

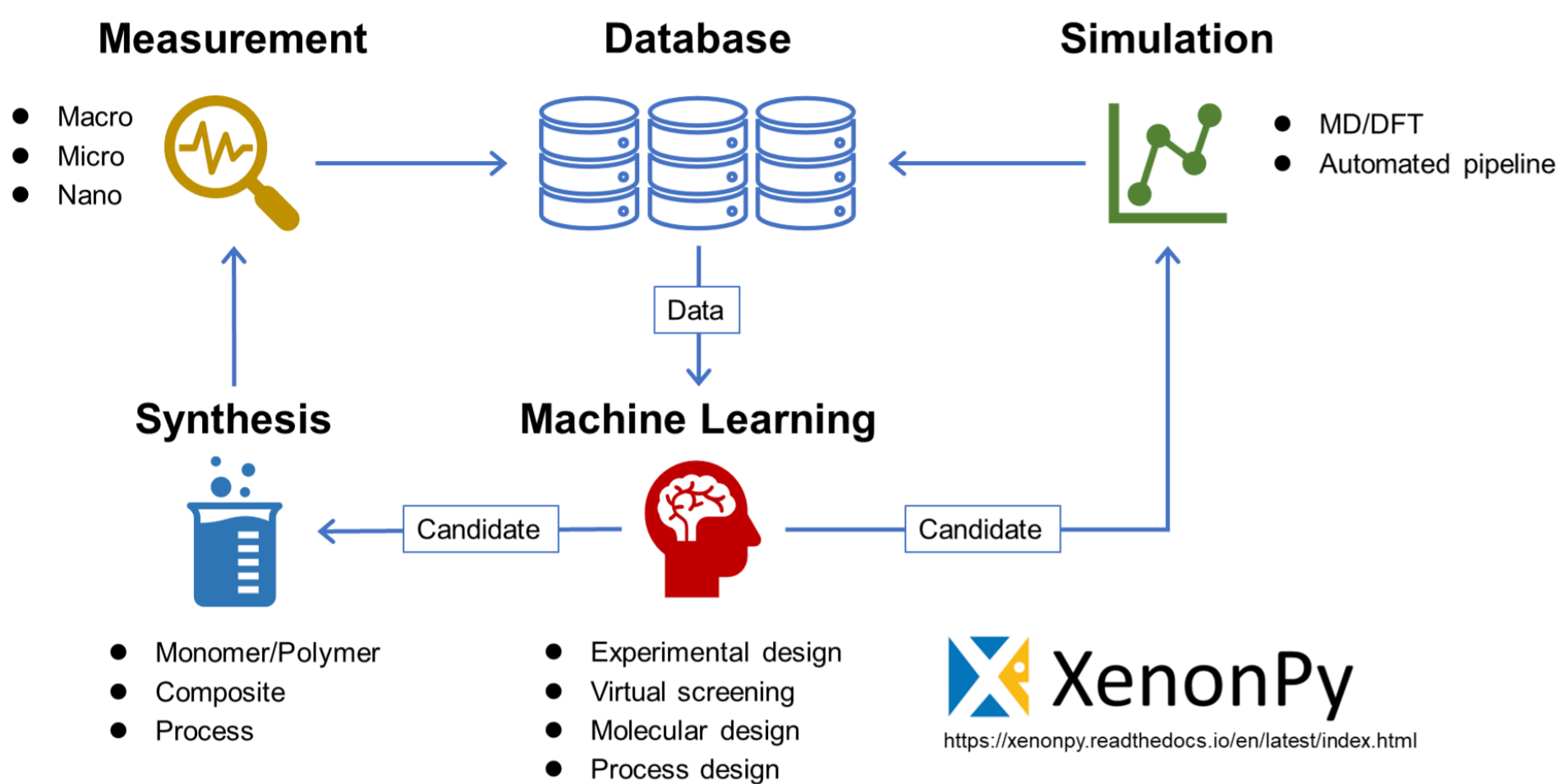
機械学習によるデータ駆動型材料研究

吉田 亮 データ科学研究系 教授

高分子熱物性のデータ駆動材料研究

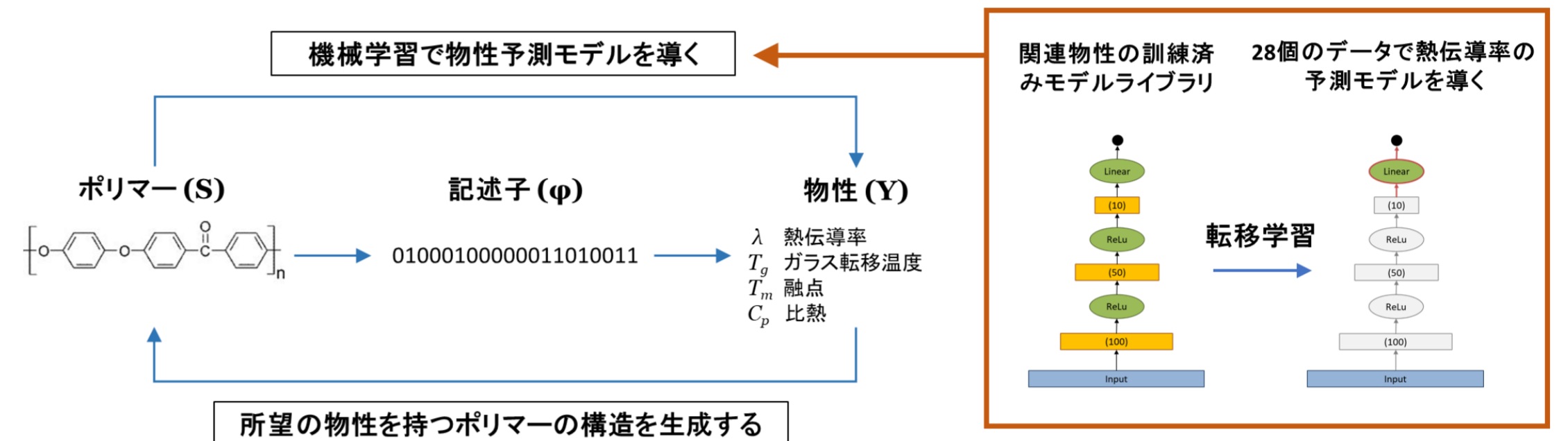
- 高分子物性の計算・実験データベースの開発
- 高分子の物性予測・構造設計を行う機械学習アルゴリズム
- 高熱伝導高分子の発見

JST-CREST 熱制御「高分子の熱物性マテリアルズインフォマティクス」
代表：森川淳子（東工大） 機械学習・計算科学グループ：吉田亮

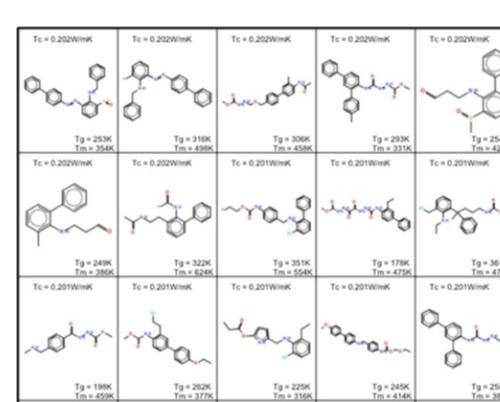


計測・合成・量子化学・分子動力学計算から創出されるマイクロ・ナノスケールの熱物性データと機械学習の先進技術を融合し、革新的な特性を有する新規の高分子材料を発掘する。

材料研究のパラメータ空間は極めて広大である。例えば、有機化合物のケミカルスペースには 10^{60} を超える候補物質が存在すると言われている。材料インフォマティクスの問題は、このような広大な探索空間から所望の特性を有する埋蔵物質を発掘することに帰着する。我々が開発したiQSPR-X [1, 2] は、所望の特性を持つ化学構造を設計する機械学習アルゴリズムである。高分子物性の実験データをiQSPR-Xで解析し、高熱伝導率を持つと予想される1000個の候補分子を設計した。さらに、三種類の芳香族ポリアミドを選定・合成し、熱伝導率0.41 W/mKを達成する新しい高分子を発見した [3]。これは典型的なポリアミド系高分子と比較して約80%の性能向上に相当する。この研究では、少数の物性データから予測モデルを導くために、転移学習と呼ばれる解析技術を駆使して問題解決を図った [4, 5, 6]。

