

DỰ BÁO MỨC ĐỘ CHẤN ĐỘNG TRONG HOẠT ĐỘNG KHAI THÁC KHOÁNG SẢN VỚI TRÍ TUỆ NHÂN TẠO

Phạm Thị Thanh Loan, Đào Hiếu
Trường Đại học Mở - địa chất

Tóm tắt: Nổ mìn khai thác là phương pháp sử dụng năng lượng của thuốc nổ để phá vỡ cấu trúc vật chất phục vụ hoạt động xây dựng, khai thác khoáng sản. Sau khi phá vỡ các cấu trúc lân cận điểm nổ, phần năng lượng còn lại tạo thành sóng chấn động lan truyền, tác động tới môi trường xung quanh. Trong kỹ thuật nổ hiện đại, tổng lượng thuốc nổ được chia thành nhiều nhóm nhỏ và được kích nổ lần lượt sau những khoảng thời gian khác nhau. Nhờ đó, hiệu quả nổ được cải thiện đồng thời mức độ chấn động giảm. Quá trình đo mức độ chấn động là việc xác định vận tốc dao động của hạt đất đá tại điểm đo. Kết quả đo là cơ sở để đánh giá mức độ nguy hại tới các cấu trúc, công trình lân cận vùng khai thác. Trong bài báo này, tác giả đề xuất phương pháp phân tích dữ liệu để xác định yếu tố quan trọng tác động đến vận tốc dao động của hạt đất đá. Một số mô hình trí tuệ nhân tạo được phát triển để dự báo mức độ chấn động do nổ mìn áp dụng cho các mỏ lộ thiên tại Việt Nam. Dữ liệu được lấy trực tiếp từ cảm biến trong thời gian nổ thực nghiệm tại mỏ than Núi Béo, tỉnh Quảng Ninh

Từ khóa: Sóng chấn động, thời gian vi sai, tiền xử lý dữ liệu, Random Forest, XGBoots, KNN.

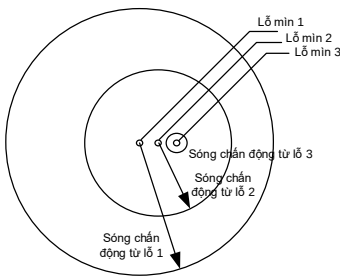
I. MỞ ĐẦU

A. Chấn động nổ mìn

Sóng chấn động nổ mìn sinh ra khi năng lượng nổ không còn đủ khả năng làm phá vỡ đất đá. Chấn động nổ mìn là một tác động tiêu cực có thể phá huỷ cấu trúc các công trình xung quanh khu vực nổ, gây nứt nẻ, đổ sập nhà cửa, mất an toàn ổn định tầng và bờ mỏ. Tốc độ di chuyển của sóng chấn động tại một khu vực nhất định gần như không đổi, nằm trong khoảng 305 – 6100 m/s. Sóng chấn động lan truyền đi từ mỗi điểm nổ là độc lập với nhau (Hình 1). Số lượng điểm nổ trong một vụ nổ càng nhiều thì lượng thuốc của một lần nổ càng nhỏ, kéo theo mức độ chấn động càng giảm. Do đó, mức độ chấn động chỉ phụ thuộc vào lượng thuốc nổ một lần nổ mà không phụ thuộc vào lượng thuốc của cả bãi [1][2].

Nổ mìn vi sai là quá trình thực hiện các đợt nổ một cách tuần tự giãn cách nhau bằng những khoảng thời gian trễ phù hợp tính bằng mili giây (ms), được gọi là thời gian vi sai. Các đợt nổ liên nhau giúp tạo mặt tự do mới, tạo sự giao thoa sóng ứng suất, cùng với hiện tượng đá văng, va

đập lẫn nhau làm tăng khả năng phá vỡ. Khoảng thời gian vi sai là một tham số quan trọng để tận dụng tối đa những hiệu ứng đó.



Hình 1. Sự lan truyền sóng ứng suất trong nổ mìn vi sai

Tham số này cần được lựa chọn dựa trên đặc tính cơ lý riêng của đất đá tại mỗi khu vực nổ. Tuy nhiên, các thông tin đó lại không được xác định một cách chính xác gây khó khăn cho hoạt động thiết kế nổ mìn. Thực tế, các kỹ sư nổ mìn thường phải sử dụng các công thức và hệ số thực nghiệm tại tương ứng với mỗi loại đất đá và cấu trúc địa chất vùng khai thác. Công thức xác định thời gian vi sai theo thực nghiệm như sau [3][4]:

$$t = k \cdot R \text{ (ms)} \tag{1}$$

Trong đó: k (hệ số nền) là hệ số phụ thuộc tính chất của đất đá (ms/m), đối với đá rất cứng $k = 3$, đá cứng $k = 4$, đá cứng vừa $k = 5$, đá nứt nẻ mềm yếu $k = 6$; R là đường kháng chân tầng (m).

Bảng 1. Thời gian vi sai khi nổ nhiều hàng

Loại đất đá	Độ cứng f	Đường kháng chân tầng R(m)				
		1,5-3	3-4,5	4,5-6	6-8	8-10
Cứng và rất cứng	12-20	12-15	19-21	25-31	31-37	37-44
Cứng trung bình	8-14	19-21	25-31	31-37	37-40	43-50
Dính kết và mềm	4-8	25-31	31-37	37-40	43-50	50-65

Ảnh hưởng của thời gian vi sai tới hiệu quả nổ mìn đã được nghiên cứu và chứng minh bằng thực nghiệm trong nhiều nghiên cứu trước đây [5][6][7][8]. Một số kết quả đã được khẳng định như: mức độ đập vỡ đất đá giảm khi nổ đồng thời (thời gian vi sai bằng 0); mức độ đập vỡ tăng dần khi thời gian vi sai tăng lên. Mức độ đập vỡ tối ưu khi

Tác giả liên hệ: Phạm Thị Thanh Loan,
Email: thanhloanbkhn@gmail.com
Đến tòa soạn: 08/2024, chỉnh sửa: 09/2024, chấp nhận đăng: 10/2024.

tỷ lệ “thời gian vi sai/đường kháng chân tảng” nằm trong khoảng từ 4ms/m đến 10ms/m. Thời gian vi sai lớn hơn ngưỡng cho phép thì hạt đất đá lại trở nên thô hơn.

