

情報のユーザ側の委員からのご意見

p2-4 奥野委員（医療ライフ分野）

p5-6 来住委員（教育分野）

p7-22 栗原委員（化学・材料分野）

p23 津田委員（バイオ分野）

ライフ・医療系視点からの意見書

京都大学大学院医学研究科 奥野恭史

1. ライフ系研究者と情報系研究者のマインドセットの違い

多くのライフ系研究者の目的は、自らの研究対象の解明であって IT 技術を用いることでより早く目的達成を行うことに主眼が置かれている。一方で、情報系研究者は新規技術や既存技術よりも精度や性能のすぐれた技術開発に主眼が置かれていることが多い。このようなお互いの興味の違いからライフ系と情報系のコラボレーションがうまくいかないケースはこれまでも多くみられている。例えば、ライフ系研究者としては自らの研究対象の解析が出来るのであれば既存のどの方法を用いても良い訳で、コラボレーターの情報系研究者に独自の方法を開発してほしいとは思っていない（既存手法がない場合はもちろん話は別である。）。IT ツールの独自性、新規性よりも目的の解析をより早く実施されることを望んでいる。一方、情報系研究者の立場ではその逆の場合も多く、ライフ系研究者から相談を受ける解析テーマについて、十分なデータ数がないケースや表計算ソフトの利用で事足りるようなケースも少なくはない。そのため、情報技術、情報基盤がエンドユーザーに十分に利活用される真のデータ駆動型・AI 駆動型社会を実現するためには、このようなエンドユーザー側（ドメイン側）と情報系側のマインドセット、目的の根本的な違いを十分に理解し、両者が Win-Win 関係になる体制作り、人材育成が、今後の持続的発展の鍵であると考えます。

必要とする人材像について

次の 3 種に大きく分けて、組織構成と人材育成の必要がある。

- I. ドメイン側の情報インフラを支える SE や技術補佐系人材（ライフ系の場合、ライフ系研究者が自ら既存ソフトを利用して解析することも多く、それらを支援する人材）
- II. ドメイン側の目的に応えるデータ解析、手法や AI の開発、さらには目的達成のための研究デザインも提案できるデータサイエンティスト
- III. ドメイン側の目的に応えることのできる方法論や AI 技術が既存手法には無く、本質的な技術開発をしなければならないケースに対して、理論研究、アルゴリズム開発、プログラム開発ができる情報科学系研究者

ドメイン側でのデータ駆動型・AI 駆動型体制の構築に向けて

ここ数年、国研、大学に新設された AI・データサイエンスセンター（総合型データサイエンスセンターと呼ぶ）は、広範囲な複数のドメイン領域を対象にした組織となっており、ドメイン側の要求の多さに応えるだけのマンパワーとしてはまだまだ足りない。今後、Society5.0 の実現のためにも、ドメイン側に AI・データ駆動型サイエンス拠点を設置し、

応用に特化した研究開発体制を構築することが最重要である。ここで重要なのは、ドメイン側に設置する AI・データサイエンス拠点では太刀打ちできない本質的な研究課題を総合型データサイエンスセンターにフィードバックし、理論研究、アルゴリズム開発、プログラム開発が共同で行える流れをつくることである。上記の人材像のアサインとしては、総合型 AI・データサイエンスセンターが人材Ⅲを対象とし、ドメイン側の拠点が人材ⅠとⅡを対象にした組織構成や人材育成を行うことが想定される。また、ドメイン側の AI・データサイエンス拠点ではエンドユーザーとの共同研究の窓口、コーディネートする研究コンサルや、エンドユーザーのデータ管理サポート、ELSI などデータの取り扱いの専門機能も具備する必要がある。

2. 医療系のデータシェアリングについて

現在、健康医療分野では、個人のライフコースのデータ収集とそれによるデータヘルスが注目されている。健康医療におけるデータベースについては、健康状態（病気になる前）の記録である PHR（Personal Health Record）と病気の症状、治療（病気になった後）の記録である電子カルテなどの EMR(Electronic Medical Record)/EHR（Electronic Health Record）があり、PHR と EMR/EHR の両方が個人ごとに連結することでライフコースデータとしてはじめて成り立つものとなる。EMR/EHR については病院間の電カル仕様の違い（ベンダー間の違い）や医療業界の派閥が弊害となって、現在でも統合化は困難な状況であるが、AMED 事業などで学会レベル、研究者コミュニティレベルでの統合化を進めているところである。一方、PHR については、我が国では、いくつかの自治体（神奈川、神戸など）が最近になってシステム化に取り組みは始めているが、このままでは自治体独自の個別の PHR システムが出来上がり、統合化に難航する EMR/EHR と同じことになる。個々の病院主導、個々の自治体主導での EMR/HER, PHR の導入を早急に見直し、ナショナルセンターを中心とした国主導で統合化をすすめるべきである。

3. 分野ごと（ドメインごと）のデータストレージについて

SINET 整備に加え、分野ごとのデータストレージの配備も重要と考える。特に、ライフ系の場合、ゲノムデータ、画像データなど生体試料から得られた貴重なデータであるため、そのデータ保存は非常に重要である。現状は研究者、現場任せになっており、データが散逸している状況となっている。公的資金で取得されたデータは論文化されないものも含めて登録、保存する仕組みをつくるべきである。少なくとも学術団体ごとでの対応を強く推進し、しかるべきナショナルセンターで DB を統括すべきである。（ただし、研究者にデータ登録を義務化することで、当事者の雑務（過度の登録作業や必要書類の作成等）が増えることを避けなければならない。）

4. 標準化データのみでの登録ではなく、生データの登録も必要

同じデータセットを対象とした解析であっても、さまざまなアウトカムが想定されるため、特定のコミュニティが標準化をしたフォーマットのみで登録するのは良くない。いったんデータレイクに生データを登録（データの取得条件など背景情報も必ず付与）し、生データから解析目的、アウトカムに応じた標準フォーマットに変換する API 開発やキュレーションを行う仕組みを構築するやり方が望ましい。

