



Journal of Mining and Earth Sciences

Website: <http://jmes.humg.edu.vn>



Real-time prediction of formation lithology using drilling parameters: an example from Ca Tam oilfield



Duong Hong Vu ¹, Hung Tien Nguyen ^{1,*}, Vinh The Nguyen ¹, Anh Tuan Nguyen ²

¹ Hanoi University of Mining and Geology, Hanoi, Vietnam

² Joint Venture Vietsovpetro, Vungtau, Vietnam

ARTICLE INFO

Article history:

Received 10th Jan. 2024

Revised 28th Apr. 2024

Accepted 19th May 2024

Keywords:

Ca Tam oil field,
Drilling,
Machine learning,
Real-time data,
Stratigraphic column.

ABSTRACT

Construction of stratigraphic column is an important stage in minerals exploration and researching the historical development of geological processes. Besides, determining and identifying the boundaries of lithological layers also helps a lot in minimizing the risk of drilling complications and incidents as well as increasing efficiency in drilling. In this study, the authors focus mainly on applying machine learning algorithms to classify lithology and identify stratigraphy directly from the real-time drilling data of 02 wells in the Ca Tam oil field. The proposed model has high accuracy, this result demonstrates the great superiority and effectiveness of applying this method. The model using the Fuzzy c-means algorithm has predicted and identified relatively accurately the three main lithological groups in the study area: sandstone, claystone, and clay. The study's encouraging findings demonstrate the need for further focus and funding on this new strategy in the future to raise the effectiveness of oil and gas well drilling in Vietnam.

Copyright © 2024 Hanoi University of Mining and Geology. All rights reserved.

*Corresponding author

E - mail: nguyentienhung.dk@humg.edu.vn

DOI: 10.46326/JMES.2024.65(3).06



Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <http://tapchi.humg.edu.vn>

Phân loại thạch học trực tiếp từ tài liệu khoan theo thời gian thực, áp dụng cho các giếng khoan dầu khí tại mỏ Cá Tầm

Vũ Hồng Dương¹, Nguyễn Tiến Hùng^{1,*}, Nguyễn Thế Vinh¹, Nguyễn Anh Tuấn²

¹ Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam

² Joint Venture Vietsovpetro, Vũng Tàu, Việt Nam

THÔNG TIN BÀI BÁO

Quá trình:

Nhận bài 10/01/2024

Sửa xong 28/4/2024

Chấp nhận đăng 19/5/2024

Từ khóa:

Công tác khoan,

Địa tầng thạch học,

Học máy,

Mỏ Cá Tầm,

Tài liệu khoan theo thời gian thực.

TÓM TẮT

Xây dựng cột địa tầng là một nhiệm vụ thiết yếu trong bài toán nghiên cứu, khôi phục lịch sử phát triển kiến tạo địa chất và giúp ích rất nhiều trong công tác tìm kiếm thăm dò tài nguyên khoáng sản, đặc biệt là dầu khí. Bên cạnh đó, việc xác định, nhận diện được ranh giới của các tầng thạch học sẽ giúp Người điều hành khoan nhanh chóng lựa chọn được các thông số khoan tối ưu cho từng công đoạn khoan, giảm nguy cơ phức tạp, sự cố cũng như gia tăng hiệu suất trong công tác khoan. Qua bài báo này, nhóm tác giả bước đầu nghiên cứu một hướng tiếp cận mới đó là sử dụng các kỹ thuật học máy để xác định thạch học và nhận diện trực tiếp địa tầng khoan qua từ tài liệu khoan theo thời gian thực, áp dụng thực tế cho 02 giếng khoan khu vực mỏ dầu khí Cá Tầm. Kết quả của mô hình khi áp dụng thuật toán Fuzzy c-means đã dự báo và nhận diện tương đối chính xác 03 nhóm thạch học chủ đạo tại khu vực nghiên cứu là cát kết, bột kết và sét. Kết quả ban đầu của nghiên cứu là tương đối khả quan, qua đó có thể thấy hướng tiếp cận mới này cần được quan tâm và đầu tư hơn trong tương lai, giúp hỗ trợ và cải thiện hiệu quả quá trình khoan các giếng dầu khí tại Việt Nam.

© 2024 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

*Tác giả liên hệ

E - mail: nguyentienhung.dk@humg.edu.vn

DOI: 10.46326/JMES.2024.65(3).06

1. Mở đầu

Việc xác định, nhận diện ranh giới của các tầng thạch học trực tiếp trong quá trình khoan, đặc biệt là tại các giếng khoan thăm dò giúp các kỹ sư có những điều chỉnh hợp lý và nhanh chóng về các thông số công nghệ khoan, giúp giảm thiểu sự cố và nâng cao hiệu quả công tác khoan. Trước đây, việc nhận diện ranh giới các tầng thạch học trong quá trình khoan thường được căn cứ vào cột địa tầng dự kiến và kinh nghiệm của Người điều hành khoan. Để xây dựng bức tranh địa chất và lên kế hoạch khoan cho những giếng khoan mới, các kỹ sư sẽ dựa trên việc minh giải và tổng hợp thông tin địa chất các giếng lân cận, có cùng độ nghiêng và phương vị, có độ sâu các tầng đất đá gần khớp với giếng đang khoan (chiều sâu đã khoan qua). Điều này không khả thi trong trường hợp các mỏ mới phát triển (chưa khoan nhiều giếng lân cận trước đó), cột địa tầng dự kiến thường có sai số tương đối lớn, đặc biệt là tại các mỏ có đặc điểm cấu trúc địa chất bất đồng nhất và phức tạp. Bên cạnh đó, việc không thể nhận diện ranh giới các địa tầng trong thời gian khoan thực tế dẫn đến gia tăng xác suất xảy ra sự cố, đặc biệt khi khoan qua các địa tầng có những tính chất phức tạp khác nhau; hiệu suất khoan không được tối ưu hoá ngay lập tức tại thời điểm khoan qua ranh giới các địa tầng. Có thể thấy, việc đề ra các phương pháp dự báo thạch học, nhận diện địa tầng trực tiếp dựa trên tài liệu khoan theo thời gian thực của giếng là cấp thiết và thời sự.

