Novel machine learning-based cluster analysis method that leverages target material property

August 6 2024

Target Material Property-Dependent Cluster Analysis of Inorganic Compounds
Sato et al. (2024) | Advanced Intelligent Systems

This new method identifies promising material groups for specific applications and key factors for desirable target properties. Credit: Tokyo Tech

In materials science, substances are often classified based on defining factors such as their elemental composition or crystalline structure. This classification is crucial for advances in materials discovery, as it allows researchers to identify promising classes of materials and explore new ones with similar functions and properties.

A recent *Advanced Intelligent Systems* study led by Researcher Nobuya Sato and Assistant Professor Akira Takahashi from Tokyo Institute of Technology developed a new machine learning-powered clustering technique. This technique groups similar materials by taking into account both their basic characteristics and target properties.

Advances in machine learning have made the classification process significantly less tedious and also opened up efficient ways of predicting materials with interesting properties based on basic features of chemical compositions and crystal structures. Cluster analysis, a commonly used machine-learning technique, uses these basic features to not only categorize materials and summarize similarities between them but also provide information regarding relationships between materials belonging to the same group.

While this represents significant progress toward discovering new materials with unique functionalities, conventional clustering techniques often fail to consider target material properties, such as band gaps and dielectric constants, which are related to these basic features.

But why is it important to include target properties for cluster analysis of materials?

Takahashi explains, "If we try to categorize semiconductors as per width of the band gap and investigate the chemical characteristics of respective categories, analyzing only with the target property wouldn't provide the complete picture. Clustering in terms of the band gap may gather materials into a cluster where some gaps are determined by electronegativity while others are determined by features relevant to covalency.

"Conversely, using only basic features might not cluster materials that are similar in the property of interest. Hence, we need an approach that considers the relationship between basic features and target properties."

To ensure the simultaneous inclusion of basic features and target properties, the researchers input the latter information into the clustering model by the random forest (RF) regression—a supervised learning algorithm that learns the relationship between the inputs and outputs to improve itself. The researchers trained the RF regression model to predict a given targeted property.

Following this, the basic features were transformed into z-vectors—information based on the paths taken by the RF model. And finally, cluster analysis was performed on the transformed z-vectors.

This allowed the researchers to categorize more than 1,000 oxides into material groups based on their basic features like composition and crystal structure alongside target properties such as the formation energy, band gap, and electronic dielectric constant. While this study focused on only single target property cases, the researchers suggest that this new technique could be extended for grouping material based on multiple target properties.

"Our method provides a unique viewpoint for clustering which emphasizes understanding and learning from the relationship between the target property and basic features thus providing unforeseen promising materials group and key factor for desirable material function, and accelerate discovery of new materials with fascinating properties," concluded Takahashi.

Provided by Tokyo Institute of Technology

Citation: Novel machine learning-based cluster analysis method that leverages target material property (2024, August 6) retrieved 7 August 2024 from https://phys.org/news/2024-08-machine-based-cluster-analysis-method.html