

NGHIÊN CỨU ÁP DỤNG CÁC THUẬT TOÁN KHAI PHÁ DỮ LIỆU KHÔNG GIÁM SÁT ĐỂ LỰA CHỌN GIẢI PHÁP NÂNG CAO HỆ SỐ THU HỒI DẦU CHO CÁC MỎ DẦU ĐANG SUY GIẢM SẢN LƯỢNG

Phạm Quý Ngọc, Đoàn Huy Hiền, Hoàng Long

Viện Dầu khí Việt Nam

Email: ngocpq@vpi.pvn.vn

<https://doi.org/10.47800/PVJ.2020.12-04>

Tóm tắt

Phương pháp nâng cao thu hồi dầu (Enhanced Oil Recovery, EOR) là giải pháp có thể gia tăng sản lượng dầu thu hồi, đặc biệt cho các mỏ dầu có sản lượng khai thác bắt đầu suy giảm và độ ngập nước tăng. Nghiên cứu này tổng hợp nhiều dự án EOR đã áp dụng thành công trên thế giới và áp dụng phương pháp phân tích nâng cao (như phân tích thành phần chính (PCA) và kỹ thuật phân cụm K-means để “học kinh nghiệm” từ các dự án này), từ đó tìm ra các tiêu chí và giải pháp EOR phù hợp cho các mỏ dầu đang suy giảm tại Việt Nam.

Từ khóa: Nâng cao thu hồi dầu, phân tích dữ liệu, thuật toán nâng cao, PCA, K-means.

1. Giới thiệu

Các mỏ dầu đang hoạt động ở thềm lục địa Việt Nam đều được phát triển và khai thác trong đá chứa cát kết thuộc địa tầng Miocene, Oligocene và đá móng phong hóa nứt nẻ trước Đệ tam. Các mỏ dầu này chủ yếu được thiết kế khai thác ban đầu theo chế độ năng lượng tự nhiên đàn hồi của dầu và khí hòa tan. Để nâng cao hệ số thu hồi dầu, giải pháp bơm ép nước duy trì áp suất vỉa ở cả 2 đối tượng cát kết Miocene, Oligocene và đặc biệt móng nứt nẻ trước Đệ tam đã được áp dụng. Công nghệ khai thác thứ cấp - bơm ép nước - cho phép gia tăng đáng kể thu hồi dầu so với khai thác sơ cấp đơn thuần - chỉ sử dụng năng lượng tự nhiên của vỉa.

Phương pháp nâng cao thu hồi dầu có thể áp dụng ngay từ giai đoạn đầu đối với các mỏ có hệ số quét tự nhiên thấp, không đủ tạo ra dòng chảy đến giếng khai thác. Bơm ép khí, bơm ép hóa chất hay gia nhiệt... là phương pháp cơ bản được thế giới áp dụng. Ở Việt Nam, các giải pháp nâng cao thu hồi dầu bằng bơm ép thử nghiệm chất hoạt động bề mặt, vi sinh, hóa lý đã được tiến hành cho đối tượng cát kết mỏ Bạch Hổ.

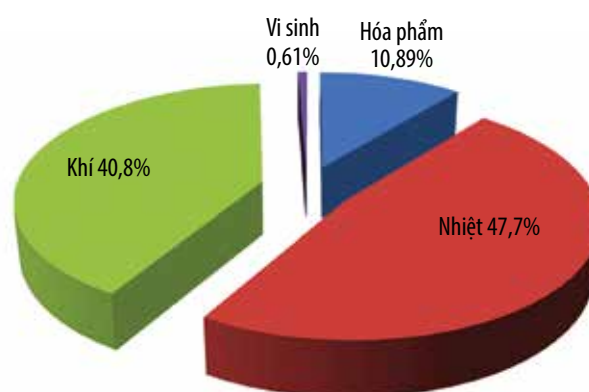
Để lựa chọn được phương pháp áp dụng và công nghệ phù hợp mang lại hiệu quả kinh tế cao nhất cần

phân tích các dự án EOR đã áp dụng thành công trên thế giới.

Bài báo giới thiệu kết quả nghiên cứu áp dụng thuật toán khai phá dữ liệu không giám sát - gồm kỹ thuật phân tích thành phần chính (Principal Component Analysis, PCA) và kỹ thuật phân cụm dựa vào trị số trung bình (K-means) - vào các dự án EOR thành công trên thế giới để tìm ra các phương pháp EOR phù hợp cho các đối tượng mỏ khác nhau tại Việt Nam. Nghiên cứu này phát triển các module tính toán bằng ngôn ngữ lập trình mã nguồn mở R và Python.

2. Tài liệu và phương pháp nghiên cứu

Dữ liệu phân tích trong nghiên cứu được thu thập từ báo cáo khảo sát các phương pháp EOR (EOR survey) đã



Hình 1. Tỷ lệ các dự án đã áp dụng EOR thành công trên thế giới



Ngày nhận bài: 19/11/2020. Ngày phản biện đánh giá và sửa chữa: 19 - 24/11/2020.

Ngày bài báo được duyệt đăng: 15/12/2020.