Tại Việt Nam, nhiệm vụ xác định thông tin địa chất và nhận diện các địa tầng trong giếng khoan thường được giải quyết bởi các phương pháp cổ điển, đó là dựa trên việc phân tích mẫu lõi hoặc minh giải tài liệu carota. Năm 2009, Nguyễn Xuân Nhạ đã giới thiệu chương trình nhận diện các địa tầng tự động cho các giếng khoan từ tài liệu carota tại vùng nghiên cứu nằm giữa Sông Tiền và Sông Hậu (Nguyễn, 2009). Sử dụng phương pháp thực nghiệm, bài báo đã đề xuất được hàm quan hệ giữa các nhóm thạch học với giá trị tỷ phần sét và cường độ phóng xạ gamma của đất đá. Qua đó, tác giả đã biểu diễn được thông tin phân bố các tầng đất đá cho giếng khoan, kết quả phân biệt chính xác được các tầng cát và sét, mặc dù vậy mô hình này chưa phân biệt được rõ ràng các nhóm thạch học có kích cỡ hạt nhỏ hơn như cát bột hoặc sét bột. Cũng tại Việt Nam, Trần Thị Thanh Thủy,

Nguyễn Tiến Thịnh và các cộng sự tại Viện Dầu khí đã có những công bố trên tạp chí Dầu khí về nghiên cứu nhận dạng các nhóm thạch học, dự báo các tính chất thạch học của đá chứa Pliocene khu vực trung tâm bể Sông Hồng (Trần và nnk., 2019) và đá chứa Carbonate tuổi Devonian mỏ Bắc Oshkhotunskoye, Liên Bang Nga (Trần và nnk., 2021). Các nghiên cứu này đều đã phân chia được tầng đá chứa thành các loại thạch học khác nhau theo chiều sâu dựa trên việc sử dụng kết quả phân tích mẫu và minh giải tài liệu carota. Trong một công bố vào năm 2020, Vũ Hồng Dương và Kiều Duy Thông đã sử dụng thuật toán KNN và mạng Neural nhân tạo để phân loại các tầng đất đá theo chiều sâu dựa trên tài liệu đo địa vật lý của bộ giếng khoan 1143 (ODP) tại biển Đông (Vũ & Kiều, 2020). Nhóm tác giả đã phân chia được chi tiết theo chiều sâu giếng khoan 03 loại thạch học chủ đạo là cát, sét và cát pha sét với sai số thấp hơn nhiều so với các phương trình xác định hàm lượng sét cổ điển. Có thể thấy, đa phần các công bố trên chỉ sử dụng nguồn tài liệu dùng lại ở mẫu lõi và tài liệu đo carota, các dạng tài liệu này đều được thu thập sau khi công đoạn khoan đã diễn ra, giá thành thi công cao và đòi hỏi lượng thời gian tương đối lớn. Tuy vậy, tại Việt Nam chưa có nghiên cứu nào theo hướng tận dụng nguồn tài liệu khoan theo thời gian thực (real time drilling data) để trực tiếp nhận diện các tầng thạch học ngay trong quá trình khoan, giúp lựa chọn được các phương pháp, thông số công nghệ khoan phù hợp, hỗ trợ cải thiện hiệu quả quá trình khoan.

Trên thế giới, các công nghệ mới trong ngành Dầu khí đã có những phát triển vượt bậc, khối lượng tài liệu đo ghi được trực tiếp khi khoan ngày càng lớn, chi tiết và đa dạng; từ đó đặt ra bài toán tận dụng nguồn tài liệu này để tối ưu hoá công tác khoan, rút ngắn thời gian và giảm giá thành thi công. Trong những năm gần đây, rất nhiều công bố trên thế giới đề cập tới bài toán này; các nhà khoa học đã nghiên cứu ứng dụng các thành tựu về AI cùng các thuật toán học máy để phân loại thạch học, nhận diện địa tầng và xây dựng được cột địa tầng trực tiếp từ tài liệu khoan theo thời gian thực với độ hiệu quả cao. Một hướng tiếp cận đang là xu thế và đã khẳng định được độ hiệu quả đó là ứng dụng mạng nơron nhân tạo (ANN) để thiết lập các hàm toán học phi tuyến, khi mối quan hệ qua lại giữa các thông số kỹ thuật khoan rất phức tạp và không dễ xác định. Điểm mạnh của phương pháp

