

グラフを用いた構造識別子の厳密性と頑強性に関する研究

(東大院工¹⁾) ○谷本 拓¹・村岡 恒輝¹・中山 哲¹

Discriminability and robustness of crystal structure identifiers using graphs

(¹*Graduate School of Engineering, The University of Tokyo*) ○Taku Tanimoto,¹ Koki Muraoka,¹ Akira Nakayama¹

In recent years, numerous materials databases have been constructed, aggregating results from experiments and first-principles calculations. However, each database uses its own unique identifiers, making it challenging to utilize data across multiple databases. To integrate these distributed data, the development of structural identifiers is desired.

We have developed Graph ID, which assigns a unique identifier to crystal structures, enabling rapid identity determination. Graph ID converts a crystal structure into graph structures, repeatedly embedding coordination information into nodes. This method is less susceptible to numerical errors in atomic positions and symmetry breaking, and it can accurately determine topological differences in various crystal structures. We conducted a detailed accuracy verification of Graph ID using broad datasets, evaluating its precision and robustness. As a result, we identified rare pairs of structures that Graph ID could not distinguish. By applying these insights, we developed and validated multiple new structural identifiers with different levels of precision and robustness.

Keywords : *Graph; Materials informatics; Crystal structure; Graph isomorphism*

近年、実験や第一原理計算の結果を集約した材料データベースが多数構築されている¹⁾⁵⁾。しかしながら材料データベースはそれぞれ独自の ID を用いているため、複数のデータベースを横断してデータを利活用することは容易ではない。これらの分散したデータを統合するために、構造識別子を開発することが求められている。

当研究室では結晶構造に一意の識別子を与えることで、同一性を高速に判定する Graph ID を開発した⁶⁾。Graph ID は、結晶構造をグラフ構造に変換し、その頂点の配位情報を情報として埋め込むことで得られる。原子位置の数値的な誤差や対称性の崩れの影響を受けにくく、生成が高速であるという好ましい特徴があり、さまざまな結晶構造のトポロジー的差異を正確に判定することができる。本研究では、Graph ID の精度検証をより幅広いデータセットで詳細に行い、その厳密性と頑強性を評価した。その結果、Graph ID が見分けることのできないごく稀な構造ペアを見出した。その知見を応用することで、異なる厳密性と頑強性を持った複数の新規構造識別子を開発し検証と応用を行った。

1) Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013). 2) Zagorac, D. *et al.* Recent developments in the Inorganic Crystal Structure Database: theoretical crystal structure data and related features. *J. Appl. Crystallogr.* **52**, 918-925 (2019). 3) Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **65**, 1501–1509 (2013). 4) Scheidgen, M. *et al.* NOMAD: A distributed web-based platform for managing materials science research data. *J. Open Source Softw.* **8**, 5388 (2023). 5) Calderon, C. E. *et al.* The AFLOW standard for high-throughput materials science calculations. *Comput. Mater. Sci.* **108**, 233–238 (2015). 6) <https://matfinder.net>