áp dụng thành công trên thế giới từ năm 1998 - 2014 của "The Oil and Gas Journal" [1 - 8] và các công trình nghiên cứu các dự án EOR trên thế giới lưu trữ tại cơ sở dữ liệu của Hiệp hội Kỹ sư Dầu khí (SPE) OnePetro [9 - 11]. Các tài liệu trên đã thống kê các thông số của mỏ/vĩa chứa với tổng cộng 25 trường dữ liệu như: vị trí địa lý, loại vỉa chứa, độ sâu, chiều dày vỉa, số lượng giếng khai thác, bơm ép, thông số địa chất mỏ, thông số vật lý của chất lưu, độ bão hòa, sản lượng khai thác, sản lượng gia tăng và các thông số khác của hơn 1.000 vỉa chứa.

Hình 1 mô tả tỷ lệ các phương pháp EOR đã áp dụng thành công trên thế giới. Trong đó 3 phương pháp chính là phương pháp gia nhiệt, bơm ép khí và bơm ép hóa phẩm, còn lại là phương pháp vi sinh.

Các phương pháp gia nhiệt, gồm đốt cháy vỉa (combustion) và bơm ép hơi nước nóng để gia nhiệt cho vỉa chứa là phương pháp cho sản lượng gia tăng nhiều nhất. Tuy nhiên, nghiên cứu này không đề cập phương pháp gia nhiệt vì không phù hợp với điều kiện ở Việt Nam: các vỉa chứa có độ sâu khai thác lớn nên nhiệt độ vỉa khá cao, từ 80 - 120 °C với trầm tích lục nguyên. Phương pháp này chỉ phù hợp với các mỏ nông (độ sâu khai thác nhỏ) và ở các vùng lạnh, độ nhớt dầu vỉa cao.

Hình 2 thể hiện sản lượng dầu gia tăng khi áp dụng các phương pháp EOR.

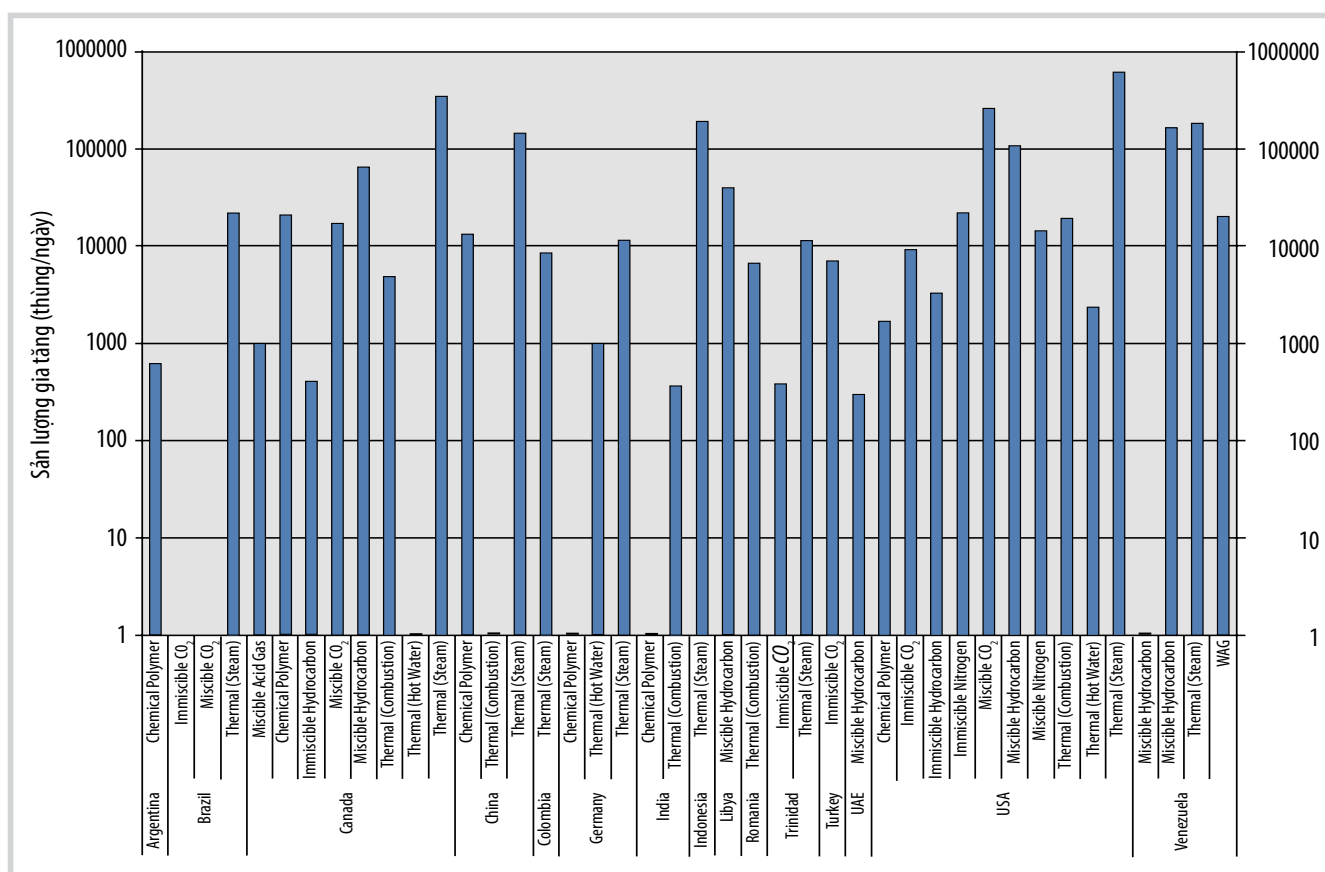
Trong số các phương pháp phân tích dữ liệu bậc cao, phương pháp phân tích thành phần chính PCA và phân nhóm K-means có thể áp dụng cho cơ sở dữ liệu lớn và cấu trúc phức tạp, đáp ứng được yêu cầu đặt ra.

