

2024.5.23 「統計数理を活用して未踏物質空間を切り拓く」産学連携シンポジウム QA集

	質問	回答
1	ポリマーの合成過程を予測する方法はありますか？	<p>反応予測と合成経路設計の機械学習モデル研究はかなり進んでおりますが、高分子への適用はまだ難航な状況だと思います。</p> <p>我々の一つの成果としては、株式会社ダイセルとの共同研究で SmiPoly(Ref: https://pubs.acs.org/doi/full/10.1021/acs.jcim.3c00329) という生成器を開発しました。この高分子の生成器から提案される候補は全て合成経路の提案も付いています。</p>
2	材料候補の選定やMIの問題設定を自動化することは可能でしょうか？	<p>可能だと思います。</p> <p>特に最近大規模言語モデルの発展によって、専門家の知見をAIに入れることは容易になったため、ユーザーの要望を答えるUser-friendlyなAI interfaceを作れると思います。</p> <p>ただし、完全に新しい問題設定には自動化が難しいと予想します。</p>
3	完全closed-Loopな材料探索は可能でしょうか？	<p>実験あるいは計算機実験を全自動化できる系であれば可能です。</p>
4	無機系材料におけるMIの研究課題は何でしょうか？	<p>未知の結晶構造をいかに予測するか、界面や表面の解析、非周期系の解析、動的特性など、課題は山積しています。</p>
5	MIに生成AI（GPTなど）を有効に活用する方法についてご意見を伺えますでしょうか？	<p>分子設計や結晶構造生成には生成AIの活用がかなり研究されています。</p> <p>近年は大規模言語モデルのMI応用研究も注目され、LLMによる分子生成、自動データ収集など様々な研究成果が発表されています。</p> <p>参考文献として以下の論文を紹介します： https://pubs.rsc.org/en/content/articlelanding/2023/dd/d3dd00113j</p>
6	異なる実験環境で取得された実験データや計算データをどのように突合して活用可能なDBをつくるのでしょうか？	<p>RadonPyの計算データに関しては、プリセットとよばれる標準プロトコルを定めており、データ取得はこのプリセットを用いています。そのため、計算条件に関しては統一されています。</p> <p>実験データ側の環境の差異ですが、転移学習により補正してやるというコンセプトです。</p> <p>例えば、アフィン転移学習 (https://arxiv.org/abs/2210.09745) を用いることで、統一条件で取得した計算データを用い、実験固有のプロセス因子を陽に考慮した転移学習により、様々な実験環境の違いを考慮した予測を行うことを計画しています。</p>
7	競合するプレイヤーからデータを統合するにはセキュリティの問題を克服する必要があると思います。その点をどのように対処されていますでしょうか？	<p>現在のRadonPyコンソーシアムでは、基礎的な物性のシミュレーション値について完全にオープンにするという前提での協力関係となっております。</p> <p>また、転移学習のターゲットタスクは各企業が自社内で個別に実施しています。そのため、セキュリティに関する対処は特に行っていません。</p>

	質問	回答
8	Sim2Realの転移学習で精度はどの程度担保されているのでしょうか？	<p>精度の担保と言われると様々な要因（データ数、計算精度、実験側の誤差、転移学習に用いる記述子、機械学習モデルのサイズ、etc）によって左右されるので、この質問に対して一概にお答えすることは難しいです。</p> <p>RadonPyの計算データからPoLyInfoの実験値へのSim2Real転移学習においては、PoLyInfoの実験値を直接学習した場合よりも精度が向上したということが今回の数値実験の結果です。</p>
9	精度の高いDFT計算は計算コストが高いともいますが、計算を高速化する研究はありますか？	<p>我々のグループでは行っていませんが、必要なアウトプットがDFT計算のポテンシャルエネルギーのみで良い場合はmachine learning potentialとよばれる研究が多数あります。</p> <p>第一原理MDのポテンシャル計算をDFTでなく機械学習に置き換えて高速化します。</p> <p>電子状態なども必要な場合は、電子密度とエネルギーの対応関係をニューラルネットワークに置き換える方法（https://pubs.acs.org/doi/10.1021/acs.jctc.1c00568）などが提案されています。</p>
10	ガラス転移温度の予測誤差は最終的にどの程度のところに落ち着きそうでしょうか	<p>現在観測した範囲のスケールリングカーブは、対数スケール上でほぼ直線的に汎化誤差が低下しており、シミュレーションデータを増やせば増やすだけ誤差0に向けて低下していくという結果になっています。</p> <p>実際には誤差は残ると思いますが、さらにデータ数を追加しないとカーブの収束先を予測できるほどの観測はできていません。</p>
11	三菱C様、JSR様ともにどのような材料領域でMIを適用研究されていますでしょうか？	三菱ケミカルにおける注力市場、新規事業分野で幅広く適用しています。
12	内挿と外挿をどう定義したのでしょうか？	<p>訓練データの入力変数の分布の密度関数が小さい（裾）あるいはゼロの領域を外挿と考えています。</p> <p>訓練・テストにおける分布シフトまでは基本的には想定していません。</p>
13	<ul style="list-style-type: none"> 生分解性プラスチックの内セルロース誘導体に着目された理由について教えてください。 その他の生分解性プラスチックへ展開は検討されていますか。 	<p>ご質問ありがとうございます。</p> <p>私たちが参画している産官学連携プロジェクトのCOI-NEXTの「再生可能多糖類植物由来プラスチックによる資源循環社会共創拠点」が多糖類植物由来のプラスチックの開発を目指しているためセルロース誘導体を計算しています。</p> <p>その他の生分解性プラスチックへの展開については、今のところはまだありませんが、将来的には検討するかもしれません。</p>
14	林さまのご講演に関して：スケールリング則が効く物性（ガラス転移温度）と熱伝導率のようにスケジューリング則が成り立たない物性があるようでしたが、目的の物性にスケールリング則が効くかどうか、あらかじめ知る方法はあるのでしょうか？	<p>一般的に計算-実験間のギャップが小さいほど（相関性が高いほど）、スケールリング則は強く観測されると考えられます。</p> <p>しかし、計算-実験間で例えばどれくらいの相関性があればスケールリング則が観測されるかという定量的な評価はまだ行っていません。</p> <p>やってみないとわからないというのが実情です。</p>

