root causes of drift problem in a specific scenario simulation. Therefore, our key contribution to support maintenance tool machine for increasing life-cycle engineering.

Keywords: Control production; Bayes learning; Complex system; Diagnostics process; Corrective Maintenance.

1. GIỚI THIỆU

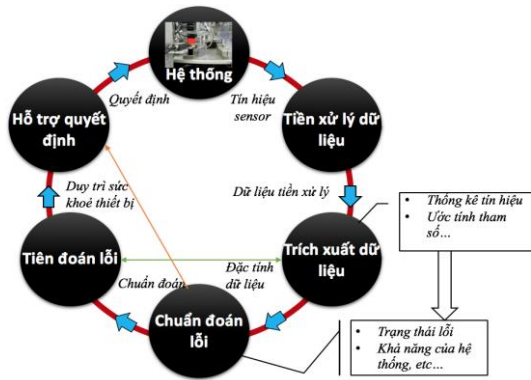
Ngày nay, quá trình công nghiệp hoá đẩy các doanh nghiệp vào một sự cạnh tranh khốc liệt. Trong bối cảnh đó, sự thành công phụ thuộc vào chất lượng sản xuất lẫn chất lượng sản phẩm. Do đó, các nhà sản xuất phải tối đa hoá hiệu suất và chất lượng sản xuất thông qua ba tiêu chí: vòng đời sản phẩm, giá và hiệu quả [1]. Để đạt được những điều đó, họ thường tập trung vào việc tăng khả năng sẵn sàng của thiết bị sản xuất

thông qua một kế hoạch bảo trì bảo dưỡng chặt chẽ và hiệu quả.

Tuy nhiên, thực tế là không phải tất cả các khâu trên thiết bị sản xuất đều được trang bị cảm biến để đo và phản hồi kết quả vì lý do kỹ thuật lẫn lý do đầu tư sản xuất. Do vậy, quá trình sản xuất luôn luôn tồn tại rất nhiều rủi ro của việc không thể quan sát được các nhiễu loạn hoặc trôi dạt mà ở đó máy móc không hoạt động đúng như chúng ta yêu cầu, tạo nên những ảnh hưởng tới chất lượng sản phẩm. Trong khi đó thiết bị sản xuất gần như

không có một cơ chế nội tại nào để xác định vấn đề đến từ đâu nếu sản phẩm đưa ra không chất lượng. Do đó, trên thực tế rất nhiều sự trôi dạt hoạt động (sự sai lệch hoạt động khiến cho thiết bị sản xuất không thể đem lại được kết quả chất lượng sản phẩm như dự kiến) không thể tránh khỏi trong quá trình sản xuất.

Trong bối cảnh đó, vấn đề này lại trở nên nghiêm trọng hơn khi mà thời đại công nghiệp 4.0 tạo nên một khối lượng đồ sộ dữ liệu sản xuất (con người, cảm biến, công thức, vật liệu...). Quy trình phân tích khối dữ liệu đó để trích xuất những thông tin cần thiết cho quá trình giám sát và theo dõi thiết bị được mô tả tổng quát như hình 1.



Hình 1. Mô hình mô tả các bước của quá trình quan sát và theo dõi sức khỏe thiết bị sản xuất

Công việc trong bài viết này của tác giả tập trung vào quá trình trích xuất dữ liệu và đưa ra mô hình chuẩn đoán lỗi cho thiết bị máy móc sản xuất.

Để làm rõ được điều đó, bài viết này sẽ được tổ chức như sau: trong mục 2, tác giả trình bày một cơ sở phân tích lý thuyết xung quanh các kỹ thuật học máy (machine learning), từ đó chọn ra và áp dụng kỹ thuật học mạng Bayesian. Mục 3 trên cơ sở đó trình bày một chu trình ứng dụng mạng Bayesian vào quá trình học dữ liệu từ hệ thống sản xuất. Tiếp đó, mục 4 trình bày trường hợp ứng dụng của phương pháp này và đưa ra kết quả chuẩn đoán nguyên nhân lỗi áp dụng trong một ví dụ mô phỏng quá trình sản xuất lò hơi Tennessee Eastman. Cuối

cùng, kết luận và thảo luận của đề tài sẽ đóng lại bài viết này.

