Innovative Materials Development with Computer Science, Data Science, and HPC

Materials Informatics (MI) is an attempt to improve the efficiency and sophistication of materials development by incorporating data science methods. While the use of MI in materials development is making headway thanks to advances in machine learning, a major obstacle that remains is a lack of data about the properties of the countless different materials known to exist. To tackle this problem, a team at Sumitomo Chemical lead by Shinya Nishino is using large-scale simulations to create a materials database suitable for machine learning. In this article, we look at this new materials development approach that combines computer science with data science and high performance computing.

Shinya Nishino
Team Leader
Digital & Data Science Innovation Department,
Sumitomo Chemical Co., Ltd.

Materials Informatics (MI): Transforming How Materials are Developed

Until recently, the development of materials for various industrial uses has mostly relied on the wisdom and intuition of experienced researchers and engineers. But, as customers' needs have grown more diverse and R&D time has been curtailed, it is becoming more and more difficult to address these needs with traditional approaches.

Enter Materials Informatics (MI), which brings AI and data science to materials development to drastically shorten development times and discover innovative materials with novel properties. Shinya Nishino at materials maker Sumitomo Chemical has striven to make

Explanation 1 How MI is being used in Materials Development

Materials Informatics can improve common development tasks:

High Throughput Screening

MI can quickly find the most suitable from a group of candidate materials.

Case 1: Using machine learning to screen 1.6 billion molecules to find OLED (Organic Light Emitting Diode) materials with 22% external quantum efficiency.

— R Gómez-Bombarelli, et al., *Nature Materials* 15, 1120-1127 (2016)

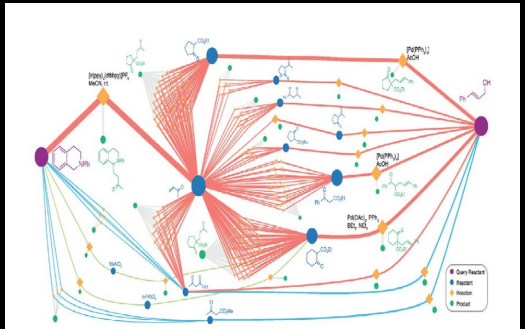
Inverse Problem Approaches

MI can start with the desired properties and work backwards to find the materials that have those properties.

Case 2: Using Bayesian inference to solve an inverse problem and successfully build a system that suggests molecular structures with the targeted properties.

— H. Ikebata, et al., *J. Computer Aided Molecular Design* (2017) 31: 379-391

Retrosynthesis

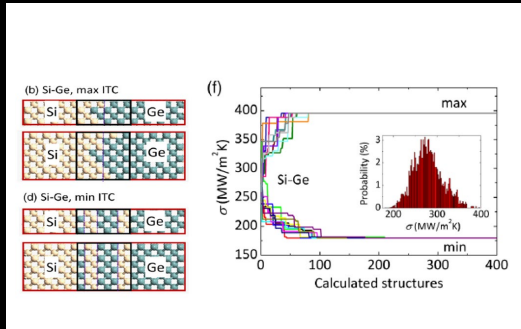


MI can learn from existing data to predict the most efficient pathways (processes) to synthesize the desired molecules.

Case 3: Learning documented synthetic pathways to predict new synthetic pathways with greater than 50% accuracy.

— M. H.S. Segler, M. P. Waller, "Modelling Chemical Reasoning to Predict and Invent Reactions", *Chem. Eur. J.* 2017, 23, 6118, Copyright Wiley-VCH GmbH

Efficient Experiment Design



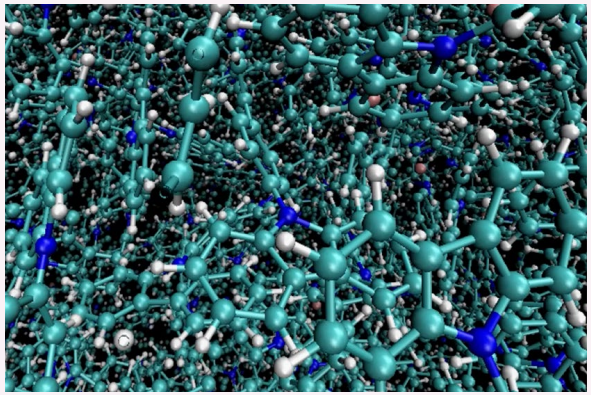
MI can help plan the most efficient experimental approach when investigating a particular property of materials.

Case 4: Using Bayesian optimization on existing data to minimize the number of physical tests in discovering optimal structures for thermoelectric materials.

— S. Ju, et al., *Phys. Rev.* X 7, 021024 (2017)

practical use of MI in the development of materials.

"When a new smartphone model comes out, for example, materials manufacturers have to provide new materials for a better-looking display or a longer battery life. Responding to these requirements with conventional approaches to materials development is time consuming, and finding new materials is becoming ever more difficult. Materials manufacturers are becoming increasingly aware of the need for a different approach to materials development, and MI is one of the answers."



