■Ngày nhận bài: 21/3/2024 ■Ngày sửa bài: 15/4/2024 ■Ngày chấp nhận đăng: 06/5/2024

Phân loại ứng xử đất dựa trên kết quả thí nghiệm CPT

CPT based soil behavior classification system

> PHẠM NGUYỄN LINH KHÁNH[1,2*], TRẦN CAO THANH NGỌC[1,2]
[1]Khoa Kỹ thuật và Quản lý xây dựng, Trường ĐH Quốc tế - Đại học Quốc gia TP.HCM
[2]Đại học Quốc gia TP.HCM; *Email: pnlkhanh@hcmiu.edu.vn

TÓM TẮT

Thí nghiệm xuyên tĩnh (CPT) đang dần được áp dụng phổ biến trong thí nghiệm hiện trường nhờ vào khả năng mô tả chính xác ứng xử của đất (SBT) trong thực tế. Ngoài ra, kết quả CPT còn được sử dụng để phân loại ứng xử đất với độ tin cậy cao. Sự gia tăng dữ liệu đất đòi hỏi việc cập nhập thường xuyên các hệ thống phân loại ứng xử đất. Bên cạnh các phương pháp truyền thống, bài báo này ứng dụng giải thuật học không giám sát (unsupervised learning) để nâng cao khả năng nhận dạng các nhóm đất có cùng đặc điểm ứng xử dựa trên bộ dữ liệu bao gồm 1339 CPT với các nhóm đất khác nhau.

Keywords: CPT; ứng xử của đất ; Unsupervised learning.

ABSTRACT

Cone penetration test (CPT) has been routinely applied in field investigation owing to its ability to describe soil behavior (SBT) in practice accurately. Besides, CPT results are also used to classify soil behavior with high reliability. The increase in soil data requires regular updating of soil behavior classification systems. In addition to traditional methods, this article applies unsupervised learning algorithms to improve the ability to identify soil groups with similar behavioral characteristics based on a data set including 1339 CPTs with various soil types.

Keywords: CPT; soil behaviors; Unsupervised learning.

MỞ ĐẦU

Hệ thống phân loại đất cung cấp cơ sở để chia tách đất thành các nhóm có tính chất cơ lý tương tự nhau. Phân loại dựa trên thành phần hạt đang được sử dụng phổ biến trong hầu hết các hoạt động xây dựng và địa kỹ thuật. Bên cạnh các ưu điểm tiện lợi và kinh tế, cách phân loại này giới hạn khả năng nhận biết ứng xử của đất vì không xét đến ảnh hưởng của cấu trúc và trạng thái đất trong tự nhiên. Gần đây, hệ thông phân loại ứng xử đất (SBT) dựa trên kết quả thí nghiệm xuyên tĩnh (CPT) [1, 2] đã trở nên phổ biến hơn nhờ khả năng mô tả chính xác các đặc tính cơ lý của đất tại hiện trường.

Các hệ thống phân loại ứng xử đất dựa trên kết quả CPT (CPT-SPT) hiện nay được xây dựng dựa trên quan sát thực nghiệm, trong đó ranh giới giữa các nhóm đất được ước lượng thông qua một số lượng mẫu giới hạn. Khi dữ liệu đất ngày càng gia tăng, dẫn tới yêu cầu cập nhập các giới hạn phân chia nhóm đất.

Sự ra đời giải thuật khai thác dữ liệu cùng với sự xuất hiện các bộ dữ liệu địa chất tạo điều kiện phát triển lời giải có độ tin cậy cao cho các bài toán trong lĩnh vực xây dựng và địa kỹ thuật. Những nghiên cứu gần đây chứng minh sự hiệu quả của giải thuật học không giám sát (unsupervised learning) trong phân nhóm dữ liệu, kết quả thu được cung cấp góc nhìn dữ liệu nâng cao độ hiểu biết trong bài toán [3].

