

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

NGUYỄN NGỌC THÁI

NGHIÊN CỨU BÀI TOÁN PHÁT HIỆN ĐỘNG ĐẤT SỬ
DỤNG DỮ LIỆU CẢM BIẾN GIA TỐC

Ngành: Công nghệ Kỹ thuật Cơ điện tử

TÓM TẮT KHOÁ LUẬN TỐT NGHIỆP

Hà Nội - 2025

MỞ ĐẦU

Với sự tiến bộ vượt bậc của khoa học và công nghệ, đặc biệt là trong lĩnh vực trí tuệ nhân tạo AI và Internet vạn vật (IoT), việc triển khai các hệ thống tự động để phát hiện sớm thiên tai đang trở thành xu thế tất yếu trong công tác phòng chống và giảm nhẹ rủi ro. Trong bối cảnh Việt Nam và nhiều quốc gia khác ngày càng đối mặt với rủi ro thiên tai từ động đất, việc phát hiện sớm động đất là một trong những ứng dụng quan trọng và cấp thiết.

Các hệ thống cảm biến hiện đại ngày nay có khả năng ghi nhận tín hiệu rung chấn với độ chính xác và tần suất cao tại hàng nghìn điểm quan trắc. Khi được kết hợp với AI và học máy, các hệ thống này không chỉ phát hiện rung chấn bất thường trong thời gian thực, mà còn có thể phân biệt được giữa động đất thật và nhiễu từ bên ngoài môi trường, việc đưa ra cảnh báo sớm chỉ trong vài giây điều có thể quyết định giữa sự sống và cái chết cho hàng ngàn người.

Xuất phát từ tầm quan trọng đó, cùng với sự hướng dẫn tận tình của tiến sĩ Hoàng Văn Mạnh và sự đồng ý của Khoa Cơ kỹ thuật và Tự động hóa, em đã thực hiện đề tài “***Nghiên cứu bài toán phát hiện động đất sử dụng dữ liệu cảm biến gia tốc***” với mục tiêu xây dựng một mô hình học máy có khả năng phân biệt hiệu quả tín hiệu động đất và nhiễu nền, góp phần hỗ trợ các hệ thống cảnh báo sớm trong tương lai.

CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

1.1: Giới thiệu về động đất

Động đất là sự rung động của một khu vực trên vỏ Trái Đất dưới ảnh hưởng của những nguyên nhân nằm trong lòng Trái Đất (nội sinh) hay nguyên nhân từ bên ngoài Trái Đất (ngoại sinh), thậm chí còn do con người tạo ra (nhân sinh).

Thang đo độ mạnh yếu của động đất, người ta dùng hai đơn vị đo là: cấp độ và cường độ.

1.2: Các mô hình học máy áp dụng

Học máy - Machine Learning là một nhánh của trí tuệ nhân tạo, cho phép máy tính học từ dữ liệu để thực hiện các nhiệm vụ như dự đoán hoặc phân loại. Dựa trên cách tiếp cận với dữ liệu, học máy được chia thành ba nhóm chính: học có giám sát, học không giám sát và học tăng cường. Trong phạm vi đề tài, các mô hình học máy được áp dụng bao gồm: Logistic Regression, SVM, KNN, Decision Tree, Random Forest,... với mục tiêu phân loại tín hiệu gia tốc thành tín hiệu động đất và nhiễu nền.

CHƯƠNG 2: PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Quy trình nghiên cứu

Quá trình 1: Mô tả dữ liệu sử dụng.

Dữ liệu được thu thập từ hệ thống K-NET dữ liệu của Kyoshin Network - một mạng lưới quốc gia về quan trắc gia tốc nền đất mạnh ở Nhật Bản, bao gồm các file tín hiệu gia tốc ba trục x, y, z và với tần suất lấy mẫu 100 Hz. Dữ liệu được chia thành hai nhóm bao gồm: 2327 tập EQ và 80 tập nhiễu từ các hoạt động bên ngoài như động cơ, đi bộ, xe bus,[8]

Trực quan hoá dữ liệu theo: miền thời gian, miền tần số FFT và miền mật độ phổ công suất PSD.

Quá trình 2: Tiền xử lý dữ liệu.

Sau khi thu thập, dữ liệu được tiền xử lý thông qua các bước loại bỏ các dòng chứa giá trị thiếu (NaN), outlier, cắt thành các đoạn tín hiệu ngắn theo cửa sổ sibling_window để kiểm tra trực tiếp sự khác biệt giữa EQ và Noise.

Quá trình 3: Trích xuất đặc trưng.

