

三菱ケミカルグループ株式会社
情報・システム研究機構 統計数理研究所

**高分子相溶性を予測・理解するための量子化学計算・深層学習統合解析
オープンソースプラットフォームを開発**

三菱ケミカルグループ^{※1}（以下「三菱ケミカル」）と統計数理研究所（以下「統数研」）との共同研究部門「ISM-MCC フロンティア材料設計研究拠点」^{※2}の研究チームは、ポリマー・溶媒系の相溶性^{※4}を高精度に予測する新たな手法を開発し、この研究成果をまとめた論文が米国時間7月10日に *Macromolecules* 誌（アメリカ化学会）に掲載されたことをお知らせいたします（掲載論文の URL = <https://doi.org/10.1021/acs.macromol.2c02600>）。

統数研と三菱ケミカルは、ポリマーと溶媒の相溶性を表す相互作用パラメータ (χ (カイ) パラメータ) を高精度かつ迅速に予測する量子化学計算・機械学習統合プラットフォームを開発しました。研究チームは、実験系由来の系統バイアスを持つ限られた実験データから高精度な予測モデルを得るために、三菱ケミカルが保有するハイパフォーマンスコンピューターを用いて生成した量子化学計算の大量のデータを利用しました。マルチタスク学習という手法を用いてこれらのタスクを同時に学習することで、従来の機械学習のモデルに比べてより広範囲のポリマー・溶媒に適用可能な予測モデルの構築に成功しました。さらに、開発されたモデルは、量子化学計算に基づく従来法に比べて約40倍の速さで χ パラメータを計算できることが確認されました。

ポリマーの溶媒への溶解は、プラスチックのリサイクルのほか、ポリマーの合成、精製、塗装、コーティングなど材料開発の様々な場面で欠かせないプロセスです。したがって、本研究成果は産学の様々な課題解決に貢献できることが期待されます。また、マテリアルズ・インフォマティクス分野におけるオープンイノベーション・オープンサイエンスの促進に貢献するため、開発したソースコードとデータの一部を無償で公開しました（公開 URL = https://github.com/yoshida-lab/MTL_ChiParameter）。

※1 三菱ケミカルグループは、三菱ケミカルグループ株式会社とそのグループ会社の総称です。

※2 https://www.m-chemical.co.jp/news/2019/_icsFiles/fieldfile/2020/02/18/190610-2.pdf

以上

お問合せ先

三菱ケミカルグループ株式会社

コーポレートコミュニケーション本部

メディアリレーション部

TEL : 03-6748-7140

大学共同利用機関法人

情報・システム研究機構 統計数理研究所

運営企画本部企画室 URA ステーション

TEL: 050-5533-8580、E-mail: ask-ura@ism.ac.jp

研究の背景

ポリマーの溶媒への溶解は、プラスチックのリサイクル、ポリマーの合成、精製、塗装、コーティングなど、材料開発において欠かせないプロセスの一つです。例えばリサイクルでは、異種プラスチックが混ざったプラスチックゴミに対して、溶媒を添加することで特定の材料だけを選択的に分離します。あるいは溶媒を相溶化剤と呼ばれる材料として用いて高機能なポリマーブレンドを製作します。このように、ポリマーと溶媒の相溶性を予測することは学術的にも産業的にも重要な課題です。

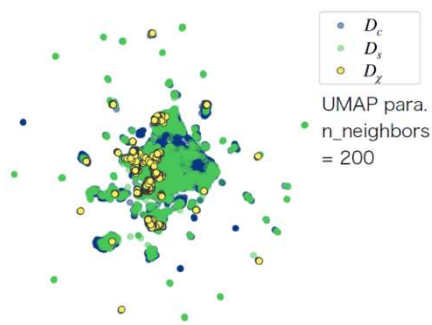
現在の計算化学の技術では様々なポリマー・溶媒系の相挙動を正確に予測することは難しいことが知られています。高分子溶液の熱力学的性質を記述するフローリー・ハギンズ理論^{※3}によれば、温度・体積分率・分子鎖長が与えられると、高分子溶液の混合自由エネルギー^{※4}は、ポリマー・溶媒間の相互作用を表す χ パラメータと呼ばれる量により決定されます。 χ パラメータの予測にはポリマーと溶媒の溶解度パラメータの距離に基づく経験的な予測手法が最も広く用いられています。例えば、ハンセン溶解度パラメータ（HSP）は、分子を分散項、極性項、水素結合項からなる3次元ベクトルで表します。ポリマーと溶媒の相溶性は、HSPベクトルの間の距離に基づいて推定されます。様々な分子の溶解度パラメータが実験的に測定されてきましたが、溶解度パラメータが未確定の分子に対しては、原子団寄与法^{※5}のような経験モデルを適用して溶解度パラメータを推定します。しかしながら、このような経験モデルは、特定の分子種以外の予測精度が非常に低いことが知られています。また、量子化学計算を用いたCOSMO-RS法^{※6}という分子シミュレーションでも χ パラメータを推定できますが、量子化学計算は計算時間がかかるため例えば溶媒候補分子の大規模スクリーニング等に適用することは困難です。また、予測精度も高くはありません。さらに機械学習による予測モデル構築のためには学習データセットが量的に不足し、実験系の性質上重大な偏りを持つことが知られています。このような課題を解決するために、ISM-MCC フロンティア材料設計研究拠点の研究チームは、マルチタスク学習と呼ばれる手法を用いて、量子化学計算の大量のデータと限られた実験データを統合的に解析することで、広範囲なポリマー・溶媒の組に適用可能な高精度予測モデルの構築を試みました。

