

ITALIANO

Argomento/Titolo

Sviluppo di modelli di apprendimento automatico per ottimizzare l'analisi di dataset geologici multidimensionali e ricostruire processi geologici complessi

Proponente (Tutor)

Prof.ssa Ester PIEGARI

Proposta di ricerca (MAX 600 PAROLE)

La principale domanda scientifica su cui la proposta di dottorato si fonda è se e in che modo le tecniche di machine learning possono ottimizzare le informazioni ricavabili dai carotaggi marini profondi e contribuire alla ricostruzione di fenomeni geologici complessi quali l'evoluzione della circolazione di acque profonde e l'interazione con processi sedimentari dei margini continentali.

L'indiscutibile aumento del volume e della complessità dei dati geologici oggi disponibili evidenzia l'importanza delle tecniche di machine learning, non solo per estrarre quante più informazioni utili possibili, ma anche per ottenere nuove intuizioni dai dati (Bergen et al., 2019). Alcuni algoritmi di ML possono essere utilizzati per automatizzare l'analisi di grandi dataset, facilitando i compiti di classificazione, mentre altri permettono una comprensione più profonda dei processi scoprendo nuove relazioni e pattern nascosti all'interno dei dati (Bhattacharya & Di, 2022; Uddin et al., 2022). In particolare, l'apprendimento non supervisionato è particolarmente adatto per l'analisi esplorativa dei dati e la loro visualizzazione in spazi multidimensionali. Diversi approcci di clustering hanno dimostrato il potenziale di raggruppare tipi diversi di dati, identificando pattern e strutture senza etichettatura preventiva e facilitando così l'individuazione di

caratteristiche geologiche (La Marca & Bedle, 2022; Piegari et al., 2023; Kumar et al., 2024). Le tecniche di riduzione della dimensionalità applicate ai cluster in spazi ad alta dimensionalità semplificano dataset complessi mantenendo le informazioni essenziali, suggerendo al tempo stesso gerarchie nei parametri da attenzionare. Poiché gli approcci di apprendimento non supervisionato possono rivelare connessioni inesplorate all'interno dei dati, il loro utilizzo è cruciale e promettente per identificare teleconnessioni, dato che questi pattern si manifestano spesso come correlazioni tra diverse località geografiche e periodi temporali. Lo scopo del progetto è appunto quello di esplorare approcci non supervisionati per migliorare l'analisi di dataset multidimensionali di grandi dimensioni relativi in particolare a dati scientifici di performance oceanica per ottimizzare l'estrazione di informazioni utili alla ricostruzione di processi geologici (Rodrigues et al., 2022). In particolare, i principali obiettivi che il progetto si propone di perseguire possono essere sintetizzati nei seguenti punti:

- Integrazione dei dati: raccogliere e integrare diversi set di dati geologici derivanti da attività scientifica di perforazione oceanica, comprendenti informazioni litologiche, geofisiche e geochemiche.
- Sviluppo di modelli: realizzare modelli di machine learning volti a classificare le caratteristiche geologiche e a ricostruire la storia geologica attraverso l'identificazione di pattern e correlazioni nei dati.
- Valutazione delle prestazioni: analizzare l'efficacia di diversi algoritmi di machine learning in termini di accuratezza e di efficienza computazionale.

Bibliografia

Bergen KJ et al (2019) Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363, 6433

Bhattacharya S & Di H (2022) *Advances in Subsurface Data Analytics*, Elsevier

(2022)

Kumar PC et al (2024) Unsupervised learning approach for revealing subsurface tectono-depositional environment: A study from NE India. Journal of Applied Geophysics, 229, 105478

La Marca, K & Bedle H (2022) Deepwater seismic facies and architectural element interpretation aided with unsupervised machine learning techniques: Taranaki Basin, New Zealand. Marine and Petroleum Geology, 136, 105427

Piegari E et al (2023) A machine learning-based approach for mapping leachate contamination using geoelectrical methods. Waste Management, 157, 121 – 129

Rodrigues, S, Hernández-Molina, FJ, Larter, RD, Rebesco, M, Hillenbrand, CD, Lucchi, RG, Rodríguez-Tovar, FJ (2022) Sedimentary model for mixed

depositional systems along the Pacific margin of the Antarctic Peninsula:

Decoding the interplay of deep-water processes,

Marine Geology, 445, 106754

Uddin S., et al. (2022) Machine learning in project analytics: a data-driven framework and case study. Sci Rep 12, 15252

Programma di ricerca

I° Anno

Raccolta e integrazione dati: stato dell'arte; acquisizione e organizzazione di dataset geologici multidimensionali provenienti da perforazioni oceaniche, includendo dati litologici, geofisici e geochimici; preprocessing dei dati per garantire omogeneità e qualità dei dati.

Formazione e sviluppo iniziale di modelli di ML: studio e implementazione di algoritmi di machine learning, con particolare attenzione all'apprendimento non supervisionato; sperimentazione preliminare su dataset di prova per validare metodologie.

Partecipazione a conferenze e networking: partecipazione, ed eventuale presentazione di un poster e/o talk, ad almeno una conferenza nazionale o internazionale; inizio della pianificazione del periodo di ricerca all'estero.

II° Anno

Sviluppo di modelli ML e analisi dati: sviluppo di modelli di ML per classificare caratteristiche geologiche e identificare pattern e correlazioni nei dati multidimensionali.

Periodo di ricerca all'estero: soggiorno presso un istituto di ricerca o università all'estero con specializzazione in machine learning applicato alle Geoscienze, per collaborazioni e formazione avanzata.

Prima pubblicazione scientifica: redazione e sottomissione del primo articolo su rivista peer-reviewed, focalizzato su metodologie e primi risultati ottenuti con i modelli ML applicati ai dataset utilizzati.

III° anno

Valutazione delle prestazioni e ottimizzazione: analisi comparativa dell'efficacia degli algoritmi utilizzati; ottimizzazione dei modelli in base ai risultati e feedback ricevuti.

Seconda pubblicazione scientifica e diffusione dei risultati: preparazione e sottomissione di un secondo articolo scientifico; presentazione dei risultati a conferenze internazionali.

Stesura della tesi: completamento della scrittura della tesi di dottorato con integrazione di tutte le fasi di ricerca e i risultati ottenuti.