Higher-Order Pattern Recognition for Materials Informatics using Explainable Artificial Intelligence

Amanda S Barnard[1], Tommy Liu[1]
[1]School of Computing, Australian National University, 145 Science Road, Acton, Australia

amanda.s.barnard@anu.edu.au

The combination of rational machine learning with creative materials science makes materials informatics a powerful way of discovering, designing, and screening new materials. However, moving from a promising prediction to a practical strategy often requires more than just an instructive structure/property relationship. Understanding how a machine learning method uses the information captured in the data to predict the target properties can often be critical. Explainable artificial intelligence (XAI) is an emerging field in computer science based in statistics that can augment materials informatics workflows. XAI can be used as a forensic analysis technique to understand the consequences of data, model, and application decisions, or as a model refinement method capable of distinguishing important information [1]. This approach is often applied to the feature space to explain the how the structural characteristics of materials contribute to the property prediction, using tools such as feature rankings to identify useful or nuisance variables. However, an alternative approach is to apply similar methods to the instance space, using different tools to identify influential or unproductive data points. In this presentation we investigate these opportunities, by exploring high-dimensional patterns structured (tabular) materials data sets in behavioral space instead of the feature space. Behavioral vectors are used to represent the contribution of an instance to an interpretable quantity (such as a material property), which complement more conventional interpretations of machine learning models. We make use of Shapley values to decompose well-studied summary statistics of the data to give rise to different interpretable clusterings and modalities compared to the original data. Our approach is efficiently demonstrated qualitatively and quantitatively over three material data sets and uncovers hidden characteristics that may aid the data analysis process. An example of projections into three behavioral spaces is shown in Figure 1. Once sets of influential instances (materials) have been identified, we also decompose the residuals of regression with respect to the data instances, to determine the effects of each individual instance on the model and each other [2,3]. This concept is illustrated in Figure 2. This provides a model-agnostic method of identifying instances of interest, which can determine the appropriateness of the model and data in the wider context of a given study, as well as providing unique material insights.

References

[1] T. Liu, A. S. Barnard, Cell Rep. Phys. Sci., 4 (2023) 101630.

[2] T. Liu, A. S. Barnard, Proceedings of the 40th International Conference on Machine Learning, 202 (2023) 21375.

[3] T. Liu, Z. Y. Tho, A. S. Barnard, Digital Disc., 3 (2024) 422-435.

Figures

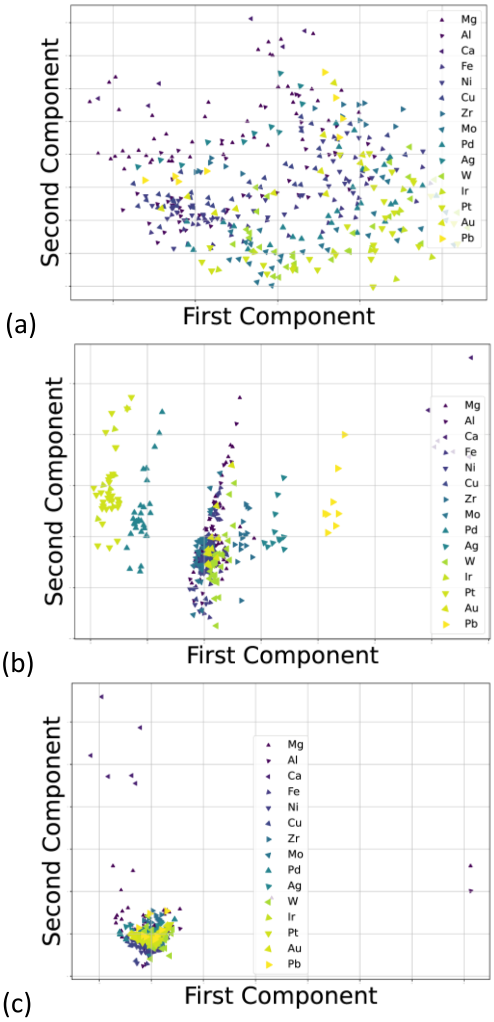

Figure 1. Materials in the Dilute Solute Data set showing the (a) raw data in “mean-space”, and projections in (b) “variance space”, and (c) “skewness-space”.

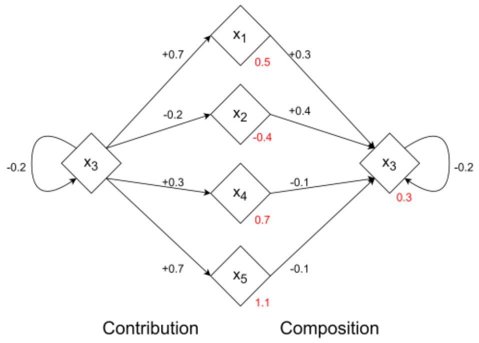

Figure 2. Contribution and composition framework for five samples. Red values indicate the residual value the model produces for each instance, black indicates the effects the instance has upon the others.