Cùng với sự phát triển của khoa học công nghệ, nổ mìn khai thác gần như được kiểm soát hoàn toàn trong những năm gần đây. Các thông số trong thiết kế vụ nổ đều được xét tới như lượng thuốc nổ, số lượng điểm nổ, kích thước lỗ nổ, đặc tính đất đá, vận tốc sóng chấn động hay thời gian trễ nổ [9]. Dữ liệu của các vụ nổ được lưu trữ một cách có hệ thống giúp các nghiên cứu về mô hình hoá, mô phỏng có tính ứng dụng cao khi các thông số nổ được xác định một cách nhanh chóng và tối ưu [10][11][12].

Một hướng nghiên cứu khác sử dụng dữ liệu này nhằm dự báo mức độ chấn động nổ, mức độ đập vỡ và mức độ nứt nẻ ảnh hưởng kết cấu của vụ nổ. Các tác giả sử dụng mô hình mạng nơron nhân tạo, mô hình học máy khác nhau để nâng cao độ chính xác trong dự báo. Cơ sở dữ liệu bao gồm các thông tin về độ cứng, độ xốp đất đá, độ ngậm nước, mật độ đất đá, lượng thuốc nổ, vị trí, thời gian kích nổ...[13][14][15][16]. Tuy nhiên, tham số thời gian vi sai trong nổ mìn vi sai chưa được đề cập đến. Bài báo này tập trung phân tích vai trò của thời gian vi sai đối với mức độ chấn động thông qua các quá trình tiền xử lý và phân tích dữ liệu.

B. Tiền xử lý dữ liệu trong học máy

Tiền xử lý dữ liệu trong quy trình học máy là một bước quan trọng, chiếm nhiều thời gian và công sức của người thiết kế mô hình. Tiền xử lý dữ liệu giúp nâng cao chất lượng và đảm bảo tính nhất quán cho bộ dữ liệu, từ đó cải thiện hiệu suất mô hình dự báo. Quy trình này bao gồm các bước cơ bản từ việc làm sạch dữ liệu đến chuẩn hoá và giảm chiều dữ liệu [17][18][19][20].

1) Xử lý thiếu dữ liệu

Các dữ liệu bị thiếu sau khi thu thập và lưu trữ có thể được loại bỏ (nếu tỷ lệ thiếu là rất nhỏ), hoặc được thay thế bằng các giá trị trung bình, trung vị, hoặc các giá trị xuất hiện nhiều nhất tùy theo tính chất đặc điểm của các đặc trưng.

Đối với bài toán cần sự chính xác cao hơn, giá trị thiếu được bù vào theo phương pháp KNN (K-Nearest Neighbors) hoặc Regression Imputation. Giải pháp này sử dụng giá trị của k điểm dữ liệu gần nhất hoặc một mô hình hồi quy để ước lượng, dự đoán giá trị còn thiếu. Phương pháp dựa trên mô hình (Model-based Imputation) tạo ra nhiều tập dữ liệu bằng cách thay thế các giá trị bị mất bằng các ước lượng từ một phân phối xác suất, sau đó tổng hợp các tập dữ liệu này để có kết quả cuối cùng.

2) Chuẩn hoá dữ liệu

Dữ liệu được quy đổi về giải giá trị [0,1] hoặc [-1,1] theo công thức (2), hoặc được chuyển đổi sao cho nó có giá trị trung bình bằng 0 và độ lệch chuẩn bằng (Z-score Scaling):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

Trong đó: x là một đặc trưng; μ , σ là giá trị trung bình và độ lệch chuẩn của tập dữ liệu.

$$x' = \frac{\bar{x} - \mu}{\sigma}$$

3) Biến đổi dữ liệu

Áp dụng hàm log, hàm mũ hoặc hàm căn bậc hai để giảm thiểu tác động của các giá trị lớn:

$x' = \log(x+1)$ (áp dụng khi dữ liệu có phân phối lệch phải)

$x' = e^x$ (dữ liệu có phân phối lệch trái)

4) Mã hoá dữ liệu

Chuyển đổi các nhãn phân loại thành các giá trị số nguyên (Label Encoding) hoặc thành vector nhị phân (One-hot Encoding). Mỗi giá trị phân loại được đại diện bởi một vector có giá trị 1 tại vị trí tương ứng và 0 tại các vị trí khác. Phương pháp này loại bỏ mối quan hệ giữa các giá trị, nhưng có thể tạo ra ma trận dữ liệu lớn nếu số lượng giá trị phân loại nhiều. Phương pháp Binary Encoding kết hợp Label Encoding và One-hot Encoding. Các giá trị phân loại được gán các số nguyên, sau đó các số nguyên này được chuyển thành biểu diễn nhị phân. Mỗi phương pháp có ưu, nhược điểm riêng, việc lựa chọn phương pháp phù hợp phụ thuộc đặc tính của dữ liệu và yêu cầu cụ thể của từng bài toán.