The Bottleneck is a Lack of Data on Materials

In Materials Informatics, a prediction model is created through machine learning. That model is then used to predict the synthetic pathways and other properties of candidate materials. What is important here is that the prediction model can replace the intuition and experience of researchers and engineers, and that a database of material properties can be used to create the model. According to Nishino, however, the availability of such materials databases lags far behind those for other fields, such as drug discovery.

“Big data approaches are currently focused on fields with data that is more readily available. Unfortunately, there is very little data in the field of materials. If you try to find materials data with the features and process conditions you want for a particular development target, you may find that there are only 100 data sets at most. This lack of materials data is the biggest hurdle to implementing Materials Informatics.”

To address this problem, Nishino had the idea of running large-scale computer simulations to automatically create a database of materials and their predicted properties. First, they run computer simulations to understand the mechanism behind a particular property of a material. Next, they map that property to a numeric value known as a “descriptor”. Those descriptors can then be used to predict the properties of other materials at a lower computational cost. And then machine learning is applied to the materials data generated from those descriptors.

“If large-scale simulations on high performance computers can be used to generate materials data that was previously unavailable, we can get a competitive advantage in development. Combining Materials Informatics with that data can make it much more valuable. I expected that a combination of computer science, data science, and HPC could revolutionize materials development.”

A Combination of Large-scale Computer Simulations and Machine Learning

Nishino applied this “materials data from simulations” approach to research on organic light emitting diode (OLED) materials, which are commonly used to illuminate the displays of modern smart-phones and televisions. An OLED material’s ability to emit light is a property known as its “luminescence”. The luminescence is determined by how certain electrons move or “hop” within the molecular structure of a given material, an important phenomenon known as “charge transfer”. To better understand how charge transfer occurs in various OLED materials, Nishino ran large-scale simulations of quantum wave packet dynamics in collaboration with

Takeo Hoshi, an Associate Professor at the Tottori University Graduate School of Engineering. Because these calculations require advanced algorithms and powerful computing resources, they used the Oakforest-PACS supercomputer at the Japan Center for Advanced High Performance Computing, which was the fastest computer available in 2019.

These were “numerical simulations”, mathematical models of physical systems that allowed Nishino to analyze how electrical charges move around inside OLED materials. He discovered that the electrical charge transfers in an OLED material can be predicted using a “physical quantity” called the “inverse participation ratio” (IPR) as a molecular descriptor that indicates the material's properties. Using this IPR descriptor to sort the materials according to their charge transfer property allowed the research team to generate a materials database suitable for use as input for machine learning.

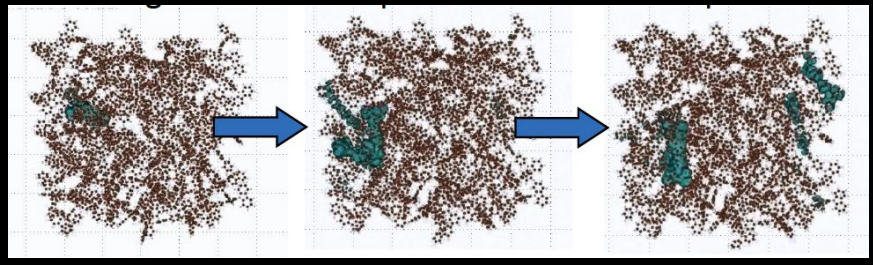
“It’s often better to use a particular property of a material as the machine learning input, rather than just the material's molecular structure. This is because models based on an underlying principle are generally more versatile. At this point, it's important to know what kind of descriptors to put into the model, which is hard to know with pure data science alone. It requires a knowledge of physics and chemistry, or experience working with the material. With that kind of knowledge and experience, it's easier to recognize that a particular section can be reduced to a feature (a physical quantity suitable for machine learning), and that it can be quantified.”

Explanation 2 Large-scale quantum simulations of charge transfers in OLED materials

The so-called “hopping” transfer of electrical charge within a molecule is the crucial physical phenomenon that determines the luminescent properties of a given OLED material. In this study, Nishino used numerical simulations of quantum wave packet dynamics to investigate how electrical charge propagates through the molecules of more than a dozen OLED materials with different combinations of host molecules and guest molecules, known as “host-guest complexes”. These numerical simulations were run by Tottori University Graduate School of Engineering Associate Professor Takeo Hoshi, an expert in large-scale high-performance computing simulations.

The figure below illustrates a charge transfer that occurs in OLED molecules as revealed by these numerical simulations. The simulations showed how the wavefunctions of two complexes overlap when a charge jumps from one complex to the other. This research demonstrated that the charge transfer property for a given OLED material can be predicted by a physical quantity called the “inverse participation ratio” (IPR). Calculating this IPR descriptor is much less computationally expensive than running quantum simulations, thereby lightening the computational load for predicting a material's properties. Nishino says his familiarity and experience with condensed matter physics theory was very helpful in discovering the IPR's usefulness a descriptor.

This series of images taken from the wave packet dynamics simulations shows how the wave packets move. The blue regions show the “hole” wave packets, the location of the charge. We can see that the charge jumps from place to place.



The IPR descriptor that Nishino found in this research is a physical quantity that is covered in the basic theory of condensed matter physics. But he says it has rarely been used in the field of materials science.