5. データの取り扱いに関するレギュレーション (ELSI) について

AMED 等の研究費申請では、データマネージメントプラン作成、データ管理に加え、倫理 (ELSI) 専門家を研究メンバーに入れることが要件にされているが、そのために、医療に精通した IT 家、ELSI 専門家が複数の研究チームにまたがって参画する過度なエフォート参画を余儀なくされている。また、申請段階だけでなく、実際の業務においても過度なエフォートが発生しており、質担保の問題や時間的ボトルネックになっている。

これらの専門家人材は今後急に人数が増えるとは思えないため、まずは専門家を集約した拠点化をすべきである。専門家を個々の研究チームに含むことを研究費の申請要件にするのではなく、専門家を集約した拠点が各研究プロジェクトをサポートできる仕組みや体制を考えるべきである。

また、個人情報の取り扱いやレギュレーション策定については、IT 技術面からの精査も必要である。IT 技術におけるセキュリティ、安全性が考慮されずに、法律家や医療人によるレギュレーション決定がなされているように見受けられる。

以上

データサイエンス教育および発展に関する要望

2019年9月17日

津田塾大学 来住伸子

情報学の学会だけではなく、情報学のユーザ的な立場の学会からも意見を集めたいという依頼に対し、個人として、データサイエンス教育やデータ活用の面から参考意見を述べさせていただきます。

まず、情報学のユーザ的な立場の学会の活動を眺めると、最近ニュースでとりあげられたブラックホールの可視化など、最先端物理から社会学における情報収集まで、ほとんどの学会で何らかの形でデータサイエンス的な手法を試みています。情報処理学会と土木学会、電子情報通信学会等が連携してシンポジウムを開催するなどの試みもあります。このように多様な学問分野で、大量のデータを収集、解析、利用という一連の流れが社会活動、学問を進める上で、大変重要であるという認識は、アカデミア、学会間で共有されつつあります。

一方、データを活用する、ということ普及する、教育するという面から考えますと、実際のデータ、現実世界を理解するために、データを新しく収集し、そこから何かを学ぶという面が海外とくらべて十分に発達していません。すでに、膨大なデータがあり、それを解析する技術の研究、データを共有するインフラ整備は大変重要です。それに加えて、これからの社会における問題を見つけ、解決するために、さらに新しいデータを記録すること、集めることが重要であり、専門分野や学会にとらわれず、社会全体で、そのようなことが重要という認識を広めていくべき、と考えます。オープンデータ、誰でも利用できるデータの存在、また、プラットフォーム、データ解析の結果として再現性のある環境を提供することは、今後のデータサイエンス教育において重要な役割を持つことは間違いありませんが、さらに、新しく多様なデータを収集する、ということに社会全体で取り組んでいく必要があります。

アメリカでデータを利用して、大きな産業を生み出した背景には、データを重視する、数字を重視する、という長年の文化があると思います。たとえば、国勢調査はアメリカ憲法で10年に1回実施することと定められおり、1790年から実施しています。1880年の国勢調査の集計に7年かかり、1890年の国勢調査には13年かかると見込まれたため、政府が新しい技術を募集し、採用されたのがホレリスのタビュレーティングマシーンでした。国勢調査という課題のために、新しい技術を発明し、機械を製造するというを19世紀末には行ったこととなります。ホレリスはその後、会社をつくり、この会社がほかの会社と合併し、のちにIBMになります。

さらに、データサイエンスの教科書では、国勢調査をはじめ、アメリカの公的機関が公開し

た、さまざまなデータが例として使われています。たとえば、社会保障局、**Social Security Administration** は、1880 年から現在までの毎年、アメリカの赤ちゃんに付けられた名前、ファーストネームのリストを提供しています。これを利用すると、毎年の出生数を実際に合計でき、男女の名前の比較や分析が可能です。日本の赤ちゃんの名前ランキングは、日本の保険会社が保険契約者について集計したランキングで、データ教育に利用できる情報にはなっていません。この例から分かるように、アメリカでは実際のデータ、学生本人が親しみやすいデータ例を使って学ぶことが中等教育レベルから可能です。また、データを操作して、犯罪を解決したり、犯したりするテレビドラマも作られています。2015 年のエミー賞を受賞した **Mr. Robot**、日本で公開されたものでは「ナンバーズ、天才数学者の事件ファイル」などがあります。

日本でも、学生が親しみをもてる実際のデータでデータサイエンスを学べるようになり、データサイエンス研究者が、現実社会に深く関わり、一般の方にわかりやすい問題に取り組みやすくなるよう、新しい、多様なデータを収集、蓄積、公開することが重要であると思います。情報学の専門家がプラットフォーム等を率先して提供するとともに、情報ユーザとしてのさまざまな分野の学会に協力のもと、オープンサイエンスの一環として研究者対象のデータ、初中等教育で興味深く学べるデータなど、多様なデータの公開、新しい活用方法の普及に取り組むことが望まれます。

化学・材料分野と情報技術

栗原和枝（東北大学）

化学・材料科学分野においても、現在の Society 5.0 の推進のもとに情報科学への関心は高く、先行事例の創出や今後の展開への議論が開始されている。本メモでは、当分野での科学技術研究への情報技術の貢献について先行的な成功例や関心についてまとめ、さらに科学技術コミュニティへの関心の広がりとして、最近、当分野で積極的に開催されている研究会やシンポジウムについて例を紹介する。

I. 材料・化学分野の研究開発の新技术としての情報技術

材料・化学分野の主要なテーマは、新規物質の創成と特性の評価・解明である。研究開発の新技术としての情報技術は、広く分けると、(1) データベースならびに人工知能を用いこれらの研究を加速しようとするもの、(2) 従来の計測やシミュレーションに加えて情報技術を研究手段として用い、従来のアプローチでは困難な課題に新しい展開を図ろうとするものとなる。また、当分野では対象が複雑で、特定の性質についてデータが十分でない場合も多く、データや研究の状況により、両者の中間に位置づけられるようなアプローチが工夫されている。以下に、気づいた先行例を含め私見も含めて方向性をまとめる。

1. データベースならびに AI を用いる研究の加速

(1) 材料探索・創成における利用

特定の機能を持った新規材料の創成は、材料科学並びに化学の中心課題である。データベースの特定の基礎物性と機能の相関を調べ、材料の候補を絞って試料を調製し、優れた特性を確認した例が報告され始めているが、まだ方法論の提示の場合も多い。具体的な例として、

