マテリアルズインフォマティクス概説

（情報・システム研究機構 統計数理研究所[1]・物質・材料研究機構[2]）○吉田 亮[1,2]
Overview of Materials Informatics ([1]*The Institute of Statistical Mathematics, Research Organization of Information and Systems*, [2]*National Institute for Materials Science*) ○Ryo Yoshida[1,2]

Keywords: Machine Learning, Forward and Inverse Problems, Small Data, Molecular Design, Synthetic Route Design

In general, the parameter space of materials science is quite huge. For example, the chemical space of small organic molecules is known to consist of more than 10^{60} candidates. In addition, the dimensionality of the parameter space explodes with the addition of other design parameters such as process, additives, and solvent selection in practical material development. Most problems that we face come down to the identification of parameters that exhibit desired properties from such a large search space. This is a multi-objective optimization problem. The essential difference to general industrial design lies in the peculiarity and high dimensionality of the parameter space. Conventional materials research has involved a time-consuming, resource-intensive cycle of parameter design based on experience and intuition of human experts, property assessments made by simulation and experiments, and revision of design guidelines. However, there are barriers that cannot be overcome by such traditional approaches. By introducing advanced technologies of data science into such a circulation process, we aim to innovate the way of materials research. This is the mission of the interdisciplinary field called materials informatics (MI). In this talk, I present an overview of MI and some key technologies of machine learning along with the concept of representation, learning, and generation of various materials. Various applications (e.g., polymer design based on Bayesian inference [1,2,3], machine learning for synthetic route planning [4], the integration of computational chemistry and machine learning technology within the framework of adaptive design of experiments, transfer learning to overcome limited data [5,6,7], prediction and computational design of microstructure using deep generative models [8], etc.) will be introduced, covering topics such as forward and inverse problems, representation and generation of materials structure, transfer learning, interpolation and extrapolation, and laboratory automation based on machine learning for adaptive experimental design strategy.

マテリアルズインフォマティクスの問題の多くは, 順問題と逆問題の形式に帰着する．順問題の目的は, 系の入力に対する出力の予測である．物性予測の文脈では，入力は物質（分子，組成，結晶等），出力は物性（エネルギー，電子状態等）に相当する．これに対し，逆問題では文字通り逆方向の予測を行う．すなわち，出力の目標値を設定した上で, それを達成する入力の状態を予測する. データ科学の観点において, これらの計算は，物質・材料の“表現・学習・生成”というタスクに相当する．記述子と呼ばれる特徴ベクトルを用いて物質の構造を“表現”し，データのパターンから構造から物性の数学的写像を“学習”する. さらに, 計算で所望の物性値を有する物質を“生成”し，有望な候補物質や設計パラメータを炙り出す．対象となる入力は，分子，組

成，結晶，混合物，プロセス，合成経路など，問題に応じて多様な形式をとりうる．本講演では，物質・材料の表現・学習・生成というコンセプトに沿って MI の諸問題と要素技術を概説する. 特に, 様々な適用例（ベイズ推論によるポリマー設計[1, 2, 3], 機械学習による合成経路の計画策定[4]，適応的実験計画に基づく計算化学と機械学習の融合，転移学習[5,6,7]，深層生成モデルによる材料組織の予測と設計[8]など）を紹介しながら，順問題と逆問題，表現と生成，転移学習，内挿と外挿，実験計画に基づくラボラトリオートメーションなどの話題を取り上げる．

[1] Ikebata et al. Bayesian molecular design with a chemical language model. J Comput Aided Mol Des. 31(4):379--391 (2017).
[2] Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. npj Comput Mater. 5:66 (2019).
[3] Wu et al. iQSPR in XenonPy: a Bayesian inverse molecular design algorithm. Mol Inform. 39(1-2):1900107 (2020).
[4] Guo et al. Bayesian algorithm for retrosynthesis. J Chem Inf Model. 60(10):4474–4486 (2020).
[5] Yamada et al. Predicting materials properties with little data using shotgun transfer learning. ACS Cent Sci. 5(10):1717-1730 (2019).
[6] Ju et al. Exploring ultrahigh lattice thermal conductivity crystals via feature-based transfer learning. ChemRxiv (2019).
[7] Minami et al. A general class of transfer learning regression without implementation cost. Proc of AAAI (2021).
[8] Banko et al. Predicting structure zone diagrams for thin film synthesis by generative machine learning. Commun Mater 1, 15 (2020).