

CBI 学会 2019 年大会 フォーカストセッション FS-12

「医療と薬づくりの AI、新しい課題は何か？」

補足解説（内容は当日まで、加筆修正される可能性があります）

## 「医療と薬づくりの AI、新しい課題は何か？」

### 要約

CBI 学会は、いわゆる第 2 次 AI ブーム以前の 1970 年代末から構想され、1980 年代から活動が開始された。そこでは統計手法、パターン認識（機械学習）、エキスパートシステム、分子計算と 3D 分子グラフィクス、化合物や配列データベースなど、今日話題になっている情報計算技法のほとんどの活用が考えられていた。このセッションでは、今日話題になっている AI が CBI 学会にどのような衝撃を与えつつあるのか、新しい課題は何か、それらにどのように取り組むべきかを討議することを目的としている。

課題 1. 計算化学 Computational Chemistry と情報化学 Chemoinformatics の本質的な融合: これまで計算科学と情報化学は、独立に発展してきたが、機械学習が計算化学に浸透し、さらに量子計算とのつながりが認識されるようになっている。

課題 2. AI 論理の透明性と説明可能性、Explainable AI (XAI): FDA が AI 内臓の機器を認可し始め、内部の計算技法の論理性、信頼性、説明可能性への懸念と関心が急激に高まっている。

課題 3. 製薬企業の「パイプライン」再考: NIH が Translational Research の司令塔として設立した NCATS の所長であるオースチン (C. Austin) は、ビッグファーマの研究者と製薬企業の「パイプライン」再考の会合を開催して、点分子およびバイオ医薬ごとの「パイプライン」の再考案を公開し、議論を呼びかけた。

課題 4. 人材（とくに D2K サイエンスの）育成: NIH は、データサイエンスを基盤とした組織 (Enterprise) に進化すると 2018 年に宣言した。我が国でも、データサイエンス（あるいは D2K サイエンス）の人材育成が焦眉の急になっている。

.....

詳細説明（以下の資料は作成中であり、修正追加を予定しているが、参加者の便宜のために公開する。）

## 課題 1. 計算化学 Computational Chemistry と情報化学 Chemoinformatics の本質的な融合

### 計算化学

計算化学の源流は、量子力学の構築に寄与したディラックが述べた、「量子力学ができたことで、電子の集合の動きから説明される化学は、量子力学の応用計算問題になった」という意味の宣言にある。それは 1920 年代の終わり、今から約 90 年前のことである。そうしたアプローチは、計算機の進歩によって、広く試みられるようになった。その典型例は Gaussian と MD 計算であり、我が国で開発された FMO 計算もその流れを汲む。同様な計算としては原子核の構成要素である核子(フェルミオン)や半導体の中の電子ガスモデルなどがある。これは物理学では多体問題 Many-body Problems と呼ばれている。その最も単純化された近似計算では、粒子同士が相互作用のない状態であると考える (Hartree-Fock 近似)。だが、実際にはそれらの粒子の間に相互作用が働いている。いわゆる量子計算では、そうした粒子の相互作用を量子纏れ (Quantum Entanglement) と呼んでいる。

### 情報化学

薬づくりの化学 (Medicinal Chemistry) の仕事は、薬となりうる化合物のデザインとそれを生成する合成化学、とくに有機合成である。その基盤となるのは、化合物の機能と構造との (相関) 関係や、薬候補化合物と標的となる生体分子と結合の解析 Docking Study である。その関係をしらべる研究方法のひとつが構造活性相関 (QSAR あるいは単に SAR) である。一方で、ある構造をもった化合物を実際に合成するには、優れた化学者の経験的な知識が必要になるが、そうした複雑な経験知識を論理的な規則に整理して、計算機で扱えるようにして、合成を支援するという合成支援や、分子のスペクトルから構造を予測するというようなシステムは、1960 年代から研究されていた。分子構造を 3D で表現するグラフィックスや化合物とその計測値、例えば各種のスペクトルデータのデータベースなどの集積と、それらに関係した各種の情報計算技法は、Chemoinformatics と呼ばれるようになった。その流れは広がり、よく知られている Gasteiger らによる教科書でも多彩な技法が紹介されている。