・断熱コーティング膜の開発(物材機構・NIMS、図 1)^{1,2)}

界面熱抵抗の DB の作成と記述子最適化と、機械学習界面熱抵抗予測モデルを用いて抽出した候補組成、専門家の選択の組み合わせより候補をさらに絞り、独自性の高い多成分・ナノ構造精密制御成膜装置を用いて(成膜条件制御)、構造最適化を行い世界最小熱伝導率向き薄膜を創成。

データ、予測モデル、専門家の知見、材料調製技術の協働の成果である。特徴的なデータがあれば小規模でも活用できる例でもある。(NIMS まであり's eye に紹介ビデオがある。)

尚、NIMS では材料への AI 活用での三つのアプローチとされる第一原理計算と機械学習を組み合わせる物質探索を中心とするマテリアルズ・インフォマティクス、プロセス・構造・特性・さらに時間に依存する特性をつなぐための統合的な計算材料工学であるマテリアルズインテグレーション、材料データの収集と管理を行うマテリアルズ・インフラストラクチャーへの取り組みが進んで

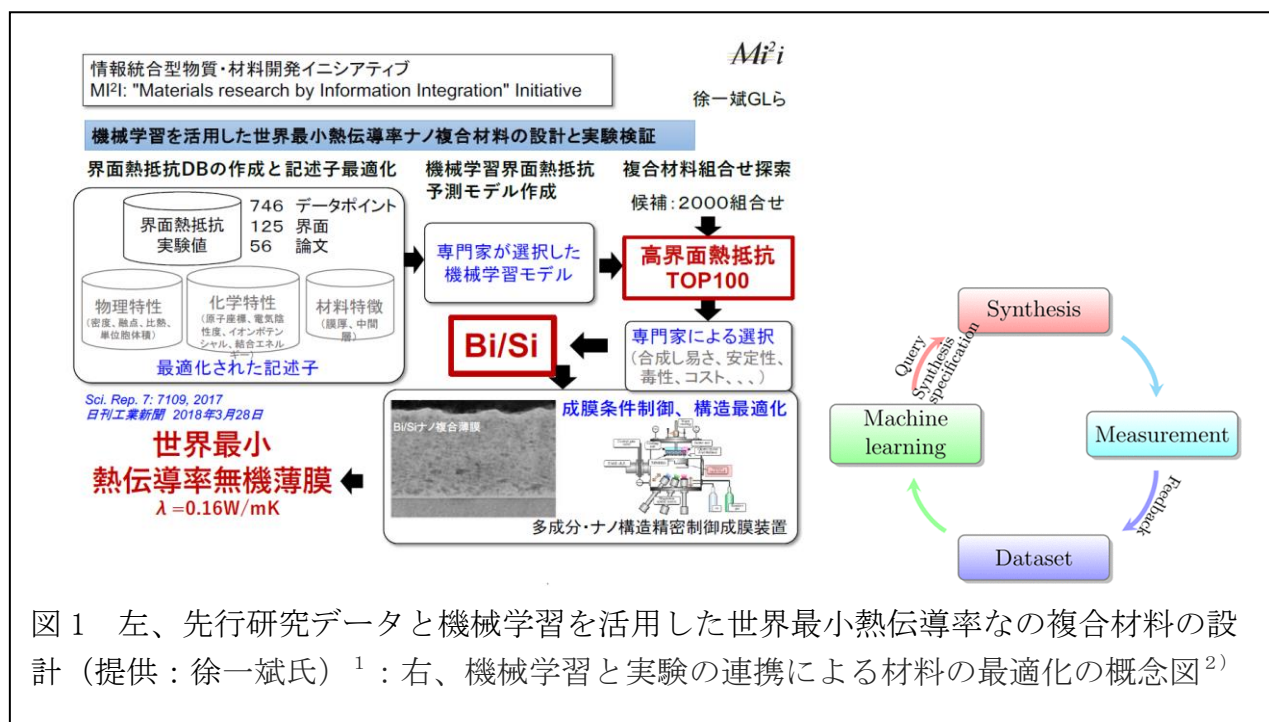
いる。上記の成果はその一部である¹⁾。

・データ科学による有機太陽電池のための高分子設計(阪大、JST)³⁾

文献からのデータを集め、合成の容易さなど実験的視点も加える、データ科学の活用による高分子設計のための方法論の提示。

・データ科学によるスピン流熱電変換素子の性能向上(NEC、東北大、JST)⁴⁾

コンビナトリアルな多くの材料調製とデータ科学を合わせて、スクリーニングし1年で100倍性能が向上。研究の大きな加速。



(2) 合成手法探索における利用

化学分野の中心課題は新規化合物の合成であり、そのための膨大な努力がなされ、様々な合成経路は例えば Organic Synthesis というような膨大なデータブック (1922 年より雑誌として刊行されている。現在はすべてオープンアクセス) にまとめられている。このような新規化合物合成の分野は、触媒の開発研究において野依や鈴木のノーベル賞受賞に示されるように、日本は非常に高いレベルにある。

有機合成分野においても、現在データベースを用いる新規合成経路探索、そして新規化合物合成は高い関心を得ている。その背景には、化学産業の高い関心があり、例えば、新規化合物の合成時間短縮、また少量合成から工業的合成までのスケールアップの時間が短縮できれば、大きな研究の加速、ひいては産業の競争力の向上にもつながると考えられている。そのため、後半に記すように、学会ベ

ースでの勉強会なども開催されている。

コンピュータによる合成経路の推定は早くから試行されていたが、ハードウェアやデータベースも十分でなく、実用には至らなかった。しかし、最近、コンピュータ支援有機合成 (CAOS) が再び注目を集め、Merck 社では CAOS として Synthia™ というシステムの運用が始まっているようである (付録資料、有機合成協会研究会第 4 回勉強会プログラム)。