5) Giảm chiều dữ liệu

Số chiều của dữ liệu được giảm xuống khi loại bỏ các đặc trưng không có ảnh hưởng lớn theo phương pháp PCA (Principal Component Analysis) hoặc LDA (Linear Discriminant Analysis). PCA là kỹ thuật giảm chiều dựa trên toán học, giúp chuyển đổi các đặc trưng gốc thành một tập đặc trưng mới không tương quan xếp theo thứ tự giảm dần của phương sai. Trong khi đó, LDA là kỹ thuật giảm chiều có giám sát có nhiệm vụ tìm các vector trong không gian đặc trưng để tách biệt các lớp dữ liệu tốt nhất.

6) Phát hiện dữ liệu ngoại lai (Outlier)

+) Z-score (Standard Score): phương pháp này đo khoảng cách của một điểm dữ liệu so với giá trị trung bình theo đơn vị độ lệch chuẩn:

$$Z = \frac{x - \mu}{\sigma} \quad (3)$$

$|Z| > 3$ được coi là ngoại lai

+) Interquartile Range (IQR): Đo lường độ phân tán của dữ liệu giữa phân vị thứ nhất (Q1) và phân vị thứ 3 (Q3). Dữ liệu ngoại lai là các điểm nằm ngoài khoảng $[Q1 - 1.5IQR; Q3 + 1.5IQR]$; $IQR = Q3 - Q1$

+) Biểu đồ hộp (Boxplot) hoặc biểu đồ phân tán (Scatter Plot): là công cụ trực quan để phát hiện dữ liệu ngoại lai. Các điểm dữ liệu nằm ngoài “râu” của hộp được coi là ngoại lai.

+) Phương pháp học máy: Isolation Forest là thuật toán học máy không giám sát, hoạt động bằng cách xây dựng các cây quyết định ngẫu nhiên và cô lập các điểm dữ liệu. Điểm dữ liệu dễ bị cô lập là ngoại lai.

+) Một số phương pháp khác như phương pháp dựa trên mô hình, phương pháp dựa trên khoảng cách hoặc phương pháp tổng hợp.

7) Chia tách dữ liệu (Data Splitting)

Dữ liệu được chia thành các tập huấn luyện, kiểm tra, kiểm định để đánh giá hiệu quả mô hình.

C. Mô hình trí tuệ nhân tạo dự báo chấn động nổ mìn

Mô hình dự báo được phân loại theo bản chất của bài toán hoặc dựa trên cách thức học của mô hình. Có 3 cách

phân loại chính: Phân loại dựa trên loại dữ liệu đầu ra; Dựa trên phương pháp học; Dựa trên kiến trúc mô hình [19][20][21][22].

Bảng 2. Phân loại mô hình

Dữ liệu đầu ra	Giá trị liên tục	Linear Regression, Decision Tree, K-nearest Neighbors, Random Forest, Neural Network,...
	Nhãn phân loại	Logistic Regression, K-nearest Neighbors, Decision Tree, Random Forest, Neural Network,...
	Chuỗi thời gian	ARIMA, LSTM, Neural Networks, ..
Phương pháp học	Có giám sát	Linear/Logistic Regression, Decision Tree, Random Forest, Neural Network,...
	Không giám sát	Phân cụm K-means, PCA..
	Học tăng cường	Q-Learning, Deep Q-Networks (DQN), ...
Kiến trúc mô hình	Tuyến tính	Linear/Logistic Regression
	Phi tuyến	Decision Tree, Neural Network
	Kết hợp	Random Forests, Gradient Boosting Machines (GBM)

1) K-nearest Neighbors

KNN là một thuật toán đơn giản sử dụng trong bài toán phân loại và hồi quy. Giả định trong thuật toán là các điểm dữ liệu gần nhau có xu hướng giống nhau. Số lượng k điểm lân cận gần nhất thường được chọn dựa trên khoảng cách giữa điểm cần dự đoán các điểm trong tập dữ liệu huấn luyện với khoảng cách là nhỏ nhất. Các phương pháp tính khoảng cách phổ biến bao gồm:

+) Khoảng cách Euclid:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

+) Khoảng cách Manhattan:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (5)$$

+) Khoảng cách Minkowski:

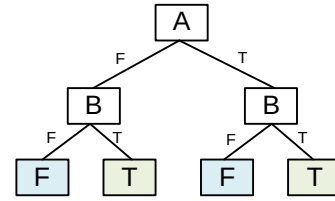
$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (6)$$

Trong đó: x, y là hai điểm trong không gian n chiều; x_i, y_i là thành phần thứ i của x và y ; p là tham số điều chỉnh của khoảng cách Minkowski.

Giá trị dự đoán chính là giá trị trung bình của k điểm lân cận (hồi quy), hoặc bằng nhãn xuất hiện nhiều nhất trong k điểm lân cận (phân loại). Mô hình được đánh giá thông qua các tham số: độ chính xác (accuracy), sai số trung bình bình phương (mean square error) ... KNN có ưu điểm về tính đơn giản, linh hoạt tuy nhiên hiệu suất chậm với tập dữ liệu lớn, hiệu suất phụ thuộc mạnh mẽ vào k

2) Decision Tree

Cây quyết định được sử dụng cho cả bài toán phân loại và hồi quy. Đây là một kỹ thuật phổ biến chia nhỏ dữ liệu thành các tập con dựa trên các đặc trưng, sau đó lặp lại quá trình này cho từng tập con tạo thành một cây quyết định.



Hình 2. Cấu trúc của cây quyết định

Cây quyết định bao gồm các nút quyết định (nút B) và các nút lá (T, F). Nút gốc (A) là điểm bắt đầu của cây quyết định. Tại đây dữ liệu được chia dựa trên một đặc trưng tốt nhất và một ngưỡng. Dữ liệu tiếp tục được chia nhỏ tại các nút quyết định với một thuộc tính và một ngưỡng mới được so sánh. Nút lá là nút cuối cùng (T, F) chứa giá trị dự đoán, thường là giá trị trung bình của các giá trị mục tiêu trong tập dữ liệu.