3. Phương pháp phân tích và xử lý số liệu

3.1. Phương pháp phân tích thành phần chính (PCA)

Phân tích thành phần chính PCA là kỹ thuật thường được sử dụng khi làm việc với các cơ sở dữ liệu nhiều chiều, được sử dụng để xác định một không gian mới với số chiều nhỏ hơn không gian gốc, thường là 2 hoặc 3 chiều để có thể dễ dàng trực quan hóa số liệu. Các trục tọa độ của không gian mới sẽ được xây dựng sao cho độ biến thiên trên mỗi trục tương ứng với phương sai của dữ liệu trên đó là lớn nhất có thể.

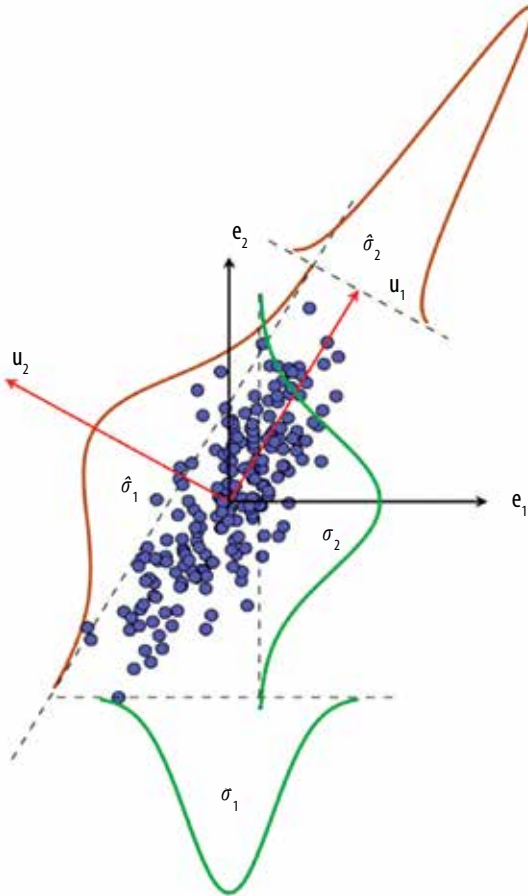
Các mục đích chính khi áp dụng kỹ thuật PCA gồm: (1) Giảm số chiều của dữ liệu; (2) Trực quan hóa dữ liệu với các dataset có nhiều chiều thông tin; (3) Chuyển đổi dataset ban đầu với số chiều (biến) lớn thành dataset mới số chiều ít hơn nhưng vẫn đảm bảo giữ lại nhiều thông tin nhất,



Hình 2. Sản lượng dầu gia tăng của các nước tương ứng với các phương pháp áp dụng

không ảnh hưởng đến độ chính xác của các mô hình dự báo; (4) Xây dựng các thành phần chính mới bằng cách tổ hợp tuyến tính các biến ban đầu; (5) Các thành phần chính không có mối tương quan tuyến tính với nhau; (6) Khám phá những thông tin mới mà có thể bị các chiều thông tin cũ che mất.

Về mặt hình học, PCA là phương pháp đi tìm hệ cơ sở trực chuẩn bằng một phép xoay sao cho trong hệ mới này



Hình 3. Không gian dữ liệu gốc (e_1, e_2) với phương sai tương ứng là σ_1 và σ_2 ; không gian mới (u_1, u_2) với phương sai tương ứng là σ_1' và σ_2' [12, 13]

phương sai theo một chiều nào đó là rất nhỏ, có thể bỏ qua được mà không ảnh hưởng nhiều đến thông tin dữ liệu (Hình 3).

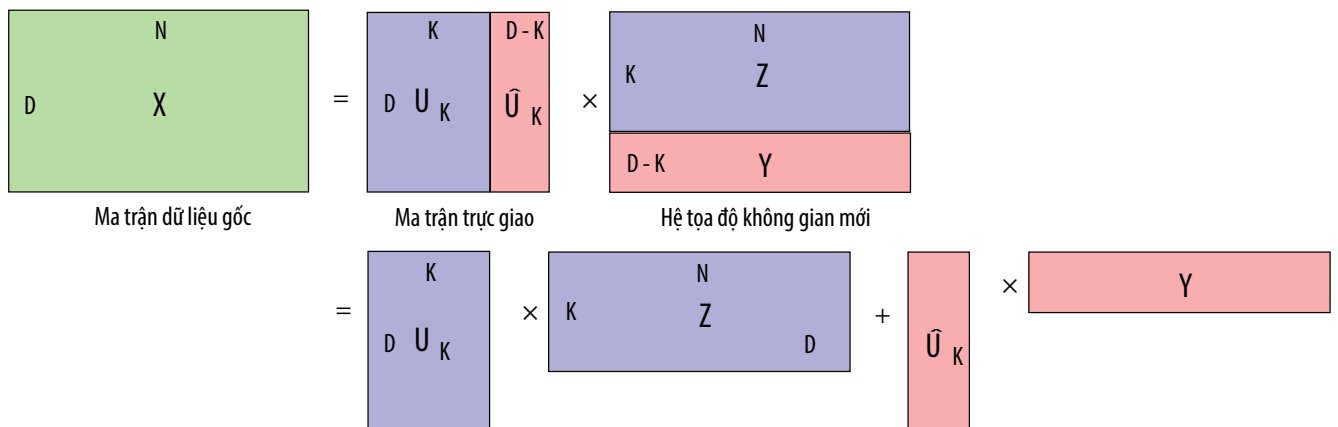
Ma trận dữ liệu gốc X có D hàng và N cột tương ứng với D đối tượng (quan sát) và N trường dữ liệu (biến). Phép xoay trực liên hệ trực quan với hệ trục chuẩn và ma trận trực giao như mô tả ở Hình 4. Mục đích của PCA là đi tìm ma trận trực giao U sao cho phần lớn thông tin được giữ lại ở phần màu tím UKZ và phần màu đỏ UKY sẽ được lược bỏ và thay vào bằng một ma trận không phụ thuộc vào từng điểm dữ liệu.

