Knowledge and Data-Driven Framework for Multi-Principal Element Alloys: Automated Knowledge Acquisition, Representation, and Rediscovery

Guangxuan Song[1], Dongmei Fu[1,*], Dawei Zhang[2,*], Lingwei Ma[2]

[1]*Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*

[2]*National Materials Corrosion and Protection Data Center, University of Science and Technology Beijing, Beijing 100083, China*

sguangxuan@ustb.edu.cn

Abstract Machine learning algorithms are advancing material research into a fourth paradigm[1]. However, their application in material science faces significant hurdles, including the expensive acquisition of high-quality data and data scarcity[2]. The introduction of prior knowledge can mitigate these limitations, but the scattered and unstructured nature of material science publications complicates knowledge aggregation. This issue is especially acute in the study of multi-principal element alloys (MPEA)[3], where the complexity of their compositions and structures complicates performance predictions[4]. Efficiently collecting material knowledge and data and incorporating complex interactions within machine learning models to enable new material discoveries presents a significant scientific challenge.

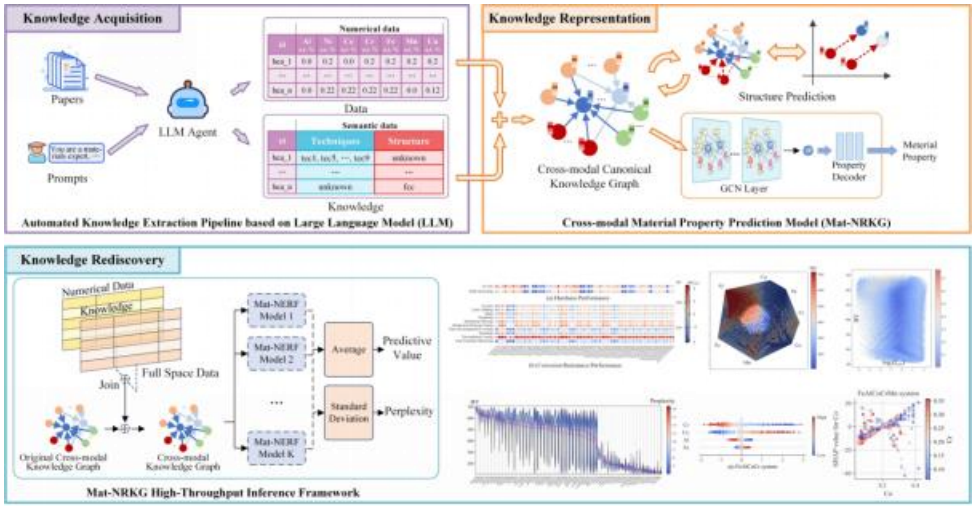

Fig 1. Knowledge and Data-driven Framework for Multi-Principal Element Alloys

To address the aforementioned issues, we propose a knowledge and data-driven framework (Fig 1) to explore MPEA, integrating automated knowledge acquisition, representation, and rediscovery. Initially, we established an automated extraction pipeline based on ChatGPT, guided by material experts, to extract component, property data, and information on processes and structures from MPEA literature, addressing data dispersion and building a diverse dataset of alloy hardness and corrosion. Subsequently, we proposed a cross-modal material property prediction model, Mat-NRKG, that utilizes a cross-modal knowledge graph to integrate experimental data and process information, achieving high-precision performance predictions of MPEA, exceeding the baseline by 18.5% and 13.7% on MSE metrics in hardness and corrosion resistance predictions, respectively. This method enables precise analysis of complex material property using small, real-world datasets. Finally, using the Mat-NRKG model, we performed high-throughput predictions, expediting knowledge discovery via SHAP[5], spatial mapping, and visualization techniques.

Through the "acquisition-integration-utilization-discovery" process, we seamlessly integrate knowledge and data, fully utilizing existing experimental results to deepen our understanding of material performance. This framework not only improves our comprehension of material properties but also expands opportunities for knowledge-driven research in material science, showing significant application potential.

Keywords Multi-Principal Element Alloys, Machine Learning, Knowledge Graph, Material Performance Prediction, Automated Knowledge Engineering

Reference

[1] Himanen L, Geurts A, Foster A S, et al. Data‐driven materials science: status, challenges, and perspectives[J]. Advanced Science, 2019, 6(21): 1900808.

[2] Wei J, Chu X, Sun X Y, et al. Machine learning in materials science[J]. InfoMat, 2019, 1(3): 338-358.

[3] Huang W, Martin P, Zhuang H L. Machine-learning phase prediction of high-entropy alloys[J]. Acta Materialia, 2019, 169: 225-236.

[4] Zhang J, Cai C, Kim G, et al. Composition design of high-entropy alloys with deep sets learning[J]. npj Computational Materials, 2022, 8(1): 89.

[5] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J]. Advances in neural information processing systems, 2017, 30.