12th International Kimberlite Conference Extended Abstract No. 12IKC-47, 2024

Machine learning classification on the chemical compositions of lithospheric diamonds and their inclusions

Jiali Lei, J ZhangZhou, ZJU Earth Data Group

Key Laboratory of Geoscience Big Data and Deep Resource of Zhejiang Province, School of Earth Sciences, Zhejiang University, Hangzhou, China,

Introduction

The classification of diamonds based on their chemical data is a pivotal aspect of gemological and geological research. Diamonds, formed under extreme conditions within the Earth's mantle, carry within them a record of the geochemical processes of their environment. This study utilizes a comprehensive dataset compiled from 304 publications, consisting of 12,138 entries for diamonds and 12,080 entries for mineral inclusions. By applying machine learning techniques, particularly the Light Gradient Boosting Machine (LightGBM) (Ke et al., 2017), this research aims to explore and classify the geochemical signatures of diamonds from various cratons worldwide, providing insights into their formation and provenance.

Data Compilation and Preprocessing

Broadly, our dataset was compiled from two primary sources: 1) the research of Stachel et al. (2022a, 2022b) draws from 138 publications, mainly doctoral dissertations. This subset comprises 6,520 entries for diamonds and 8,108 for associated inclusions. 2) Data from 166 published journal papers on natural diamonds, which comprises 5,618 entries for diamonds and 3,972 for associated inclusions.The dataset encompasses a wide range of chemical data, including nitrogen concentrations and isotopic ratios of carbon and nitrogen Additional information such as geographic coordinates, the ages of host rocks, and types of diamond inclusions (Peridotite-type & Eclogite-type, or P-type & E-type) were also compiled. The data was rigorously verified for consistency with original publications by the ZJU Earth Data Group.

Machine Learning Model Results

Given the prevalence of missing values—a common issue in geological datasets—LightGBM was chosen for its efficacy with sparse data. This algorithm utilizes decision trees as base learners and builds models sequentially, aiming to correct errors in previous iterations, thus optimizing the predictive accuracy.

The primary objective of the machine learning model was to predict the cratons of origin based on the chemical data of diamond inclusions. The features included major elements like SiO_2, TiO_2, and MgO, among others. The model training involved dividing the dataset into an 80% training subset and a 20% testing subset, with stratified sampling to maintain representation of all classes.

Model optimization was conducted using a grid search technique, which fine-tunes the algorithm parameters to enhance model performance. Performance metrics such as Recall and Precision were employed to evaluate the effectiveness of the model. Recall measures the model's ability to correctly identify true instances of a class, while Precision assesses the accuracy of these identifications.
The LightGBM model demonstrated substantial recall and precision in the training subset, though these metrics were slightly lower in the testing subset, indicating challenges in generalization. The analysis of

the confusion matrix revealed specific insights into the predictive accuracy per craton, particularly highlighting that diamonds from the Kaapvaal Craton were most frequently correctly predicted.

Moreover, the study also included a correlation analysis, establishing a similarity model among different cratons. This model compared the 'statistical centers'—akin to the core of an apple—and 'inner structures' or the textural context of the surrounding flesh. This metaphor illustrates how diamonds from similar cratons share common features in their elemental makeup, much like how the texture and core of apples might vary subtly but discernibly between different varieties.

Statistical and Correlation Analysis Insights

The correlation analysis was instrumental in further dissecting the relationships within the chemical data. For example, a strong positive correlation was observed between MgO and NiO contents in inclusions, suggesting a similarity in olivine content, which is a common mineral in the Earth's mantle. Such correlations help in understanding the mineralogical environment of the diamonds, which in turn reflects on their craton of origin.

Discussion on Craton Similarities and Differences

Significant findings from the correlation analysis indicated that diamonds from the Kaapvaal, Amazon, Siberia, and Slave Cratons showed higher similarity scores. In contrast, those from Kimberley and North China displayed considerable differences. This suggests that despite the global distribution of these cratons, there are distinct geochemical environments influencing diamond formation.

Conclusion and Future Directions

The use of machine learning in classifying diamonds by their chemical signatures has opened new avenues in the study of Earth's deep geological processes. The LightGBM model, complemented by rigorous statistical analysis, provides a framework for understanding the complex interplay of elements that form diamonds. Moving forward, enhancing the dataset with more geochronological data, expanding the range of cratons studied, and incorporating other machine learning techniques could further refine our understanding and classification of diamonds. This research not only aids in academic pursuits but also has practical implications for mining industries and gemological assessments, making it a cornerstone for future explorations into the Earth’s interior.

References

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 30.

Stachel, T., Aulbach, S., & Harris, J. W. (2022a). Mineral inclusions in lithospheric diamonds. Reviews in Mineralogy and Geochemistry, 88(1), 307-391. http://dx.doi.org/10.2138/rmg.2022.88.06

Stachel, T., Cartigny, P., Chacko, T., & Pearson, D. G. (2022b). Carbon and nitrogen in mantle-derived diamonds. Reviews in Mineralogy and Geochemistry, 88(1), 809-875. http://dx.doi.org/10.2138/rmg.2022.88.15