AN INTEGRATED MACHINE LEARNING PLATFORM FOR EXPERIMENTAL NANOMATERIALS RESEARCH: PROPERTY PREDICTION, PHENOMENA SIMULATION, AND ASSISTED REASONING

Autores: Gabriel García Zambrano[1*]; Martínez Enríquez Arturo Isaías[1]; Mondaca Espinoza Felipe[2]

Afiliación

[1]Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional - Unidad Saltillo, Doctorado en Nanociencias y Nanotecnología, Saltillo, México.
[2]Universidad San Sebastián, Facultad de Ingeniería, Concepción, Chile.

Resumen

Before synthesizing a candidate nanomaterial, a researcher usually weighs what the literature already reports, what DFT or empirical correlations predict, and what physical intuition from analogous systems suggests. These inputs are scattered across disconnected tools, slowing the decision of whether a candidate is worth pursuing. We developed a machine learning platform that connects these information streams for materials scientists working on oxide and nanocrystalline systems. The platform answers three questions that arise daily in our laboratory: what is the likely band gap or formation energy of this composition, how would a given physical property respond to a parameter change, and what has already been reported for related structures. The predictive layer combines supervised learning with quantum-chemical simulation to produce property estimates; the full model and its benchmarks are presented in a separate paper. For physical interpretation, the user explores canonical nanoscale phenomena through interactive simulations that expose the equations driving each response. A language model grounded on a curated corpus of scientific documents handles literature queries and links each answer to its sources, so a predicted property can retrieve related precedent in the same session. We deployed the platform as a web service and use it internally on oxide and nanocrystalline systems. Predicted band gaps agree with DFT benchmarks at levels detailed in the forthcoming publication. Six canonical phenomena are currently implemented. The grounded language model reduces unsupported statements relative to a retrieval-free baseline. In routine internal use, the three questions above are answered within a single session, rather than requiring separate codes, spreadsheets, and literature searches, as is

Cinvestav

Sociedad Mexicana de Materiales A.C.

typical of early-stage candidate selection. The three modules cover property estimation, physical interpretation, and literature retrieval in a single environment. Current work extends the predictive targets to additional stability-related quantities and expands the corpus to sub-fields requested by users in the group. The same architecture is being adapted for two related material families under study in our laboratory.

Palabras clave: [machine learning for materials science, nanomaterials property prediction, retrieval-augmented generation]

Agradecimientos: The authors thank SECIHTI and the Doctoral Program in Nanoscience and Nanotechnology of CINVESTAV. Support is also acknowledged from Nucleus No. 7 UCN-VRIDT 076/2020 (Scientific Modeling and Simulation Nucleus, NMSC) and the NLHPC supercomputing infrastructure (ECM-02). F.M. was supported by ANID FONDECYT Iniciación (Chile) project 11240822.

Correo del autor de correspondencia: synyg@hotmail.com