

2019 年理研シンポジウム  
RIKEN Symposium 2019

# 計算で物事を理解する予測する

～データサイエンス、自然知能、そして圏論へ～

2019 年 12 月 23 日 (月)

- 主催 国立研究開発法人理化学研究所 科技ハブ産連本部バトンゾーン研究推進プログラム 中村特別研究室、九州大学マス・フォア・インダストリ研究所
- 共催 (株)先端力学シミュレーション研究所、(株)地球快適化インスティテュート、(株)トプコン、三菱ケミカル(株)、理研 数理創造プログラム(iTHEMS)
- 協賛 応用物理学会、日本応用数理学会、日本化学会、VCADシステム研究会、分子科学会
- 実行委員 中村 振一郎、安達 泰治、杉本 学、原 正彦





# 趣 意

このシンポジウムは計算科学の大きな潮流がさらに世代を超えて充  
実し、役割を果たし続けることを願って企画しました。その中心的な  
役割とは、これからの日本社会は何に拠って立つのか、という根源的  
な問いに答えることです。そこに光明を与えるのは、学問と産業の両  
輪が活力を取り戻すことです。そのためには“最先端の理論”、“実現  
可能なモデル化と実装コード化”、そして“社会の課題解決につながる”、  
という3条件を満たした技術構築と実践適用が必須です。今回は、デー  
タサイエンス、データドリブンものづくり、量子コンピュータに加えて、  
自然知能と圏論が焦点です。現状の到達点を公開し、切磋琢磨しつつ、  
展望を浮かび上がらせることが目的です。

国立研究開発法人 理化学研究所  
科技ハブ産連本部  
バトンゾーン研究推進プログラム  
中村特別研究室

中村 振一郎



# プログラム・目次

開会の辞			
9:50-10:00	小寺 秀俊 理化学研究所理事	開会の挨拶	
趣旨説明			
10:00-10:15	中村 振一郎 理化学研究所	趣旨説明	
セッション1 データサイエンス・計算科学・量子コンピュータ			
	司会 加瀬 究 理化学研究所		
10:15-10:45	小林 紀郎 理化学研究所	データ知識化でものづくりを変える ～メタデータが拓く新たな価値創造の世界～	4
10:45-11:15	矢野 裕司 株式会社エリジオン	データ変換屋からみたデジタルエンジニアリング	6
	司会 檜貝 信一 株式会社アーク・イノベーション、AIMaP		
11:15-11:45	Gao Qi 三菱ケミカル株式会社	万量子コンピュータの AI と化学分野への応用研究紹介	7
11:45-12:15	茂本 勇 東レ株式会社	高分子計算屋が考える、インフォマティクスとの付き合い方	8
Lunch break			
セッション2 ミクロとマクロ量子論と連続体のフロンティア			
	司会 杉田 有治 理化学研究所		
13:30-14:00	熊谷 悠 東京工業大学	第一原理計算による半導体物性の計算手法開発	9
14:00-14:30	湯原 大輔 三菱ケミカル株式会社	分子動力学シミュレーション×機械学習 ～データの高速生成と解析の効率化～	10
	司会 小野 謙二 九州大学		
14:30-15:00	杉本 学 熊本大学	どうすれば計算カガクで役に立つモノを作れるか？	12
15:00-15:30	安達 泰治 京都大学	生物の形態形成の力学：分子から組織まで	14
Coffee break			
セッション3 データサイエンス自然知能から圏論へ			
	司会 甘利 俊一 理化学研究所		
15:50-16:50	田邊 國士 早稲田大学、 理化学研究所	データサイエンスにおける帰納的推論の社会的承認 ～機械学習に基づくがん診断支援装置の開発を例として～	15
Coffee break			
	司会 成瀬 誠 東京大学		
17:00-18:00	堀 裕和 山梨大学	圏論に基づく自然界の機能と知能の表現と理解	17
閉会の辞			
18:00-18:10	山口 兆 大阪大学名誉教授、理化学研究所		
懇親会			
18:15-20:00	理化学研究所 第一食堂		

## データ知識化でものづくりを変える ～メタデータが拓く新たな価値創造の世界～

小林 紀郎

理化学研究所 情報システム本部 研究開発部門 データ知識化開発ユニット  
norio.kobayashi@riken.jp

理研は自然科学の総合研究所として幅広い分野で研究が行われており、これら活動で産出される研究データは多種多様で膨大なものである。理研内外を問わず、このような研究データの利活用を促進し新たな科学的知見へと結び付けていくためには、データの集積のみならず、体系化や統合化を施して一つの知識を作り上げ、解析、検索、推論に応用できるようにする「データ知識化」が望まれる。このような知識の記述、集積、活用を含むデータ知識化をウェブ上で実現する仕組みがセマンティックウェブであり、メタデータを記述するための標準フレームワークである Resource Description Framework (RDF)や、知識をある領域の概念と概念間の関係の集合として形式的に表現するオントロジーなどの技術が使われる。ライフサイエンスはセマンティックウェブの利用が最も進んでいる分野であり、世界的にメタデータやオントロジーが数多く公開されている。理研でも理研メタデータベースと呼ぶ、ライフサイエンスデータの統合と利活用推進のためのセマンティックウェブ準拠の情報基盤を構築して2015年から運用し、2019年12月現在で理研発の62件を含む134件のデータベースの公開に成功している。さらに他分野にも対象を広げていくことでオープンサイエンスの実現に貢献すべく、さらなる技術開発に取り組んでいる。

一方で、日本のものづくりにおいては、データは秘匿されがちで流通が進んでいないのが現状である。近年のものづくり、例えば自動車産業においては、当初は部品製造や組み立て工程では多くの人間の労働力が必要であったが、機械化やデジタル化を背景に技術や科学的知識に基づく設計・製造の高度化が進んだことにより知識集約型ものづくりの方向性が見えてきたが、未だデータの利活用は十分とは言えない。そこで、この知識化の方向性をさらに強力に進めるため、材料、加工方法、設計モデルやシミュレーションなどを含む集約された知識とデジタル技術の利活用を最大限進める戦略（データ公開や共有のインセンティブづくりやビジネスモデルづくり）のもとで新たな価値を生み出すデジタルトランスフォーメーションの実現が今日の課題となろう。本講演では、理研メタデータベースのデータ知識化の知見を活用しつつ、ものづくりにおける知識化されたデータの流通と利活用に必要なデータ公開・非公開戦略、データガバナンス、エンタープライズアーキテクチャ等を含めた情報基盤を提案するとともに、その応用例を示すことで、データ駆動型ものづくりの方向性を共有したい。