	質問	回答
15	高分子のデータベースを作成中とのことですが、どのようなデータのフォーマットで公開されるか等教えて頂けると嬉しいです。	物性値などはcsv、MD計算の最終スナップショットはJSONファイルをPythonオブジェクトの形でロードすると、RadonPyやRDKitを用いて座標や力場情報にアクセスできるようにする予定です。また、いずれもPython上でAPIを使用してダウンロード可能になる予定です。
16	RadonPyで計算できないポリマーはありますか？ (例えば、特殊な原子が入ったポリマー、ブロック子ポリマーなど)	公開しているバージョンでは、使用できる原子はH,C,N,O,F,S,P,Cl,Br,Iです。 また、分岐のない線形ポリマー・アモルファス状態に限ります。 開発中の非公開のバージョン（コンソーシアム内限定公開）では、分岐ポリマー、熱硬化性樹脂、ケイ素系ポリマーなどの計算が実行可能な拡張が行われています。
17	MDシミュレーション精度の乏しさは、シミュレーションデータで不足を補う上でのデータ駆動型予測における潜在的な課題になるかと思いました。 当該課題に対するアプローチについての考えを教えてください。 MDシミュレーション精度が乏しいことはありきとして、他の実験・文献データを使って補正するようなアプローチとなるのでしょうか？	おっしゃる通り、MDシミュレーション精度が乏しいことはありきとして、他の実験・文献データを使って補正するというアプローチがメインです。 計算精度と計算コストはトレードオフの関係にあり、計算精度を上げるということは、すなわち計算データの生産数が少なくなるので、機械学習のデータ源として考えた場合はとにかく計算精度を上げれば良いというわけではないと考えています。 また、MDシミュレーションにより計算可能な空間スケールはナノメートルオーダー、時間スケールはナノ秒オーダーであり、現実系（センチメートルの空間スケール、秒・分の時間スケール）とのギャップを完全に埋めることがそもそも困難であるという事情もあります。
18	準結晶はどのようなデバイスや産業に応用できる材料でしょうか？	準結晶は、現段階では基礎研究の段階であり、まだ産業的な応用に耐えうる材料の発見には至っていないというのが正直な所だと思います。 強いて言えば、フライパンのコーティング剤に使われているのが最も実的な例でしょうか。 (https://blog.miraikan.jst.go.jp/articles/20111014102011-1.html)。
19	発見された準結晶は3つとのことでしたが、機械学習のスクリーニング段階ではもっと多くの候補があったのでしょうか。	おっしゃる通り。 機械学習のスクリーニングでは1000個くらいの候補が提案されました。 合成可能性や元素の取り扱いやすさなどを配慮して、最終的に10個の候補を選んで合成実験しました。
20	RadonPyで引張試験のような機械的特性のシミュレーションは可能でしょうか？	ヤング率に関しては現在開発中です。 破断点などの大きな変形を伴う場合は全原子MDで取り扱うことが困難になるため、粗視化MDへの拡張を将来的に考えています。
21	https://scienceportal.jst.go.jp/explore/interview/20111108_01/index.html では熱が伝わりにくく、摩擦も少ないことからフライパンのコーティングなどに応用されましたが、重量が重く、高価になるためあまり普及はしなかったようですとあります	そうですね。最も実的な例であるフライパンのコーティング剤でも、あまり普及しなかったようです。 よってまだ準結晶は実用化に至っていないと言えます。 正二十面対称性や準周期性に起因する独自の物性が将来的に見つかる可能性を見込んで、研究を進めている段階です。