

# ナノ構造解析における機械学習と第一原理計算の活用

溝口照康

東京大学生産技術研究所

## Applications of machine learning and DFT simulation for materials characterization

Teruyasu Mizoguchi

Institute of Industrial Science, the University of Tokyo

In recent years, materials informatics, which integrates machine learning and artificial intelligence with materials research, has become a highly active field. Machine learning applications are proving valuable not only in materials simulation but also in materials processing and characterizations. This presentation mainly introduces our recent researches utilizing machine learning and first-principles calculations for nanoscale structural analysis. Furthermore, the application of generative AI and data-driven methods for a materials discovery will be touched in my presentation.

### 1. はじめに

近年マテリアル研究に機械学習などの人工知能技術を利用する「マテリアルズインフォマティクス」に関する研究が盛んにおこなわれている。機械学習はマテリアルのプロセスに加え、シミュレーションや、計測にも有効である。本発表では特に電子線やX線を用いて測定される電子エネルギー損失分光 (EELS) やX線吸収分光 (XAS) などのナノ構造解析に機械学習や第一原理計算を利用した研究について紹介する。

### 2. マテリアル×機械学習＝マテリアルズインフォマティクスについて

近年マテリアル開発に機械学習などの人工知能技術を組み合わせた「マテリアルズインフォマティクス」は盛んに行われており、すでに「マテリアルズインフォマティクス」と強調することがはばかられるぐらいにマテリアル開発に自然に使用されている。マテリアルズインフォマティクスの歴史は比較的古く、2005年に Krishna Rajan

博士の“Materials Informatics”という記事が Materials Today 誌に掲載され、2006年には米国材料学会の学会誌 MRS Bulletin に“Materials Informatics”の特集号が組まれている [1]。その後、2011年にスタートした Materials Genome Initiative プロジェクトにより、「マテリアルズインフォマティクス」が特に注目を集めた。

マテリアルズインフォマティクスの勃興と同時期の2012年に、トロント大学の Alex Krizhevsky 博士、Geoffrey Hinton 博士らにより Deep Learning が発表され、人工知能技術も大きく進歩した [2]。(本業績の重要性から、今年のノーベル物理学賞を受賞している)。また、2016年には、Google 社が開発した人工知能 Alpha Go が碁の世界チャンピオンに勝利し、人工知能技術が一般的に脚光を浴びている [3]。人工知能技術の発展は著しく、Google 社は同 Alpha Go に圧勝する人工知能 Alpha Go Zero をわずか1年後に発表している [4]。人工知能技

術において世界をリードしてきた Google 社であるが、最近では同技術を原子レベルシミュレーション [5] やたんぱく質の構造決定 [6] などの材料関連研究に活用し始めている点を、我々材料関連研究者は注視すべきであると感じている（たんぱく質の構造決定に関する人工知能の活用についても今年のノーベル化学賞を受賞している）。

機械学習は分類・回帰・クラスタリング・探索・推論といった問題に対するアルゴリズムを人間が直接与えるのではなく、学習のモデルを与えてアルゴリズムを学習により獲得させることで問題を解決しようとするアプローチである。機械学習では学習に用いるためのデータが必要となるため、近年材料に関する様々なデータが広く公開されている。

データは Materials Project [7] や NOMAD [8] のような web 経由で公開しているものもあれば、Scientific data や Data in Brief のように雑誌として公開しているものもある。最近では完全結晶を対象とした材料データベースに加え、二次元化合物や触媒に関する専門のデータベースなども実験・計算を問わず（多くの場合無料で）公開されている。いずれにせよデータの価値は近年急激に増しており、いかにして意味のあるデータを、利用できる形式で所有している（もしくはアクセスできる）か、そしてそのデータを活用するための技術を持っているかどうか、研究力や開発力に直結する時代が到来している。

以上のような状況の中、筆者の研究グループは、それまではシミュレーションとナノ計測を用いた物質の原子・電子構造解析を行ってきたが、2015 年ごろから機械学習を研究に取り入れ、格子欠陥（界面や表面）

の解析 [9–16] やスペクトル解析 [17–24] に利用してきた。本稿では、機械学習や第一原理計算を活用した EELS/XAS 解析について主に紹介する。

### 3. EELS/XAS 理論計算の基礎

EELS/XAS の吸末端近傍微細構造 (ELNES/XANES) は非占有軌道の PDOS 形状を反映しているため、第一原理計算を行うことで、スペクトルから化学結合や電子構造に関する情報を得ることができる。

遷移元と遷移先の軌道の局在性や相互作用の強さから、一粒子・二粒子・多粒子を扱う理論計算法を適切に選択することで、ELNES/XANES を計算することが可能になる [25, 26]。