化合物特性は多岐にわたることから新規化合物の提案をめざすには課題は多く、また化合物の構造そのものを表現する記述法 (記述子) の開発も必要と考えられている。これらの多様な研究を網羅するのは本ノートの範囲を超えているので、下記の先行研究を挙げるにとどめる。論文には広い背景も述べられている。

・反応効率の予測への機械学習からのアプローチ (産総研、NEDO)⁶⁾

シミュレーションで得た触媒の電子状態と振動パラメータを用い、機械学習によりアルケンのタンゲステン触媒によるエポキシ化反応の反応効率を 26% の誤差で予測する方法を提案。

・Diels-Alder 反応生成物の立体構造予測⁷⁾

どのような特徴量を取り上げるのが良いかを議論している。

(3) 分析データの解釈

データベースを利用し、赤外振動スペクトル、可視紫外電子スペクトルの同定をすることはルーチンの作業であるが、複雑な分子や高分子では必ずしも容易ではない。データベースを利用し、同定を容易にする方向はすでに示されている。さらに、たんぱく質など複雑な分子の構造同定にも展開が進むと考えられ、非常に基本的な点であるが、今後の装置共用などの計画には注意すべき観点であろう。

2. シミュレーションとの連携、シミュレーションの高度化

(1) データベース補完としてのシミュレーションデータ

様々な分野でデータ活用が求められている中、すべての分野でデータベースの実データ取得が充実させられるかは現実的には困難な可能性がある。そのような場合の補完、あるいはより先進的な取り組みとして、シミュレーションによるデータ創出さらに目的物質予測・提案が期待されている。

一方、シミュレーションによる物性予測の信頼性を上げるためにデータと機械学習を用い、対象の絞り込みの加速を行う実践も行われている。

現在進行中のポスト「京」(富岳) 開発プロジェクトでは、様々な分野のアプリケーションソフトウェアの開発が進められており、デバイス・材料科学分野も重点課題の一つである。その中で、これらの統合アプローチの試行が進んでおり、まだ実際の材料創成には至っていないが、富岳による大量の計算データを用いた成果が期待されている。現在試行中の課題と磁石材料に関する試行例を図2に示す。

ポスト「京」重点課題7



永久磁石の新材料開発 (ポスト「京」重点課題7 サブ課題D)

マテリアルズインフォマティクスの研究成果

- (A) 未知材料の仮想スクリーニングのため、物性値予測に適した汎用的な記述子である軌道場行列を考案。4,220種類の遷移金属化合物の生成エネルギーや658種類の遷移金属化合物の局所磁気モーメントで有効性を検証。
- (B) 階層クラスタリングを用いて重要な記述子を特定するサブグループ関連性解析を開発し、希土類遷移金属合金のキュリー温度データに適用。
- (C) ベイズ最適化を用いた効率的な結晶構造探索手法を提案。

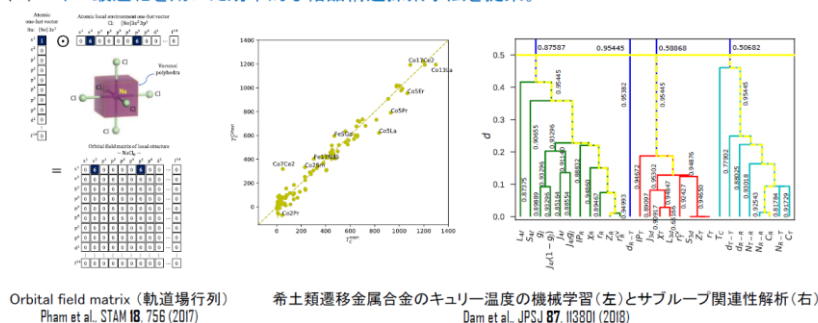


図2 東京大学常行真司教授提供

(2)シミュレーションの高度化のための情報技術

コンピュータの大型化に伴い、計算できるシステムのサイズは大きくなってきたが、まだ材料の計算には十分でない点も多く、様々な粗視化が必要である。

これらに情報技術は貢献できよう。特に時間発展については十分でない。慶応大学の泰岡顕治らは、シミュレーションの時間発展を予測する手法 (MD-GAN) にディープラーニングを適用している⁸⁾。このような新たな情報技術利用の展開も重要であろう。

3. 現象の階層をつなぐ情報技術

現象の階層性については、地震や津波の発生の地層レベルの変化から大規模現象、場合によれば被害予測までをつなぐ情報技術が期待されている。材料科学・化学分野にも様々な現象の階層性があり、その間をつなぐのは課題である。

例えば合成化学における実験室レベルの合成から工場プラントでの製造には、条件の設定に様々なギャップがあるとされている。これらの開発を、初期の少量スケールでのデータを活用して連続的にできれば、開発が大きく加速できよう。

また、材料の分子／原子レベルの構造や特性が、最終的なマクロ機能／特性の発現までどのように続いているのかには幾つかの階層があり、十分につながっているとは言えない。よく知られた例として、土井正男 (東京大学名誉教授、開発時は名古屋大学) を中心に開発された高分子シミュレーションソフト OCTA⁹⁾ があり、特にメゾ領域の計算に優れているとされているが、幾つかの階層に区切り、下の階層で計算したパラメータを移すような手法である。このような階層を超えた接続を情報技術を用いてできれば有効性は高く、またナノレベルの実験データをマクロ現象につなぐ場合にも適用できる可能性があると考えている。

II. 材料科学・化学分野でのデータベース

よく知られまた活用されているデータベースには国外のものとして、SciFinder (米国化学会、2 億以上の物質情報)、PubChem (NIH 傘下の NCBI 提供、1 億近くの物質情報)、Cambridge Structural Database (100 万余の有機物結晶構造)、Protein Data Bank (PDB, 15 万近い蛋白構造に関するデータ) などがある。前記した Organic Syntheses もデータベースといえよう。日本では、主に無機物質・材料・高分子についての NIMS のデータベース MAT NAVI が多くのデータを網羅している。PDB には、これまでに X 線結晶構造解析、NMR およびクライオ電子顕微鏡などで構造解析された蛋白質の立体構造がすべて網羅されており、日本は PDBj (日本蛋白質構造データバンク) が世界 4 拠点の一つとしてアジア地区からのデータ登録などを行っている。