Có nhiều cách để triển khai cây quyết định, nổi bật là phương pháp C5.0, CART (phân loại và cây hồi quy) với thuật toán Gini Index, Variance Reduction:

$$Gini(S) = 1 - \sum_{i=1}^c p_i^2 \quad (7)$$

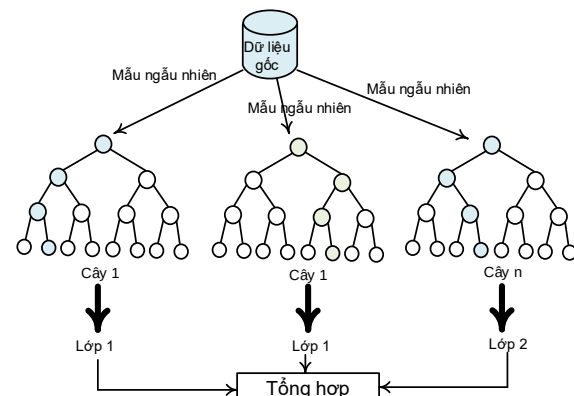
$$Variance(S) = \frac{1}{|S|} \sum_{i=1}^{|S|} (y_i - \bar{y})^2 \quad (8)$$

Trong đó: p_i là xác suất của lớp i trong tập hợp S ; y_i là giá trị mục tiêu của mẫu i và $\bar{y}$ là giá trị trung bình của các giá trị trong tập S .

Gini Index sử dụng cho bài toán phân loại, đo lường mức độ không thuần nhất của một tập hợp. Variance đo mức độ giảm biến động sau khi chia tập hợp dựa trên một đặc trưng. Quá trình chia nhỏ sẽ dừng lại khi cây đạt đến độ sâu tối đa và số lượng mẫu trong một nút ít hơn số lượng mẫu tối thiểu của một nút lá.

Cây quyết định có các quy tắc đơn giản, hoạt động hiệu quả đối với tập dữ liệu nhỏ và lớn. Nhược điểm của mô hình này dễ gây quá khớp, có hiệu suất thấp đối với bài toán phân loại nhiều lớp.

3) Random Forest



Hình 3. Cấu trúc mô hình rừng ngẫu nhiên

Random Forest là kỹ thuật mở rộng của Decision tree, trong đó nhiều cây quyết định được xây dựng và các dự đoán của chúng được lấy trung bình để cải thiện độ chính xác và giảm hiện tượng quá khớp. Số lượng thuộc tính trong mỗi tập con thường được lấy là $\sqrt{m}$, với m là tổng số đặc trưng của bộ dữ liệu.

$$\hat{y} = \frac{1}{B} \sum_{i=1}^B \hat{y}_i \quad (9)$$

Với $\hat{y}$ là giá trị dự đoán cuối cùng; $\hat{y}_i$ là giá trị dự đoán của mỗi cây; B là số lượng cây quyết định.

Random Forest cho độ chính xác cao hơn so với một cây quyết định đơn lẻ, khả năng chống nhiễu tốt do việc lấy trung bình các dự đoán nhưng nó đòi hỏi khả năng tính toán và mức độ phức tạp cao của thuật toán.

4) XGBoosting

XGBoost bao gồm nhiều cây quyết định được thiết kế theo tuần tự. Các cây này không độc lập như trong Random Forest mà được xây dựng dần với mục tiêu giảm thiểu hàm lỗi bằng phương pháp Gradient Boosting.

+) Hàm mất mát được tính theo công thức:

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (10)$$

Trong đó: $L^{(t)}$: hàm mất mát tại bước t ; l : là hàm mất mát; $\Omega(f_t)$ là hàm regularization

+) Gradient và Hessian:

$$g_i^{(t)} = \frac{\partial l(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}; \quad h_i^{(t)} = \frac{\partial^2 l(y_i, \hat{y}_i^{(t-1)})}{\partial (\hat{y}_i^{(t-1)})^2} \quad (11)$$

+) Cập nhật dự đoán:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (12)$$

XGBoost là một công cụ mạnh mẽ và linh hoạt, mang lại hiệu suất cao. Nó được tối ưu hóa để chạy nhanh và hiệu quả trên bộ dữ liệu lớn. XGBoost có khả năng tự động xử lý giá trị thiếu mà không cần tiền xử lý đồng thời giảm hiện tượng quá khớp nhờ kỹ thuật regularization.

Chỉ tiêu đánh giá độ chính xác của các mô hình được thể hiện qua giá trị RMSE (Root Mean Squared Error): độ lệch chuẩn của sai số dự báo; MAE (Mean Absolute Error): độ lớn trung bình của các sai số dự báo; MSE (Mean Square Error): giá trị trung bình của bình phương sai số dự báo; R-square (R^2): tỷ lệ phương sai của biến mục tiêu được giải thích bằng mô hình dự báo. R^2 càng gần 1 thì mô hình càng tốt.

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ MSE &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ R^2 &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \end{aligned} \quad (13)$$

Trong đó: n là số lượng quan sát; y_i là giá trị thực tế tại quan sát thứ i ; $\hat{y}_i$ là giá trị dự báo của quan sát thứ i ; $\bar{y}$: giá trị trung bình của tất cả các giá trị thực tế y_i .