## ものづくり産業連携への応用: 国際競争に勝つ設計力強化基盤の構築

- **売れる・壊れない製品製造の最重要工程は基本設計**。新材料、新加工法を共有するデータ駆動型企業間連携を実現させ、**設計全体の品質向上と大幅期間短縮を狙う**
- **AIから利用できる技術の知識化、データ化の手法とものづくりデータ基盤を理研が提案**
- **知財、リスクに配慮したデータガバナンスを通じた、企業の意思決定、オープン・アンド・クローズ戦略に基づいたデータ流通を実現。公正な競争による新設計の健全的強化を実現**

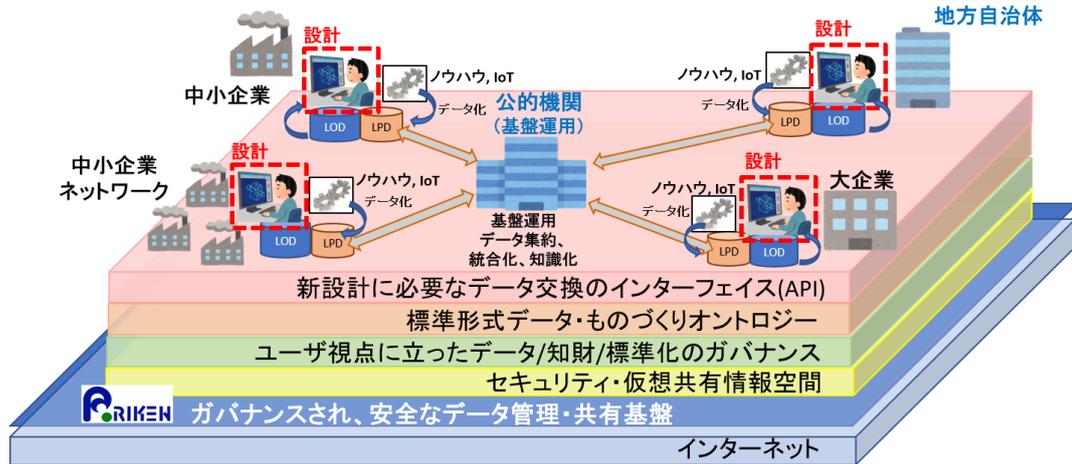


図 1. 設計力強化を目指した、ものづくり産業連携基盤の概念図

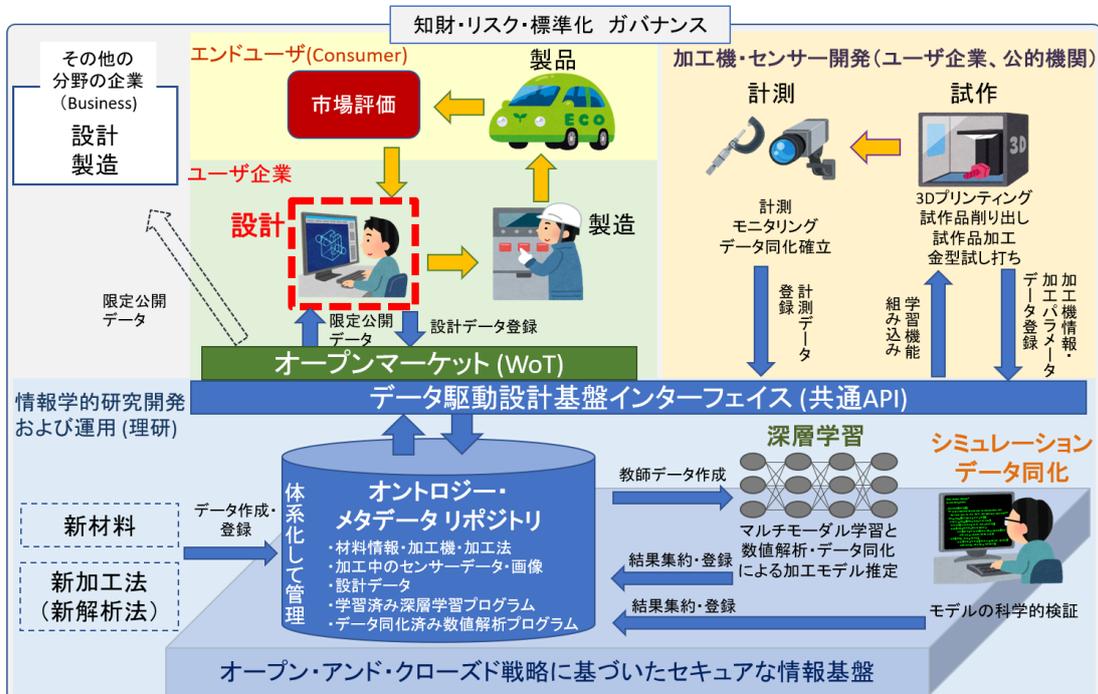


図 2. 多様なステークホルダーとの協調による、データ駆動型ものづくりの実施形態



# 万能量子コンピュータの AI と化学分野への 応用研究紹介

高 玘

三菱ケミカル株式会社 Science & Innovation Center

近年 IBM、Google、Microsoft、Alibaba などの IT 大手が量子コンピュータへの取り組みを強化している。また米国、欧州、中国等の世界の国において、国家レベルでの開発プロジェクトが推進していて、量子コンピュータの開発競争が過熱している。そのなかで慶応義塾大学は2018年5月に IBM の量子コンピュータを使いつつ、産官学共同で実用問題を研究する「IBM Q Hub」をオープンした。三菱ケミカルは同年6月に Q Hub に参画し、IBM や慶応義塾大学の研究者たちと共同で量子コンピュータの化学分野への応用研究を開始した。講演では現在進行している“AI”と“量子化学”分野の研究プロジェクトの内容を紹介する。