実験結果のみならず、計算科学も組み込んだデータベース化も進められている。2011年、実験・データ科学・計算科学を連携させた手法により、多数の物質・材料についての膨大なデータを構築し、新材料や新機能の開拓を加速することを目指したアプローチが、米国において、Materials Genome Initiative として始まっている。

最近のマイクロデバイス分析化学の発展により、多くの試料を条件を変えて調製し性能を測定し性能向上を図る、溶液条件を変えてたんぱく質の結晶化の進行を調べさらに構造解析に進むというような研究法が多くなっている。この

ようなデータ創出の機会は貴重なので、それらを小さなデータベースとし、機械学習により目的の物質や機能を得るといったようなアプローチも、対象により有効であろう。

III. 学術界での関心、学会活動など

日本学術会議並びに学会での議論も活発になされている。

日本学術会議では、化学委員会化学企画分科会のもとに“情報科学との融合による新化学創成 小委員会”が作られ、提言“化学・情報科学の融合による新化学創成に向けて”の提出（今年度）を目指して現在、議論を進めている。現状や課題を示し、AIによる分野横断型の新しい化学の創製、情報科学を活用した化学教育の変革、化学データの戦略的収集と戦略的創出、情報科学を活用した化学学術・生産の連携システム構築などの観点を捉え、人材育成から技術課題まで広く議論がなされている。

学会におけるシンポジウムや研究会などの活動を添付資料に例示する。具体的な利用を目指し議論する活動が始まっていることがわかる。例示したのは、

1. 日本鉄鋼協会の第70回白石記念講座 「AI（人工知能）、ビッグデータが拓く 鉄鋼の未来を考える」（2018年10月）プログラム
2. 日本有機合成協会のAIと有機合成化学研究部会の第3回公開講演会、第4回勉強会のプログラム、

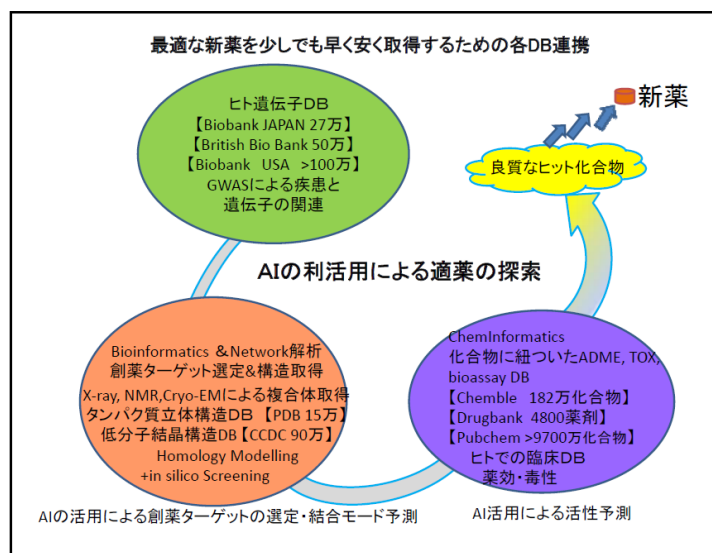


図3 上村みどり氏提供

3. 情報計算化学生物学会（CBI 学会）の2019年大会（10月22日～24日）「構造生物学と情報科学の真の融合を目指して～AI時代の創薬化学～」の開催趣

旨と 2011 年からの年会のテーマリスト

である。CBI 学会は 1981 年に設立されたそうであるが、近年、時機を得て活発な活動がされている。図 3 は当学会の期待の一例である。

IV まとめ

化学・材料科学分野の情報技術の活用について、目に触れる範囲でまとめた。期待は大きく、先行的な成果もあるが、対象が複雑なため、課題も多い。データを単に情報処理するようなアプローチでは難しく、実験、シミュレーションとの協働が必要であり、それをどう進めるか、同時にどのようなデータベースを作成するのか（しないのかも含め）などが課題であろう。また、急速に発展している部分に対する迅速な人材育成も課題である。

<謝辞>

次の方に資料を提供いただいた：物質・材料研究機構統合型材料開発・情報基盤部門 出村雅彦副部門長、東京大学理学系研究科 常行真司教授、北里大学 菅原洋子名誉教授、帝人ファーマ（株）生物医学総合研究所 上村みどり 上席研究員。

参考文献

1. 出村雅彦、“AI を活用した材料開発における我が国の取り組み～NIMS を例として” 第70回白石記念講座「AI (人工知能)、ビッグデータが拓く鉄鋼の未来を考える」(日本鉄鋼協会編, SMS-ISIJ 70, ISSN: 1344-0942)、2018年11月26日
2. Z. Hou, Y. Takagiwa, Y. Shinohara, Y. Xu, K. Tsuda. Machine-Learning-Assisted Development and Theoretical Consideration for the $\text{Al}_2\text{Fe}_3\text{Si}_3$ Thermoelectric Material. *ACS Applied Materials & Interfaces*. **11** [12] 11545-11554 (2019)
3. S. Nagasawa, E. Al-Naamani, A. Saeki, Computer-Aided Screening of Conjugated Polymers for Organic Solar Cell: Classification by Random Forest, *J. Phys. Chem. Lett.*, **9**, 2639–2646 (2018).
4. https://jpn.nec.com/press/201802/20180209_04.html
5. S. Szymkuc, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, Computer-Assisted Synthetic Planning: The End of the Beginning, *Angew. Chem. Int. Ed.*, **55**, 5904-5937 (2016)
6. A. Yada, K. Nagata, Y. Ando, T. Matsumura, S. Ichinoseki, K. Sato, Machine Learning Approach for Prediction of Reaction Yield with Simulated Catalyst Parameters, *Chem. Lett.* **47**, 284-287 (2018).
7. W. Beker, E. P. Gajewska, T. Badowski, B. A. Grzybowski, Prediction of Major Regio-, Site-, and Diastereoisomers in Diels–Alder Reactions by Using Machine-Learning: The Importance of Physically Meaningful Descriptors, *Angew. Chem. Int. Ed.* **58**, 4515-4519 (2018).
8. K. Endo, K. Tomobe, and K. Yasuoka, *Thirty-Seond AAAI Conference on Artificial Intelligence*, New Orleans, USA, 04-09, Feb., 2018
9. octa.jp/jp/
10. www.advancesoft.jp/product/advance_octa/advanceocta_04.html