Phương pháp PCA được áp dụng để giảm số chiều của tập dữ liệu gồm một số lượng lớn của các biến liên quan. Các biến kém quan trọng được giảm bớt bằng cách chuyển đổi dữ liệu vào không gian mới của các biến, các thành phần chủ yếu (principal component, PC) trực giao nhau (không tương quan) và được sắp xếp theo thứ tự tầm quan trọng giảm dần.

Kỹ thuật PCA xác định được các giá trị riêng và phương sai từ ma trận tương quan của các biến ban đầu. Thành phần chính là các biến không tương quan, thu được bằng cách nhân các biến tương quan ban đầu với hệ số tải nhân tố. Vì vậy, các thành phần chính được kết hợp tuyến tính của các biến ban đầu. PC cung cấp thông tin về các thông số có ý nghĩa nhất, trong đó mô tả toàn bộ dữ liệu thiết lập dựng hình với việc giảm thiểu các thông tin ban đầu. Đây là kỹ thuật mạnh áp dụng cho mô hình giải thích sự thay đổi của một tập lớn các tương quan biến và chuyển đổi thành một tập hợp nhỏ hơn của các biến độc lập (thành phần chính).

3.2. Phương pháp phân cụm K-means

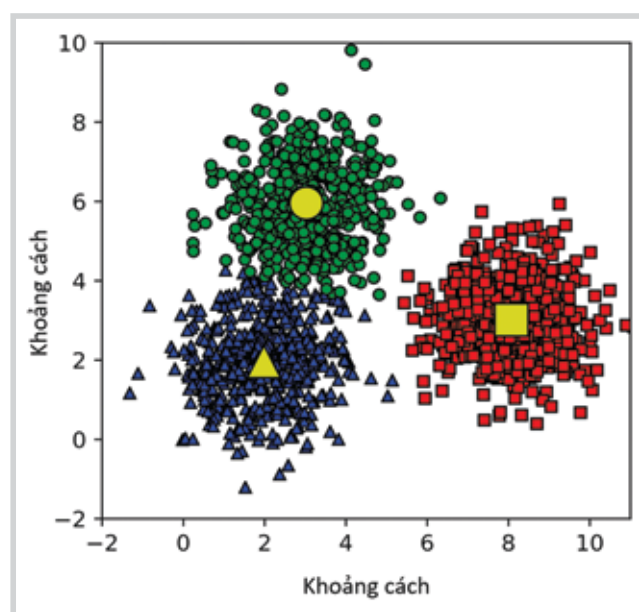
Trong phương pháp K-means, nhãn của từng điểm dữ liệu là không xác định. Phương pháp này sẽ phân tách



Hình 4. Kỹ thuật biến đổi ma trận trong PCA để xác định một không gian mới với các thành phần quan trọng nhất là K thành phần đầu tiên của ma trận [12]

dữ liệu thành các cụm (cluster) khác nhau sao cho dữ liệu trong cùng một cụm có những tính chất giống nhau.

Một định nghĩa đơn giản của nhóm/cụm là tập hợp các điểm có các vector đặc trưng gần nhau. Việc đo khoảng cách giữa các vector thường được thực hiện dựa trên khoảng cách giữa các vector trong một không gian xác định, trong đó khoảng cách Euclidean, tức khoảng cách trong không gian 2 chiều được sử dụng phổ biến hơn cả.



Hình 5. Phân tách nhóm K-means thành 3 cụm dữ liệu trong không gian 2 chiều [14]

