材料研究における逆問題と統計的機械学習

（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]
Statistical Machine Learning for Inverse Problems in Materials Research ([1]*The Institute of Statistical Mathematics, Research Organization of Information and Systems*, [2]*National Institute for Materials Science*) ○Ryo Yoshida[1,2]

Data-driven science is advancing rapidly and transform the world around us. As such, machine learning has received considerable attention as a key driver to the next frontier of materials science, enabling us to reap substantial time and cost savings in the discovery and development of innovative materials. The objective of designing materials is to identify a set of design variables that exhibits desired properties. Here, we conduct a two-stage workflow that consists of forward and backward predictions. The objective of the forward problem is to predict the properties of any given design variable. The task of the inverse problem is to identify promising design candidates that exhibit desired properties by solving the inverse mapping of the forward model. This talk describes a basic concept and key technologies of machine learning for the inverse materials design, such as, probabilistic inference using generative machine learning models, explainable machine learning, and adaptive design of experiments. Without going into technical details, the essence of these machine learning techniques is illustrated with various applications taken from polymer design [1, 2, 3], computational design of synthetic reaction routes [4], prediction and discovery of new quasicrystals [5], integration of computational chemistry and machine learning based on adaptive design of experiments, transition learning [6, 7, 8, 9], microstructure prediction for composite materials using deep generative models [10]. In particular, I discuss use cases and strategies for obtaining human-interpretable knowledge from black-box models of machine learning.
Keywords :Materials Informatics; Inverse Design; Explainable Machine Learning; Ammonium Iodide

　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．物性探索の文脈では，入力は物質（分子，組成，結晶等），出力は物性に相当する．これに対し，逆問題では，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基づいてマテリアルズインフォマティクスの諸問題と要素技術を概説する．特に，様々な適用例（ポリマー設計[1, 2, 3], 機械学習による合成経路探索[4], 機械学習による新しい準結晶の予測と発見[5], 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[6,7,8,9]，深層生成モデルによる複合材料の微細組織の予測と設計[10]など ※時間の制約上，一部を抜粋）を紹介しながら，順問題と逆問題，表現と生成，転移学習，内挿と外挿などの話題を取り上げる．特に、機械学習のブラックボックスモデルから人間が解釈可能な知識を得るための戦略を議論する．

[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] Liu et al. Machine learning to predict quasicrystals from chemical compositions. Advanced Materials. 33(36):2102507 (2021).

[6] Yamada et al. Predicting materials properties with little data using shotgun transfer learning. ACS Cent Sci. 5(10):1717-1730 (2019).

[7] Ju et al. Exploring diamondlike lattice thermal conductivity crystals via feature-based transfer learning. Physical Review Materials. 5:053801 (2021).

[8] Minami et al. A general class of transfer learning regression without implementation cost. Proceedings of the AAAI Conference on Artificial Intelligence. 35(10):8992-8999 (2021).

[9] Torres et al. Descriptors of intrinsic hydrodynamic thermal transport: screening a phonon database in a machine learning approach. J. Phys. Condens. Matter, in press.

[10] Banko et al. Predicting structure zone diagrams for thin film synthesis by generative machine learning. Commun Mater 1, 15 (2020).
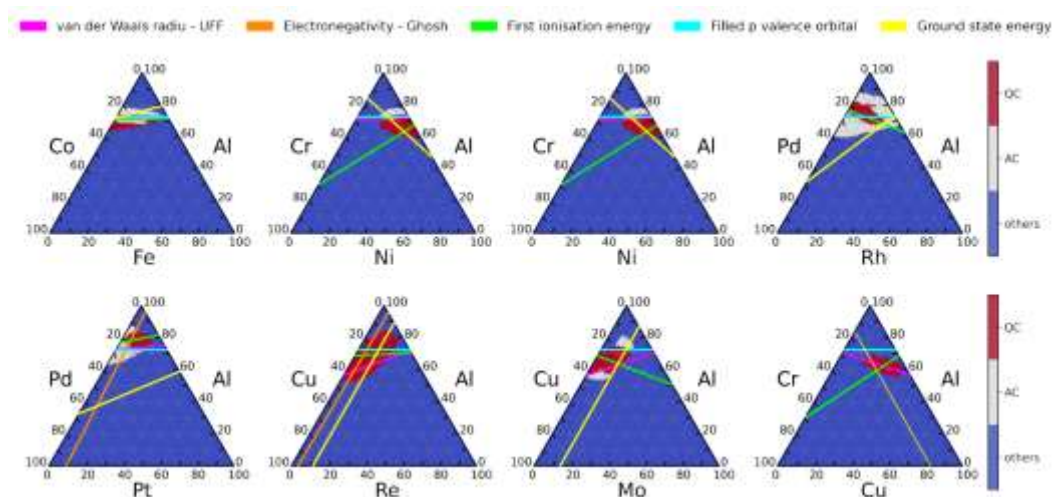
機械学習による準結晶の組成予測（三元系相図）．逆問題を解くことで，機械学習のブラックボックスモデルに内在する入出力のルールを抽出し，熱力学的に安定な準結晶相の形成する五つのルール法則を明らかにした．発見されたルールは，人間が解釈可能な五つの数式で表現される．