AI 分野に関しては IBM が Nature 誌 (Nature 567, 209 (2019)) で発表した世界初の量子コンピュータ実機上で実行可能な AI 手法 (「quantum kernel estimator」法) の予測精度の向上に関する研究内容を紹介する。具体的には量子空間に埋め込んだ学習データの分布様子を調べる手法を開発した。その手法を用いて、適切にデータの埋め込みに使われている関数 (feature map 関数) が選択できることによって、高精度の学習モデルの作成ができた。ベンチマーク計算では4種類の非線形パターンのデータに対して学習を行い、適切な feature map 関数を用いることによって、学習精度は 0.95 以上に達成できた。

量子化学分野に関しては量子コンピュータ上でリチウム空気電池の充電プロセスの初期反応を計算する研究内容を紹介する。具体的にはまず量子コンピュータシミュレータ上で VQE 法によりリチウム空気電池の充電プロセスの初期反応の計算結果が古典コンピュータの Full CI の結果と一致することが確認できた。さらに実機上の計算も行ったが、実機のエラーによって計算結果は Full CI の結果と乖離があった (約 10kcal/mol の差であった。)。その乖離を改善するために、IBM が開発したエラー緩和のアルゴリズムを取り入れた。その結果、乖離は約 3kcal/mol 程度まで低減できた。さらに慶應 Q Hub が独自に考案したエラー緩和アルゴリズムを取り入れることによって、乖離は約 1kcal/mol まで改善できた。

## 高分子計算屋が考える、インフォマティクスとの付き合い方

東レ（株）先端材料研究所 茂本 勇

高分子材料を含む化学工業・プラスチック工業は、自動車と並んで現在でも国際競争力を保持し続けている産業である。その存在感は、2017年度の工業統計表<sup>1)</sup>において製品出荷額の12.9%、付加価値額では15.5%を占めるという統計にも現れている。

産業競争力を支える重要な要素の1つが、新製品開発に向けた研究・開発であることも広く理解されているであろう。筆者が長年取り組んできた計算化学の分野においても、すでに1980年代には欧米の大手化学メーカーを中心に、触媒反応の量子化学計算や高分子鎖の分子動力学（MD）シミュレーションへの取り組みが始まっていた。<sup>2)</sup>

その後、日本ではバブル崩壊とともに雌伏の時期が続いたが、状況を変えるきっかけになったのが、古くは1998年からNEDOプロジェクトで開発されたOCTA<sup>3)</sup>であった。さらに2011年にスパコン性能ランキングで世界一を記録した「京」は計算機シミュレーション全般への社会の関心を高めることになる。2015年以降のAIブームをきっかけに、マテリアルズ・インフォマティクス（MI; AI材料設計）への期待も一気に膨らんだ。産業界における高分子計算科学への期待は、これまでになく高まっている。年内に発行予定のCICSJ Bulletin<sup>4)</sup>では、ゲスト編集者として、若手の研究者に高分子計算科学のstate-of-the-artを語ってもらった。ご参考になれば幸いである。

ところで、MIについては正直なところ態度を決めかねている。MIが過度なブームに陥ったことは間違いない。かと言って、単に流行り物と決めつけてしまうには、転移学習の応用<sup>5)</sup>といった魅力的な新技術が次々に登場しており、適切な技術を適切な課題に適用する見極めが付けば、明るい将来が待っていそうにも見える。

ただ1つだけ確実なのは、高分子のように複雑な階層構造を持つ材料の設計は、1つの新技術で一気に進むほど簡単ではないということだ。過度な期待は、過度の軽視に容易に反転しうる。これまで着実に実績を重ねてきた計算機材料設計全体がMIバブル崩壊の道連れになるような事態は、避けなければならない。

### 参考文献

- 1) 経済産業省 工業統計表, <https://www.meti.go.jp/statistics/tyo/kougyo/result-2.html>
- 2) 茂本勇; 化学工業, 2018年1月号, 40-47 (2018).
- 3) OCTA web site, <http://octa.jp/jp/>
- 4) CICSJ Bulletin (日本化学会ケモインフォマティクス部会誌) 特集『高分子計算科学』, 37 (4) (2019).
- 5) H. Yamada, C. Liu, S. Wu, Y. Koyama, Ss Ju, J. Shiomi, J. Morikawa, R. Yoshida, *ACS Cent.Sci.*, **5**, 1717-1730 (2019).

# 第一原理計算による半導体物性の計算手法開発

東京工業大学科学技術創成研究院フロンティア材料研究所 熊谷 悠

## 1. 研究目的

近年、計算機性能の向上と計算手法の改善により、実験結果に依存しない第一原理に基づいた材料データベース構築が盛んにされるようになってきた。しかしながら、世界中で開発されている Materials Project(<https://materialsproject.org>)や Aflow(<http://afloplib.org>)、OQMD(<http://oqmd.org>)などの計算材料データベースでは、結晶構造やエネルギー、単純な電子構造などの情報に限られており、高精度なバンドギャップや光吸収係数、点欠陥特性などの半導体物性に関するデータが欠如している。そこで近年、我々は、バンドギャップを有する非金属材料を対象とした、多様な半導体物性値を含んだ計算材料データベースの構築を進めてきた。

## 2 研究結果

計算材料データベースの構築には、数百万回以上の第一原理計算を実行する必要があるため、計算に必要な入力ファイルの生成、エラーが生じた場合のハンドリング、得られた計算結果の解析など従来手動で行ってきた作業を全自動化する必要がある。そこで本研究では、(i)点欠陥計算に必要な入力ファイルの生成と出力結果の解析を行う `pydefect` コード、(ii)第一原理計算の入力ファイルの生成、エラーハンドリング、出力結果の解析を行うための `vise` コード、(iii)自動計算ワークフロープログラム `Fireworks`(Jain *et al.*, 2015)および `atomate`(Mathew *et al.*, 2017)を拡張した `titas` コード、の3つのプログラムを継続的に開発してきた。これらのプログラムは `python` を用いて構築されており、総計おおよそ3万行の長さののぼる。