2. CƠ SỞ LÝ THUYẾT

Trong bối cảnh mà hệ thống sản xuất được đặc trưng bởi sự phức tạp cao và sự không chắc chắn bởi nhiều yếu tố rủi ro. Sản xuất công nghiệp thậm chí còn phức tạp hơn với nhiều chu trình sản xuất cùng chạy trên một dây chuyền sản xuất với khối lượng các bước thực hiện sản xuất rất lớn (đặc biệt trong môi trường bán dẫn và công nghiệp dược phẩm) và thời gian sản xuất cũng kéo dài (có khi lên đến 8 đến 10 tuần sản xuất). Do đó, môi trường sản xuất trong công nghiệp hàm chứa nhiều yếu tố không chắc chắn nhất định (sự sai lệch hoạt động của thiết bị sản xuất, sai sót của người vận hành...) có thể ảnh hưởng đến quá trình điều khiển và bối cảnh bảo trì.

Do đó, tác giả phân tích những phương pháp học máy phổ biến dựa trên các tiêu chuẩn được định ra để phù hợp với bối cảnh sản xuất công nghiệp 4.0:

- Thích hợp sử dụng với đa dạng đặc tính của tập hợp tham số (rời rạc, liên tục, định tính và định lượng). Ví dụ: thời gian, số đo kỹ thuật số, số lượng...).
- Thích hợp sử dụng với nhiều lớp tham số, dữ liệu từ thiết bị sản xuất (cảm biến, động cơ...) và sản phẩm.
- Thích hợp sử dụng với đa dạng nhiều loại biến (biến quan sát được và không quan sát được).
- Thể hiện được mối tương quan (mối quan hệ nguyên nhân – hệ quả) giữa các biến và các sự kiện.
- Xử lý được các vấn đề thiếu/mất một phần dữ liệu và hoặc dữ liệu không chắc chắn (trong các trường hợp dữ liệu học đầy đủ và dữ liệu học không đầy đủ).
- Khả năng thích ứng: được định nghĩa như là khả năng sử dụng đa mục đích của phương pháp mô hình hoá từ dữ liệu, chẳng hạn: có thể sử dụng để chuẩn đoán, dự đoán, tiên lượng lỗi, truy xuất nguyên nhân lỗi...

Dựa trên các tiêu chí này, tác giả tiến hành phân tích ba loại phương pháp học máy phổ biến nhất: Cây quyết định (Decision tree), mạng Nơ ron và mạng Bayesian.

2.1 Cây quyết định

Cây quyết định là một công cụ phổ biến nhất để phân biệt các lớp (biến, tham số hoặc sự kiện). Ưu điểm chính của cây quyết định là chúng có thể dễ dàng được chuyển đổi thành các quy tắc dễ hiểu. Do đó, đường dẫn (logic) dẫn cây đến một quyết định rất rõ ràng đối với người dùng. Nếu thuộc tính là nhị phân, chúng ta có hai quyết định có thể, trong khi nếu thuộc tính có phương thức k , chúng ta có k quyết định có thể. Do đó, mặc dù việc quyết định được thực hiện nhanh chóng, việc xây dựng mô hình cây tốn nhiều thời gian hơn. Do đó, khả năng quản lý các lớp tham số, Cây quyết định là không lớn và khó khả thi.

Cây quyết định vẫn còn một số hạn chế: Đầu tiên trong số đó là chúng không thực sự hỗ trợ các giá trị liên tục. Luôn luôn có thể phân biệt chúng nhưng điều này sau đó đặt ra vấn đề phân biệt tối ưu (mất thông tin so với biến ban đầu). Ngoài ra, cây quyết định rất nhạy cảm với ồn nhiễu trong dữ liệu. Điều này có nghĩa là luôn cần một bộ lọc dữ liệu cho phương pháp này.

2.2 Mạng Nơ ron

Mạng nơ-ron nhân tạo, còn được gọi là mạng nơ-ron thần kinh, là một kỹ thuật phi tuyến tính để dự đoán dữ liệu và thực hiện kết nối một số tế bào thần kinh làm phát sinh một mạng lưới thần kinh. Mạng có các đầu vào từ bên ngoài được kết nối với một số nơ-ron nhất định và sau đó mạng cung cấp một hoặc nhiều đầu ra ra bên ngoài (đầu ra của một hoặc nhiều nơ-ron trong mạng). Do đó, một tế bào thần kinh làm cho nó có thể mô hình hóa một số lượng đáng kể các hành vi theo các trọng số. Tuy nhiên, mạng lưới thần kinh hầu như chỉ xử lý các biến liên tục.