添付資料(学会における研究会など)

1. 日本鉄鋼協会

第70回白石記念講座「AI(人工知能)、ビッグデータが拓く 鉄鋼の未来を考える」

講座の視点

最近のセンシングや計算機能力の向上とともに、ビッグデータや AI(人工知能)を活用した技術が目覚ましい進歩を遂げ、様々な分野で適用が広がっている。鉄鋼業界においても、操業支援や労働生産性向上などを目的に、これらの技術を導入する機運が高まっている。本講座では、ビッグデータや AIに関する最新の知識や技術を学習し、技術革新による鉄鋼業の進歩や発展の可能性について考える。最初に、根幹となる AI 技術を中心にコア技術について概説する。引き続いて、材料開発とプロセス技術を主要テーマとして、それぞれの要素技術の最新情報や材料研究、生産インフラの保全、AIソリューション、自動運転など、様々な分野での最先端の研究開発の取組状況などについて紹介するとともに、鉄鋼業での材料開発や生産・物流プロセスへの応用可能性について議論する。

協賛: (公社)応用物理学会、(公社)化学工学会、(公社)計測自動制御学会、(一社)資源・素材学会、(50音順) (一社)システム制御情報学会、(一社)情報処理学会、(一社)人工知能学会、(一社)電気学会、(一社)電子情報通信学会、(公社)土木学会、(一社)日本機械学会、(公社)日本技術士会、(公社)日本金属学会、(一社)日本建築学会、(公社)日本材料学会、(一社)日本塑性加工学会、日本中性子科学会、(一社)日本熱処理技術協会、(公社)日本分析化学会、(一社)表面技術協会、(公社)腐食防食学会、物質・材料研究機構、(一社)溶接学会

1. 日時・場所 2018年11月26日(月) 9:30~16:30 受付時間 9:00~15:30

早稲田大学 西早稲田キャンパス 63号館 2階会議室(東京都新宿区大久保 3-4-1)

2. 講演題目・講演者、司会者

司会者: 玉置久(神戸大学)

1) 9:30~10:30 [人工知能技術の発展と展望](#)

産業技術総合研究所 情報・人間工学領域 人工知能研究センター 副研究センター長 麻生 英樹

2) 10:30~11:30 [AIを活用した材料開発における我が国の取り組み~NIMSを例として~](#)

3) 12:20～13:20 [材料開発ツールとしてのマテリアルズインテグレーション](#)

東京大学 大学院工学系研究科 マテリアル工学専攻 教授 榎 学

司会者: 出村雅彦(物質・材料研究機構)

4) 13:20～14:20 [人とAIの協調による新たな社会価値の実現に向けて](#)

日本電気株式会社 データサイエンス研究所 所長 広明 敏彦

5) 14:30～15:30 [超成熟社会に向けたクルマの知能化](#)

神奈川工科大学 創造工学部 自動車システム開発工学科 教授、
先端自動車技術開発研究所所長, 自動車工学センター長 井上 秀雄

6) 15:30～16:30 [適応的エリアセンシング手法を用いた知能化設備異常診断](#)

神戸大学 大学院システム情報学研究科 教授 玉置 久

3. 講演内容

1) 人工知能技術の発展と展望 麻生 英樹

人工知能技術は、1950年代から研究が続けられてきたが、2000年代になって、インターネット上の検索、推薦、SNS等のサービスを通じて収集された大規模データに、深層学習を始めとする統計的な機械学習技術を適用することで高度化し、様々な形で実用化され、社会の知的インフラ技術の一つとなった。さらに近年は、IoT やロボット技術とも結びついて、実世界でのものづくりやサービスへと社会実装の場を拡げつつある。本講座では、まず、人工知能技術の研究の歴史と現在の到達点について概説し、次に、現在重要になっている「実世界に埋め込まれる人工知能」の課題や基盤技術について最近の研究開発事例を紹介し、最後に、今後の方向性についても考察したい。

2) AIを活用した材料開発における我が国の取り組み～NIMSを例として～ 出村 雅彦

AIを活用した材料開発手法刷新への取り組みが世界的に活発化している。囲碁や画像認識、医療診断等で伝えられるAIの可能性に期待する一方で、複雑な材料の課題にどこまで適用できるのか疑問といったあたりが、現状の認識と思われる。確かに、材料の性能を支配する因子の組み合わせは無限と思えるほど広大で、これに比して材料データは不十分であり、また、質の評価を含め、材料データを集める方法論も未だ試行錯誤の段階である。本講演では、当該分野の世界的な動向を俯瞰しつつ、我が国における取り組みと其中で現在得られつつあるいくつかの先行的な事例を紹介し、材料専門家の閃きを引き出し、直感を磨く、そういうAI活用の可能性について議論したい。

3) 材料開発ツールとしてのマテリアルズインテグレーション 榎 学

構造材料開発においてはそれらの性能評価のために多くの時間が必要となり、開発速度を律速する原因となっている。したがって、組織や性能の予測がある程度の精度で可能となることは材料開発において非常に意義がある。これまで個々の特性や性能に関しては、理論的な考察に基づく理論式や、多くの実験結果から得られた経験則が提案されている。一方、近年のコンピューティングの飛躍的な発達により、種々のマルチスケールでの計算手法を組み合わせることが可能となってきている。構造材料の組織と時間依存の性能の予測を、理論や経験則、数値計算、データベース構築、データ駆動型アプローチなどの融合により可能とするマテリアルズインテグレーション(MI)システムについて概説する。

4) 人と AI の協調による新たな社会価値の実現に向けて 広明 敏彦

NEC は半世紀にわたって AI 技術「NEC the WISE」に取り組み、現在は IoT 技術との融合により、様々な業種のお客様と共にビジネス変革の可能性を追求している。人を超える繊細さとスピードで様々な兆しを捉える自動認識、大規模で複雑なシステムのホワイトボックス型自動最適化など、世界 No.1/Only one の AI 技術で実現するデジタルトランスフォーメーションの可能性を事例と共に紹介しながら、皆様と共に鉄鋼業の生産プロセスイノベーションなどへの応用可能性を探る。