図1上に、これらの自動計算プログラムを用いて得られた電子系および格子系誘電定数とバンドギャップの関係性を示す。これを見ると、電子系誘電定数の最大値は、バンドギャップと反比例の関係にあることが見てとれる。一方、格子系誘電定数は、バンドギャップとの強い相関を示さず、電子系と比べて大きな値になる傾向がある。このことから、高誘電体材料の探索には、格子系誘電定数に注目することが重要であると言える。またランダムフォレストを用いた機械学習を行ったところ(図3下)、電子系および格子系誘電定数の双方で、実用的な精度で予測可能であることがわかった。

次に、図2上に、2,183個の酸素分子を基準とした酸素空孔形成エネルギーを示す。これから、酸素空孔形成エネルギーは、4.64 eVを平均として、0.2 eVから7.5 eVと幅広く分布している様子がわかる。また誘電定数と同様、ランダムフォレストを用いた機械学習を用いることで、0.3 eV以下の誤差で酸素空孔形成エネルギーを予測することができた。

本講演では、第一原理計算自動化手法の詳細について紹介する。

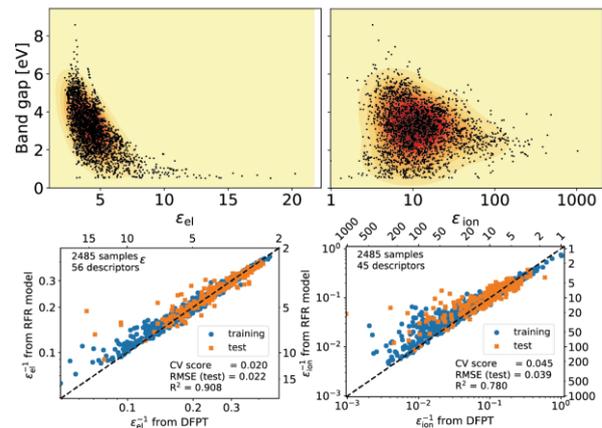


図1. (左)電子系及び(右)格子系誘電定数の、(上)バンドギャップ依存性と(下)その機械学習による予測。

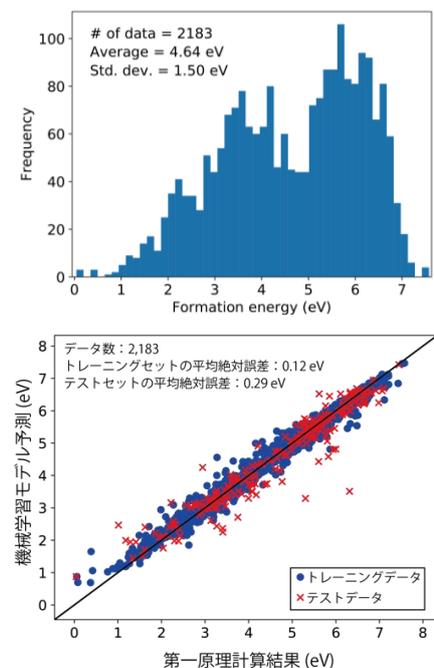


図2. (上) 酸素空孔形成エネルギーの分布図と(下)機械学習による酸素空孔形成エネルギーの予測。

# 分子動力学シミュレーション×機械学習 ～データの高速生成と解析の効率化～

(三菱ケミカル株式会社)○湯原 大輔  
(慶應義塾大学)遠藤 克浩, 泰岡 顕治  
email yuhara.daisuke.ma@m-chemical.co.jp

分子動力学(MD)シミュレーションは、分子のふるまいを調べるための便利で強力な手法である。様々な現象を理解するためには、巨大な系で長時間のシミュレーションを行う必要があるが[1]、現在の計算機を用いてもその計算時間はあまりに膨大であり、適用可能な空間・時間の範囲はとても狭い。また、例え計算が成功したとしても、膨大な数の分子のふるまいから系を特徴づける重要なふるまいだけを抽出するのは極めて困難であるため、解析も大変なプロセスとなる。そこで本研究では機械学習を導入し、分子動力学シミュレーションの高速化、ならびに解析の効率化を図る。

これまで、MDシミュレーションの高速化においては様々な工夫がされてきた。巨大な系のシミュレーションでは、空間を分割して超並列で計算を行う手法[2]や、分子の粗視化[3]がよく用いられる。一方、長時間シミュレーションの高速化は容易ではない。まず、時間の分割による並列化は原理的に不可能なため、上のような手法は適用できない。長時間シミュレーションの高速化を可能にする一般的な手法は、力の種類によって時間ステップの幅を変える多時間刻み幅法(Multi time step method)であり、様々なアルゴリズムが提案されてきた。しかし、これらを用いたとしても「時間」という壁は未だ絶大であり、革新的な高速化手法の開発が期待されている。

MDシミュレーションにおける時間方向の高速化を達成するため、本研究グループでは「機械学習による分子系の新たな時間発展法の獲得」を目指し、MD-GANを開発した[4]。MD-GANは、短時間のMDシミュレーションデータから部分系の時間発展法を獲得し、部分系の長時間のふるまいを予測するモデルである。ここで部分系とは、所望の物理量を求めるのに必要な情報を含む小さな時系列データであり、例えば分子の拡散係数を知りたい場合は、各分子の重心座標の短時間時系列を選択する。部分系の時系列予測が可能になれば、全体系の時系列を獲得する(=MDシミュレーション)必要がなくなり、大幅な高速化が期待できる。本モデルはGenerative Adversarial Nets (GANs)の一種である、Wasserstein GANs (WGAN) [5]に基づいて構築された(モデルの詳細については参考文献[4]を参照されたし)。ポリエチレン系のMDデータに本モデルを適用したところ、拡散係数ならびに末端間ベクトルの回転相関関数を精度良く予測することに成功した。