Mặc dù số lượng tế bào thần kinh trong lớp đầu vào và đầu ra được áp đặt bởi số lượng đầu vào cho hệ thống cũng như mã hóa của các lớp khác nhau (đối với lớp đầu ra), không có quy tắc toán học nào để xác định số

lượng các lớp ẩn và số lượng tế bào thần kinh trong mỗi lớp này. Tuy nhiên, một thực tế phổ biến là người ta sử dụng một lớp ẩn duy nhất được tạo thành từ khoảng một nửa số nơ-ron trong hệ thống cộng với một. Vì vậy, để xử lý dữ liệu bị thiếu là cần thiết nhất định có điều kiện. [2]

Ưu điểm chính của việc xác định bởi một mạng thần kinh là cho phép mô hình hóa các quá trình đa biến phi tuyến tính. Thật vậy, sức mạnh của mạng nằm ở cấu trúc của nó và trên các trọng số được áp dụng cho các kết nối của nó, nhưng điều này do đó không cho phép "tính thích ứng" được tính đến.

2.3 Mạng Bayesian

Mạng Bayes là một hệ thống đại diện cho kiến thức và giúp tính toán xác suất có điều kiện cung cấp giải pháp cho các loại vấn đề khác nhau. Cấu trúc của loại mạng này rất đơn giản: một biểu đồ trong đó các nút biểu thị các biến ngẫu nhiên và các cung (do đó là đồ thị được định hướng) kết nối mạng sau có liên quan đến xác suất có điều kiện [3]. Biểu diễn đồ họa trực quan nhất về ảnh hưởng của một sự kiện độc lập hoặc một sự kiện với một biến số khác, liên kết nguyên nhân với hiệu ứng bằng một mũi tên. Các biến này có thể rời rạc hoặc liên tục (phân phối bình thường), có thể quan sát hoặc không quan sát được, phát hiện hoặc không phát hiện ...

Trong trường hợp chung, tập hợp biến $X = \{X_1, X_2, \dots, X_n\}$, hàm phân phối chung $P(X)$ được viết như sau:

$$P(X) = \prod_{i=1}^n P(X_i / \text{Parents}(X_i)) \quad (1)$$

Có thể nói, chỉ có một công cụ có thể xử lý cả trường hợp các biến rời rạc và biến liên tục: mạng Bayesian. Bởi vì, chúng ta có thể xây dựng mạng Bayesian với biểu đồ của mô hình trên dữ liệu của các biến rời rạc hoặc các bảng xác suất của từng biến liên tục.

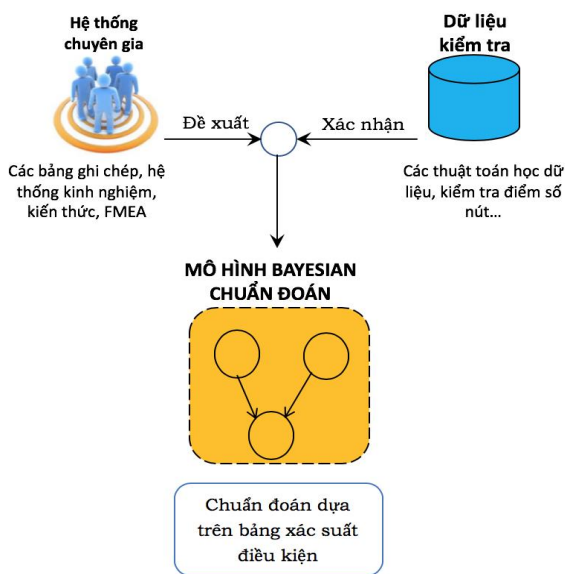
Mạng Bayesian có thể đại diện cho nhiều biến số (theo nút) và các lớp phân cấp. Việc tính toán xác suất được thực hiện từ thuật toán ước tính tối ưu (Tối đa hóa khả năng và tối đa hóa kỳ vọng) có thể biểu thị mối tương

quan giữa các nút (sẽ được trình bày ở mục sau). Do đó, lợi thế lớn của mạng Bayesian là khả năng thích ứng. Bên cạnh đó các mạng Bayes cho phép tính đến khía cạnh thời gian nhờ vào các mạng Bayes động [3]. Thật vậy, cấu trúc và tham số của nó không thay đổi theo thời gian, nhưng có thể biểu diễn mạng Bayesian với các mối tương quan thời gian.