II. XỬ LÝ DỮ LIỆU TRONG DỰ BÁO CHẤN ĐỘNG TẠI MỎ THAN NÚI BÉO

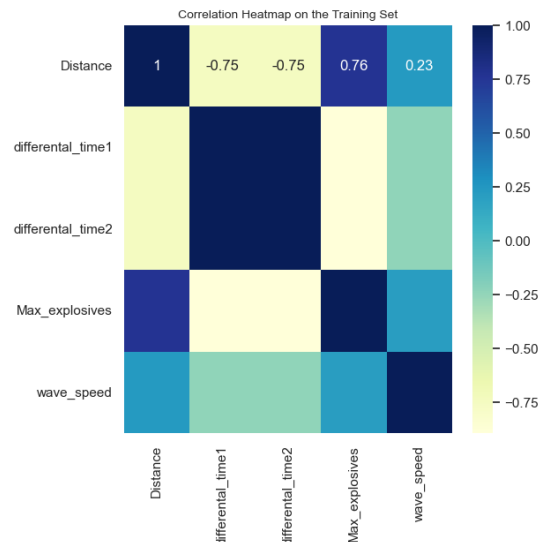
Tập dữ liệu thực nghiệm thu thập được tại mỏ than Núi Béo bao gồm 1115 giá trị cùng 5 đặc trưng: lượng thuốc nổ lớn nhất tại một điểm nổ; thời gian vi sai, vận tốc sóng lan truyền. Các đặc trưng này được sử dụng để dự báo vận tốc của hạt đất đá.