MDシミュレーションで生成された座標や速度のデータから、我々は様々な物性値を計算する。そして、その物性の違いに貢献している分子のふるまいについて詳しく解析する。しかし、大量の分子のふるまいの中から特徴的な分子のふるまいを抽出するのは容易ではない。そこで、本研究グループは「系の違いを理解し、その違いに貢献している分子のふるまいを特定する」ための機械学習モデルを開発した[6]。まず、何かしら(例えば、温度、濃度、分子種等)が異なる分子系を複数用意し、シミュレーションを行う。本モデルは、

各系内の分子のふるまいがとる確率分布の違いを Wasserstein 距離で表現し、低次元空間に埋め込む。Wasserstein 距離の計算は、ディープニューラルネットワークを用いることで効率的に計算できる。各系間の距離の埋め込みを獲得すると、各系の違いが明確になる。埋め込み空間内で最も離れている系同士で分子のふるまいを比較すれば、系の違いに貢献している分子のふるまいが特定できる。Wasserstein 距離の計算で用いたディープニューラルネットワークは分子のふるまいを入力としているため、これを利用すればどの分子のふるまいが系の違いに貢献しているかを簡単に特定できる。本モデルにアミノ酸 20 種類の溶液を適用し、水分子のふるまいの違いを調べた。すると、モデルはアミノ酸と相互作用している水分子が系の違いを表しているという答えを出した。また、各系間の距離の埋め込みは 1 次元に並んでおり、それらの順番は各系内の水分子の回転運動における緩和時間の大小の並びと一致していた。実際にシミュレーションを用いた研究では、アミノ酸と相互作用する水分子の回転運動の違いがよく議論されており、本モデルはそれを自動的に抽出したことになる。今後は、明らかになっていない物性発現メカニズムの解明に本モデルを応用するつもりである。

#### 参考文献

- [1] Y. Shibuta, S. Sakane, E. Miyoshi, S. Okita, T. Takaki and M. Ohno, *Nat. Commun.*, **8**, 2017.
- [2] S. Ayuba, D. Suh, K. Nomura, T. Ebisuzaki and K. Yasuoka, *J. Chem. Phys.*, **149**, 2018.
- [3] K. Kremer and G.S. Grest, *J. Chem. Phys.*, **92**, 1990.
- [4] K. Endo, K. Tomobe and K. Yasuoka, *Proc. Conf. AAAI Artif. Intell.*, 2018.
- [5] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, *Adv. Neural. Inf. Process. Syst.*, 2018.
- [6] K. Endo, D. Yuhara, K. Tomobe and K. Yasuoka, *Nanoscale*, **11**, 2019.

# どうすれば計算カガクで役に立つモノを作れるか？

熊本大学大学院先端科学研究部 杉本 学

(E-mail: sugimoto@kumamoto-u.ac.jp)

計算の道具・コンピューターは、今や確実に我々の生活の一部となり、それを支えている。この道具を使って、役に立つモノ、すなわち役に立つ物質、役に立つデバイス（複合物質）、役に立つシステム（物質の連鎖）、あるいは役に立つ者（人物）をうまく生み出すにはどのような工夫が必要か、を考えたい。なお、ここで言う「役に立つ」とは、「社会の役に立つ」ことでも、「個人に役立つ」ことでもよいし、「自然環境に役立つ」ことでもよい。時空を超えて「未来に役立つ」ことも意識したい。

コンピューターの真骨頂は数値データの処理にある。それが我々の生活を支える道具になっていることには、三つの原因があると思われる。一つは、我々の生活に必要な情報が数値データで表現されること。もう一つは、原子分子の振る舞いが量子力学で記述できること。そしてもう一つは、我々の脳や神経系がシナプスを介した電気信号で制御されていることである。第一の原因については、例えば真空中の光が周期的な波であり、数学的に記述されることを考えるとよく理解できる。例えば荘厳なレンブラントの絵画を構成するのは絵具である。画家の技巧はその個性を発揮する上で欠かせないが、絵具のもつ光吸収挙動が、その技巧を具現化している。従って、描かれた絵画は数値データの集合体となっており、それが画家の個性を数値的に表現する。

現時点では、コンピューターを使う計算は万能に思えるが、実際にはそうではない。例えば、現在店頭に並んでいる有機 EL テレビを考えよう。このテレビの商品化では、積層された有機薄膜による電界発光(Electroluminescence: EL)の発見が礎となった。EL は単一物質で実現されているのではなく、複数の材料からなるデバイスとして実現されている。そしてそれを製造するための装置や、動作に必要な回路設計なども必要である。単純に発光層の材料を予測する計算化学研究 [1]や、ホール輸送層のキャリア輸送特性の重要因子を解明する計算化学研究[2]だけでは、「役立つ」という観点では全く不十分であった。類似のデバイスであるペロブスカイト太陽電池について、実験に先駆けた予測を行うような格子欠陥制御に関する計算化学研究[3]もなされているが、その実用化は、計算だけでは全く達成できない。

ライフサイエンスでは、創薬や化合物の毒性予測が重要な研究課題となっている。創薬については、いわゆる ADME (Absorption, Distribution, Metabolism, Excretion) 特性と言われる人体内での薬物動態（体内での薬分子の動きや変化）も含めた創薬が求められる。体内には多数の臓器と腸内細菌があり、それが関わる全ての化学過程を計算で調べ尽くすことは絶望的に思われる。このような問題に計算カガクで立ち向かいにはどうすればよいであろう

うか？物質に関するものであるから分子に関する「計算化学」、凝縮系材料に関する「計算物理」で十分であろうか？

デバイスやシステム全体となると、一部の構造、時にはその構成成分が何かさえわからないことが多い。このような状況では、全てを計算で解き明かそうと企てても、全く不可能である。このような場合に活用できるのが、実験データを教師データとする「教師あり機械学習」である。これによって推論が可能になる。実際に、量子力学（量子化学）に基づく計算データと実験データを活用すれば、抗がん剤の構造活性相関[4]や有機系分子の毒性予測[5]が可能となることが最近の研究でも示されている。