### 参考文献

- E. J. Corey, et al., Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics, *J. Am. Chem. Soc.* 94, 421-430 (1972).
- J. Gasteiger, Chemoinformatics: Achievements and Challenges, a Personal View, *Molecules* 2016, 21, 151

- ・ J. Gasteiger の講演スライド、Solved and Unsolved Problems in Chemoinformatics。
- ・ I.W. Davies, The digitization of organic synthesis, *Nature*, 570: 175-181, 2019.
- ・ H. Chen et al. Rise of deep learning in drug discovery, *Drug Discovery Today*, 23(6): 1241-1250, 2018.
- D. Schwalbe-Koda and R. Gómez-Bombarelli, *Generative Models for Automatic Chemical Design*, 2019.
- ・ 芹沢貴之、Medicinal Chemistry への AI の活用、*ファルマシア*、55(10): 919-923, 2019.

#### 新しい潮流

構想や準備を含めると CBI 学会の歴史は、1970 年代末に遡るが、それから最近まで、量子化学や古典的なニュートン力学を基礎にした MD (Molecular Dynamics) 計算、FMO 法などの計算化学と情報化学の研究者たちは、少なくとも仕事の上では、ほとんど独立に交流なく活動してきたように思われる。ところが最近、機械学習あるいは深層学習の技法を計算化学に組み込む試みがなされるようになってきた。この流れは CBI 学会の歴史からみると、本質的に新しい流れであるように思われる。

この新しい潮流に参加するためには、量子化学の研究者が深層学習や機械学習を学ぶ必要がある。その基本には、物理学の多体問題と同じ近似の問題がある。そこには量子計算の基礎になる量子纏れ (Quantum Entanglement) の問題がある。一方で、現在の機械学習や深層学習を使って「望みの化合物をこれまでよりずっと早くまた効率的に合成してみせる」ことを売り物にする、Kebotix のような野心的な若い企業も出現している。こうした状況は、量子化学の研究者が、単に Gaussian のようなパッケージを使って計算できるだけでなく、機械学習や深層学習の技法を理解することで計算化学の精度を高めたり、量子化学の計算技法が量子計算の原理と深く関係していることを理解したりするような、教育の機会が必要なことを示唆しているように思われる。

#### 参考文献

- ・ N. Nosengo, The material code, *Nature*, 533: 22-25, 2016.
- ・ A. Chandrasekaran et al. Solving the electronic structure problem with machine learning, *Computational Materials*, 5:22 2019.
- ・ L. Cheng et al. A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules, *J. Chem. Phys.* 150: 131103, 2019.
- ・ R. Zubatyuk et al. Accurate and Transferable Multitask Prediction of Chemical

Properties with an Atoms-in-Molecule Neural Network, ChemRxiv. 2018

- S. Amabilino et al. Training Neural Nets To Learn Reactive Potential Energy Surfaces Using Interactive Quantum Chemistry in Virtual Reality, *J. Phys. Chem. A*, 123: 4486–4499, 2019.
- J. G. Freeze, H. Ray Kelly, and V. S. Batista, Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists, *Chem. Rev.* 119: 6595–6612, 2019.
- T. Nudajima et al. Machine-learned electron correlation model based on correlation energy density at complete basis set limit, *J. Chem. Phys.* 151: 024104, 2019.

## 課題 2. AI 論理の透明性と説明可能性、Explainable AI (XAI)

現在の AI の象徴的な技法である深層学習 (Deep Learning, DL) のアキレス腱は、よい結果を出してもその理由が説明できないことにある。このことは画像認識や将棋のような判定や判断の結果がわかりやすい場合は問題ないが、医薬品の研究開発や医学的な診療など、技法の良否の判定や判断が難しい場合は、実用化への大きいな障害になると思われる。