Trong bài báo, tác giả thực hiện một phần phân tích theo các yếu tố tiêu chí đã được trình bày để làm rõ hơn về sự chọn lựa mạng Bayesian trong cách xây dựng mô hình chuẩn đoán dựa trên học dữ liệu, chi tiết về phương pháp tiếp tục được giới thiệu trong mục 3.

3. QUÁ TRÌNH THIẾT LẬP MẠNG BAYESIAN CHO VẤN ĐỀ CHUẨN ĐOÁN LỖI

Trong phần này, tác giả mô tả quá trình thiết lập mạng Bayesian cho vấn đề chuẩn đoán bao gồm hai bước chính được mô tả như hình 2.



Hình 2. Quy trình phương pháp chuẩn đoán lỗi bằng mạng Bayesian

3.1 Phân tích quan hệ nhân quả

Phân tích chế độ lỗi và hệ quả (Failure mode and effects analysis – FMEA) là phương pháp thường được sử dụng để xác định một danh sách các loại lỗi và nguyên nhân lỗi bởi các chuyên gia. Trên thực tế, bảng phân tích chế độ lỗi và hệ quả này được

thể hiện thông qua một bảng dẫn báo cáo tổng hợp từ nhiều kỹ sư vận hành trực tiếp đến giám sát hệ thống sản xuất. Nó được xác định bởi sự phân loại định tính các chế độ lỗi và các kết quả bảo trì thực hiện bởi các chuyên gia dựa trên kinh nghiệm và sự hiểu biết của họ [4]. Ứng dụng phương pháp này vào mạng Bayesian, bảng phân tích lỗi và hệ quả của nó được chuyển hoá thành các nút mạng (node) và các đường biểu thị nguyên nhân kết quả (arc) trong mô hình mạng dưới dạng như mô tả trong hình 4.

Trong bài viết này, tác giả sử dụng phân tích chế độ lỗi và hệ quả FMEA được sử dụng tối ưu bởi kinh nghiệm từ người vận hành và giám sát hệ thống sản xuất đã được thiết lập ở hầu hết các hệ thống sản xuất thực tế, trong khi đó cơ sở dữ liệu thể hiện các phân tích nguyên nhân - hệ quả chưa thực sự được quan tâm ở nhiều cơ sở sản xuất.

3.2 Mô hình hoá

Trong mục này, tác giả đề xuất một cơ chế học dữ liệu để kiểm tra lại mối quan hệ nhân quả đề xuất bởi các chuyên gia (và có thể tìm ra mối quan hệ mới). Cơ chế tự học tự động được thực hiện bởi việc trích xuất từ dữ liệu lịch sử sản xuất. Mục tiêu của giai đoạn học tập là tìm ra cấu trúc của mạng Bayes (học cấu trúc) và ước tính xác suất có điều kiện (học các tham số). Có thể chia ra hai loại dữ liệu học tập: (i) với dữ liệu đầy đủ, tất cả các biến được quan sát, không có biến tiềm ẩn. Phương pháp đơn giản và được sử dụng rộng rãi nhất là ước tính thống kê, bao gồm ước tính xác suất của một sự kiện theo tần suất xuất hiện của sự kiện trong cơ sở dữ liệu. Cách tiếp cận này, được gọi là khả năng tối đa (Maximum Likelihood Estimation) [4]:

$$P(X_i = x_k | Pa(X_i) = x_j) = \theta_{i,j,k} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \quad (2)$$

Trong đó $N_{i,j,k}$ là số lượng sự kiện trong cơ sở dữ liệu mà biến X_i ở trạng thái x_k và cha mẹ của nó nằm trong cấu hình x_j .

(ii) với dữ liệu không đầy đủ: Trong các ứng dụng thực tế, cơ sở dữ liệu rất thường không đầy đủ. Sau khi nhận thấy sự tồn tại của các loại dữ liệu không đầy đủ khác nhau, chúng

tôi sẽ tiếp cận hai trường hợp có thể được xử lý tự động và sau đó tập trung vào một trong các thuật toán được sử dụng nhiều nhất là thuật toán cực đại hóa kỳ vọng (Expectation-maximization algorithm) [5].