これらの研究と関連して、最近分子の香りを予測する手法の開発が進められている[6]。この研究では、既知化合物に関する数値計算の情報と香りの分類に関する言語情報が利用されている。これは、我々の行う計算カガク研究では、単に数値データだけ、言語を用いて行なっていることに着想を得たものである。我々は言語を用いて数値計算の結果を解釈したり、位置付けを考え、論理を構築して結論を導く。従って、数値情報と同様に、言語で記述された情報を活用する計算カガクも今後益々重要となる。

我々の生活との双対空間をなす数値情報空間の発展が「社会の役に立つ」ことはもはや疑いようもない。では、今後どのような計算科学技術が必要であろうか？そこでは「物質の研究は計算化学で」というような偏狭な発想ではなく、異分野融合あるいは境界領域の研究[7]に挑戦し続ける必要がある。この挑戦では、様々な分野の計算技術を俯瞰することが極めて重要であり、そのような機会[8]は大変貴重である。我々計算カガク者は、専門化された世界（マイクロな世界）を眺める虫の目、全体を大局的に見る鳥の目、そして、実験・理論・計算カガクの流れを感じる魚の目、の三つの目を同時に持って、研究を行う必要がある。

本講演では、計算カガクを用いて役立つモノづくりへの貢献を目指してきた演者の研究と、それを通じて得た問題意識、およびその解決法について議論する。できれば、今後の計算カガクに必要な国の政策についても議論したい。これらの議論が、我が国を支える産業・医療の発展や、科学技術分野の人材育成のお役に立てば幸いである。

[1] M. Sugimoto et al., *Appl. Phys. Lett.*, **79**, 2348 (2001)

[2] M. Sugimoto et al., *J. Phys. Chem. A*, **103**, 5551 (1999).

[3] M. Sugimoto et al., *Chem. Sci.*, **9**, 3341 (2018).

[4] M. Sugimoto et al., *J. Comp. Aid. Chem.* in press.

[5] M. Sugimoto et al., *Mol. Info.*, in press.

[6] M. Sugimoto et al., *J. Comp. Chem. Jpn.*, **16**, 155 (2017).

[7] M. Sugimoto et al., *AIP Conference Proceedings*, **1702**, 090038 (2015)

[8] 中村振一郎, 牧野内昭武, 安達泰治, 杉本 学 (監修), 「23 の先端事例でつなぐ 計算科学のフロンティア –計算で物事を理解する予測する–」, 近代科学社(2020).

# 生物の形態形成の力学：分子から組織まで

安達 泰治

京都大学 ウイルス・再生医科学研究所

生物の形態と機能との関連を力学的に探る研究の歴史は古く、力学的な機能をもつ骨をはじめとして、数多くの器官・組織のバイオメカニクス研究において、現在もなお興味を尽きない課題の一つである。生物の発生過程における形態形成においては、例えば、上皮組織が力学的に折り畳まれて様々な形が作り出される。ここでは、時間・空間的に制御された上皮組織の折り畳みにおいて、細胞収縮による引張力や細胞増殖による圧縮力が、重要な役割を果たしている。本研究では、分子レベルの力感知とそのフィードバックが、組織レベルの形態形成に果たす役割を多階層バイオメカニクスに基づく実験と数値研究を組み合わせた *in vitro/in silico* 実験の融合的アプローチにより明らかにする。

まず、上皮組織の細胞間接着結合部位において張力センサー分子として知られる $\alpha$ カテニンの張力作用下におけるメカノフィードバックの一つの要素過程を原子間力顕微鏡 (AFM) を用いて一分子蛍光観察した(図1)<sup>2)</sup>。次に、分子レベルのメカノフィードバックが、組織レベルの形態形成に及ぼす影響を数値モデル化し、そのダイナミクスをVertex動力学シミュレーション<sup>3)</sup>により解析した(図2)。さらに、組織の収縮と成長を連続体力学の枠組みで表現し、有限要素法を用いた組織形態形成シミュレーションへと適用した<sup>4)</sup>。

まず、AFMを用いた *in vitro* における一分子実験では、 $\alpha$ カテニンが張力作用下においてその構造を変化させ、ビンキュリンの結合を誘導することをリアルタイムで観察することができた(図1)<sup>2)</sup>。次に、 $\alpha$ カテニンのメカノフィードバックを考慮したVertexモデル<sup>3)</sup>により、組織の成長過程における滑らかな形態維持において、このフィードバックが重要であることを示した(図2)。さらに、連続体モデルを用いた多細胞組織の形態形成シミュレーションにより、形態形成の多様性と頑健性をエネルギー地形の描出により理解する手法を提案した<sup>4)</sup>。このような分子から組織形態形成に至る複雑なふるまいを多階層バイオメカニクスの手法を用いて力学的に理解することが可能となりつつある。

本研究は、一部、AMED-CREST(メカノバイオロジー)、および、科研費新学術領域(脳構築)の助成により行った。また、京都大学井上康博氏、亀尾佳孝氏、牧功一郎氏、竹田宏典氏、金沢大学奥田覚氏らとの共同研究によるものである。記して謝意を表す。

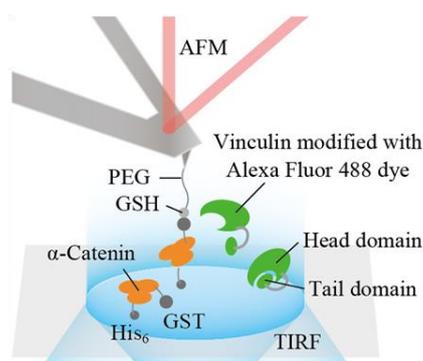
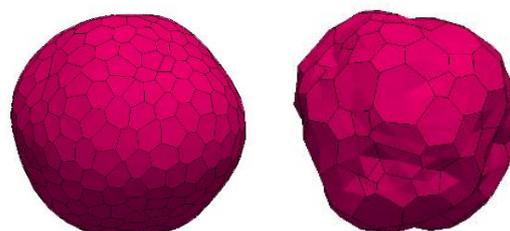


図1：張力作用下における $\alpha$ カテニン分子のメカノフィードバックのAFM観察。



(a) フィードバック有 (b) フィードバック無

図2：成長する上皮小胞組織の形態形成におけるメカノフィードバックの影響

## 参考文献

- (1) 安達泰治, 日本機械学会バイオエンジニアリング部門報, No. 48, pp. 3-5 (2019).
- (2) Koichiro Maki, *et al.*, Scientific Reports, 8(1): 1575 (2018).
- (3) Satoru Okuda, *et al.*, Science Advances 4(11): eaau1354 (2018).
- (4) Hironori Takeda, *et al.*, Biomechanics and Modeling in Mechanobiology (2019).