ANN là các mô hình này có thể tự học và điều chỉnh lại các tham số, giúp đưa ra kết quả dự báo sát với tài liệu đo thực tế và không chịu ảnh hưởng của ý kiến chủ quan. Với tính ưu việt trên, việc nghiên cứu phương pháp ANN giải quyết bài toán dự báo thạch học, nhận diện địa tầng dựa trên tài liệu khoan theo thời gian thực, bao gồm các thông số đầu vào như: chiều sâu thẳng đứng, tốc độ quay chông, tải trọng lên chông, mô men xoắn tại chông, lưu lượng dung dịch khoan, tỉ trọng dung dịch khoan,... đã được rất nhiều các tác giả khác nhau đề xuất (Moazzeni & Haffar, 2015; Yunxin và nnk., 2018; Jian và nnk., 2019; Chen, 2020). Năm 2022, Romy và cộng sự đã nghiên cứu các mô hình vận dụng SGD (Stochastic Gradient Descent), RMS PROP (Root-meansquare propagation) nhằm cải thiện tính chính xác của mạng ANN nhận diện thạch học (Romy và nnk., 2020). Tài liệu khoan theo thời gian thực của 4 giếng khoan với bộ thông số: tốc độ cơ học, kích cỡ mũi khoan, tải trọng lên chông, áp suất bơm, tỷ trọng dung dịch khoan,... được dùng làm thông tin huấn luyện của mạng ANN, đầu ra của mô hình sẽ là thông tin nhận diện 5 loại đất đá khác nhau theo chiều sâu giếng. Thông tin phân loại thạch học từ các mô hình đề xuất của nhóm tác giả có độ trùng khớp khá cao (từ 81÷92%) khi so sánh với thông tin mẫu lõi. Cũng trong một công bố gần đây vào tháng 4/2022, Mikkel Leite và cộng sự đã vận dụng mô hình Deep Neural Networks (DNNs) (là một loại mạng neural nhân tạo sâu, phát triển từ mạng neural nhiều tầng (multilayer neural networks)) để xác định được độ sâu các tầng đất đá khác nhau dựa trên tài liệu khoan theo thời gian thực và thông tin mẫu lõi của 05 giếng khoan, kết quả dự báo có độ trùng khớp lên tới 94% (Mikkel và nnk., 2022). Bên cạnh đó, các phương pháp học máy có giám sát cũng được sử dụng hiệu quả trong các nghiên cứu liên quan đến bài toán nhận diện thạch học. Trong một công bố vào năm 2022, Tian -Tai Li và cộng sự đã sử dụng không chỉ một mà tới 3 thuật toán học giám sát khác nhau bao gồm Decision Tree, SVP (support vector machine), K-Nearest Neighbor (KNN) trong nghiên cứu của mình để phân tích và dự báo thạch học từ tài liệu khoan với các thông tin về: tải trọng lên chông (WOB), tốc độ cơ học khoan (ROP), tốc độ quay chông (RPM), áp suất tại vòi phun (SPP), mô men tại chông (TORQUE),... (Li và nnk., 2022). Các nhóm thạch học được phân loại từ mô hình đều có

độ trùng khớp cao so với tài liệu thực tế (từ 77,7÷91,7%). Trong một công bố gần đây vào tháng 9/2023, Khalifa và cộng sự đã sử dụng rất nhiều các thuật toán học máy khác nhau để dự đoán thạch học trực tiếp từ tài liệu khoan theo thời gian thực (Khalifa và nnk., 2023). Mô hình ML được áp dụng thông qua ứng dụng web "GeoVision", được đào tạo bằng cách tích hợp chín thông số chế độ khoan từ bộ dữ liệu mỏ dầu Volve nằm ở phía nam biển bắc Na Uy. Mô hình đã đạt được hiệu suất vượt trội trong giai đoạn đào tạo với độ chính xác là 98%. Nghiên cứu này đã nhận diện thành công độ sâu các tập đá sét, Macnơ và cát kết khi thử nghiệm trên dữ liệu mới với độ trùng khớp là 95%. Như vậy, qua các công bố mới đây trên thế giới, có thể thấy tính ưu việt khi áp dụng các thành tựu về AI để giải quyết nhiệm vụ dự báo và nhận diện các địa tầng thạch học từ tài liệu khoan theo thời gian thực. Kết quả các công bố nêu trên đã khẳng định độ chính xác và hiệu quả đạt được là rất triển vọng khi tiếp cận theo hướng nghiên cứu mới này, việc dự báo được chính xác độ sâu các tầng thạch học sẽ hỗ trợ cho việc lên phương án và điều chỉnh các thông số chế độ khoan hợp lý, giảm thiểu thời gian, chi phí và các rủi ro có thể gặp phải trong công đoạn khoan. Qua đó, có thể khẳng định đây cũng là hướng nghiên cứu đáng được đầu tư và ứng dụng tại Việt Nam.

2. Khu vực và tài liệu nghiên cứu

Mỏ Cá Tầm nằm trong lô 09-3/12, Đôn Nam bể Cửu Long, thuộc quản lý của Liên doanh Vietsovpetro (Hình 1). Cơ sở dữ liệu trong nghiên cứu này là tài liệu khoan của 02 giếng A và B tại mỏ Cá Tầm. Theo tài liệu mẫu và master log, có 03 nhóm thạch học chính được phát hiện tại các giếng khoan mỏ Cá Tầm là các lớp bột kết, cát kết xen kẹp, đôi khi xuất hiện các lớp sét mỏng (Hình 2).