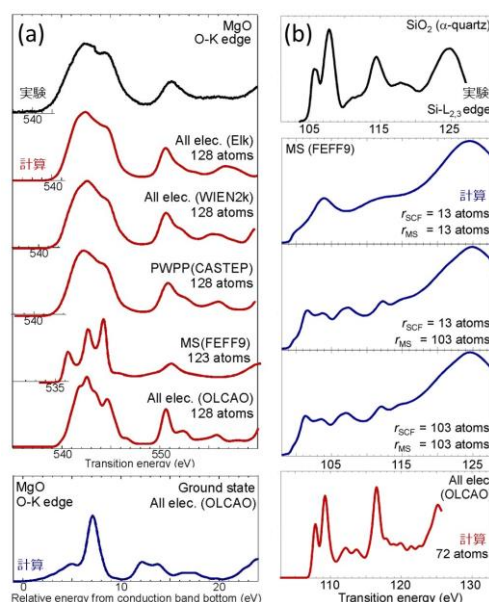


図 1 (a) MgO O-K 端の実験および計算スペクトル、(b) SiO<sub>2</sub> の Si-L<sub>2,3</sub> 端の実験および計算スペクトル

例えば、図 1(a) には酸化マグネシウム (MgO) の O-K 端の実験スペクトルと計算スペクトルを示す。内殻空孔を含まない基底状態 (Ground state) の計算においては実験スペクトルを再現することはできない (図 1(a) 下部)。内殻空孔を考慮し、内殻空孔間の相互作用を小さくするために十分に大きなスーパーセルを用いることで再現することができる。図 1(a) では、128 原子程度のモデルを用いた計算スペクトルを示す。計算には、第一原理全電子バンド計算

法 (All ele.) の WIEN2k [27], elk [28], OLCAO [29], 第一原理擬ポテンシャル計算法 (PWPP) の CASTEP [30], さらに第一原理多重散乱計算法の FEFF9 [31]を用いた。すべての計算コードにおいて実験スペクトルを再現できていることがわかる。

一方で, 図 1(b)には,  $\alpha$ -quartz 型二酸化シリコン ( $\text{SiO}_2$ ) の Si-L<sub>2,3</sub> 端の実験スペクトルと計算スペクトルを示す。計算には主に, 多重散乱計算手法の FEFF9 の結果を示している。FEFF9 では, 多重散乱 (MS) と自己無頓着 (SCF) を考慮する原子数を指定することができる。MS と SCF とともに 100 原子程度考慮しても, 実験スペクトルが再現できていないことがわかる。これは, FEFF9 コードで使用されている球対称ポテンシャルが影響している。FEFF9 は, 球対称ポテンシャルが良い近似になる MgO のような物質には利用可能であるが,  $\text{SiO}_2$  のように結合の異方性が高い物質のスペクトルには不向きである。

#### 4. 機械学習を利用したスペクトル解析の例: スペクトルを記述子とした物性予測

次に, 有機分子の炭素 K 端スペクトルを, 機械学習の入力 (つまり記述子) として利用し, 分子物性予測した研究について紹介する [21]。

まず, 有機分子の構造・物性データベースとして知られる QM9 データベース [32] のうち 22,155 分子中の非等価な 117,337 炭素サイトそれぞれについて炭素 K 吸収端スペクトルを計算した。作成した炭素 K 端についてはデータベースとして公開され, フリーでダウンロードすることができる [33]。

スペクトル形状からの情報抽出のため, 各分子スペクトルを励起エネルギーで揃え強度を標準化した 240 次元のベクトルを入力, QM9 データベースに収録されている 12 の物性値のうちの 1 つを出力とした順伝播型ニューラルネットワークを各物性値ごとにモデルを作成して学習させた。

学習した各モデルによる予測結果の一部

を図 2 に示す。

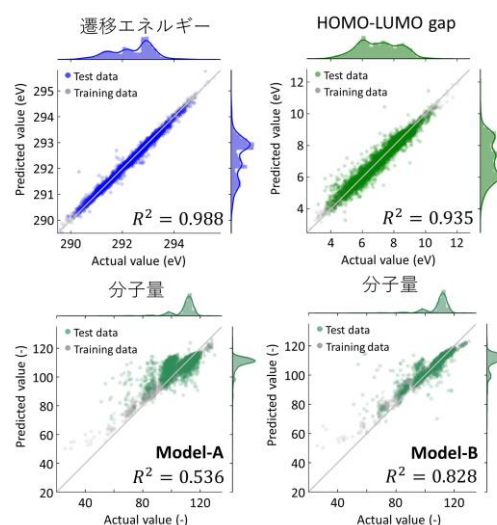


図 2 C-K 端を記述子として予測された各種分子物性

各図の右下には決定係数 ( $R^2$ ) も示している。決定係数が 1 に近いほど高精度な予測ができていることを示している。例えば, 励起エネルギーや HOMO-LUMO ギャップなどの示強性の物性値について決定係数 0.7 以上の良い予測ができた。励起エネルギーとスペクトル形状は個別に議論されることが多いが, 機械学習による予測ができていることからスペクトル形状に励起エネルギーの情報が含まれていることが示唆される。また  $\text{sp}^3$  炭素サイトの比率や零点振動エネルギーなど, 多くの分子物性を予測できることが明らかになっている [21]。