# データサイエンスにおける帰納的推論の社会的承認

～機械学習に基づくがん診断支援装置の開発を例として～

田邊國士（早稲田大学/RIKEN/統計数理研）

今日、Deep Neural Netなど種々の学習機械による様々な課題の解決が図られています。本講演では演者が提案した汎帰納推論機械 dPLRM(2001) に基づくがん診断支援装置（山梨大学/早稲田大学/島津製作所）の開発を具体例にとり、機械学習による帰納的推論の科学研究における位置付けについて考察したいと思います。

従来の科学やエンジニアリングはニュートンとデカルトによる方法論に基づいて発展してきました。すなわち事象の素因となる原始要素を**予め分節化**し、それを精密に同定・測定し、**分節化に依拠した論理的な推論**によって導かれる帰結を予測し、事象の観測データと照合して、事象を理解し制御する方法論です。実際、科学研究の現場においては原始要素の精密な観測・測定が要求されますし、エンジニアリングにおいても精密な観測を担保する機器の製作に意を用います。**因果関係における関与因子の数や相互作用の単純性を秘かに仮構**しているこのパラダイムは無意識の裏に現代人に刷り込まれており、科学研究はおろか社会的諸制度が要求する規制に対する適合証明（例えば薬品の有効性の証明）は、このパラダイムに則り行わねばなりません。これに基づかないものは非科学的であるとして退けられます。

しかし、機械学習の概念はニュートン-デカルト・パラダイムを覆すものです。因果関係の連鎖の同定なしに、科学的推論を可能とするものが学習機械です。変数の間に措定されるべき機序の**先験的な分節化**という従来人間が行うものとされてきた帰納の営為を要しません。帰納の過程を機械化することによって科学研究過程とエンジニアリングの方法論にもう一つパラダイムをもたらしているのです。そのひとつの顕著な特徴として、学習機械における推論には**精密な測定データは必ずしも必要ありません**。個々の対象に付随する**モードの異なる数多くの変数の観測データの組み合わせの情報**から学習機械は帰納的推論を実行することが出来ます。その場合個々の変数の観測データの**高い精密性は必要がなく、多数の変数の組み合わせがもたらす情報の方が機械学習による推論過程に大きな比重を占める**ことがあります。このため機械学習は、単純な因果関係の措定が困難な生体現象のように、異なる時間スケールで相互作用しながら変化する多種な要素が結合した多自由度を持つ対象を探る上では格好の方法論です。

がん診断における判断材料としてがんに特異的に検出されるバイオマーカーとよばれる生化学的分子の量の測定が広く行われています。多くの医学研究者や医療関係会社は各種がんに対応するバイオマーカーの発見に多くの資源を投入してい

ます。ニュートン-デカルト・パラダイムに従えばこのバイオマーカーの探索という研究開発の方針は当を得たものでしょう。しかし、個々のガンに対して1, 2のバイオマーカーを探索するのは賢明なことでしょうか？ 膨大な費用が掛かるだけでなく、**検証すべき仮説の数に比して必要な検証用データの収集は容易ではありません**。ガンは遺伝子上の突然変異によって引き起こされますが、それがもたらす細胞内外での代謝過程で生成される数多くの生化学的分子群の代謝経路は非常に異種多様です。同じ急性骨髄性白血病と診断されるものでも、その遺伝子の変異の仕方は多種異形です。ガンの発生に関与する代謝分子は多重に組み合わせさせてガンを発現しているのです。一般にガンの病名は臨床的所見にあるいはガン細胞の形態学的観察による命名に過ぎず、**実体的に定義されるものではありません**。ガンの物理モデルはないという論文もあります。従ってその代謝物質のひとつ二つをバイオマーカーとして特定することを試みるよりも、代謝分子群全体の中の組み合わせを観測して診断の方が遙かに理にかなっています。筆者らはこの点に着目して、肝臓ガン、腎臓ガン、大腸ガン、胃ガンの細胞から検出される極微量の液滴の質量分析データに学習機械を適用してガン診断支援装置を開発しました。しかも学習機械を適用するに当たって、精度の粗い質量分析データを用いて代謝分子群の**個々の分子を同定することなく**診断に成功しています。

#### 参考

1. Tanabe, K., *Penalized logistic regression machines: New methods for statistical prediction 2*, (2001), Proceedings of 2001 Workshop on Information-Based Induction Science (IBIS2001), 71-76, 2001.
2. Tanabe, K., *Penalized Logistic Regression Machines and Related Linear Numerical Algebra*. 京都大学数理解析研究所講究録. No. 1320, 239-249, (2003)
3. 回通園士, 帰納推論機械PLRMとdPLRM—方法論, モデル, アルゴリズムおよび応用, システム/制御/情報j 第51巻第2号, 87-95, (2007).
4. Tanabe, K., et al, *dPLRM, Revisited: Towards Cancer Diagnosis with PESI-Mass Spectrometry Data*, Poster presented at the 61st Annual Conference on Mass Spectrometry, Tsukuba, Japan, Mass Spectrometry Society of Japan, (2013).
5. Matsui, T. and Tanabe, K., *dPLRM-Based Speaker Identification with Log Power Spectrum*, Proc. Interspeech, 2017-2020, (2005).
6. Matsui, T. and Tanabe, K., *Comparative Study of Speaker Identification Methods: dPLRM, SVM and GMM*, IEICE E89-D, 3, 1066-1073, (2006).
7. Birkenes, O., Matsui, T, Tanabe, K. et al., *Penalized Logistic Regression with HMM Log-Likelihood Regressors for Speech Recognition*, IEEE Trans. Audio. Speech and Language, Vol. 18, No. 6, 1440-1454, (2010).
8. 竹田扇, 吉村健太郎, 出水秀明, 平岡賢三, 谷畑博司, 田進国土, 中島宏樹, 堀裕和, 質量分析法と統計的学習機械を組み合わせた新規がん診断支援装置の開発, 島津評論. Vol. 69 (3・4) 203-210, (2013.)
9. Kentaro Yoshimura, et al, *Real-time diagnosis of chemically induced hepatocellular carcinoma using an nvel mass spectrometry -based technique*, Analytical Biochemistry, 441, 32-37, 20 (2013).
10. Kentaro Yoshimu, et al, *Analysis of Renal Cell Carcinoma as a First Step for Developing Mass Spectrometry-Based Diagnostics*, Journal of American Society for Mass Spectrometry, (2012).
11. 特許6189587号, 質量分析装置及該装置を用いた癌診断装置. (2017)tanabe