Bảng 3. Thông số của các đặc trưng

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1115 entries, 0 to 1114
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Distance               1115 non-null   float64
1   differential_time1     1115 non-null   int64
2   differential_time2     1115 non-null   int64
3   Max_explosives         1115 non-null   float64
4   wave_speed             1115 non-null   int64
5   partical_speed         1115 non-null   float64
dtypes: float64(3), int64(3)
memory usage: 52.4 KB
```

A. Khảo sát mối tương quan giữa các đặc trưng

Độ chính xác của mô hình phụ thuộc vào bộ dữ liệu quan sát. Sự tương quan lớn giữa một cặp biến đầu vào có thể dẫn tới giảm hiệu quả mô hình. Bản đồ nhiệt tương quan (Correlation heatmap) là một công cụ trực quan giúp phát hiện mối quan hệ tuyến tính giữa các biến từ đó loại bỏ được đặc trưng phụ thuộc.

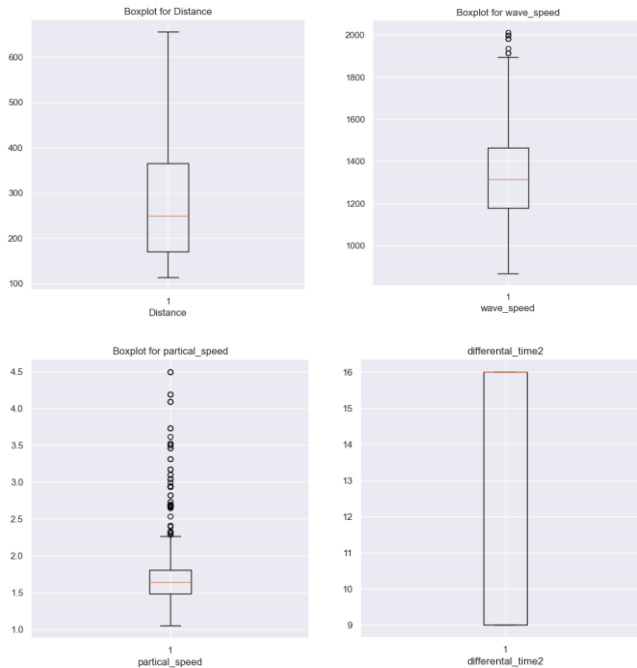


Hình 4. Mối tương quan giữa các đặc trưng

Thời gian vi sai 1 và thời gian vi sai 2 có mối tương quan lớn ≈ 1 nên không thể sử dụng đồng thời cả hai đặc trưng này để dự báo vận tốc hạt đất đá. Đặc trưng thời gian vi sai 1 sẽ được loại bỏ khỏi tập dữ liệu đầu vào. Đặc trưng lượng thuốc nổ lớn nhất và thời gian vi sai có mối liên hệ thấp nhất, khoảng 25%, trong khi tương quan giữa khoảng cách và thời gian vi sai xấp xỉ 70%. Do đó cả 4

đặc trưng còn lại đều được sử dụng như đầu vào để dự báo vận tốc hạt.

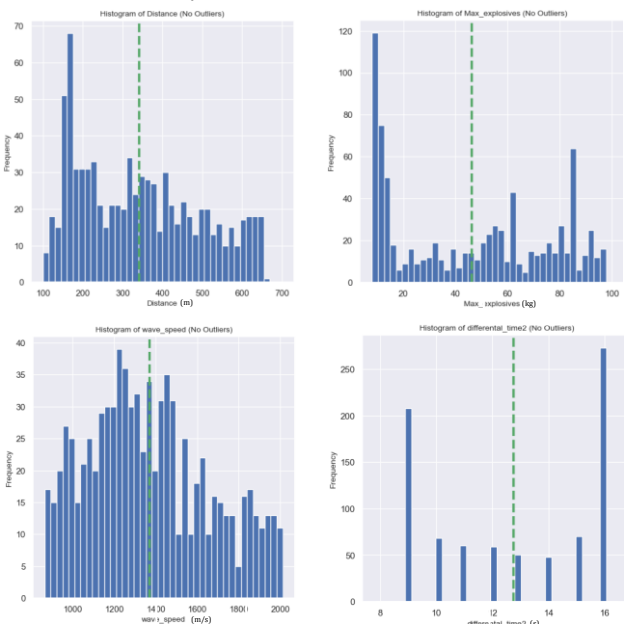
B. Phát hiện dữ liệu ngoại lai (Outlier)



Hình 5. Biểu đồ Boxplot

Phân phối dữ liệu khoảng cách và thời gian vi sai 2 cho thấy có sự biến động khá lớn trong giá trị của hai đặc trưng này. Tuy nhiên, không tồn tại giá trị ngoại lệ trong khoảng cách và thời gian vi sai 2. Dữ liệu về tốc độ lan truyền sóng và tốc độ hạt đất đá ít có biến động. Vận tốc hạt dao động chủ yếu quanh các giá trị 1.5-1.75m/s. Có một số ngoại lệ ở vận tốc hạt và vận tốc lan truyền sóng. Các giá trị ngoại lệ này sẽ được loại bỏ theo phương pháp IQR.

C. Biến đổi dữ liệu



Hình 6. Xác suất phân phối dữ liệu

Tập giá trị của các đặc trưng được phân bố lệch phải trên biểu đồ tần suất (Hình 66). Giá trị khoảng cách [120 -

630m] giữa các điểm nổ xuất hiện nhiều nhất trong biến Distance, chiếm 90% dữ liệu. Đặc trưng thời gian vi sai và lượng thuốc nổ lớn nhất được phân bố không đều, bao gồm một số dữ liệu cao bất thường (9kg đối với lượng thuốc nổ và 16s đối với thời gian vi sai). Để đưa dữ liệu về dạng phân phối chuẩn, tác giả triển khai phương pháp chuẩn hoá Z-score và biến đổi Yeo_Johnson. Phương pháp này áp dụng cho cả dữ liệu có giá trị dương và giá trị âm. Công thức biến đổi như sau:

$$y \geq 0: \quad y' = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y+1) & \lambda = 0 \end{cases} \quad (14)$$

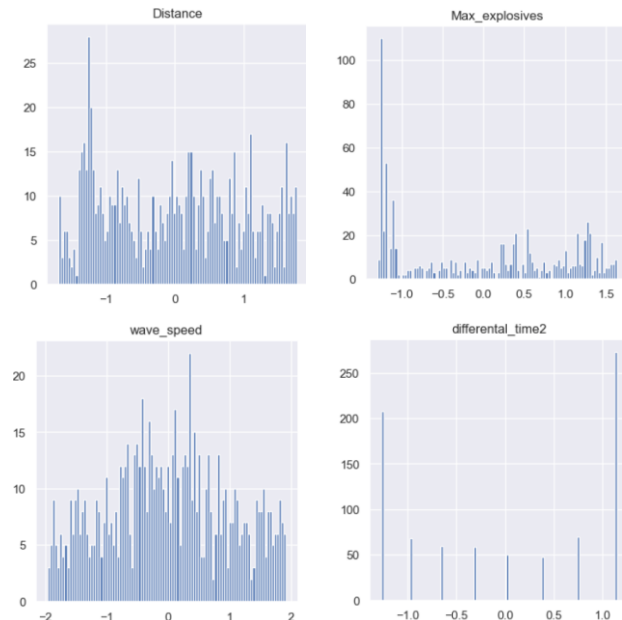
$$y < 0: \quad y' = \begin{cases} -\frac{(-y+1)^{2-\lambda} - 1}{2-\lambda} & \lambda \neq 2 \\ -\ln(-y+1) & \lambda = 0 \end{cases}$$

Trong đó: λ là tham số điều chỉnh quyết định cách biến đổi dữ liệu.

Bảng 4. Phương pháp biến đổi dữ liệu

Giá trị λ	Phương pháp biến đổi
0	Biến đổi logarit
1	Giữ nguyên dữ liệu
Khác 0 và 2	Biến đổi theo công thức Yeo_Johnson
Ước lượng	Tối ưu hoá sự biến đổi

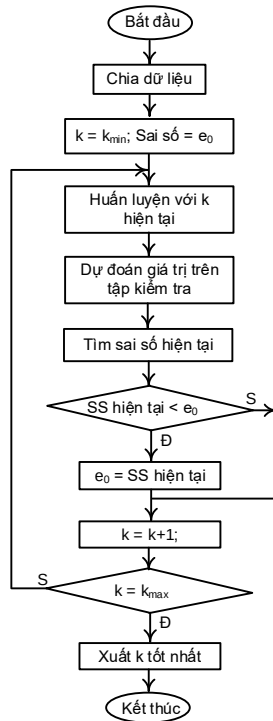
Dữ liệu sau khi chuẩn hoá được đưa về dạng gần phân phối chuẩn (Hình 7). Điều này giúp nâng cao độ chính xác các cho mô hình dự báo nhạy cảm với phân phối dữ liệu.



Hình 7. Dữ liệu được chuẩn hoá theo phương pháp Z-Score

III. TRIỂN KHAI CÁC MÔ HÌNH DỰ BÁO MỨC ĐỘ CHẤN ĐỘNG NỔ Mìn

Dự báo mức độ chấn động nổ mìn của mỏ than Núi Béo thông qua vận tốc hạt đất đá có xét đến tác động của lượng thuốc nổ lớn nhất tại một điểm nổ thời gian trễ nổ, vận tốc sóng lan truyền là bài toán dự báo hồi quy, đối tượng phi tuyến mạnh.



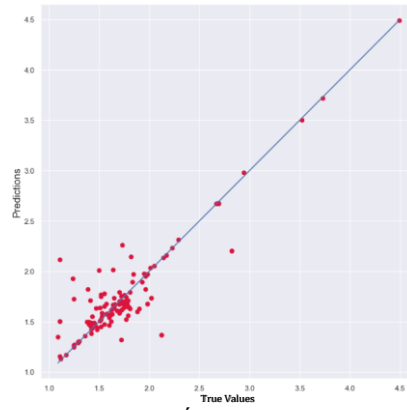
Hình 8. Thuật toán xác định k

Mỗi mô hình dự báo có các ưu, nhược điểm riêng và phù hợp với từng nhóm đối tượng cụ thể. Đối với mục tiêu dự báo chấn động nổ mìn, nhóm tác giả đề xuất triển khai một số mô hình: KNN, Decision tree, Random Forest; XGBoost.

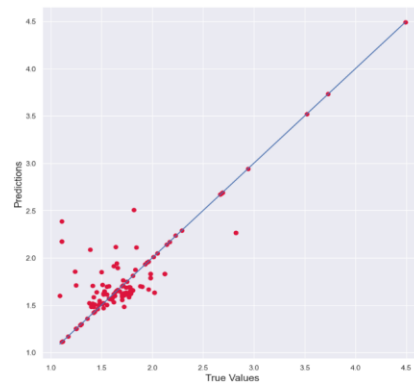