5) 超成熟社会に向けたクルマの知能化 井上 秀雄

日本は、超成熟社会を迎え、多くの課題を抱えている。自動車産業においても、高齢社会、交通渋滞、交通事故などの課題を解決し、豊かな活力ある社会を築いていくことが重要である。又、ICT、ITS などの技術発展も目覚しく、自動車にも、情報通信技術、エレクトロニクス技術はなくてはならないものになっている。このような背景をもとに、単なる自動運転ではなく、安全の進化の観点から、クルマの知能化の方向性に言及する。特に「情報モデルと物理モデルを融合したリスク予測」「人とクルマの関係に革新を齎す Shared Control」等の運転知能技術について、そのプロジェクトと技術内容を示す。

6) 適応的エリアセンシング手法を用いた知能化設備異常診断 玉置 久

国内製鉄所の老朽化進行に対して、基盤インフラ設備は極めて数多くかつ広域にわたって存在するため、部分的なモニタリングと人に依存した判断だけでは設備の安定稼働を支えるに十分ではなく、メンテナンス不足に起因したトラブルは根絶できていないのが現状である。このような背景のもと、計測・制御・システム工学部会で推進中の研究会「適応的エリアセンシング手法を用いた知能化設備異常診断」では、製鉄所の安定稼働・レジリエンス向上を目指し、高度なセンシング画像をベースとした大量データをもとに、人が見つけるのが困難な異常部位・異常状態の診断システムをテーマに研究開発を進めている。本講演では、この研究会の狙いと進捗・成果を中心に活動内容を紹介する。

2. 有機合成化学協会

第3回公開講演会「AIと有機合成化学」

日時

2019年1月31日(木)13:00~17:30(17:45~19:00 情報交換会)

場所

早稲田大学国際会議場[東京都新宿区西早稲田 1-20-14]交通:東京メトロ東西線「早稲田駅」下車徒歩10分

主催

有機合成化学協会

共催:早稲田大学理工学術院総合研究所



有機合成化学

第3回

公開講演会

公益社団法人有機合成化学協会におきましては、平成30年2月より「AIと有機合成化学」研究部会を設立、有機合成におけるAIの利用を学術・技術の両面から検討する試みを始めました。

今後も定期的な講演会・研究会を開催し協会員同士の情報交換を図り、これからの社会に広く貢献していく有機合成化学の未来のかたちを探ります。今回も有機合成化学におけるAIの利用にいち早く取り組まれている先生方にご講演をお願いし、「AIと有機合成化学」第3回公開講演会を開催いたします。有機合成化学協会の会員の皆様だけでなく広い分野の方々のご参加をお待ちいたしております。

プログラム

※下記の時間は質疑応答を含む / 敬称略

【第I部】話題提供・講演(於:国際会議場 井深大ホール)

開会挨拶 13:00~13:05

- 挨拶:(東レ株式会社医薬研究所副所長)林 亮司

「創薬研究の最前線」(13:05-14:05)

- (帝人ファーマ株式会社生物医学総合研究所上席研究員)上村みどり
- 「適切な薬剤を適切な量、適切な時間」に使うことはまさに、自分や家族が病気になった際に患者が望む理想の医療をうけるという意味で非常に大切な概念である。そのためにも、臨床現場におけるビッグデータをもとにした AI を利用するさまざまな取り組みが我が国においても始まっている。ドラッグデザインにおいても、より効率的に既存の手法をリストラクチャリングすることで臨床での薬効予測を高めることでコストの削減を可能とする。

「人工知能(AI)と統計モデル」(14:10-15:10)

- (株式会社ワイハット代表取締役社長)梶山昌之
- 人工知能(AI)の本質はどのようなものとお考えでしょうか。また、AIを有機合成化学に活用する方法を検討中の方も多いたと思います。しかしながら、AIと機械学習はなにが違うのか、機械学習と統計モデルの違いは何かについて疑問を持っている方も多いたのが現状です。また、どのような課題に対してAIが適しているのかは良く知られていません。本講演では、これらの疑問について解説します。さらに、AIの活用事例を紹介し、AIによりもたらされる未来について考えます。

一休憩&ブレイク※コーヒーをご用意しております (15:15-16:00)

「データ駆動による新材料の発見」(16:00-17:00)

- (京都大学工学研究科教授)田中 功
- 近年、実験や理論計算で大量のデータの収集が可能となり、またそれらを活用するためのデータ科学手法に大きな進歩が見られている。演者らは、網羅的な第一原理計算をもとに、超低熱伝導度材料、光触媒材料等について、効率的に候補材料を見つけ出すことに成功した。また無機物質の結晶構造データベース ICSD 収録の既知化合物データをもとに、第一原理計算や記述子を使用せずに安定な未知化合物が存在する化学組成を予測する推薦システムを構築した。これらの成果について紹介する。

「AIと有機合成化学」第4回勉強会

2019年10月2日(水)14:30~16:40(※交流会=17:00~18:30)

[勉強会]: ワイム貸会議室「お茶の水」(RoomA+B)

[交流会]: 明治大学アカデミーコモン1F「カフェ パンセ」

有機合成化学協会



有機合成化学

研究部会

第4回勉強会

10月2日(水)ワイム貸会議室「お茶の水」(RoomA+B)にて、「AIと有機合成化学」研究部会—第4回勉強会を開催いたします。(研究部会員対象)

Chematica (Synthia)の開発者本人の講演となります。皆様の関心の高いシステムの能力を知る良い機会かと思えます。データ駆動有機合成の最先端です。なお、製品化しているシグマアルドリッチジャパン合同会社田中紀子氏からも現在の状況を短く紹介していただこうと思っております。

勉強会 プログラム(敬称略)

I. 「講演の部」

14:30~

趣旨説明松原 誠二郎(「AIと有機合成化学」研究部会代表幹事/京都大学大学院工学研究科教授)