Về cơ bản, phương pháp Bayesian là một phương pháp suy luận dựa trên báo cáo xác suất. Nghĩa là kết quả của quá trình học tập từ dữ liệu của mô hình mạng này là một bảng xác suất bao gồm hai loại xác suất [6,7]: xác suất biên (Marginal probability) là xác suất của một sự kiện (biến) mà không quan tâm đến các sự kiện (biến) khác và xác suất có điều kiện (Conditional probability) là xác suất của một sự kiện (biến) A nào đó, biết rằng một sự kiện (biến) B khác xảy ra.

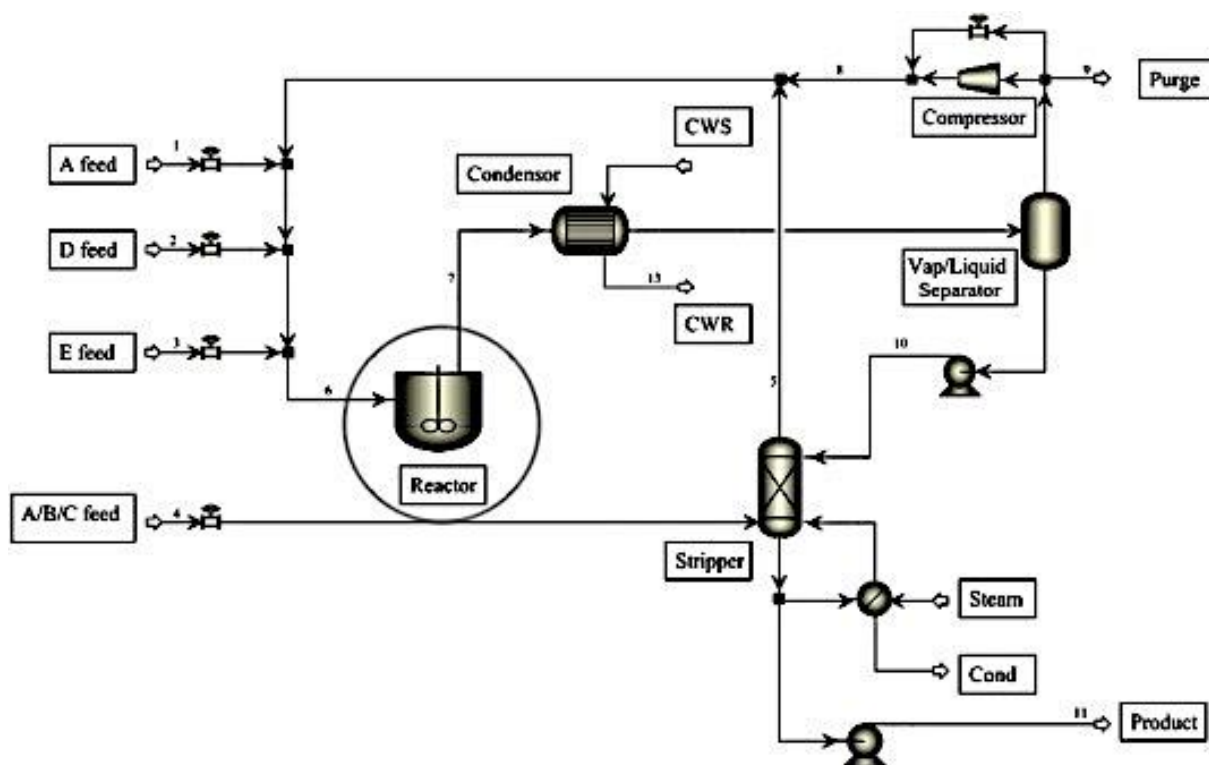
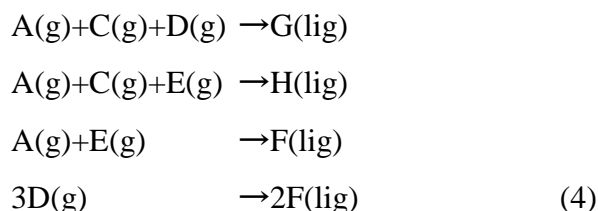
$$p(A | B) = (p(B | A) p(A)) / p(B) \quad (3)$$

Định lý của Bayes cho phép đảo ngược xác suất. Đó là nói từ nguyên nhân để suy ra hậu quả. Hay nói cách khác, bắt đầu từ một hậu quả để quay trở lại nguyên nhân, tức là chuẩn đoán. Do đó từ một sự thất bại (hoặc lỗi), một hệ thống dựa trên các mạng Bayes sẽ có thể xác định các nguyên nhân có thể xảy ra nhất dẫn đến sự cố.