Tham số k được khởi tạo ở giá trị nhỏ nhất trong giải cho phép đồng thời thiết lập sai số ban đầu e_0 lớn nhất. Dữ liệu được đưa vào huấn luyện với k hiện tại và xác định ra sai số sau huấn luyện. Quá trình tiếp tục cho đến giá trị lớn nhất của k và xác định được tham số k tương ứng với sai số nhỏ nhất (hình 8). Tham số k của mô hình KNN (KNeighborsRegressor) được chọn trong khoảng [5, 30] vì số lượng dữ liệu không lớn. Kết quả thực hiện mô hình cho thấy độ chính xác của mô hình đạt cao nhất với $k = 9$. Tại đó, độ chính xác đạt 0.95 cho tập huấn luyện và 0.8 cho tập thử nghiệm. Thuật toán xác định k cho trên Hình 8. Các tham số của các mô hình còn lại cũng được lựa chọn theo thuật toán trên. Giá trị max_depth và random_state trong mô hình DecisionTreeRegressor nằm trong dải [1, 30] với bước tính là 2. Độ phù hợp của mô hình khá cao, lên tới 91.5% khi max_depth là 9 và random_state là 3. Mô hình XGBoost cho hiệu suất cao nhất lên tới 96,6%. Chỉ tiêu đánh giá mô hình được cho trong Bảng 5.

Bảng 5. Kết quả dự báo của các mô hình

Mô hình	Coefficient	RMSE	MSE	MAE
KNN	0.952	0.0061	1.545	0.0011
Decision tree	0.915	0.160	0.025	0.131
Random Forest	0.947	0.126	0.015	0.076
XGBoots	0.966	0.019	9.35	0.018



Hình 9. Giá trị thực tế và giá trị dự đoán của mô hình XGBoost



Hình 10. Giá trị thực tế và giá trị dự đoán của mô hình KNN

Kết quả hình 9 và hình 10 cho thấy, mối quan hệ giữa giá trị thực tế và giá trị dự đoán từ mô hình có xu hướng trở nên tuyến tính 1-1, nghĩa là giá trị dự báo đã theo sát giá trị thực tế của hệ thống.

IV. KẾT LUẬN

Trong bài báo này, nhóm tác giả đã thực hiện phân tích dữ liệu chi tiết và triển khai một số mô hình cơ bản để dự báo vận tốc hạt đất đá, từ đó xác định được mức độ chấn động nổ mìn trong mỏ than Núi Béo. Kỹ thuật tìm và loại bỏ dữ liệu ngoại lai, chuẩn hoá và phân chia tập dữ liệu đã được áp dụng để đảm bảo chất lượng của dữ liệu khi triển khai mô hình dự báo. Một số mô hình dự báo hồi quy đã được thử nghiệm để tìm ra mô hình phù hợp nhất với bài toán. Kết quả cho thấy, mô hình XGBoost và mô hình KNN cho hiệu suất cao nhất với các chỉ số MSE; MAE; RMSE rất nhỏ. Nghiên cứu này cung cấp một cơ sở quan trọng cho việc dự báo mức chấn động nổ mìn. Từ đó xác định ra thời gian vi sai tối ưu phục vụ công tác thiết kế nổ mìn trong các mỏ than tại Việt Nam.

TÀI LIỆU THAM KHẢO

- [1] Calvin J. Konya, Edward J. Walter (1991), Rock blasting and overbreak control, National Highway Institute, Publication No. FHWA-HI-92-001.
<https://ieeexplore.ieee.org/document/8244882>
- [2] Langefors U, Kihlstrom B, Westerberg H (1958), Ground Vibrations in Blasting. Water Power, September, pages 335-38. October, pages 390-95, November, pages 421, 424.
- [3] Hồ Sĩ Giao, Đàm Trọng Thắng, Lê Văn Quyền, Hoàng

- Tuân Chung (2010), Nô hóa học – Lý thuyết và thực tiễn, NXB Khoa học và Kỹ thuật, Hà Nội.
- [4] Rossmann H. P. (2003), “The Mechanics and Physics of Electronic Blasting”, Proceedings of the 33rd Conference on Explosives and Blasting Technique, ISEE, Nashville, TN, 2003.
- [5] Johansson, Daniel and Ouchterlony, Finn (2013), Vibration Wave Interactions in Rock Blasting: the Use of Short Delays to Improve Fragmentation in Model-Scale”, Rock Mechanics and Rock Engineering, January 2013, Volume 46, Issue 1, pp 1-18
- [6] Katsabanis P.D Tawadrous A., Braun C., Kennedy C (2006), “Timing Effects on Fragmentation”, Proceedings of the 32nd Conference on Explosives and Blasting Technique, ISEE, Dallas, TX, 2006
- [7] Katsabanis P. ., Omid O., Rielo O., and Ross P. (2014), “A Review of Timing Requirements for Optimization of Fragmentation”, Proceedings of the 40th Conference on Explosives and Blasting Technique, ISEE, Denver, CO, 2014
- [8] Sjoberg J., Schill, M. Hilding, D., Yi, C., Nyberg, U., and Johansson, D. (2012), Computer Simulations of Blasting with Precise Initiation. Eurock, Stockholm, Sweden, 2012
- [9] Iau Teh Wang “Field Experiments and numerical analysis of the ground vibration isolation of seismic wave propagation under explosion seismic loading”, MDIP Journal, 11/2019, doi:10.3390.
- [10] Jinxing Lai, Haobo Fan, Jianxun Chen, Junling Qiu, Ke Wang (2015), “Blasting vibration monitoring of undercrossing railway tunnel using wireless sensor network”, International Journal of Distributed Sensor Networks, volume 2015, article Id 703980