Khi so sánh cột địa tầng của 02 giếng khoan A và B, có thể nhận thấy sự phân bố các lớp đất đá ở giếng B là phức tạp hơn rất nhiều, các lớp đất đá tại giếng B có xu hướng nằm xen kẹp nhau với chiều dày tương đối mỏng. Qua đó có thể đưa ra nhận định mặc dù các giếng khoan này đều nằm cùng một khu vực mỏ nhưng đặc điểm cột địa tầng của chúng có rất nhiều khác biệt, khi áp dụng phương pháp truyền thống để dự đoán phân bố các địa tầng cho giếng khoan mới từ cột địa tầng của các giếng đã khoan sẽ mang đến sai số rất lớn.

Để nâng cao độ chính xác khi dự báo thông tin các tầng thạch học, nhóm tác giả thiết lập mô hình để dự báo trực tiếp được các tầng đất đá đang khoan qua dựa trên tài liệu khoan theo thời gian thực của 02 giếng khoan A và B; tài liệu này bao gồm các phép đo các thông số khác nhau như tốc độ khoan, tải trọng lên chông, mô-men xoắn, áp suất, lưu lượng bơm dung dịch,... (Bảng 1).

Dựa theo tài liệu mẫu gồm 2510 điểm của 02 giếng A và B, các tầng đất đá ở đây được phân thành 03 loại thạch học chính: bột kết chiếm chủ đạo (59%), cát kết (24%) bên cạnh đó còn có các lớp mỏng sét chiếm tỷ phần tương đối nhỏ (17%) (Hình 3). Qua phân tích bộ dữ liệu trên, nhóm tác giả quyết định sử dụng các thuật toán học máy nhằm dự báo và phân loại các tầng thạch học theo chiều sâu thành 03 nhóm tương ứng dựa trên tài liệu bao gồm bộ thông số chế độ khoan của 02

giếng A, B khu vực mỏ Cá Tầm.

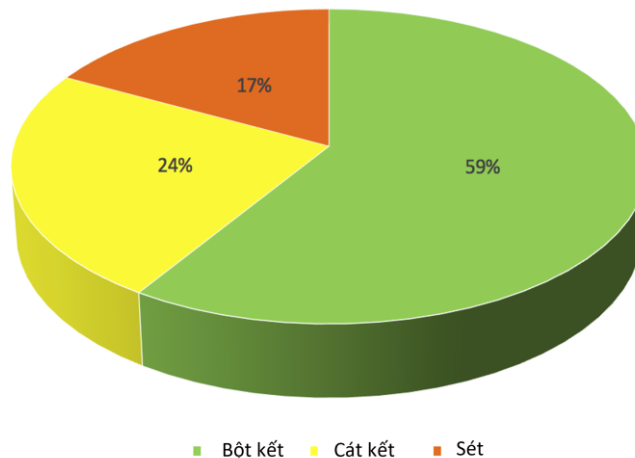
3. Phương pháp nghiên cứu

3.1. Lựa chọn các thông số làm dữ liệu huấn luyện mô hình phân loại thạch học:

Có một nguyên tắc khi huấn luyện các mô hình học máy đó là học từ toàn bộ các thông tin dữ liệu đầu vào được cung cấp. Do đó nếu bộ dữ liệu sử dụng làm tập huấn luyện bao gồm nhiều và những thông tin vô nghĩa thì tính chính xác của kết quả đầu ra thu được sẽ không cao. Để huấn luyện được mô hình dự báo thạch học, nhóm tác giả đã sử dụng bộ số liệu lớn bao gồm 06 thông số chế độ khoan của 02 giếng A, B và rất có khả năng một phần trong bộ số liệu này sẽ không có ảnh hưởng hay đóng góp nhiều vào độ chính xác của kết quả.

Bảng 1. Thống kê chi tiết bộ tài liệu thông số chế độ khoan 02 giếng A và B.

Giếng		Thông số					
		ROP (m/h)	WOB (tấn)	RPM (số vòng/phút)	TQR (kG.m)	FR (l/s)	SPP (atm)
Giếng A (1137 điểm đo)	Giá trị lớn nhất	96,59	6,76	141	2226,74	39,02	201,2
	Giá trị nhỏ nhất	6,21	1	129	1249,23	26,89	86,9
	Giá trị trung bình	39,35	4,02	135,53	1739,99	37,88	169,85
	Độ lệch chuẩn	20,65	1,09	4,58	204,57	1,316	22,66
Giếng B (1373 điểm đo)	Giá trị lớn nhất	126,57	8,34	125	3841,92	55,34	195,4
	Giá trị nhỏ nhất	0,71	2,16	79	2746,63	15,73	40,9
	Giá trị trung bình	27,44	4,58	118,66	3162,46	42,51	163,54
	Độ lệch chuẩn	13,42	0,99	6,24	203,62	11,43	28,88
Tổng hợp 2 giếng (2510 điểm đo)	Giá trị lớn nhất	126,57	8,34	141	3841,92	55,34	201,2
	Giá trị nhỏ nhất	0,71	1	79	1249,23	15,73	40,9
	Giá trị trung bình	32,84	4,32	126,30	2518,01	40,41	166,41
	Độ lệch chuẩn	18,08	1,07	10,07	737,15	8,8	26,43



Hình 3. Tỷ lệ mẫu thạch học từ bộ tài liệu 02 giếng A, B.