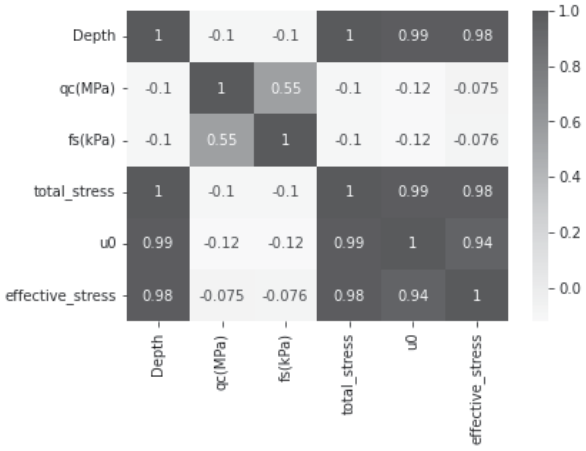
Bài báo này ứng dụng mô hình Gaussian mixture, một trong những giải thuật unsupervised learnign phổ biến hiện nay, để phân nhóm đất dựa trên 1339 mẫu CPT [4], thu thập tại các khu vực khác nhau. Kết quả phân loại được so sánh với đồ thị phân loại Robertson [1].

1. DỮ LIỆU CPT

Bộ dữ liệu được công bố bởi Rauter et al. [4] với 1339 kết quả CPT khác nhau (CPT, CPTu, SCPT, SCPTu) thu thập tại nhiều khu vực ở Áo và Đức. Để đơn giản hóa, bài báo này chỉ sử dụng dữ liệu CPTu và SCPTu để thành lập mô hình phân loại. Bảng 1 trình bày mô tả thống kê các thông số thuộc bộ dữ liệu. Hình 1 mô tả mức độ tương quan giữa các thông số trong bộ dữ liệu. Dễ dàng nhận thấy, sự tương quan cao giữa ứng suất hữu hiệu, ứng xuất tổng và áp lực nước lỗ rỗng với độ sâu được lý giải bởi ý nghĩa vật lý của các thông số này. Trong khi đó, các cặp thông số còn lại không cho thấy độ tương quan đáng kể.

Bảng 1: Mô tả thông số thống kê cho các mô hình thuộc bộ dữ liệu

Thông số	Trung bình	Độ lệch chuẩn	Min	Max
Độ sâu (m)	12.4	10.43	0.01	103
q_c (MPa)	5.57	8.48	-8.61	122.9
Hệ số ma sát dọc f_s (kPa)	64.56	254.2	-100	47436
Áp lực nước lỗ rỗng u_0 (kPa)	114.86	101.48	0	1010.43
Ứng suất hữu hiệu σ_{vo}' (kPa)	120.83	10.43	0.09	946.57



Hình 1. Độ tương quan giữa các thông số trong bộ dữ liệu.

Trước khi thiết lập mô hình, bộ dữ liệu được xử lý lọc nhiễu do sai xót trong quá trình ghi chép hoặc thí nghiệm. Các bước xử lý dữ liệu bao gồm: (1) loại bỏ các dữ liệu nhiễu, (2) chuẩn hóa bộ dữ liệu trong khoảng 0-1 nhằm nâng cao khả năng làm việc của mô hình Gaussian mixture.

2. MÔ HÌNH GAUSSIAN MIXTURE

Mô hình Gaussian mixture (GMM) là một phương pháp phân cụm phổ biến nhằm mô hình hóa dữ liệu dưới dạng hỗn hợp các phân bố Gaussian. Việc phân cụm được thực hiện bằng cách gán phân bố xác suất cho từng nhóm dữ liệu, cho phép phân cụm linh hoạt và chính xác hơn các phương pháp khác. GMM sử dụng phân phối chuẩn (công thức 1) mô tả dạng tổng quát của hàm mật độ xác suất cho các nhóm dữ liệu.

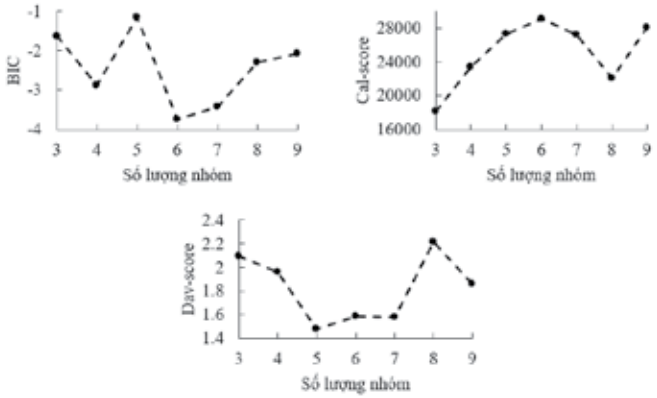
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