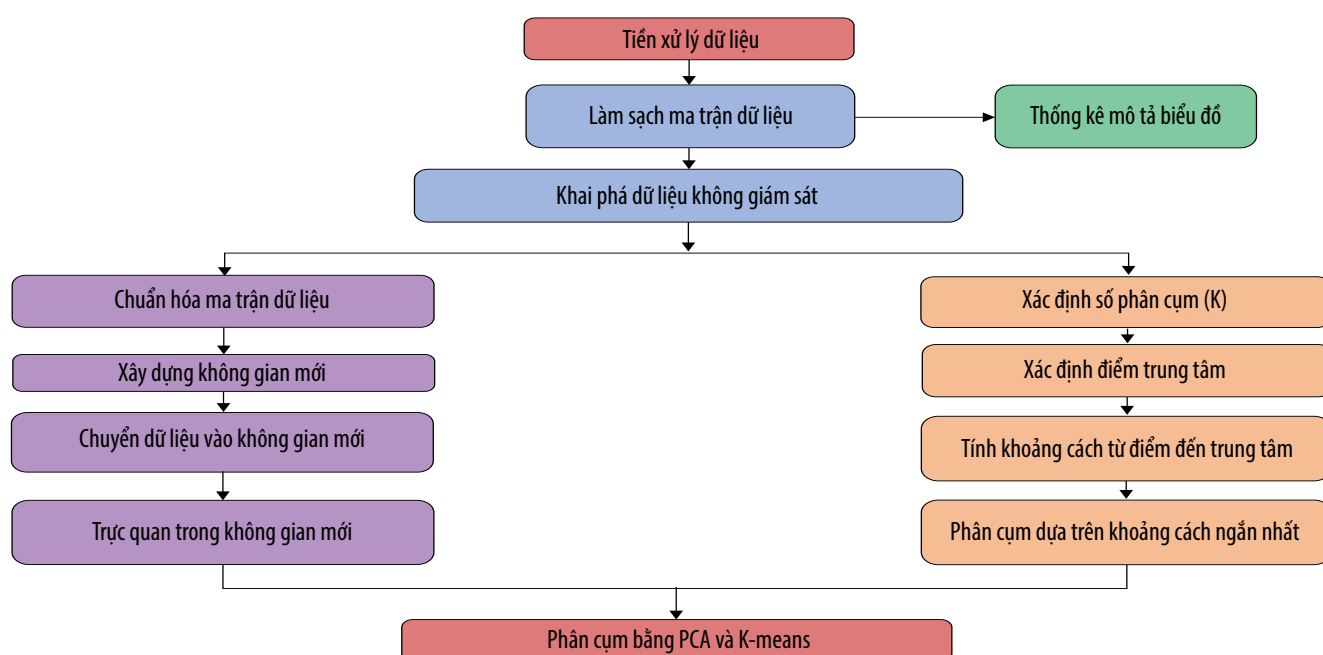
Hình 5 là ví dụ về dữ liệu được phân tách vào 3 nhóm. Giả sử mỗi nhóm có một điểm tâm nhóm đại diện (centroid) màu vàng và nhóm của mỗi điểm được xác định qua việc gần với điểm đại diện nào nhất trong 3 điểm.

K-means là thuật toán quan trọng và được sử dụng phổ biến trong kỹ thuật phân cụm. Mục đích chính của thuật toán K-means là tìm cách phân nhóm các đối tượng đã cho vào K cụm (K là số các cụm được xác định trước, K nguyên dương) sao cho tổng bình phương khoảng cách giữa các đối tượng đến tâm nhóm là nhỏ nhất. Khoảng cách giữa các điểm thường dùng là Euclidean, từ các điểm tới tâm có thể dùng khoảng cách Manhattan.

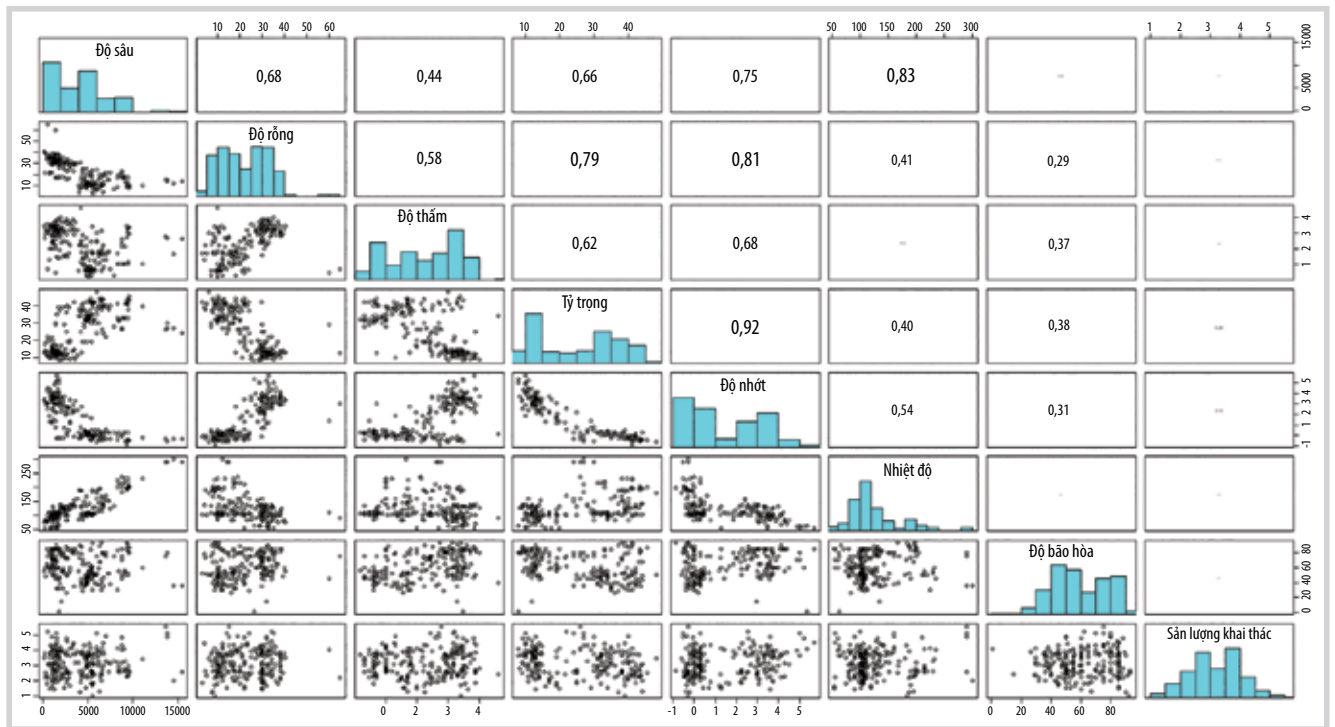
$$d(i, j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)} \quad (1)$$