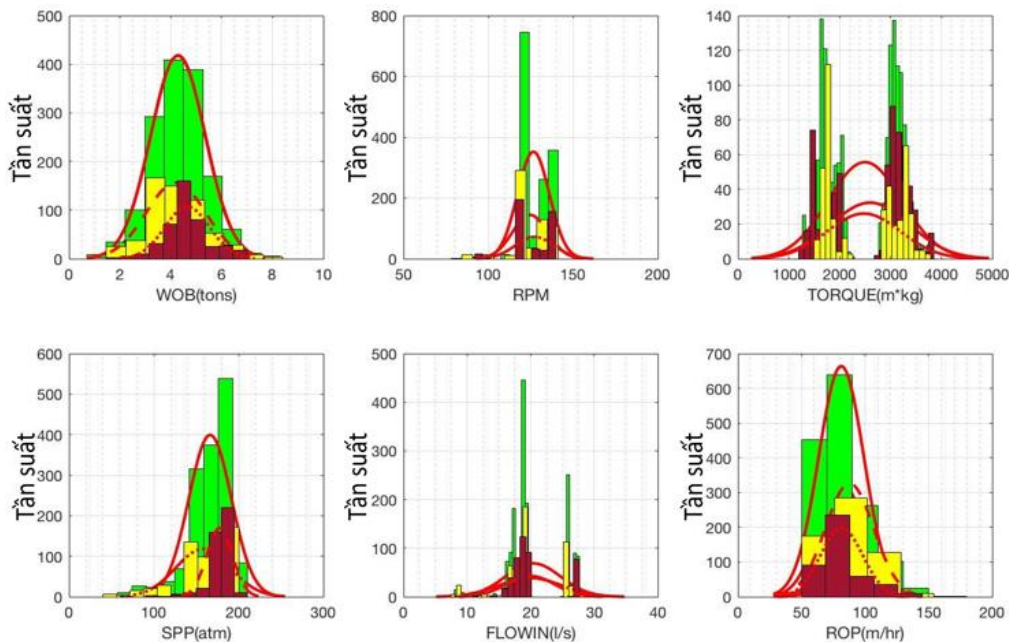
Ngoài ra, việc sử dụng quá nhiều thông số đầu vào cũng có thể làm chậm quá trình huấn luyện, việc có quá nhiều các dữ liệu rác cũng sẽ dẫn đến độ sai lệch lớn của kết quả đầu ra. Do đó trước khi thiết lập mô hình dự báo thạch học, cần có công đoạn phân tích số liệu để lựa chọn các thông số chế độ khoan phù hợp để sử dụng làm tập huấn luyện cho mô hình (Hình 4).

Hình 4 trình bày biểu đồ tần suất của 06 thông số chế độ khoan, những biểu đồ này cho ta thông tin trực quan về đặc tính phân bố của tập tài liệu. Tọa độ ngang (trục X) minh họa phạm vi giá trị của các thông số khoan, tọa độ dọc (trục Y) mô tả tần số điểm dữ liệu trong một phạm vi. Trong biểu đồ, trục y định lượng số lần lặp lại của các phạm vi giá trị cụ thể trong tập dữ liệu. Việc trực quan hóa toàn diện các thông số này đóng vai trò quan trọng trong việc lựa chọn được bộ thông số phù hợp cho việc đào tạo mô hình dự báo. Có thể thấy ở cả 06 thông số chế độ khoan, biểu đồ tần suất đều có dạng đường cong phân phối chuẩn hình chuông thường thấy trong xác suất thống kê, điều này cho thấy trạng thái ổn định của bộ dữ liệu, không xuất hiện những giá trị ngoại lai có thể làm giảm độ chính xác của kết quả dự báo (Choudhury, 2014). Ở mỗi biểu đồ tần suất, tập giá trị đã được phân thành 03 nhóm màu đại diện cho 03 nhóm thạch học riêng biệt: bột kết, cát kết và sét. Để dễ dàng nhận thấy ở cả 06 thông số chế độ khoan, khoảng phân bố giá trị của các nhóm thạch học đều có dấu hiệu

bao trùm lên nhau, không có nhóm thạch học nào có khoảng giá trị tách biệt hoàn toàn; điều này khẳng định rằng không thể phân biệt các nhóm thạch học dựa trên một thông số riêng biệt nào. Do đó, việc kết hợp tất cả các thông số này để tạo nên bộ dữ liệu huấn luyện mô hình dự báo thạch học là cần thiết và hợp lý.

3.2. Phương pháp nghiên cứu và kết quả

Mục tiêu đặt ra của nhóm tác giả là thử áp dụng và đánh giá hiệu quả kỹ thuật học máy khi xử lý bài toán dự báo các tầng thạch học trực tiếp trong khi khoan các giếng khoan tại mỏ Cá Tầm. Với bộ tài liệu bao gồm 06 thông số chế độ khoan cùng thông tin phân bố các nhóm thạch học (Bột kết, cát kết và sét) theo chiều sâu thẳng đứng từ tài liệu master log của 2 giếng A và B, đây có thể được xếp vào bài toán phân loại (classification), là một dạng bài toán điển hình có thể được xử lý hiệu quả bởi kỹ thuật học máy có giám sát (supervised machine learning). Để giải quyết nhiệm vụ đề ra, nhóm tác giả áp dụng thuật toán fuzzy c-means để phân chia theo chiều sâu giếng thành 3 nhóm khác nhau dựa vào sự tương đồng về tính chất của các thông số chế độ khoan, sau đó sẽ gán nhãn 3 nhóm này tương ứng với 3 nhóm thạch học xuất hiện trong tài liệu master log của 2 giếng A và B (Bột kết, cát kết và sét) và đánh giá độ trùng khớp với tài liệu mẫu (Bezdek và nnk., 1984).