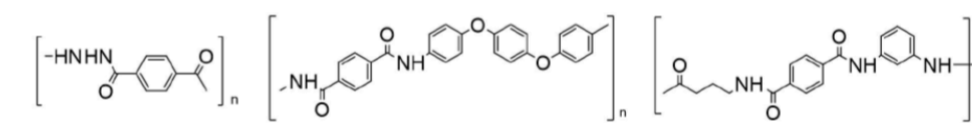


仮想ライブラリの生成

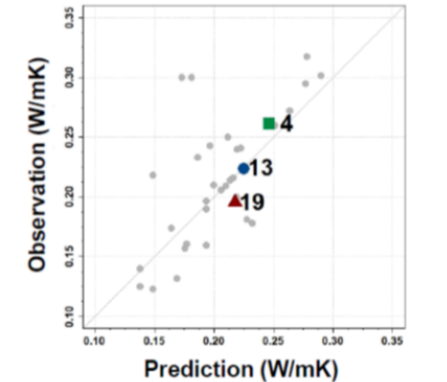


三種類の新規ポリマーを合成、超高速熱分析による熱物性の検証

熱伝導率：従来比最大80%増
高耐熱性
有機溶媒への溶解性
フィルム加工の容易性



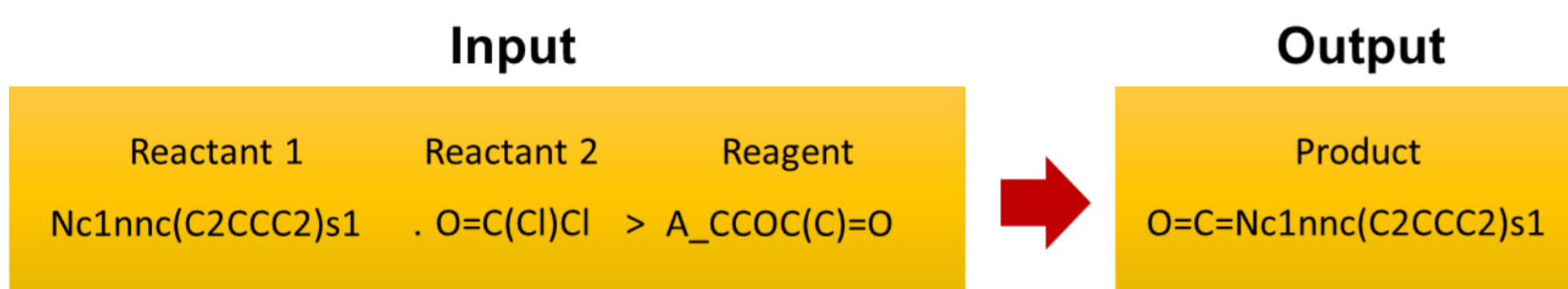
転移学習による熱伝導率の予測



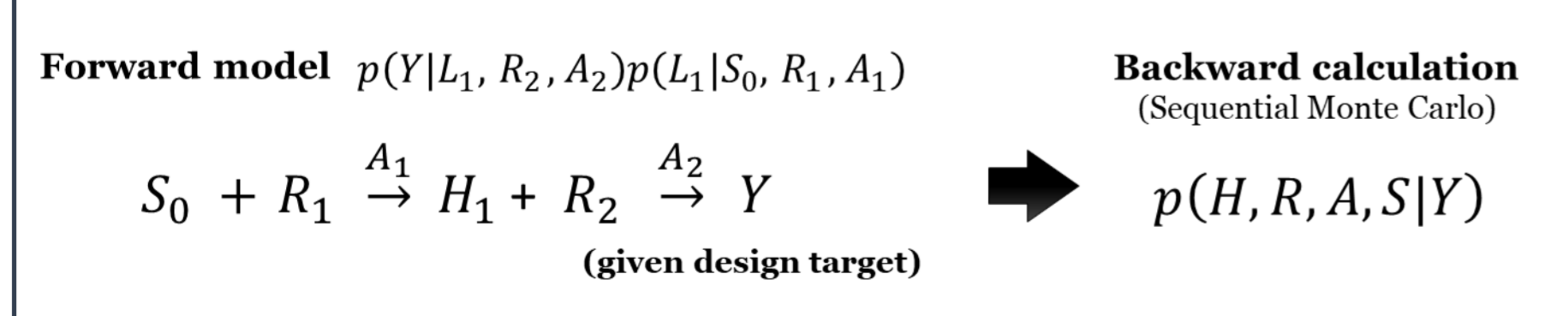
ベイズ推論による合成経路の自動設計

合成経路の設計における順問題と逆問題を定式化する [7]。順問題では、反応物の組Sから生成物Yの予測モデル $Y=f(S)$ を導く。一方、逆問題では、生成物のターゲット $Y=Y^*$ が与えられたもとでその逆写像 $S=f^{-1}(Y^*)$ を求める。

