Goldschmidt 2025 Abstract
https://doi.org/10.7185/gold2025.31508

Integrating Geoscience Text Data into Mineral Prospectivity Mapping Using BERT-Based Natural Language Processing

NYAH BAY[1,2], MOHAMMAD PARSA[2] AND ANDREI SWIDINSKY[1]

[1]University of Toronto
[2]Geological Survey of Canada

Regional geochemical datasets are necessary for delineating metallogenic provinces and guiding mineral exploration. However, integrating geochemical data with geological knowledge remains a challenge due to issues such as machine-readability of text data found in geoscience reports, field notes, exploration assessment reports, and drill core logs. These documents contain valuable contextual information about mineral systems, yet their unstructured format limits direct application in downstream geoscience machine learning applications, such as data-driven mineral prospectivity mapping (MPM). Recent advancements in natural language processing (NLP) provide new opportunities to extract, structure, and utilize this knowledge in geoscience. This study proposes the use of Bidirectional Encoder Representations from Transformers (BERT) to process and analyze geochemical text data. We will develop a domain-adapted BERT model trained on Canadian geoscience and geochemistry literature, enabling the embedding of geochemical terminology into high-dimensional vector spaces that capture semantic relationships. This model will extract meaningful patterns associated with mineralized systems by identifying co-occurrence patterns and contextual linkages between element concentrations, lithological units, and alteration styles. Our methodology will integrate network analysis, principal component analysis, and cosine similarity measures to uncover the link between geochemical and geological trends relevant to exploration targeting. Additionally, we will assess the utility of these text-derived embeddings in constructing mineral prospectivity maps by linking them with spatial geochemical datasets. By using transformer-based NLP models, this study aims to establish a framework for incorporating geochemical text data into mineral exploration workflows.