FDA は、すでに多くの AI 内臓製品を認可しているが、診療の結果に責任を負うのは医師など、国家資格のあるサービスの提供者である\*。計算機がどう判定しているか、その計算法 Algorithm を、そのような専門家が理解可能なように表現し、理解可能にしておくことは、そのような機器を活用するためにも、また、専門家の教育や学習のためにも有用と思われる。AI の透明性や説明可能性への努力は、Explainable AI (XAI) と呼ばれる。医薬品の研究開発においても臨床診療においても専門家の判断は、多層であり、多段階であり、入れ子構造であり、要素は結合され複雑かつダイナミックなネットワークを構成している。

そのような複雑な推論過程を分析し、計算の視点から再構築するには、多くの専門分野にまたがる専門家と、幅広い情報計算の専門家との密接な協力作業が必要だ。そのような方向への努力の現状を知ることが必要である。

## 参考文献

- Eric J. Topol, High-performance medicine: the convergence of human and artificial intelligence, *Nature Medicine*, 25 : 44-56, 2019.
- E. H. Shortliffe, Artificial Intelligence in Medicine: Weighing the Accomplishments, Hype, and Promise, *IMIA Yearbook of Medical Informatics*, 2019.
- A. Nguyen et al. Deep Neural Networks are Easily Fooled: High Confidence

Predictions for Unrecognizable Images, CVPR2015. Open,

・ R. Guidotti et al. A Survey of Methods for Explaining Black Box Models, ACM Comput. Surv. 51, 5, Article 93 (August 2018):ブラックボックス人工知能への懸念.

・ A. Kurakin et al. ADVERSARIAL EXAMPLES IN THE PHYSICAL WORLD, Workshop track - ICLR 2017; 機械学習システムへの悪意のある攻撃による誤り発生の可能性.

### 課題 3. 製薬企業の「パイプライン」再考

もし、薬づくりに AI を活用しようとするれば、必然的に上で述べたような薬づくりやそれが使われる医療領域全体における、判断や意思決定に関わる過程を見直し、より効果的かつ効率的に仕事を進めていく仕組みをつくる必要がある。このことは、当然、製薬企業が今まで磨き上げてきたパイプラインと呼ばれる仕事の仕組みの見直しにつながる。これに関して神沼は、「創薬のひろば」の4号の「薬づくりの"パイプライン"は見直されうるか」で、最初の問題提起をしている\*。

昨年末、Translational Research (TR) を先導する NIH 傘下の研究機関 NCATS の活動に関する報告が2編発表された。それらは、“Drug Discovery, Development and Deployment Maps (4DM)”に関する NCATS の先導理念を表現した地図（経路網図）の試案に関するものだった。この地図は NCATS の責任者であるオースチン（C. P. Austin）が、製薬企業やアカデミア、学術研究 NGO などの専門家を集めて協議して作成した提言資料である。そうしたパイプラインの見直しを、AI やデータサイエンス、あるいは D2K サイエンスなど、ICT の活用を考慮した視点から、さらに見直しする必要があるように思われる。

#### 参考文献

・ K. Baxter et al. An End to the Myth: There Is No Drug Development Pipeline, Science Translational Medicine, 5(171), 2013.

\* 神沼二眞、薬づくりの“パイプライン”は見直されうるか？ 創薬のひろば、vol.4: 4-9, 2016.

・ J. A. Wagner, et al. A dynamic map for learning, communicating, navigating and improving therapeutic development, Nat. Rev. Drug Disc. 17(2):150, 2017.

・ J. A. Wagner, et al. Application of a dynamic map for learning, communicating, navigating and improving therapeutic development. Clin. Transl Sci. 11: 166-174, 2018.