Hình 4. Biểu đồ tần suất (histogram) của 06 thông số chế độ khoan của 2 giếng A, B.

Ưu điểm của phương pháp so với thuật toán K-means là khi áp dụng thuật toán fuzzy c-means các đối tượng sẽ được xếp vào nhiều nhóm với hệ số khác nhau. Quy trình phân nhóm này được mô tả bởi phương trình dưới đây:

$$\Phi_{FCM} = \sum_{j=1}^N \cdot \sum_{k=1}^C \cdot u_{jk}^q \cdot \|z_j - v_k\|_2^2 \quad (1)$$

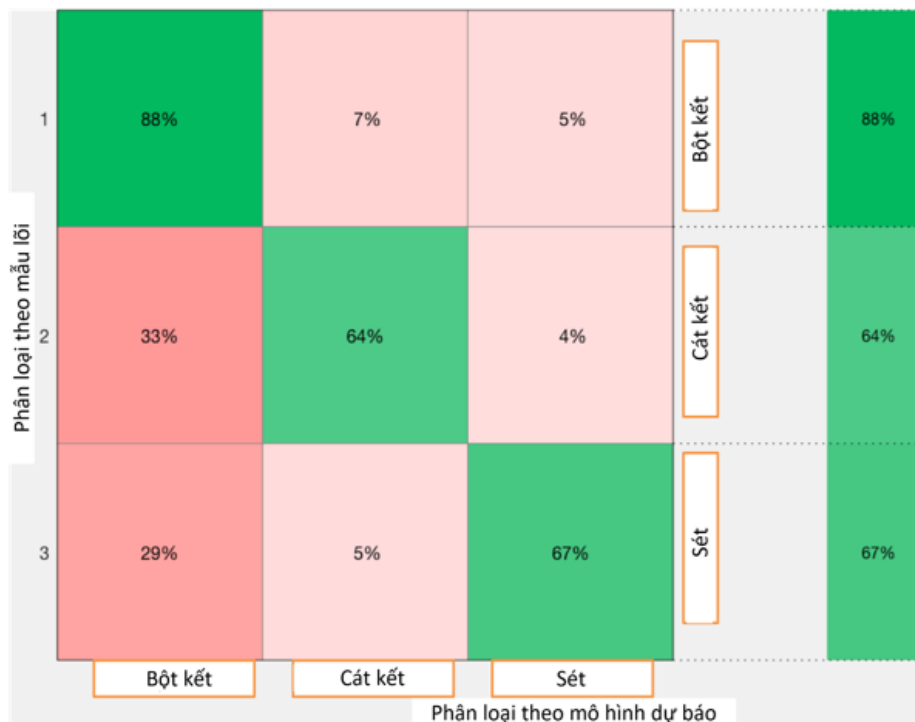
Trong đó: N - số điểm của tập dữ liệu z; C - số nhóm; q (q>1) - tham số 'mờ' (fuzziness) thường được chọn bằng 2; v_k - giá trị trung tâm của nhóm thứ k, u_{jk}^q - hệ số thành viên của phần tử thứ j đối với nhóm thứ k.

Hình 5 biểu diễn kết quả dự báo và phân nhóm thạch học từ bộ dữ liệu thông số chế độ khoan khi sử dụng thuật toán fuzzy c-means; kết quả dự báo có độ trùng khớp tổng khi so với tài liệu mẫu là 78% và độ trùng khớp của từng nhóm thạch học lần lượt là bột kết (88%); cát kết (64%) và sét (67%). Có thể nhận thấy mô hình này có khả năng nhận diện chính xác các tập bột kết khi chỉ dự báo nhầm lẫn 7% nhóm này với cát kết và 5% với sét. Ở 2 nhóm thạch học cát kết và sét, độ chính xác khi dự báo dùng lại lần lượt ở ngưỡng 64% và 67%, sai số trong dự báo 2 loại thạch học này chủ yếu là nhầm lẫn với bột kết (lần lượt là 33% và 29%); dưới đây là một số nguyên nhân có thể dẫn đến sai số trên:

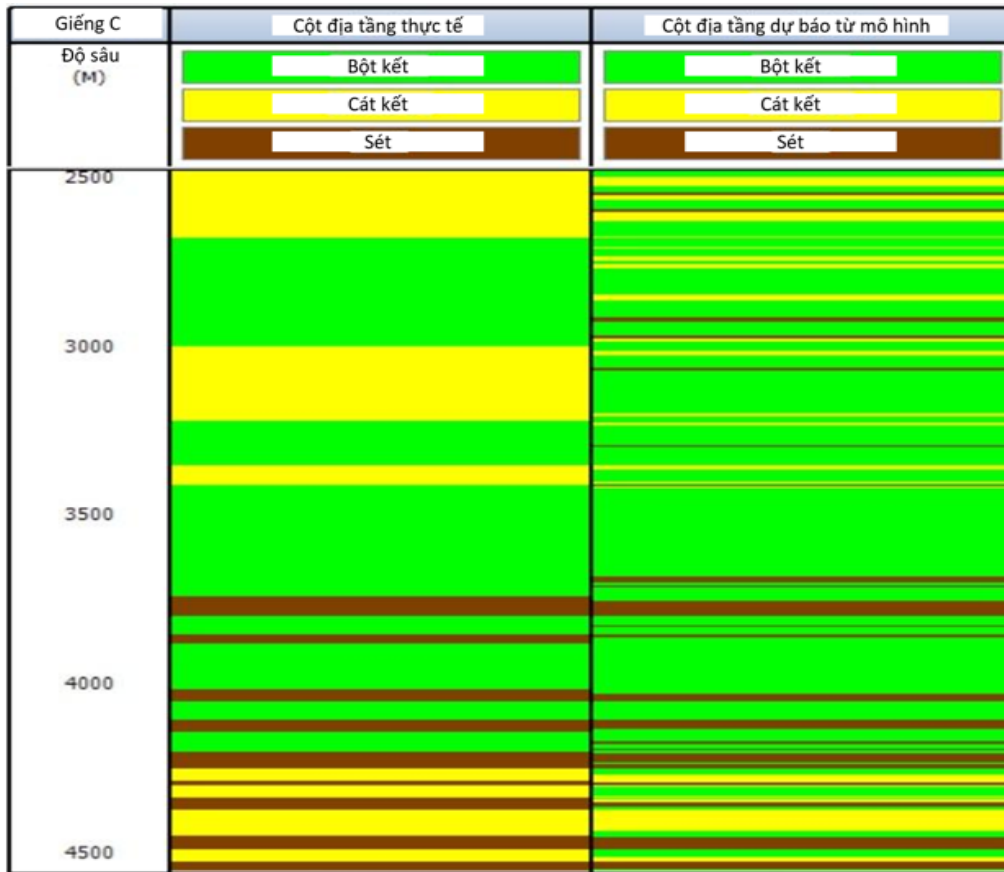
- Trong bộ tài liệu mẫu bao gồm 2510 điểm có tỷ lệ mẫu bột kết là vượt trội so với 2 nhóm thạch học còn lại nên xác suất để mô hình nhận diện đối tượng là bột kết sẽ cao hơn.

- Cơ sở dữ liệu dùng để huấn luyện gồm 06 thông số chế độ khoan, giá trị các thông số này bị chi phối chủ yếu bởi độ cứng và độ gắn kết của các tầng đất đá do đó các nhóm thạch học có độ cứng và độ gắn kết càng tương đồng nhau thì khả năng mô hình dự báo nhầm lẫn giữa chúng sẽ càng cao. Trong khi đó, cát kết có độ cứng từ 6÷7, sét có độ cứng từ 2÷3 còn bột kết có độ cứng nằm ở giữa, trong khoảng từ 3,5÷4; do đó mô hình dự báo sẽ có xu hướng dự báo nhầm 2 nhóm thạch học cát kết và sét thành bột kết.

Để đánh giá khắt khe hơn độ chuẩn xác và ưu việt của phương pháp dự báo, ngoài việc ghi nhận độ trùng khớp giữa kết quả dự báo khi so sánh với tài liệu 02 giếng A và B; mô hình này còn được dùng để phân loại các tầng đất đá cho 01 giếng khoan khác trong mỏ Cá Tầm là giếng C (tài liệu khoan của giếng này chưa được dùng làm dữ liệu huấn luyện). Hình 6 cho thấy độ trùng khớp giữa cột thạch học dự báo bởi mô hình và cột thạch học thực tế của giếng khoan C là 66,3%. Mô hình xác định được độ sâu các lớp bột kết tương đối trùng khớp với thực tế, sai số chủ yếu chỉ xảy ra ở 2 nhóm thạch học cát kết và sét với cùng lý do như đã trình bày bên trên.



Hình 5. Kết quả mô hình dự báo thạch học.



Hình 6. So sánh cột thạch học dự báo bởi mô hình và cột thạch học thực tế giếng C.

4. Kết luận

Qua kết quả thu được, nghiên cứu này đã tiếp cận một phương pháp mới giải quyết nhiệm vụ xác định các tầng đất đá trực tiếp từ tài liệu khoan theo thời gian thực nhằm nâng cao hiệu quả công tác khoan cho các giếng khoan dầu khí. Mô hình xây dựng được từ nhóm tác giả có khả năng dự báo thạch học bằng cách sử dụng thuật toán fuzzy c-means với tập huấn luyện là bộ tài liệu bao gồm 06 thông số chế độ khoan của 02 giếng A và B khu vực mỏ Cá Tầm. Mô hình đưa ra kết quả phân nhóm thạch học có độ chính xác tương đối cao (78%) khi so sánh với với cột thạch học thực tế của 02 giếng A và B. Đáng chú ý khi áp dụng mô hình này để xây dựng cột thạch học từ tài liệu khoan của 01 giếng khoan khác trong khu vực nghiên cứu, kết quả thu được cũng là rất khả quan (độ chính xác 66,3 %). Sự sai lệch trong kết quả dự báo có thể được lý giải do các tầng đất đá khu vực mỏ Cá Tầm có độ cứng và độ gắn kết khá tương đồng, các yếu tố này quyết định giá trị đo được của

các thông số khoan (đây chính là tập dữ liệu để huấn luyện mô hình dự báo). Dựa trên các kết quả bước đầu, nhóm tác giả có một số kết luận và đề xuất như sau:

- Công tác xác định, nhận diện ranh giới của các tầng thạch học trong khi khoan theo thời gian thực là nhiệm vụ thiết thực và quan trọng, đòi hỏi có các hướng tiếp cận mới để cải thiện hiệu quả và nâng cao tính chính xác.