“I used to do research in theoretical condensed matter physics, which is how I determined the IPR could be used as a descriptor. The team members on this project include specialists in materials development, computer science, and data science. Plus, we have support from RIST, which is a leading organization for HPCI work. I think that diversity of expertise helped lead to this achievement.”

From Materials Informatics to the Cyber-Physical Systems Loop

Nishino says he wants to use HPCI resources like Fugaku to further apply this new “computer science meets data science meets HPC” approach to materials development.

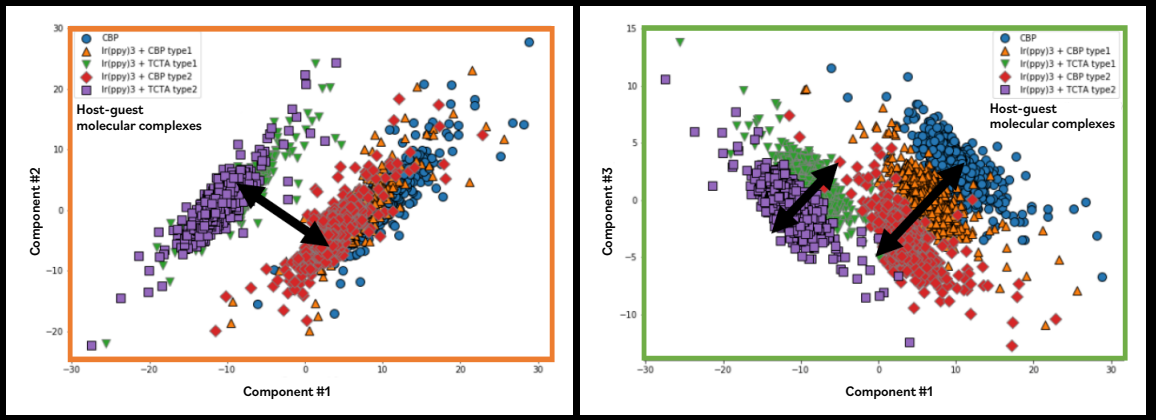
“It's increasingly vital to create data that others can't get, and to understand mechanisms that haven't been understood thus far. So, I think advanced simulations using HPC and data generation via those simulations will become even more important. What I want to work on in the future is a way to create machine learning models for the kinds of materials that are relatively easy to simulate, and then using those models even in areas where it's difficult to acquire data. This is called 'transfer learning'. It's often used in image analysis. But I think it could also be of use in materials development.”

Nishino hopes to create a more advanced and data-driven approach to materials development, one that combines theory, experimentation, simulations, and data science.

Explanation 3

Verifying the categories of materials produced with the IPR descriptor

The figures below show the results of tests to verify whether using the IPR as a descriptor accurately categorized the OLED materials simulated. The IPR is a multi-dimensional parameter, so the “principal component analysis” technique was used to reduce the number of dimensions when verifying the results. The left figure shows changes in the molecular host-guest pair complexes. The right figure shows changes in the concentration of these complexes. Both results show distinct differences for materials with different host-guest complex compositions. This suggests that a machine learning model that uses the IPR descriptor data to categorize materials should be effective in predicting the luminescence properties of OLED materials.



“At this stage, we've combined theory, computer science, and data science. But our goal is to add experimentation to this, generate new theories and new models based on the findings of those experiments, and then generate new data through simulations based on those models. The goal for the next iteration of Materials Informatics is realizing a materials development cycle called the 'cyber-physical systems loop'. And I think my role is in building a system that can sustainably create value by developing human resources who can master MI and the cyber-physical systems loop.”

About the

Researcher

Shinya Nishino says he studied the basic theory of physical properties at university, and worked in materials research which combined that basic theory with computer simulations as a graduate student and post-doctoral researcher. At that time, he and Takeo Hoshi at Tottori University developed the software that was used for these numerical simulations.

Nishino first became interested in Materials Informatics as a postdoctoral researcher. When running numerical simulations of materials for lithium-ion batteries at the request of a Japanese auto maker, he saw how Materials Informatics was being used by competitors abroad to examine candidate materials, and sensed the great potential of this field.

Nishino's comprehensive experience with materials theory, large-scale computer simulations, and actual materials development in practice is likely the key to his success as a pacesetter in Materials Informatics.

Associated Research Projects

- “Organic Semiconductor Material Design using Large-Scale Quantum Simulations and Machine Learning” (hp190066)
- “Design of Organic Semiconductor Materials by Machine Learning Model built from Big Data Generated by Large Scale Quantum Chemical Calculations” (hp200049)

HPCI magazine

富岳百景
FUGAKU HYAKKEI

HPCI magazine FUGAKU HYAKKEI vol.4

Registered Institution for Facilities Use Promotion /
The Representative for HPCI Operation

**Research Organization
for Information Science and Technology**

1-5-2, Minatojimaminami-machi, Chuo-ku, Kobe, Hyogo, 650-0047 JAPAN
Tel: +81-78-599-9511

HPCI Portal Site
https://www.hpci-office.jp
/folders/english

HPCI PR site
https://fugaku100kei.jp/e/

Issued: September 2021