Trong đó: μ trung bình của phân phối, σ là độ lệch chuẩn

Hai phương pháp phổ biến được sử dụng trong xác định số lượng nhóm dữ liệu bao gồm: đo lường bên ngoài và đo lường nội bộ. Việc đo lường bên ngoài dựa trên tính đồng nhất, tính hoàn chỉnh, và chỉ số rand. Tuy nhiên, phương pháp này đòi hỏi dữ liệu được phân loại sẵn để so sánh, đánh giá, do đó sẽ có độ chính xác kém hơn cũng như sai lệch nhiều hơn [1]. Đo lường nội bộ đánh giá dựa trên các chỉ số cơ bản: Calinski-Harabasz (Cal-score), Davies-Bouldin (Dav-score), và Bayesian Information criterion (BIC). Phương pháp này dựa trên sự gắn kết và tách biệt của từng nhóm dữ liệu nên có xu hướng ít sai lệch và chính xác hơn.

3. KẾT QUẢ

Hình 2 thể hiện tương quan số lượng nhóm và sự biến thiên các tham số đo lường (BIC, Cal và Dav). Căn cứ kết quả đạt được, nghiên cứu này lựa chọn số lượng nhóm đất tối ưu là 6. Bảng 2 trình bày so sánh sự tương ứng giữa các nhóm đất phân loại theo Gaussian mixture và phương pháp đồ thị Robertson.



Hình 2. Kết quả đánh giá số lượng nhóm tối ưu mô hình Gaussian mixture

Bảng 2: So sánh sự tương đồng giữa các nhóm đất phân loại theo Gaussian mixture và Robertson [1]

Gaussian mixutre	Ứng xử đất dựa trên CPT	Ứng xử đất dựa trên CPTu	Ứng xử đất dựa trên SCPT	Ứng xử đất dựa trên SCPTu
Nhóm 1	Claylike Contractive (CC)	Claylike Contractive, Claylike Contractive Sensitive (CC, CCS)	Sandlike Dilative (SD)	Sandlike Contractive (SC)
Nhóm 2	Transition Contractive and Transition Dilative (TC, TD)	Sandlike Contractive, Transition Contractive (SC, TC)	Claylike Contractive Sensitive (CCS)	Claylike Contractive, Claylike Contractive Sensitive (CC, CCS)
Nhóm 3	Claylike Contractive (CC)	Claylike Dilative (CD)	Claylike Contractive (CC)	Transition Contractive (TC)
Nhóm 4	Sandlike Contractive (SC)	Sandlike Dilative, Transition Dilative (SD, TD)	Claylike Dilative, Transition Dilative (CD, TD)	Sandlike Dilative (SD)
Nhóm 5	Sandlike Dilative (SD)		Sandlike Contractive, Transition Contractive (SC, TC)	Claylike Dilative, Transition Dilative (CD, TD)
Nhóm 6	Claylik Contractive Sensitive (CCS)			

4. KẾT LUẬN

Bài báo giới thiệu ứng dụng giải thuật Gaussian mixture trong phân loại ứng xử đất dựa trên dữ liệu CPT bao gồm nhiều loại đất khác nhau. Kết quả phân loại dựa trên Gaussian mixutre có nhiều tương đồng so với phương pháp đồ thị Robertson hiện đang được sử dụng phổ biến. Ưu điểm của phương pháp đề suất là khả năng xử lý nhanh, cho phép phân loại cơ sở dữ liệu lớn, nâng cao khả năng ứng dụng trong thực tế. Qua đó, hướng tới xây dựng một hệ thống xử lý, phân loại và chia sẽ dữ liệu CPT với quy mô lớn.

TÀI LIỆU THAM KHẢO

1. Robertson PK (2016) Cone penetration test (CPT)-based soil behaviour type (SBT) classification system - An update. Can Geotech J 53:1910-1927. https://doi.org/10.1139/cgj-2016-0044.

2. Robertson PK (1990) Soil classification using the cone penetration test. Can Geotech J 27:151-158. https://doi.org/10.1139/t90-014.

3. Sresakoolchai J, Kaewunruen S (2022) Railway defect detection based on track geometry using supervised and unsupervised machine learning. Struct Heal Monit 21:1757-1767.

4. Rauter S, Tschuchnigg F (2021) Cpt data interpretation employing different machine learning techniques. Geosci 11. https://doi.org/10.3390/geosciences11070265.