Dữ liệu sau tiền xử lý sẽ có dạng đầu vào là ma trận đặc trưng cho mỗi cửa sổ, dùng để trích xuất các đặc trưng thời gian và tần số phục vụ huấn luyện mô hình phân loại EQ và non-EQ. Các đặc trưng sử dụng để huấn luyện là: IQR, ZC, dominant Frequency, mean, độ lệch chuẩn, ...

Quá trình 4: Chuẩn hoá dữ liệu.

Thực hiện áp dụng phương pháp Z-score (Standard Scaler) giúp đưa những dữ liệu về cùng thang đo, giúp cho mô hình học tốt hơn, hội tụ nhanh hơn và giúp mô hình không bị thiên lệch trong quá trình học.

Quá trình 5: Áp dụng mô hình học máy

Tập dữ liệu được sẽ chia thành hai phần:

- Training set: chiếm 80% tổng số mẫu.
- Test set: chiếm 20% tổng số mẫu.

Áp dụng những mô hình học máy Machine Learning có sẵn trong thư viện sklearn như Logistic Regression, SVM, KNN, Decision Tree, Random Forest và hai mô hình Deeplearning: ANN và CRNN để phân biệt giữa tín hiệu EQ và noise.

Quá trình 6: Đánh giá mô hình học máy.

Sau bước huấn luyện mô hình, em có thử nghiệm với nhiều phương pháp khác nhau như train/test split và Kfolder. Qua quá trình đánh giá dựa trên những metrics sau: Accuracy, Recall, F1-score, Confusion Matrix,... Các mô hình được so sánh dựa trên những chỉ số này để xác định mô hình nào tối ưu nhất.

Quá trình 7: Triển khai mô hình.

Cuối cùng, mô hình được tích hợp vào một giao diện người dùng đơn giản được xây dựng bằng PyQt6. Giao diện này cho phép người dùng tải lên một file tín hiệu bất kỳ, thực hiện dự đoán và hiển thị kết quả phân loại là động đất hay nhiễu.

CHƯƠNG 3: THỰC NGHIỆM VÀ KẾT QUẢ

3.1. Thiết lập thực nghiệm

Sau khi biến chuyển đổi đơn vị từ giá trị thô sang đơn vị gia tốc gal thì áp dụng cửa sổ sibling_window với during_time bằng 10 giây và overlap bằng 50%. Tạo ra được một bộ dữ liệu mới với kích thước tập EQ là (87177, 33) và tập nhiễu là (9859, 33).

3.2. Kết quả huấn luyện và đánh giá mô hình

Bảng 3.1. Thống kê các kết quả đánh giá mô hình huấn luyện.

Mô hình	Accuracy	Precision	Recall	F1
DT	0.9279	0.85	0.7	0.75
KNN	0.91	0.98	0.97	0.98
Naïve Bayes	0.923	0.79	0.79	0.79
LR	0.9468	0.91	0.77	0.83
Random Forest	0.9967	0.99	0.99	0.99
SVM	0.9883	0.97	0.96	0.96
ANN ^[9]	0.976	0.8935	0.8646	0.975
CRNN ^[9]	0.9989	1.0	0.99	0.99

3.2.2. Dự đoán độ lớn của trận động đất

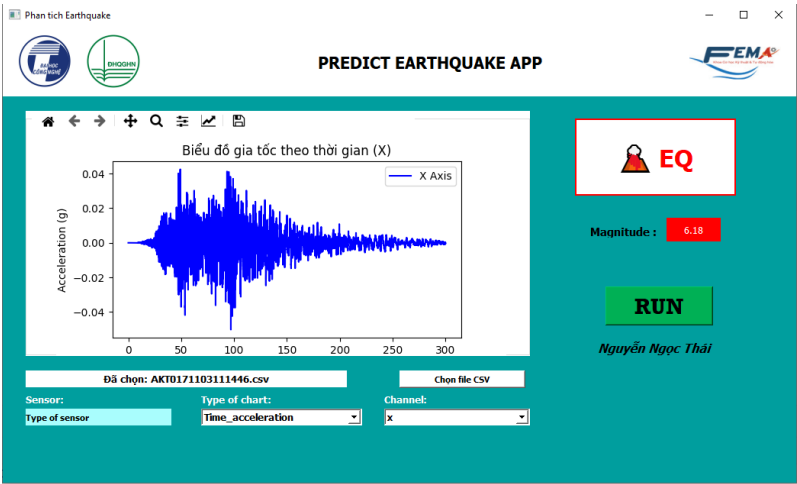
Bảng 3.2. Thống kê kết quả các mô hình phán đoán độ lớn trận động đất.