4. ỨNG DỤNG CHUẨN ĐOÁN TRÊN QUY TRÌNH SẢN XUẤT TENNESSEE EASTMAN

4.1 Quy trình Tennessee Eastman

Quy trình Tennessee Eastman là một quy trình được phát triển bởi Công ty Hóa chất Eastman nhằm cung cấp một mô phỏng của một quy trình công nghiệp thực sự để thử nghiệm các phương pháp kiểm soát và / hoặc giám sát quá trình [8]. Quy trình này bao gồm năm thành phần chính: lò phản ứng, thiết bị ngưng tụ, máy nén tái chế, thiết bị phân tách và bộ tháo gỡ. Thuốc thử khí (gas) A, C, D, E và khí trơ B được đưa vào lò phản ứng. Thành phần G và H là hai sản phẩm khí mong muốn tạo ra (product), trong khi F là sản phẩm phụ không mong muốn. Các phản ứng hóa học của quá trình được đưa ra bởi hệ phương trình (4).



Hình 3. Mô hình mô phỏng quy trình công nghiệp Tennessee Eastman [9]

(g) được ký hiệu chỉ khí gas

(lig) được ký hiệu chỉ chất lỏng

Trong khuôn khổ của bài báo, chúng tôi chỉ ứng dụng mô hình Bayesian chuẩn đoán lỗi cho lò phản ứng trong chu trình sản xuất Tennessee Eastman. Tại lò phản ứng, các thuốc thử và các loại khí được đưa vào để bắt đầu quá trình sản xuất, được mô tả phân hoạch tròn như hình 3.

Tất cả các phản ứng là không thể đảo ngược, tỏa nhiệt và thứ tự được quy định bởi hệ phương trình (4). Tốc độ phản ứng được biểu thị như là một hàm của nhiệt độ Arrhenius. Phản ứng tạo năng lượng G lớn hơn kích hoạt tạo ra H, dẫn đến độ nhạy nhiệt độ cao hơn.

Quy trình Tennessee Eastman bao gồm hơn 20 lỗi khác nhau. Những lỗi này có nhiều tính chất khác nhau: chẳng hạn như thiếu thiết bị truyền động hay như van bị chặn... Bên cạnh đó, quy trình bao gồm 53 tham số (cảm biến, thông số thiết lập, tùy chỉnh van...). Mô tả về 20 lỗi này và 53 tham số được đưa ra trong công việc [8]. Tuy nhiên, tác giả thực hiện phân ứng dụng của mình để mô hình hoá khả năng áp dụng mô hình chuẩn đoán này với chỉ một số biến nhất định, được mô tả như bảng 1.

Bảng 1. Các biến sử dụng trong mô hình chuẩn đoán

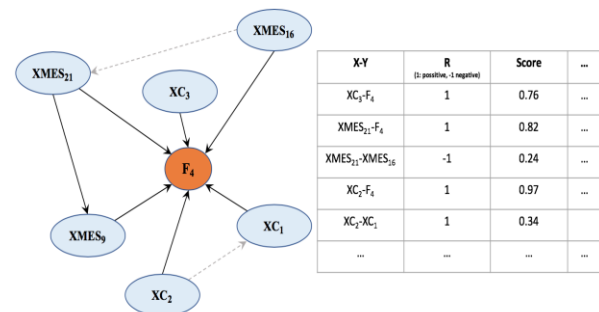
Nút	Mô tả	Đơn vị
F ₄	Lỗi báo nhiệt độ của bình ngưng tụ làm mát của lò phản ứng	°C
XMES ₉	Thông số nhiệt độ của lò phản ứng	°C
XMES ₁₆	Thông số áp lực của bộ tháo dỡ	Kpa
XMES ₂₁	Nhiệt độ của chất làm mát ở đầu ra lò phản ứng	°C
XC ₃	Lưu lượng nạp khí A (Luồng 1)	Kg/hr
XC ₁	Lưu lượng nạp khí D (Luồng 2)	Kg/hr
XC ₂	Lưu lượng nạp khí E (Luồng 3)	Kg/hr