- [11] Matus Ivanco, Romana Erdélyiová, Lucia Figuli (2019), “Simulation of detonation and blast waves propagation”, 13th International Scientific Conference on Sustainable, Modern and Safe Transport (TRANSCOM 2019), High Tatras, Novy Smokovec – Grand Hotel Bellevue, Slovak Republic, pp 1756-1363.
- [12] Nick Barton (2007), Rock Quality, Seismic Velocity, Attenuation and Anisotropy. Taylor & Francis Group, London, UK. ISBN 0-203-96445-4 Master e-book ISBN.
- [13] Hossein Moayed, , Ahmad Safuan A Rashid, Mohammed Abdullahi Muazu, Hoang Nguyen, Xuan-Nam Bui, Dieu Tien Bui (2019), “Prediction of ultimate bearing capacity through various novel evolutionary and neural network models”, Engineering with Computers, DOI: 10.1007/s00366-019-00723-2.
- [14] Jie Dou, Ali P. Yunus, Abdelaziz Merghadi, Ataollah Shirzadi, Hoang Nguyen, Yawar Hussain, Ram Avtar, Yu long Chen, Binh Thai Pham, Hiromitsu Yamagishi (2020), “Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning”, Science of the Total Environment (720), pp 137320-137350.
- [15] Nguyen Hoang (2020), Research and development of blasting vibration wave prediction model for open-pit coal mine using artificial intelligence and machine learning algorithms. Ministry level topic.
- [16] Dieu Tien Bui, Hossein Moayed, Bahareh Kalantar, Abdolreza Osooli, Biswajeet Pradhan, Hoang Nguyen and Ahmad Safuan A Rashid (2019), “A Novel Swarm Intelligence”, Harris Hawks Optimization for Spatial Assessment of Landslide Susceptibility, Sensors (19), pp 3590-3612.
- [17] Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, “Machine Learning”, Pearson India Education Services Pvt. Ltd, CIN: U72200TN2005PTC057128.
- [18] Vu Huru Tiep, “Machine Learning cơ bản”, machine learningcoban.com2020.
- [19] Larose, C., & Larose, D, “Data Science Using Python and R. Wiley”, 2019
- [20] Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V , “Introduction to data mining”, Pearson, Second Editor, 2020.
- [21] Bernd Klein, “Machine learning with Python Tutorial”, www.pyhton-course.eu, 2021
- [22] Mark Lutz, “Learning Python”, O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 2009.

SEISMIC LEVEL PREDICTING IN MINERAL EXTRACTION WITH ARTIFICIAL INTELLIGENCE

Abstract: Blasting in mining is the process of utilizing explosive energy to break down physical structures for construction and mineral exploitation activities. After breaking the rock near the explosion point, the remain energy forms a vibration wave. It propagates, affecting the surrounding geological structure and buildings. In modern blasting techniques, the total amount of explosive is divided into parts to detonate after different periods of time. That improve blasting efficiency, while reducing the level of vibration. Measuring the vibration level is determining the oscillation velocity of soil and rock particles. The results are the basis for assessing the level of danger to buildings near the mining area. In this paper, the authors propose a data analysis method to determine important factors affecting the oscillation velocity of soil and rock particles. A number of artificial intelligence models have also been developed to predict blasting shock waves applicable to open pit mines in Viet Nam. Data was taken directly from the sensor during an experimental explosion at Nui Beo coal mine, Quang Ninh province.

Keywords: Shock waves, differential time, data preprocessing, Random Forest, XGBoost, KNN.



Phạm Thị Thanh Loan tốt nghiệp đại học, cao học và nhận bằng Tiến sỹ ngành Kỹ thuật điều khiển và Tự động hóa năm 2003, 2005 và 2015 tương ứng tại trường Đại học Bách khoa Hà nội. Năm 2003 chị làm việc như một kỹ sư Tự động hóa tại Trung tâm Công nghệ cao - Viện Máy và Dụng cụ Công nghiệp. Từ năm 2004, TS Loan là giảng viên tại Bộ môn Tự động hóa – Khoa Cơ Điện – Đại học Mỏ Địa chất. Lĩnh vực nghiên cứu bao gồm: Điều khiển quá trình, Điều khiển dự báo, mô hình hóa, năng lượng tái tạo, AI.



Đào Hiếu tốt nghiệp đại học, cao học và nhận bằng Tiến sỹ ngành Kỹ thuật điều khiển và Tự động hóa năm 2005, 2009 và 2023 tương ứng tại trường Đại học Mỏ - Địa chất. Năm 2005-2007 anh làm việc như một kỹ sư Tự động hóa tại các nhà máy công nghiệp. Từ năm 2008, TS Hiếu là giảng viên tại Bộ môn Tự động hóa – Khoa Cơ Điện – Đại học Mỏ Địa chất. Lĩnh vực nghiên cứu bao gồm: Năng lượng tái tạo, hệ thống điều khiển thông minh, IoT.