Phương trình (1) là khoảng cách Minkowski tổng quát trong đó: d là khoảng cách giữa 2 điểm; x_i, x_j là 2 điểm cần tính khoảng cách; khi $q = 1$, khoảng cách Minkowski trở thành khoảng cách Manhattan và $q = 2$ là khoảng cách Euclidean.

Thuật toán K-means được thực hiện qua các bước chính sau: (1) Chọn ngẫu nhiên K tâm (centroid) cho K cụm trong đó mỗi cụm được đại diện bằng các tâm của cụm; (2) Tính khoảng cách giữa các đối tượng đến K tâm; (3) Phân tách các đối tượng vào cụm gần nhất; (4) Xác định lại tâm mới cho các cụm; (5) Thực hiện lại bước 2 cho đến khi không có sự thay đổi cụm nào của các đối tượng.



Hình 6. Sơ đồ khai phá dữ liệu áp dụng thuật toán PCA và K-means



Hình 7. Phân bố các tham số chính và quan hệ tương quan: Depth (độ sâu), Por (độ rỗng), Perm (độ thấm), API (tỷ trọng), Visc (độ nhớt), Temp (nhiệt độ), Start_Sat (độ bão hòa) và T_Prod_log (Logarit sản lượng khai thác).

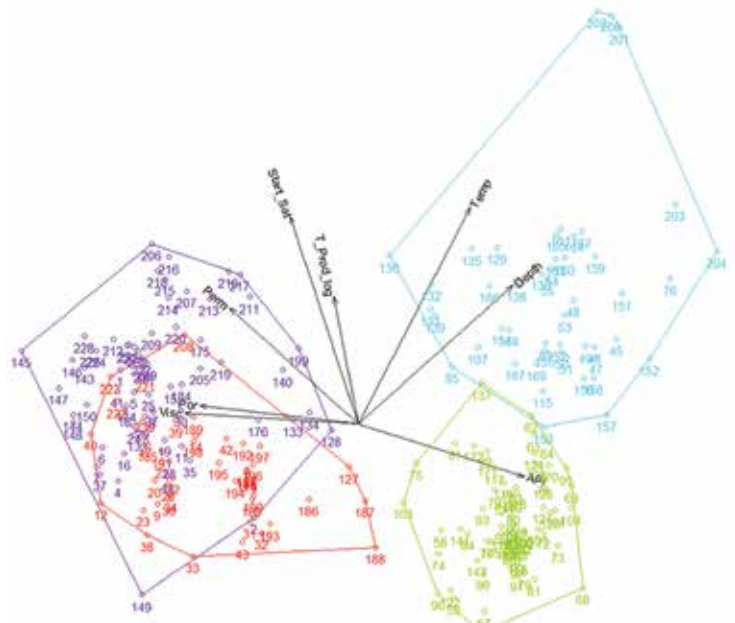
Sơ đồ các bước thực hiện khai phá dữ liệu áp dụng thuật toán PCA và K-means cho cơ sở dữ liệu EOR trên thế giới thể hiện như Hình 6.

4. Kết quả và thảo luận

Thống kê mô tả tương quan các tham số chính về bộ dữ liệu thu thập từ các mỏ áp dụng EOR thành công trên thế giới thể hiện ở Hình 7.

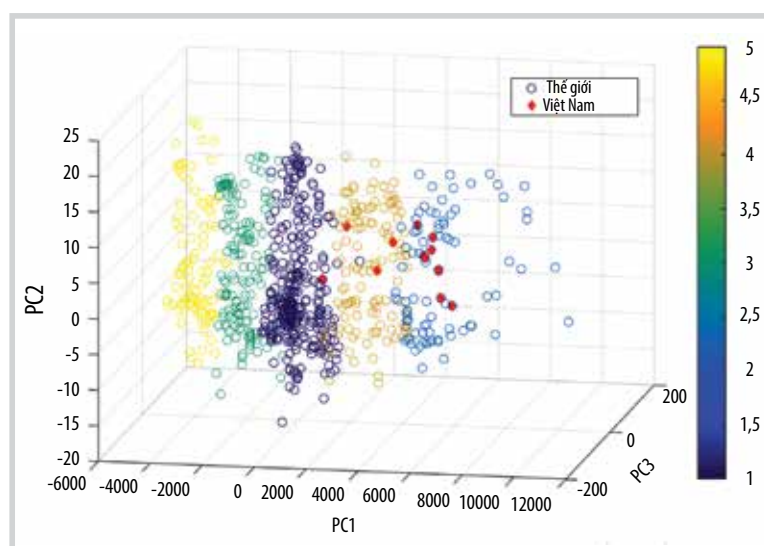
Phân bố các tham số độ sâu (Depth), độ rỗng (Por), độ thấm (Perm), tỷ trọng (API), độ nhớt (Visc), nhiệt độ (Temp), độ bão hòa (Start_Sat) và sản lượng khai thác (T_Prod_log) được thể hiện trên đường chéo của Hình 7. Phía dưới đường chéo là đồ thị mô tả quan hệ giữa các đại lượng và hệ số tương quan Pearson của các đại lượng biểu diễn ở các ô phía trên đường chéo tương ứng. Theo đó, độ nhớt có tương quan cao với tỷ trọng, hệ số tương quan là 0,92. Nhiệt độ cũng tương quan cao với độ sâu, hệ số là 0,83. Ngược lại, những đại lượng không tương quan với nhau, hệ số tương quan rất nhỏ, thì không nhìn thấy trên Hình 7.