機械翻訳モデルによる合成経路の順方向の予測。1ステップの合成反応における反応物の組SをSMILES文字列に変換し、両者をピリオドで連結する。また、生成物の化学構造YもSMILES文字列に変換する。1ステップの合成反応の予測は、文字列から文字列への写像を求めることに帰着する。

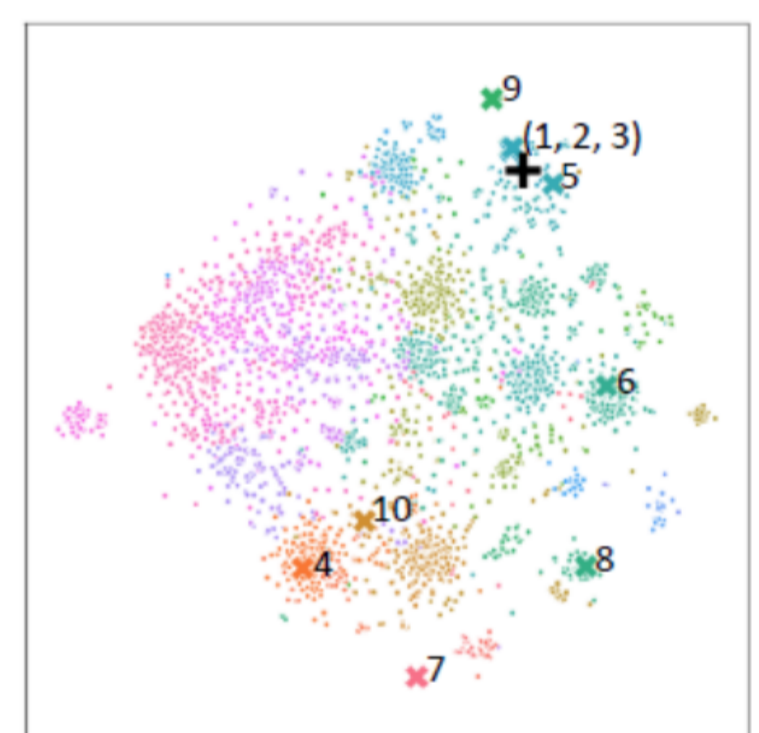


ベイズ推論に基づく合成経路の設計。ベイズの定理に基づき $S \rightarrow Y$ の予測モデルを $Y \rightarrow S$ （事後分布）に反転させる。逐次モンテカルロ法で、事後分布から反応物の組み合わせを抽出する。



1. Identification of > 6,000 candidate routes based on 10^6 purchasable compounds
2. 35-60% would be chemically valid according to judgments made by expert chemists

Candidates of synthetic route	Target	Step 1	Step 2
		1	1
		1	1
		1	1
		3	1
		1	1
		1	1
		3	1
		1	1
		1	1
		3	1



t-SNE projections of identified reactant pairs which formed nearly 98 clusters

Fig 4 in Guo et al. A Bayesian algorithm for retrosynthesis. JCIIM (2020) in press

一つの標的分子に対する2ステップの反応経路の解析例。探索空間は約 10^6 個の候補反応物から構成される。この例では、6,000以上の候補経路が発掘された。有機合成の知見に基づき合成可能性を評価し、約35%の候補経路が化学的に妥当であるという結論に至った [7]。

- [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. iQSPR in XenonPy: a Bayesian inverse molecular design algorithm. Mol Inform. 39:1900107 (2020)
- [3] Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. npj Comput Mater. 5:66 (2019)
- [4] Yamada et al. Predicting materials properties with little data using shotgun transfer learning. ACS Cent Sci. 5(10):1717-1730 (2019)
- [5] Ju et al. Exploring diamond-like lattice thermal conductivity crystals via feature-based transfer learning. Phys Rev Materials. 5053801 (2021)
- [6] Minami et al. A general class of transfer learning regression without implementation cost. AAAI 2021 (2020) in press (arXiv:2006.13228)
- [7] Guo et al. Bayesian algorithm for retrosynthesis. J Chem Inf Model. 60(10):4474–4486 (2020)
- [8] Kusaba et al. Recreation of the periodic table with an unsupervised machine learning algorithm. Sci Rep. 11:4780 (2021)
- [9] Liu et al. Machine learning to predict quasicrystals from chemical compositions. Adv Mater (2021) in press