

## AIMaP 研究集会等実施報告書

(Part 1/4) 名称・重点テーマ・キーワード等

項目	内容
名称	「MI <sup>2</sup> 新材料探索のためのデータ科学」MI <sup>2</sup> I チュートリアルセミナー(第9回) 「物質探索のための記述子設計」 <a href="http://www.nims.go.jp/MII-I/event/tutorial2018_9.html">http://www.nims.go.jp/MII-I/event/tutorial2018_9.html</a>
採択番号	2018A004
重点テーマ	記述子設計、転移学習、特徴量(構成元素、化学組成、結晶構造)
キーワード	機械学習、深層学習、データセット(事前準備)、新物質・材料探索(逆問題解析)、記述子、転移学習、訓練済みモデル、因果と相関
主催機関	主催: 国立研究開発法人物質・材料研究機構(NIMS) 後援: 情報・システム研究機構 統計数理研究所 協賛: 文部科学省委託事業 AIMaP (受託拠点 :九州大学 IMI) 協力: 科学技術振興機構
運営責任者	NIMS 統合型材料開発・情報基盤部門 情報統合型物質・材料研究拠点 MI <sup>2</sup> I 副プロジェクトリーダー 真鍋 明
開催日時(開始)	2018/12/25 13:00
開催日時(終了)	2018/12/25 16:40
開催場所	JST 東京本部別館(K's 五番町)1階ホール

(Part 2/4) 最終プログラム・参加者数

項目	内容
最終プログラム	13:00 -15:30 「物質構造の表現と学習」 講師: 吉田 亮 (統計数理研究所 教授/NIMS MI <sup>2</sup> I) 質疑・応答

参加者数	<p>15:50 -16:25 「XenonPy の紹介」</p> <p>講師: Liu Chang(劉 暢)</p> <p>(NIMS MI<sup>2</sup>I /統計数理研究所)</p> <p>質疑・応答</p>
	<p>16:25 閉会(実際には前半の講義が延長となり、16:40 に閉会)</p> <p>(添付、プログラムと要旨をご参照願います。)</p> <p>・受講者:107 名(事前登録数:121 名、欠席:14 人、当日登録:2 人)</p> <p>(数学・数理科学:10 人, 諸科学: 14 人, 産業界: 79 人, その他: 4 人)</p> <p>・主催者:10 名</p> <p>(数学・数理科学:2 人, 諸科学: 6 人, 産業界: 0 人, その他: 2 人)</p> <p>以上は名簿からの推定結果(総人数:117 名)の内訳</p>

(Part 3/4) 論点・現状・今後の展開

項目	内容
当日の論点	<p>2016 年に開始した本シリーズも今回で第 9 回となる。今年は機械学習前のデータセット準備にフォーカスする。材料の原子、分子の組み合わせ数はほぼ無限にあり、求める材料特性は多岐にわたる。そのため物質・材料のデータ探索空間は巨大である。一方、既存データは多くの場合スパースである。そのため物質・材料の構成元素や結晶構造など、特徴量を適切に表現する記述子設計が要諦となるケースが多い。今回のセミナーでは記述子設計の基礎と実例について分かり易く解説する。さらに適切な記述子により獲得した学習モデルを有効に利用する手法としての転移学習についても触れる予定である。</p>
研究の現状と課題(既にできていること、できていないことの切り分け)	<p>&lt;物質構造の表現と学習&gt;</p> <p>目的: インputデータを適切に記述することで、各材料の特性を順方向で精度よく予測することができる。この記述子(Descriptor)設計が適切であれば、発生器(Generator)による逆問題解析、すなわち新材料探索の精度が高まる。このように、データ科学を適用した新材料探索の根幹を占める記述子について、高分子材料を中心に基本的な記述子(一般論:構成元素種と化学組成、化学構造と結晶構造)から、新潮流(深層学習)における記述子の現状と概要を学習し、転移学習による実用化事例を紹介した。また後段の</p>

講演の事前情報として、記述子選択の考え方や解釈の方法等の実践的な事例紹介を行った。

現状と課題: データ数が少ない特性を予測する際に、データ量が豊富な特性との相関を利用して予測精度を高める転移学習の有効性を、具体的な新材料探索における複数の成功例を交えて概説した。データセットと学習済みモデルを含む記述子設計～新材料探索のシステムツールとして開発した XenonPy (ゼノン・パイ) の先進性と有用性について具体例を用いて概説し、そのアクセス方法等、公開情報を提供した。

今後の課題: このような有用な手法・ツールを広範囲に普及させていく必要がある一方、ケースバイケースで成功例に導く基本的かつ実践的な適用事例の積み重ねと情報共有による、ユーザへの指導情報を多数発信していく必要がある。また、ユーザとの接点を継続的に持ち続け、ツールとしての改善も必要となる。

#### < XenonPy の紹介 >

目的: 前段で紹介のあった記述子設計のシステムツール XenonPy の概要を説明し、実際にインストールから内容物の確認、プログラムを稼働させるところをデモとして紹介し、ポテンシャルユーザとしての受講者に、その利用価値を理解してもらうことを目的とした。

現状と課題: XenonPy には、大量のデータセット、有効性が確認された複数の学習済みモデル、記述子生成モジュール、容易にニューラルネットワークを構築可能な深層学習との連携プログラム、逆問題解析が可能な予測機能、解析結果を俯瞰できる各種の可視化ツールから構成されている。すでに Version1.0 は公開済みであるが、大幅に機能を追加した Version2.0 の公開が間近となっている。