4.2 Kết quả mô phỏng

Kết quả mô phỏng được tác giả thể hiện trên hai nội dung:

a) Mọi quan hệ để thiết lập nên mạng Bayesian của các điểm nút trong bảng 1 thực tế đã được đề xuất bởi chuyên gia của công ty Eastman. Tác giả đề xuất tham khảo thêm các mối quan hệ giữa các biến lỗi và thông số cảm biến tại [8]. Hơn thế nữa, trên quy trình được đề xuất ở hình 2- mục 2 trong bài báo, các mối quan hệ giữa các biến trong bảng 1 được xác nhận bởi một quá trình học dữ liệu, mà tương ứng độ chính xác của mối quan hệ đó được đánh giá bởi điểm số học từ dữ liệu (score by learning) thực hiện dựa trên thuật toán kiểm tra mối tương quan [10].

Công việc mô phỏng trong bài báo được thực hiện trên môi trường Matlab, thư viện BNT [11]. Hình 4 thể hiện mô hình Bayesian các mối quan hệ nguyên nhân- hệ quả của các biến trong ví dụ (các đường mũi tên bình thường thể hiện mối quan hệ được xác nhận, các đường nét đứt mờ thể hiện mối quan hệ không được xác nhận do điểm số thấp).

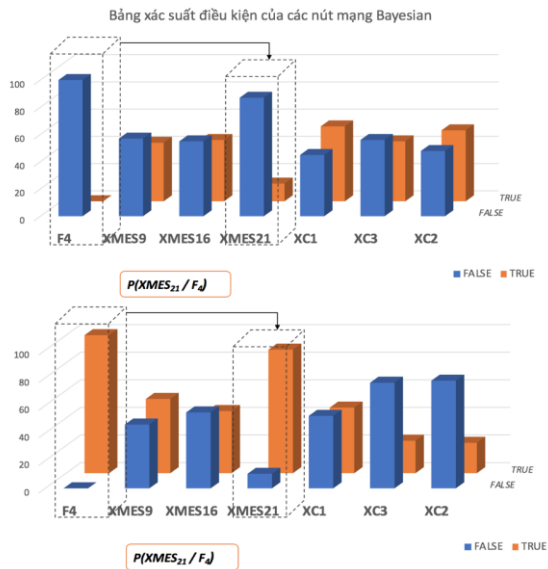


Hình 4. Mô hình chuẩn đoán mạng Bayesian

b) Sau khi mô hình Bayesian được xác lập, bước còn lại chỉ còn là vấn đề của sự tính toán. Bảng xác suất của mỗi biến trong mạng Bayesian được xác lập sau quá trình học dữ liệu trên một cơ sở lịch sử dữ liệu sản xuất mô phỏng, được cung cấp trong [12].

Xác suất có điều kiện $P(XMES_{21} | F_4)$ thể hiện xác suất của việc nhiệt độ quá cao của chất làm mát ở đầu vào của lò phản ứng gây nên lỗi báo quá nhiệt của bình ngưng tụ làm mát của lò phản ứng trong chu trình Tennessee Eastman. Tương ứng với dữ liệu lịch sử học tập và giả định do tác giả tạo nên

trong dữ liệu học tập đó. Hình 5 thể hiện rất rõ xác suất của sự thay đổi này khi phát hiện lỗi F4 trong hệ thống. Kết quả này cho thấy sự tương đồng của về mặt chuẩn đoán có từ mô hình và dữ liệu giả định của tác giả.



Hình 5. So sánh sự thay đổi của bảng xác suất điều kiện của các nút mạng khi phát hiện lỗi trong hệ thống sản xuất.

Bên cạnh đó, kết quả của bài báo này cũng thể hiện sự sàng lọc lại các mối quan hệ không chắc chắn (được thể hiện bằng mũi tên đứt ở hình 4) so với một số kết quả thực nghiệm phân loại lỗi và chuẩn đoán ở những nghiên cứu tương tự như [8]. Điều đó tạo nên sự tinh giảm hữu ích kích thước của mô hình mạng Bayesian.

