

## 生成 AI 活用による材料開発効率化のための材料データベース自動構築技術

(株式会社 日立製作所<sup>1)</sup> ○我妻 正太郎<sup>1</sup>・露木 雅文<sup>1</sup>

Automatic Construction of a Material Database for Data-Driven Materials Design Using Generative AI (<sup>1</sup>Hitachi, Ltd.) ○Shotaro Agatsuma,<sup>1</sup> Masafumi Tsuyuki<sup>1</sup>

Data-driven materials design relies on large-scale materials databases (DBs). Extracted experimental procedures from patents and papers, structured as graph-like data (experimental recipes), are valuable for predicting material properties. For constructing a DB of experimental recipes, it's best to use a flexible method that adapts to varying procedures based on the materials. This study proposes a method to automatically construct a DB by inputting literature and data structures into generative AI (GPT-4o) to extract experimental recipes (Fig. 1). To determine the optimal input format for the AI, we tested three formats: ① natural language, ② bullet points, and ③ JSON. We extracted recipes with five node types and four edge types from five patents each. Comparing F1 scores with the correct data (Fig. 2), format ③ (JSON) proved most effective, suggesting it as the preferred input format.

**Keywords :** Materials Informatics; Data-driven Materials Design; Large Language Model;

データ駆動型の材料開発には、大規模な材料データベース (DB) が必要である。特に、特許や論文から抽出した実験手順を構造化したグラフ状のデータ (以下、実験レシピ) は、物性値予測などに有用で、ニーズが高い。実験レシピに関する DB 構築には、実験手順は開発対象の材料に応じて異なるため、抽出するデータの構造を柔軟に変えられる手法を適用するのが望ましい。そこで、本研究では、文献と抽出すべきデータの構造を生成 AI (GPT-4o) に入力して実験レシピを抽出させ、DB を自動構築する手法を提案する (Fig. 1 参照)。本報告では、生成 AI に入力するデータ構造の形式を決めるべく、①自然文、②箇条書き、③JSON 形式の 3 つの形式を用い、特許 5 件の実施例から、5 種類のノードと 4 種類のエッジを持つ実験レシピを抽出した。各パターンについての F1 値を正解データとの比較により算出した結果 (Fig. 2 参照)、③が最も良く、入力データ形式は③が適切である見込みを得た。

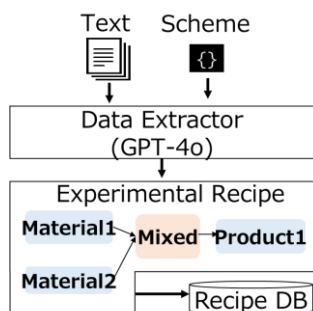


Fig. 1 Overview

|                                                 |  | Scheme (example)                                                                                                                                    | Result |           |        |      |
|-------------------------------------------------|--|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----------|--------|------|
| Scheme Pattern ①:<br>Sentence                   |  | The material,<br>process... could be<br>nodes in the graph                                                                                          |        | Precision | Recall | F1   |
|                                                 |  |                                                                                                                                                     | Node   | 0.90      | 0.78   | 0.84 |
|                                                 |  |                                                                                                                                                     | Edge   | 0.99      | 0.72   | 0.83 |
| <hr/>                                           |  |                                                                                                                                                     |        |           |        |      |
| Scheme Pattern ②:<br>Itemization                |  | Node:<br>•Process: experimental<br>process ...                                                                                                      |        | Precision | Recall | F1   |
|                                                 |  |                                                                                                                                                     | Node   | 0.94      | 0.56   | 0.70 |
|                                                 |  |                                                                                                                                                     | Edge   | 0.59      | 0.38   | 0.46 |
| <hr/>                                           |  |                                                                                                                                                     |        |           |        |      |
| Scheme Pattern ③:<br>JSON<br>(Function Calling) |  | <pre>"ProcessNode": {<br/>  "title":<br/>    "ProcessNode"<br/>  "description":<br/>    "ProcessNode represents<br/>    experimental process"</pre> |        | Precision | Recall | F1   |
|                                                 |  |                                                                                                                                                     | Node   | 0.98      | 0.77   | 0.86 |
|                                                 |  |                                                                                                                                                     | Edge   | 0.98      | 0.83   | 0.90 |

Fig. 2 Data Extraction Result

- 1) Gyoung S. Na., "Artificial Intelligence for Learning Material Synthesis Processes of Thermoelectric Materials," Chemistry of Materials 2023 35 (19), pp.8272-8280.