一方で, 示量性の物性値は予測精度が低い傾向がみられた。たとえば, 図 2 の下部に示した分子量の予測では 2 つの予測モデル (Model-A と Model-B) を用いた予測結果を示している。Model-A では, 炭素 K 端のみを記述子として作成されたモデルの予測結果であり, 決定係数が 0.5 程度と高くない。そこで, 新たに Model-B を構築した。Model-B では, 上述の炭素 K 端に加え, 組

成比 (N, O, F の C に対する比) を記述子として追加した。その結果, 分子量の予測精度が大幅に向上していることがわかる。

つまり, 炭素 K 端と組成比を用いることで, 分子量のような示量性物性の予測も可能であることがわかる。また, 実験スペクトルを入力として物性値の予測が定性的に可能であることも示している [21]。

## 5. 最後に

以上, 本稿ではスペクトルの第一原理計算と機械学習を用いた解析法について述べた。スペクトルが得られる原理は普遍であり, そこから得られる情報は物質開発において重要であることは間違いない。

一方で, マテリアル開発において有益な情報を, スペクトルから抽出することは容易ではない。最新の計算手法や解析手法, データベースのトレンドをとらえて実際の解析に利用することが, 今後のマテリアル開発を加速する上で不可欠であると考えている。

本発表では, 以上の内容に加えて, 生成 AI やデータ駆動型手法を活用した材料探索などについても触れる予定である。

## 謝辞

ここで紹介した研究は当時大学院生だった清原慎博士 (現・東北大学), 菊政翔氏, 当研究グループの川口直登氏, Poyen Chen 氏, 高原泉氏, 柴田基洋博士とともに行われたものである。また, 科研費, JST および NEDO のサポートを受けて実行された。ここに謝意を表す。

## 参考文献

- [1] K. Rajan, *Materials Today* **8**, 38 (2005).
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Adv Neural Inf Process Syst*, edited by F. Pereira *et al.* (Curran Associates, Inc., 2012).
- [3] D. Silver *et al.*, *Nature* **529**, (2016).
- [4] D. Silver *et al.*, *Nature* **550**, 354 (2017).
- [5] M. Schaarschmidt *et al.*, (2022).
- [6] M. Baek *et al.*, *Science* (1979) **373**, 871 (2021).
- [7] <https://www.materialsproject.org/>

- [8] <https://nomad-lab.eu/nomad-lab/>
- [9] S. Kiyohara and T. Mizoguchi, *Physica B Condens Matter* **532**, 9 (2018).
- [10] S. Kiyohara and T. Mizoguchi, *J Chem Phys* **148**, 241741 (2018).
- [11] H. Oda, S. Kiyohara, and T. Mizoguchi, *Journal of Physics: Materials* **2**, 034005 (2019).
- [12] R. Otani *et al.*, *Applied Physics Express* **13**, 065504 (2020).
- [13] S. Kikuchi *et al.*, *Physica B Condens Matter* **532**, 24 (2018).
- [14] H. Oda *et al.*, *J Physical Soc Japan* **86**, (2017).
- [15] S. Kiyohara *et al.*, *Jpn J Appl Phys* **55**, (2016).
- [16] S. Kiyohara *et al.*, *Sci Adv* **2**, e1600746 (2016).
- [17] T. Mizoguchi and S. Kiyohara, *Microscopy* **69**, 92 (2020).
- [18] S. Kiyohara and T. Mizoguchi, *J Physical Soc Japan* **89**, 103001 (2020).
- [19] S. Kiyohara, M. Tsubaki, and T. Mizoguchi, *NPJ Comput Mater* **6**, (2020).
- [20] S. Kiyohara *et al.*, *Ultramicroscopy* **233**, (2022).
- [21] K. Kikumasa *et al.*, *Advanced Intelligent Systems* **4**, 2100103 (2022).
- [22] S. Kiyohara *et al.*, *Journal of Physics: Materials* **2**, 024003 (2019).
- [23] S. Kiyohara *et al.*, *Sci Rep* **8**, 13548 (2018).
- [24] P. Y. Chen *et al.*, *Journal of Physical Chemistry Letters* **14**, 4858 (2023).
- [25] H. Ikeno and T. Mizoguchi, *J Electron Microsc (Tokyo)* **66**, 305 (2017).
- [26] T. Mizoguchi *et al.*, *Micron* **41**, 695 (2010).
- [27] K. Schwarz, P. Blaha, and G. K. H. Madsen, *Comput. Phys. Commun.* **147**, 71 (2002).
- [28] The Elk FP-LAPW Code, (2014).
- [29] W.-Y. Ching and P. Rulis, *J Phys Condens Matter* **21**, 104202 (2009).
- [30] S. J. Clark *et al.*, *Z. Kristall.* **220**, 567 (2005).
- [31] M. S. Moreno, K. Jorissen, and J. J. Rehr, *Micron* **38**, 1 (2007).
- [32] R. Ramakrishnan *et al.*, *Sci Data* **1**, 140022 (2014).
- [33] K. Shibata *et al.*, *Sci Data* **9**, (2022).