Tuy nhiên, cần phải lưu ý rằng việc phân loại các biến lỗi cũng như thông số của nó sẽ quyết định kết quả học tập và kết quả chuẩn đoán của mô hình. Do đó, việc xác nhận lại mối quan hệ nguyên nhân – hệ quả bằng một phương pháp tính điểm score các mối quan

hệ như trong công việc của tác giả là rất cần thiết. Kết quả của việc chuẩn đoán cũng phụ thuộc rất lớn vào chất lượng của dữ liệu sử dụng để học tập. Vì thế, kết quả chuẩn đoán dựa trên xác suất chỉ là để hỗ trợ cho quyết định can thiệp của con người.

5. KẾT LUẬN

Công việc của bài báo này đưa ra một phương pháp xây dựng mô hình chuẩn đoán lỗi dựa trên mạng Bayesian căn cứ trên những phân tích đặc tính của các phương pháp máy học hữu dụng nhất. Tác giả cũng đã thực hiện một mô phỏng ví dụ ngắn gọn trên một hệ thống sản xuất mô phỏng thực, mô hình Tennessee Eastman được áp dụng rộng rãi trong học thuật quốc tế. Kết quả từ mô hình mô phỏng của bài báo cho thấy hiệu quả của mô hình trong bước đầu hỗ trợ quyết định bảo trì bảo dưỡng của người vận hành chu trình sản xuất.

Bên cạnh đó, mô hình Bayesian bài báo xây dựng vẫn còn một số hạn chế: (i) phụ thuộc vào chất lượng của cơ sở dữ liệu có thể dẫn đến những kết quả sai lệch trong trường hợp dữ liệu học tập không đầy đủ; (ii) chưa thể hiện được những mối quan hệ nguyên nhân – hệ quả theo thời gian bởi thực tế có các những rủi ro tác động gây ra lỗi theo những thời gian khác nhau. Hạn chế này sẽ được khắc phục bởi một mạng Bayesian thời gian sẽ được tác giả tiếp tục nghiên cứu của mình trong thời gian tới.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ kinh phí bởi Trường Đại học Sư phạm Kỹ Thuật – Đại học Đà Nẵng trong đề tài có mã số T2019-06-131.

TÀI LIỆU THAM KHẢO

- [1] Kunio S., Mitsugu K., Yoshifumi K., *An Advanced step in TPM Implementation*. (pages 64-65). Paris, France, 1995.
- [2] Dreyfus, Gérard. *Réseaux de neurones: méthodologie et applications*. Eyrolles, 2004.
- [3] Jensen F.V., *Introduction to Bayesian networks*, UCL Press, London, 1996.
- [4] Tran et al., *Identification and classification protocol for complex systems*. In: 2nd European Conference of the Prognostics and Health Management Society, PHME 2014. 2014. p. 58-65.

- [5] Philippe Leray. *Réseaux bayésiens « apprentissage et modélisation de systèmes complexes »*. 2006.
- [6] Sayad Mahmoud et Abbaris Amara. *Modèle discriminant pour la classification de documents XML à l'aide des réseaux bayésiens et le noyau de Fisher*. Ecole nationale Supérieure d'Informatique (ESI). Mémoire de fin d'études. 2010.
- [7] Feller, William. *An introduction to probability theory and its applications*. Vol. 1. John Wiley & Sons,, 1968.
- [8] Par Sylvain VERRON. *Diagnostic et surveillance des processus complexes par réseaux Bayésiens*. Thèse do[ctorale d'Angers. 2007.
- [9] Lau, C. K., et al. *Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS*. Chemometrics and Intelligent Laboratory Systems 120 (2013): 1-14.
- [10] Wetzels, Ruud, and Eric-Jan Wagenmakers. *A default Bayesian hypothesis test for correlations and partial correlations*. Psychonomic bulletin & review 19.6 (2012): 1057-1064.
- [11] Murphy, K. *The BayesNet Toolbox for Matlab*. Computing Science and Statistics. Proceedings of Inference, vol.33.
- [12] Basha, Nour, et al. *Multiclass Data Classification using Fault-Detection-based Techniques*. Computers & Chemical Engineering (2020): 106786.

Tác giả chịu trách nhiệm bài viết:

TS. Trần Ngọc Hoàng

Trường Đại học Sư Phạm Kỹ Thuật – Đại học Đà Nẵng, Đà Nẵng

Email: tnhoang@ute.udn.vn