- Qua bài báo, có thể thấy việc áp dụng các thuật toán học máy vào giải quyết nhiệm vụ dự báo và nhận diện các tầng thạch học từ tài liệu khoan đem lại kết quả chính xác và rất hứa hẹn, hướng đi này cần được tập trung đầu tư và có những nghiên cứu chi tiết hơn.

- Để khắc phục sai số và nâng cao hiệu suất mô hình dự báo, cần bổ sung thêm các thông tin khác cho tập huấn luyện, ví dụ như các đường cong đo địa vật lý ngay trong khi khoan. Các đường cong này phản ánh thông tin về tính chất vật lý liên quan chủ yếu đến thành phần khoáng vật chứ không chịu ảnh hưởng nhiều bởi giá trị độ cứng và

độ gắn kết của các tầng đất đá giống như các thông số chế độ khoan.

Lời cảm ơn

Bài báo này là sản phẩm của đề tài KHCN mã số T24-11 do Trường Đại học Mỏ - Địa chất tài trợ kinh phí.

Đóng góp của tác giả

Vũ Hồng Dương - nghiên cứu phương pháp luận, phân tích dữ liệu, điều tra, khảo sát và viết bản thảo bài báo; Nguyễn Tiến Hùng và Nguyễn Thế Vinh - hỗ trợ phần phương pháp luận, đánh giá và chỉnh sửa; Nguyễn Anh Tuấn - thu thập, thống kê dữ liệu.

Tài liệu tham khảo

- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, Volume 10, Issues 2-3, 191-203.
- Chen, G. (2020). Study on Real-time Lithology Identification Method of Logging-while-drilling. *IOP Conf. Series: Earth and Environmental Science*, 546, 052007 doi:10.1088/1755-1315/546/5/052007
- Choudhury, A. (2014). A Simple Approximation to the Area Under Standard Normal Curve. *Mathematics and Statistics*, 2, 147-149.
- Jian, S., Qi, L., Mingqiang, C., Long, R., Guihua, H., Chenyang, L., & Zixuan, Z. (2019). Optimization of models for a rapid identification of lithology while drilling - A win-win strategy based on machine learning. *Journal of Petroleum Science and Engineering*, 176, 321-341. <https://doi.org/10.1016/j.petrol.2019.01.006>
- Khalifa, H., Tomomewo, O. S., Ndulue, U. F., & Berrehal, B. E. (2023). Machine Learning-Based Real-Time Prediction of Formation Lithology and Tops Using Drilling Parameters with a Web App Integration. *Eng 2023*, 13, 2443-2467. <https://doi.org/10.3390/eng4030139>.
- Li, T. T., Tong, J., Xiang, R. Y., & Yan, X. D. (2022). Research on Intelligent Lithology Identification Method Based on Real-Time Data of Drilling Wells. *RICAI 2022, Dongguan, China*, 890-895.
- Mikkel, L. A., John-Morten, G., Ole, M. A. (2022). Classification of Drilled Lithology in Real-Time Using Deep Learning with Online Calibration. *SPE Drilling & Completion*, 26-37.
- Moazzeni, A., Haffar, M. A. (2015). Artificial Intelligence for Lithology Identification through Real-Time Drilling Data. *Earth Sci Clim Change*, 6(3), 265. doi:10.4172/2157-7617.1000265.
- Nguyễn, X. N. (2009). Chương trình phân tích thạch học theo tài liệu Địa vật lý giếng khoan. *Tạp chí Khoa học và Công nghệ*, 12, 6.
- Romy, A., Aashish, M., Robello, S., & Amit, S. (2022). Real-Time Prediction of Litho- Facies From Drilling Data Using an Artificial Neural Network: A Comparative Field Data Study With Optimizing Algorithms. *Journal of Energy Resources Technology*, Vol. 144, 043003.
- Trần, T. T. T., Nguyễn, T. T., Nguyễn, H. A., Lê, M. H., Nguyễn, T. A., & Trần, X. Q. (2021). Đặc trưng vật lý, thạch học của đá chứa Carbonate tuổi Devonian mỏ Bắc Oshkhotynskoye, Liên Bang Nga. *Tạp chí Dầu khí*, 3, 11-21.
- Trần, T. T. T., Nguyễn, T. T., Nguyễn, T. T., Đỗ, Q. Đ., Nguyễn, H. A., & Nguyễn, T. T. T. (2019). Đặc trưng vật lý, thạch học của đá chứa Pliocene khu vực trung tâm bể Sông Hồng. *Tạp chí Dầu khí*, 8, 21-28.
- Vũ, H. D., & Kiều, D. T. (2020). Phân loại thạch học từ các tham số vật lý trong tài liệu giếng khoan 1143, chương trình khoan đại dương tại biển đông bằng mạng trí tuệ nhân tạo. *Báo cáo khoa học tại Hội nghị toàn quốc Khoa học Trái đất và Tài nguyên với Phát triển bền vững (ERSD2020)*, 113-116.
- Yunxin, X., Chenyang, Z., Wen, Z., Zhongdong, L., Xuan, L., & Mei, T. (2018). Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering*, 160, 182-193. <https://doi.org/10.1016/j.petrol.2017.10.028>.