（次ページへ続く）

研究内容と成果

モデルの学習には、46種類のポリマーと140種類の溶媒分子からなる1,190ポリマー・溶媒ペアの χ パラメータの実験値を用いました。データセットには、温度や組成の違いに対する χ パラメータの測定値も含まれています。データセットのポリマー・溶媒の分子種は、化学空間全体のごく限られた領域に分布しています(図1左)。またある実験系では、非相溶状態のポリマー・溶媒系の χ パラメータを測定することが困難であるため、データの分布に大きな偏りが生じます(図1右)。したがって、このデータセットのみを用いて学習されたモデルは、一般に予測の適用範囲が狭く、非相溶状態の χ パラメータの予測が不得意です。

実験データセットのポリマー・溶媒種に構造的な偏り
● 実験データ ● 計算データ (COSMO-RS) ● 良溶媒・貧溶媒データ



実験データは相溶するポリマー・溶媒ペアに偏り
一部の実験手法は相溶した状態のみに適用可能

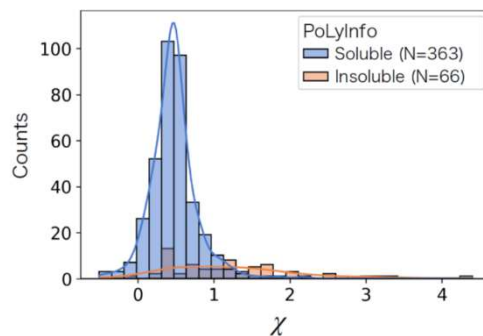
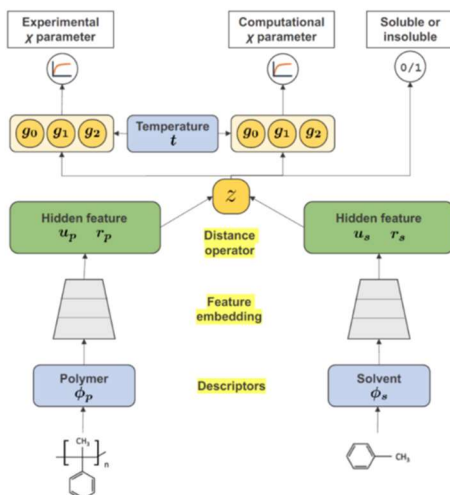


図1. χ パラメータの実験データに存在するポリマー・溶媒種には重大なバイアスが存在する。

この問題を解決するために、量子化学計算に基づくCOSMO-RS法を用いて、9,129ポリマー・溶媒ペアの χ パラメータのデータセットを生成しました。また、29,777種類のポリマー・溶媒の組み合わせについて溶媒が良溶媒であるか貧溶媒であるかの実験結果を表す二値ラベルが付与したデータセットを作成しました。これら三つのデータセットを用いて、ポリマーと溶媒の化学構造から χ パラメータの実験値とCOSMO-RS法で得られた値、ポリマーと溶媒の溶解性を表す二値ラベルを予測するディープニューラルネットワークを学習しました(図2)。この方法はマルチタスク学習と



✓ 潜在変数空間：ポリマー・溶媒の特徴 (u_p, u_s) の距離

$$z_i = (u_{p,i} - u_{s,i})^2 - r_{p,i}^2 - r_{s,i}^2 \quad (i = 1, \dots, k)$$

HSP溶解度パラメータ (分散項・極性項・水素結合項) の一般化になっている。

✓ χ パラメータの温度依存性 (経験式)

$$f_\chi(p, s, t) = g_{\chi,0} + \frac{g_{\chi,1}}{t} + \frac{g_{\chi,2}}{t^2}$$

潜在変数 z から係数 g への写像をモデリング・学習

図2. マルチタスク学習に用いたディープニューラルネットワーク

呼ばれます。マルチタスク学習では、背後に共通のメカニズムが存在する異なるタスクを統一的なモデルで同時に学習します。主タスクの χ パラメータの実験データは量的に限られており、実験系由来のバイアスも含まれています。そこで、二つの補助タスクを定めて、広範囲な分子種を包含するデータを学習に用いることで、予測モデルの適用領域を拡大することができました。また、このモデルは、量子化学計算に基づく従来法と比較して約40倍の速さで χ パラメータを計算できることが確認されました。