・ NCATS の Drug Discovery, Development and Deployment Maps は、次のサイトにある。(https://ncats.nih.gov/translation/maps)

#### 課題 4. 人材（とくに D2K サイエンスの）育成

ゲノム解読やゲノム編集、それらに伴うオミックス、細胞のリプログラミングなど、基礎生物学の猛烈な進歩を臨床に生かす橋渡し研究（Translational Research）を重視する NIH は、医療の急激なデジタル化に対応するために、ICT 領域の専門家の職を運営の中核に設け、BD2K（ビッグデータ・トゥ・サイエンス）プロジェクトの実験を経て、NIH を「データを基盤にした組織 Digital Enterprise」に進化させる、という基本戦略を昨年（2018 年に）打ち出した\*。上で述べた薬づくりのパイプラインの見直しも、そうした戦略の下で議論がなされているように思われる。

・ NIH STRATEGIC PLAN FOR DATA SCIENCE、

[https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf)

データサイエンス、D2K サイエンスへの関心の高まりは、一国のあるいは社会の基盤としての科学や技術の様相が急激かつ劇的に変わってきていることが根源にある。NIH が組織を挙げてデータサイエンスに対応するという戦略を掲げたのも、そうした国の動きに、対応しているように思われる。問題は、そうした変化に適応するための人材の育成、教育である。昨年末、米国は「STEM 教育戦略」を発表した。STEM とは、Science, Technology, Engineering, Mathematics を意味する。その骨子は、それらの分野の進歩があまりに急激であるので、継続的な学習の機会を構築する努力が必要だ、という訴えである。

人材育成は、専門分野を越え、アカデミア、企業、国、NGO/NPO のような立場を越えて、基盤のかつ取り組む必要のある、時間が掛かる課題である。我が国でもこの課題に関しては、国が主導したいいわゆる有識者が集まった会合による、提言がまとめられている。また国民受けするような夢のプロジェクトの構想も打ち上げられている。

だが日本では、我々が関心をもっている生物医学、あるいは次世代ヘルスケアと薬づくりの研究開発やサービスに関わる、データサイエンスあるいは D2K サイエンスの人材育成に関する戦略的、個別専門的、かつ具体的な仕組みづくりに向けた話し合いや、そのための“コミュニティ”の存在はまだ見えにくい。

本来そうした活動は学会に期待したいところである。1960 年代末、それまで貴族と言われていた物理学の博士号をとった若手研究者を襲った就職難の時、米国の物理学会は、状況を把握すべく会長であった南部陽一郎らが大学などへの訪問調査をしていた。若くして米国の「医学の人工知能」のグループを立ち上げ、それを現在の AMIA（米国医療情報学会）に発展させたクリコフスキーやショートリフたちは、2012 年に学会として人材教育をどうしたらよいかに関する白書を出している\*。

我が国でも、我々の関心領域におけるデータサイエンスあるいはD2Kサイエンスの人材育成には、「ICTとヘルスケアや薬づくりに関わる」学会が関心をもち、学会間の関係を探ってもらいたいと考える。また、そこで若手世代の活躍の機会を増大するような環境づくりに、とくに関心を示していただきたいと考えている。

我々が提言するコミュニティにおいて、こうした大きな流れを分析しつつ、自分たちができることを相談してみることは、有意義ではないかと考えている。最も大切なことは、この変化が、より大きな前兆であり、我々が継続してそうした加速する変化に対応することが迫られているという現状認識である。私たちは、継続して学習する機会を至急つくらねばならないのではないかと考えている。

#### 参考文献

\* NSTC, Charting a course for success: America's strategy for STEM education, 2018. <https://www.whitehouse.gov/wp-content/uploads/2018/12/STEM-Education-Strategic-Plan-2018.pdf>

\* NIH STRATEGIC PLAN FOR DATA SCIENCE, [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf)

・ C. Mura, E. J. Draizen, and P. E. Bourne, Structural biology meets data science: Does anything change? Current Opinion in Structural Biology, October 2018.

\* NLM/NIH, A Platform for Biomedical Discovery and Data-Powered Health, Strategic Plan 2017–2027, [https://www.nlm.nih.gov/pubs/plan/lrp17/NLM\\_StrategicReport2017\\_2027.pdf](https://www.nlm.nih.gov/pubs/plan/lrp17/NLM_StrategicReport2017_2027.pdf)

\* C. A. Kulikowski, E. H Shortliffe et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline, J Am Med Inform Assoc (2012). doi:10.1136.

文責、神沼二眞（NPO 法人 サイバー絆研究所）

※コメントなどは、[mail@join-ica.org](mailto:mail@join-ica.org) へお願いします。