Hình 8 biểu diễn trên không gian 2 chiều thành phần chính PC1 và PC2 với các trục là 8 biến tham số như mô tả ở trên và các cụm được phân nhóm bằng thuật toán K-means. Góc giữa

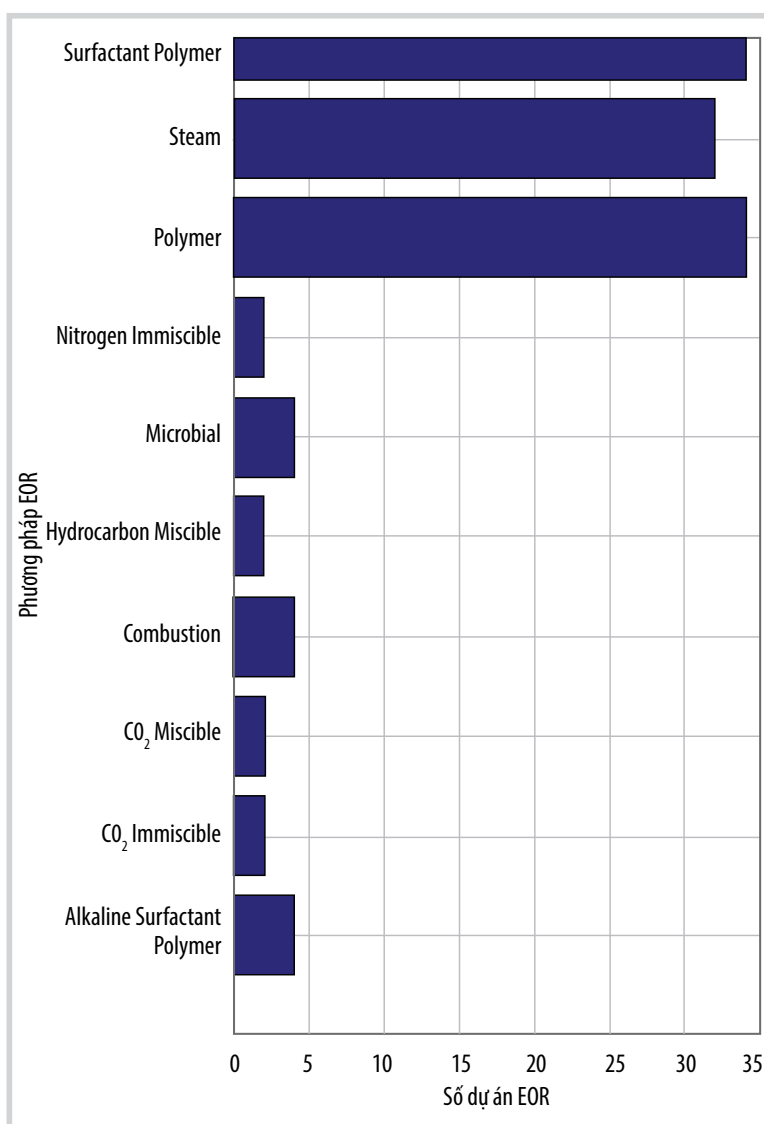


Hình 8. Trục quan dữ liệu trong không gian 2 chiều kết hợp 2 phương pháp PCA và K-means: Depth (độ sâu), Por (độ rỗng), Perm (độ thấm), API (tỷ trọng), Visc (độ nhớt), Temp (nhiệt độ), Start_Sat (độ bão hòa) và T_Prod_log (Logarit sản lượng khai thác).

các trục thể hiện mức độ tương quan giữa các biến. Trục độ nhớt (Visc) và trục độ rỗng (Por) có góc nhỏ và cùng chiều mũi tên, thể hiện 2 biến này tương quan cao và là tương quan thuận. Ngược lại trục độ tỷ trọng có chiều mũi tên ngược lại cho thấy tương quan cao với độ nhớt và độ rỗng nhưng là tương quan nghịch.



Hình 9. Áp dụng phương pháp PCA và K-means với các mỏ thế giới và Việt Nam (hình thoi màu đỏ) trong không gian 3 chiều (PC1, PC2 và PC3)



Hình 10. Các dự án EOR và công nghệ tương ứng đã áp dụng thành công trên thế giới ở phân nhóm 2

Các thông số và điều kiện của các mỏ ở Việt Nam đã được tích hợp vào ma trận dữ liệu và phân tích bằng phương pháp PCA và K-means. Hình 9 biểu diễn trực quan dữ liệu có tích hợp các mỏ ở Việt Nam trong không gian 3 chiều (PC1, PC2 và PC3). Hình 9 cho thấy các mỏ đang quan tâm tại Việt Nam (hình thoi màu đỏ) chủ yếu rơi vào phân nhóm 2 (màu xanh nước biển).