Mô hình	MAE	MSE	R ²
Logistic R	0.7816	0.7955	0.6319
Random Forest	0.3577	0.477	0.8678
SVM	0.7816	0.9541	0.4705

Nhận thấy mô hình Random Forest đưa ra kết quả tốt nhất cho bài toán dự đoán độ lớn của trận động đất khi có metric $R^2 = 0.8678$.

3.2.3. Thiết kế giao diện người dùng

Giao diện được xây dựng trên bằng phần mềm QT Designer kết hợp với ngôn ngữ lập trìnhPython.



Hình 3.11. Giao diện người dùng hệ thống

KẾT LUẬN

Trong khuôn khổ khoá luận tốt nghiệp, đề tài “**Nghiên cứu bài toán phát hiện động đất sử dụng dữ liệu cảm biến gia tốc**” đã tiến hành xây dựng một hệ thống phân biệt tín hiệu địa chấn thông qua việc ứng dụng các mô hình học máy trên dữ liệu cảm biến. Dữ liệu được xử lý, trích xuất đặc trưng và huấn luyện với nhiều mô hình khác nhau nhằm mục đích đánh giá khả năng phân loại giữa tín hiệu động đất (EQ) và tín hiệu nhiễu (Noise).

Mặc dù kết quả đạt được là khả quan, đề tài vẫn còn một số hạn chế nhất định. Dữ liệu sử dụng trong huấn luyện và kiểm thử chủ yếu được lấy từ các nguồn công khai với số lượng và phạm vi địa lý còn hạn chế. Việc triển khai hệ thống phát hiện động đất theo thời gian thực vẫn chưa được hiện thực hóa trong phạm vi nghiên cứu này. Hơn nữa, hệ thống hiện tại mới chỉ dừng lại ở bước phân loại tín hiệu, chưa tích hợp chức năng định vị chấn tâm một cách tự động và chính xác.

Trong tương lai, để nâng cao tính ứng dụng và độ tin cậy, đề tài có thể phát triển theo các hướng sau tích hợp hệ thống phát hiện theo thời gian thực định vị địa chấn và cải thiện khả năng xử lý dữ liệu tại các khu vực cụ thể.

Tóm lại, nghiên cứu này không chỉ góp phần khẳng định vai trò của trí tuệ nhân tạo, học máy trong bài toán phân tích và cảnh báo thiên tai. Đây là một bước khởi đầu quan trọng để hướng tới việc xây dựng một hệ thống giám sát và cảnh báo động đất thông minh, chi phí thấp và có khả năng mở rộng trong tương lai.

TÀI LIỆU THAM KHẢO

Tiếng Việt:

- [1] baochinhpphu.vn. “Tìm Hiểu về Động Đất.” *Baochinhpphu.vn*, 22 Apr. 2011, baochinhpphu.vn/tim-hieu-ve-dong-dat-10298355.htm. Accessed 18 Apr. 2025.
- [2] Động đất là gì? Nguyên nhân và hậu quả của động đất. “Động Đất Là Gì? Nguyên Nhân và Hậu Quả Của Động Đất.” *Phuongnam24h.com*, 2023, phuongnam24h.com/dong-dat-la-gi.html. Accessed 18 Apr. 2025.
- [3] Hoài N. (2025, March 29). Những trận động đất lớn nhất ở Việt Nam thế kỷ qua. Báo Điện Tử Tiền Phong. <https://tienphong.vn/nhung-tran-dong-dat-lon-nhat-o-viet-nam-the-ky-qua-post1729297.tpo>
- [4] Tất cả về hệ thống giám sát địa chấn - IMV CORPORATION. (n.d.). IMV CORPORATION. <https://we-are-imv.com/vi/support/library/seismograph/>

Tiếng Anh:

- [5] Abbadia, Jessica. “Understanding Earthquake Magnitudes and Their Impact.” *Mind the Graph Blog*, 27 Feb. 2023, mindthegraph.com/blog/earthquake-magnitude/.
- [6] Panchuk, Karla. “12.2 Seismic Waves and Measuring Earthquakes.” *Opentextbc.ca*, 20 Aug. 2021, opentextbc.ca/physicalgeologyh5p/chapter/seismic-waves-and-measuring-earthquakes/.

- [7] *IEEE Xplore Full-Text PDF*: (n.d.).
<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=106233>
44
- [8] *Data Download after Search for Data*. (n.d.).
https://www.kyoshin.bosai.go.jp/kyoshin/data/index_en.htm
- [9] *CrowdQuake+: Data-driven earthquake early warning via IoT and deep learning*. (2021, December 15).
IEEE Conference Publication | IEEE Xplore.
<https://ieeexplore.ieee.org/document/9671971>