14:40~16:15

1. 「Organic Synthesis, Synthetic Planning, and Reaction Design Using Computer」
Bartosz A. Grzybowski(Distinguished Professor, Ulsan National Institute of Science and Technology)

16:20~16:40

2. 「Synthiaの現在位置」田中 紀子(シグマアルドリッチジャパン合同会社)

3. 情報計算化学生物学会

2019 年会大会テーマ :

構造生物学と情報科学の真の融合を目指して ~AI 時代の新創薬化学~

開催趣旨

BDD が創薬に利用されるようになって久しくなります。しかしながら、それが日本の製薬会社における創薬研究で最大限に活用されているかといえば、現実にはそれに程遠いものと感じられます。構造生物学が創薬に十分に生かされていない理由の主なものは、実際に見たいターゲットの構造をタイムリーに取得するのが難しく、そして、構造を解析する速度が遅すぎて実際に製薬会社の求める創薬サイクルに供することができないことが挙げられると思います。これらの制限を回避するために、シミュレーションを利用する機会が多く、これまで構造生物学と情報科学は必ずしも最適な協調関係にあったとは言えません。

近年の構造生物学の進展は、この状況を一変させる可能性があります。一つはクライオ電子顕微鏡を用いた単粒子解析により、結晶を用いなくても生体分子の原子分解能の構造を得られるようになったことです。これにより、構造解析の対象は超分子複合体や染色体のような巨大で柔軟な構造を持つものへと広がりつつあります。別の興味深い動きとしては、自由電子レーザーを用いた動的構造解析が挙げられます。これはフェムト秒パルスの X 線を用いてストロボ撮影のようにタンパク質の動きを捉える技術であり、これにより、数十フェムト秒で起こる化学反応から MD のシミュレーションでは計算機の能力から難しいミリ秒や秒にいたる広い範囲の時間分解能での構造変化のその場観察が可能になりました。

これらの二つの方法に共通しているのは、その解析に情報科学が大きな役割果たしうるという点です。単粒子解析のサンプルは異なった状態の構造が混ざっており、これをどのように処理していくかは大きな問題です。これを逆手にとって、多くの粒子のイメージからタンパク質の構造変化を一網打尽にできる可能性があります。自由電子レーザーの動的解析においては、タイムポイントをつなげるため、そして同期が徐々に外れていき複数の状態が混じったものを分離するためにも計算機科学との組み合わせは不可欠であると考えられます。

本大会では、上記の新手法やこれまでの方法の新展開なども含めて、構造生物学の新しい潮流をどのように情報科学と組み合わせ、それから創薬に役立つ情報を最大限に引き出すことができるのかについて、構造生物学と計算科学の専門家が集いディスカッションできればと考えています。

CBI 学会 2019 年大会 大会長 岩田 想 (京都大学大学院医学研究科)

実行委員長 上村 みどり (帝人ファーマ株式会社)

最近の年会のテーマ

- 2018年 [「創薬と育薬のレギュラトサイエンス～AI創薬時代の新展開～」](#)
- 2017年 [「データ駆動型研究が拓く創薬」](#)
- 2016年 [「in silico 創薬の将来－生体分子シミュレーション、構造生物学、ビッグデータの連携から アカデミア創薬へ」](#)
- 2015年 [「創薬のオープンイノベーション－新領域と in silico の接点－iPS 創薬・アカデミア創薬・ビッグデータ」](#)
生命医薬情報学連合大会 [「生命情報ビッグデータ時代のバイオインフォマティクスの挑戦 ～環境から医療まで～」](#)
- 2014年 [「iPS, ion channel, in silico が拓く、新しい創薬パラダイム」](#)
生命医薬情報学連合大会 [「データ駆動型生命情報科学の挑戦 次世代型医療への扉を拓く」](#)
2014年10月2日 ≫ ≫ [CBIシンポジウム](#)
- 2013年 [「オミックス・計算 そして創薬」](#)
- 2012年 [「情報学がつなぐこれからの生命医科学」](#)
- 2011年 [「計算科学の拓く新しい生命像」](#)

2019 年9 月13 日
東京大学 津田 宏治

第三回情報委員会：バイオ分野からの論点

○多数の癌サンプルからのDNA シーケンシング、超多数の細胞の遺伝子発現量を別々に取得できる一細胞解析など、バイオ分野におけるデータ量は増加の一途を辿っており、求められるデータ解析の技術も高度化している。バイオ分野の研究を行う上で、情報分野からの支援は不可欠である。

○大学や研究機関で用意されているスーパーコンピュータやクラスタマシンには、物理シミュレーションを前提として設計されているものが多く、大量のメモリとストレージが必要なゲノム配列処理には向かないものが殆どである。また、個人ゲノムなどを扱う際には、セキュリティ面での配慮が求められるが、対応できるシステムは少ない。東京大学医科研と、国立遺伝研に、バイオ専用のスパコンがあるが、能力は不足しているのが現状であるため、新たな計算機資源の配分が必要である。

○現状、大規模な研究は、情報系研究者も含めたチームサイエンスの形で行われることが多い。チームサイエンスにおける情報系研究者の仕事（データサイエンス）は、通常の情報科学の研究とは根本的に異なっている。新たな計算手法の開発は必ずしも必要ない。ドメイン知識に基づいて、研究対象の検討と、実験データのモデリングを行い、適切なツールの組み合わせを行なって、共同研究者が解釈できる解析結果を取得する。物理・化学・生物などの学会で求められるのは、データサイエンティストであるにも関わらず、そのような人材を養成する機関が不足している。例えば、ハーバード大学には、IACS (Institute for Applied Computational Science) という機関があるが、国内でもこのような試みが必要である。

○現状の情報分野の研究者の評価システムは、新たな手法を考案するということに重点が置かれすぎていて、データサイエンティストは、その専門性を正当に評価されない傾向がある。例えば、論文において、第一著者や責任著者になることは稀であり、多数いる著者の中に埋没しがちである。研究資金の申請書に「チームサイエンスへの貢献」の欄を作るなど、データサイエンティストのキャリア形成を支援する方策が必要である。

○KEGG, MAFFT など日本発で世界標準で用いられているデータベースやソフトウェアと、それを作成・管理している研究者を高く評価すべきである。