Biểu diễn các phương pháp áp dụng thành công và số lượng các dự án đã áp dụng thành công trong phân nhóm 2 như Hình 10. Phân tích nhóm 2 cho thấy hơn 100 dự án EOR trên thế giới đã áp dụng thành công phương pháp polymer hoạt tính bề mặt, polymer và hơi nước. Như vậy, dựa trên kinh nghiệm các dự án EOR thành công trên thế giới có thể nghiên cứu để áp dụng 1 trong 3 phương pháp trên ở Việt Nam, đặc biệt là phương pháp polymer hoạt tính bề mặt hoặc polymer.

5. Kết luận

Xác định các tiêu chí và phương pháp lựa chọn cho một dự án EOR mới là công việc rất quan trọng để có thể tìm ra được các giải pháp EOR phù hợp và khả thi đối với từng đối tượng. Nghiên cứu này đã thu thập các dự án EOR thành công trên thế giới và tính chất mỏ tương đồng với điều kiện địa chất - khai thác của Việt Nam. Việc áp dụng thuật toán PCA và K-means để học hỏi từ kinh nghiệm trên thế giới sẽ là tiền đề cho các nghiên cứu sản phẩm công nghệ và triển khai áp dụng thực tiễn cho các mỏ dầu khí đang khai thác tại Việt Nam.

Lời cảm ơn

Nhóm tác giả trân trọng cảm ơn Viện Dầu khí Việt Nam, Bộ Khoa học và Công nghệ hỗ trợ nguồn lực và tài trợ kinh phí thực hiện nghiên cứu này. Bài báo là một phần kết quả nghiên cứu của đề tài khoa học và công nghệ cấp Quốc gia "Nghiên cứu lựa chọn các giải pháp công nghệ và thực nghiệm đánh giá các tác nhân nâng cao hệ số thu hồi dầu cho đối tượng trầm tích lục nguyên của các mỏ dầu thuộc bể Cửu Long" mã số ĐTĐLCN.26/19.

Tài liệu tham khảo

- [1] Anonymous, "1996 worldwide EOR survey", *Oil & Gas Journal*, Vol. 94, No. 16, pp. 45, 1996.
- [2] Anonymous, "1998 worldwide EOR survey", *Oil & Gas Journal*, Vol. 96, No. 16, pp. 59 - 77, 1998.
- [3] Anonymous, "2002 worldwide EOR survey", *Oil & Gas Journal*, Vol. 100, No. 15, pp. 71, 2002.
- [4] Anonymous, "2004 worldwide EOR survey", *Oil & Gas Journal*, Vol. 102, No. 14, pp. 53 - 65, 2004.
- [5] Anonymous, "2006 worldwide EOR survey", *Oil & Gas Journal*, Vol. 104, No. 15, pp. 45 - 57, 2006.
- [6] L. Koottungal, "2010 worldwide EOR survey", *Oil & Gas Journal*, Vol. 108, No. 14, pp. 41 - 53, 2010.
- [7] L. Koottungal, "2012 Worldwide EOR Survey", *Oil & Gas Journal*, pp. 57 - 69, 2012.
- [8] L. Koottungal, "2014 worldwide EOR survey", *Oil and Gas Journal*, 2014.
- [9] J.J. Taber, "Technical screening guides for the enhanced recovery of oil", *SPE Annual Technical Conference and Exhibition, San Francisco, California 5 - 8 October 1983*. DOI: 10.2118/12069-MS.
- [10] J.J. Taber, F.D. Martin, and R.S. Seright, "EOR screening criteria revisited - Part 1: Introduction to screening criteria and enhanced recovery field projects", *SPE Reservoir Engineering*, Vol. 12, No. 3, pp. 189 - 198, 1997. DOI: 10.2118/35385-PA.
- [11] J.J. Taber, F.D. Martin, and R.S. Seright, "EOR screening criteria revisited - Part 2: Applications and impact of oil prices", *SPE Reservoir Engineering*, Vol. 12, No. 3, pp. 199 - 206, 1997. DOI: 10.2118/39234-PA.
- [12] T.M. Mitchell, *Machine learning*. New York: McGraw-Hill Education, 1997.
- [13] S. Misra, H. Li, and J. He, *Machine learning for subsurface characterization*. San Diego: Gulf Professional Publishing, 2019.
- [14] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*. Cambridge University Press, 2004.

APPLICATION OF UNSUPERVISED DATA MINING ALGORITHMS TO SELECT EOR SOLUTIONS FOR DEPLETED OILFIELDS

Pham Quy Ngoc, Doan Huy Hien, Hoang Long

Vietnam Petroleum Institute

Email: ngocpq@vpi.pvn.vn

Summary

Enhanced oil recovery (EOR) provides a solution to increase oil production, especially in cases where the reservoirs have high water cut and declining oil production rate. This study involves the collection of numerous successful EOR projects throughout the world and application of advanced data mining techniques such as principal component analysis (PCA) and K-means clustering to learn from the experiences of these projects, and on that basis find suitable criteria and EOR solutions for depleted oil fields in Vietnam.

Key words: EOR, data analysis, advanced algorithms, PCA, K-means.