今後の課題: ユーザ向けに利用しやすさを担保できるマニュアル類の強化や API の利用環境の改善、転移学習の学習済みモデルの増強と提供、新しい転移方法や各種の材料への応用方法の発掘と提供、逆問題解析用に開発済みの解析ツール iqspr との連動を可能とするサービス機能の追加、などの継続的な改善が必要となっている。

また、MI<sup>2</sup>I 事業でコンソーシアム会員企業向けに提供している MI<sup>2</sup>I-DPF (データプラットフォーム) での利用環境の整備も必要となっている。

新たに明らかになった課題、今後解決すべきこと

- \* 幅広いユーザに実際に利用してもらい、使い勝手や新しい手法等、ユーザの声を広く集めて、より使いやすく様々な応用が可能なシステムツールへの改善。
- \* 転移学習の事例研究を積み重ね、得られる知見をもとに、転移学習のコツのようなユーザ向けの情報提供と発信。
- \* 実際にユーザが使ったときに、ヘルプデスクの様なサービス機能の充実が望まれるものと推察される。

今後の展開・ フォロー アップ	<ul style="list-style-type: none"> <li>* XenonPy (Version 2.0) 公開情報の提供を行い、幅広いユーザによる試行とフィードバックを得ることが必要。</li> <li>* 協力いただけるユーザから、可能な範囲で成功事例の実践例を提供してもらい、XenonPy への実装・共有化を進めていく必要がある。</li> <li>* これまでと同様に講義を撮影・DVD 化し、MI<sup>2</sup>を学び取る機会を提供する。</li> </ul>
-----------------------	---

(Part 4/4) 写真

項目	内容
添付写真 1	<div data-bbox="399 772 1364 1489" data-label="Image"> </div> <p data-bbox="391 1512 997 1556">後ろから: 事前登録 121 名、受講者 107 名で満席。</p> <p data-bbox="502 1568 1077 1601">キャンセル待ちの受講者も全員受講できました。</p>

添付写真 2



講師 左:吉田亮, 右:劉暢

添付写真 3



前から: 司会者による講師と聴講者とのつなぎトークをきっかけに、

## MI<sup>2</sup>I チュートリアルセミナー（第 9 回） 「物質探索のための記述子設計」

主催: 主催: NIMS 情報統合型物質・材料研究拠点 (MaDIS-CMI<sup>2</sup>)

後援: 情報・システム研究機構 統計数理研究所

協賛: 文部科学省委託事業 AIMaP (受託拠点 :九州大学 IMI) (※)

協力: 科学技術振興機構

**開催日:** 2018 年 12 月 25 日 (火) 13:00~16:30

**会場:** 科学技術振興機構 (JST) 東京本部別館 1 階ホール 東京都千代田区五番町 7 K's 五番町

2016 年に開始した本シリーズも今回で第 9 回となる。今年は機械学習前のデータセット準備にフォーカスする。材料の原子、分子の組み合わせ数はほぼ無限にあり、求める材料特性は多岐にわたる。そのため物質・材料のデータ探索空間は巨大である。一方、既存データは多くの場合スパースである。そのため物質・材料の構成元素や結晶構造など、特徴量を適切に表現する記述子設計が要諦となるケースが多い。今回のセミナーでは記述子設計の基礎と実例について分かり易く解説する。さらに適切な記述子により獲得した学習モデルを有効に利用する手法としての転移学習についても触れる予定である。

### プログラム

12:30	受付開始
13:00-15:30	<b>「物質構造の表現と学習」</b> 吉田 亮 (統計数理研究所/NIMS MI <sup>2</sup> I) 質疑・応答
15:30-15:50	休憩
15:50-16:25	<b>「XenonPy の紹介」</b> Liu Chang (統計数理研究所 /NIMS MI <sup>2</sup> I) 質疑・応答
16:30	閉会

※: AIMaP: 数学アドバンスイノベーションプラットフォーム/Advanced Innovation powered by Mathematics Platform

IMI: マス・フォアインダトリ研究所/ Institute of Mathematics for Industry

# MI<sup>2</sup>I チュートリアルセミナー（第9回） 「物質探索のための記述子設計」

## 「物質構造の表現と学習」

吉田 亮

情報・システム研究機構 統計数理研究所 データ科学研究系 教授  
情報・システム研究機構 統計数理研究所 ものづくりデータ科学研究センター センター長  
総合研究大学院大学 複合科学研究科 統計科学専攻 教授  
物質・材料研究機構 特別研究員  
〒109-8562 東京都立川市緑町 10-3  
E-mail: yoshidar@ism.ac.jp



Liu Chang (劉 暢)

国立研究開発法人 物質・材料研究機構 NIMS ポスドク研究員  
統計数理研究所 外来研究員  
〒109-8562 東京都立川市緑町 10-3  
E-mail: liu.chang@ism.ac.jp



MI 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、結晶など）、出力は物性値（エネルギー、電子状態など）に相当する。データ科学の文脈において、これらの解析は物質構造の“表現・学習・生成”を行うことに相当する。記述子と呼ばれる特徴ベクトルによって物質の構造を“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、計算機を用いて所望の物性値を有する物質を“生成”し、有望な候補物質を炙り出す。対象となる入力は、分子、組成、結晶、混合物、プロセス、合成経路等、問題に応じて多様な形式をとりうる。

本チュートリアルでは、MI における最も基本的な要素技術である記述子に焦点を絞り、分子・組成・結晶構造を対象とする基本的な記述子に関する包括的レビューを行う。さらに「転移学習に基づく記述子の自動設計技術」や「記述子の設計とモデルの解釈可能性」等について解説する。また、Python に基づき開発された記述子計算用のソフトウェア「XenonPy」の紹介を行う。