## 圏論に基づく自然界の機能と知能の表現と理解

山梨大学 大学院総合研究部 堀 裕和

現代数学の言語である圏論に基づいて複雑な自然界のダイナミクスを捉え、これにある価値観を与えることで、通常の計算や情報処理の概念を超えた多様な機能や知能を創生することが可能となる。例えば、その実態を把握することが困難な複雑な現象をもたらすある構造を、それと同等の複雑さを持つ自然界に見いだされる別の系の構造と対比し、それぞれの系が示す計測可能な量に基づいて、両者がある数学的な意味において同じとみなされる随伴関係を構築することによって、一方から他方の構造を推定したり、一方の振る舞いから他方を制御したりすることができる。これは圏論において自然変換と呼ばれる構造に対応し、私たちの脳が意思に基づく身体活動と外界の反応の観測と理解によって自然界の諸現象を把握し理解する構図でもある。目的とする価値観に応じて、何を以って物事を同じと見るかという指標が設定され、それによって随伴関係を設定すべき数学的な内容が選択される。

世界は、単に物質とその相互作用から構成されているのではなく、私たちがそれをどう捉え、そこに何を求めるのかという態度に応じて創発するものであり、それゆえ今を生きる私たちの身体的特性や思考の形態、さらには多数の私たちが調和させる社会的な側面が関与してくる。複雑に絡み合った世界の組み紐の、どの部分を解いて、それによってどの部分がより複雑に絡むことになるのかを、世界の一体性あるいは全体性に基づいて考察する必要がある。世界を把握するために諸現象を計測するという最も基本的な段階においても、測定結果が異なるという指標を設定するためには、その値が同じであるとみなされる範囲を設定し、その範囲を代表する数を定めることが必要となるとともに、それに伴って定めた代表値の関係がどう接続されるのかという、離散化された値の非自明な接続を司る調和の法則を導入することが求められることになる。世界全体をあるがままに捉えるのではなく、恣意的に分断した世界を再び調和させるという手法をとるのは、私たちが常に、何らかの価値と考える計測可能と思われる指標に対して物事を順序付け、それがあある向き付けのもとに変わっていくことを求めることに基づいている。これは、機能あるいは知能と私たちが呼ぶ物事の本質であると考えてよいのではないかと思う。

シンポジウムのタイトルである「計算で物事を理解し予測する」ということも、価値観に基づく離散化と順序付けおよび向き付けされた調和の構造の構築に関

わっている。本研究は、このような世界の理解に関わる構造がどのように構成され得るのかという問題を、圏論に基づいて考察する試みである。単に形式的に試みを展開するのではなく、具体的に私たちの価値観を基盤とした自然界の構造の随伴性と、これに基づく意思決定などの機能あるいは知能と呼べるような構造について、その実装例と圏論的表現がどのようになされるかを紹介し、「計算で物事を理解し予測する」ということの本質を明らかにしたい。

現代科学は、物質世界を取り扱うことに関して極めて高度な段階に達し、それゆえに世界の全体性や非自明性に出合ったときに、その素朴な性質をも精緻な理論体系の下に捉えようとして当惑してしまうような状況にあるのではないかと思われる。物事の構造を捉えるということは、必ずしも精緻な記述や理解を求めるものではなく、むしろ構造の同じさの緩い枠組みにおいて、大きくくりになされてこそ意味があるのではないかと考えられる。精緻さと曖昧さを兼ね備えた体系を構築するために、圏論がどのように力を発揮するのかを、いくつかの具体的な研究例に基づいて紹介する。

#### 参考図書・文献

- [1] 堀裕和, 成瀬誠, 市瀬龍太郎: 特集「自然界に見いだす数物構造を利用した知的情報処理」, 人工知能, 33 巻 5 号, 2018.
- [2] 西郷甲矢人, 能美 十三: 圏論の道案内, (日本評論社, 2019).
- [3] イヴァセン, B.: 層のコホモロジー, 前田博信訳 (丸善出版, 2012).
- [4] フルトン, W.: 代数的位相幾何学入門, 上下, 三村護訳 (丸善出版, 2012).
- [5] デービス, M.: 超準解析, 難波完爾訳 (培風館 1982).
- [6] 中村徹: 超準解析と物理学 (日本評論社 1998).
- [7] 竹内外史: 層・圏・トポス (日本評論社, 1978).
- [8] René Thom: Structural stability and morphogenesis, Engl. Trans. D.H.Fowler (Westview Press, 1988).
- [9] 西郷甲矢人: 自然知能と圏論, 人工知能 33, 553-560 (2018).
- [10] 小嶋泉: 量子場とマイクロマクロ双対性 (丸善出版, 2013).
- [11] 前野俊昭: Schubert 多項式とその仲間たち (数学書房, 2016).
- [12] 堀 裕和, 成瀬 誠: 計算で物事を理解する・予測する (本シンポジウム2018), 第 5 部 圏論とその周辺 (2019).