このモデルは、異なる三つのタスクのいずれにおいても、非常に高い予測性能を有することが実験的に確認されました。また、COSMO-RS法による量子化学計算やHSPに基づく経験的手法をはるかに凌ぐ予測力を示しました(図3)。このモデルのアーキテクチャは、HSPのコンセプトを拡張するように設計されています。HSPは、分子の潜在的溶解性は分散力、極性、水素結合の強さで決まると仮定しています。一方、機械学習のアルゴリズムは、分子の溶解性には34種類の因子が関与していることを示唆しています。そのうちのいくつかの因子は、HSPの3因子に対応していることが分かりました。このことは、ポリマー・溶媒の相溶性決定のメカニズムには、HSPでは無視されてきた未知の因子が存在することを示唆しています。

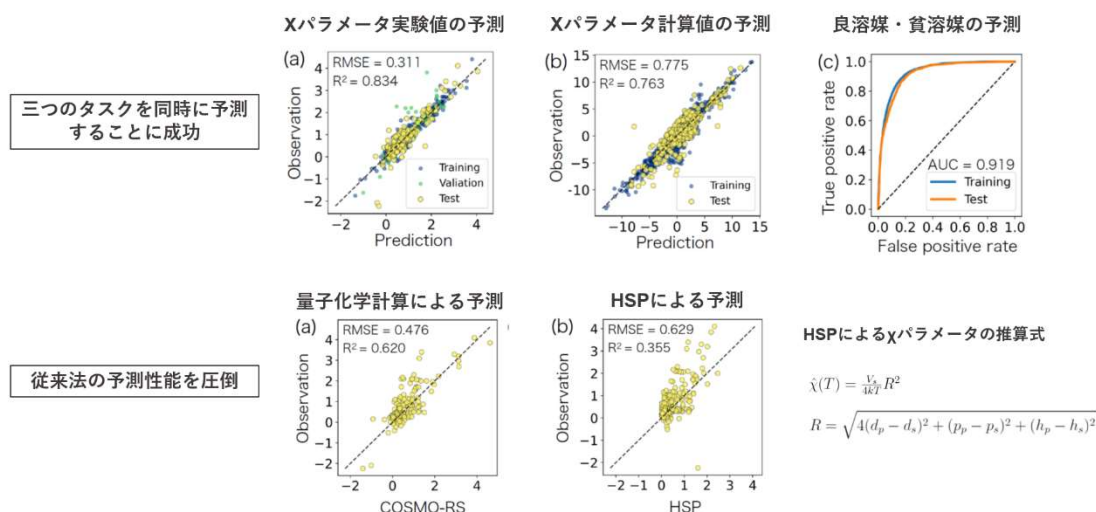


図3. マルチタスク学習の予測モデル、量子化学計算、HSPに基づく経験モデルの予測性能の比較

今後の展開

今回の研究により、ポリマー・溶媒系の相溶性予測モデルを構築するための量子化学計算・深層学習統合解析プラットフォームが完成しました。今後もデータ生産を継続しながら、モデルの予測性能を向上させていく予定です。

ポリマーと溶媒の相溶性を予測・理解することは、今後の材料開発において益々重要になっていきます。特に近年は脱炭素社会の実現に向けて廃プラスチック資源循環のための技術革新に対する期待が急速に高まっています。廃プラスチックのリサイクル比率を向上させるには、様々な異種ポリマーに対する相溶化剤の開発が必要不可欠となります。ISM-MCC フロンティア材料設計研究拠点は、今回のモデルを相溶化剤開発に実践展開していきます。また、機械学習技術のさらなる改善と拡張、マテリアルズ・インフォマティクス分野におけるオープンイノベーション・オープンサイエンスを促進するために、開発したソースコードとデータの一部を一般公開しました。

掲載論文

論文題目: Multitask machine learning to predict polymer-solvent miscibility using Flory-Huggins interaction parameters

著者: Yuta Aoki, Stephen Wu, Teruki Tsurimoto, Yoshihiro Hayashi, Shunya Minami, Okubo Tadamichi, Kazuya Shiratori, Ryo Yoshida

雑誌: Macromolecules

DOI: 10.1021/acs.macromol.2c02600

掲載日時: 2023 年 7 月 11 日 9 時 (米国時間 10 日 20 時)

用語解説

- ※3 1942 年にフローリーとハギンズにより独立に提案された格子モデルをもとにした統計熱力学理論。現在でも高分子溶液や高分子混合系の熱力学的性質を議論する際によく用いられる。
- ※4 2 成分を混合することによる自由エネルギーの変化。
- ※5 ある分子構造を、CH₃やOHなどの原子団に分けて、各原子団の寄与から、未知の構造の物性を推算する手法。
- ※6 量子化学計算により求まる表面電荷分布から、溶液中の分子間相互作用を評価し、活量係数・溶解度などの熱力学物性を推算する手法。