

スーパーコンピュータ

# 京×データサイエンス シンポジウム

2017年8月25日

[ データ meets シミュレーション ]

報告書

データとシミュレーションが出会う時——新たな価値が生まれる





## CONTENTS

シンポジウム概要、講演プログラム .....	2
主催者あいさつ .....	3
<b>基調講演:AI 機械学習の新展開 ~シミュレーション科学~</b> .....	4
上田 修功 理化学研究所 革新知能統合研究センター 副センター長	
<b>講演:医療・ヘルスケア 予測と予防の個別化医療実現に向けた データとシミュレーションの融合</b> .....	6
桜田 一洋 ソニーコンピュータサイエンス研究所 シニアリサーチャー	
<b>講演:社会経済 「未来観測所」から見える日本企業の現状と未来: 「京」をつかって</b> .....	8
高安 美佐子 東京工業大学 科学技術創成研究院 ビッグデータ数理科学研究ユニット ユニットリーダー・准教授	
<b>講演:データ同化 シミュレーションと実測データを融合する データサイエンス</b> .....	10
三好 建正 理化学研究所 計算科学研究機構 データ同化研究チーム チームリーダー	
<b>パネルディスカッション データ meets シミュレーション</b> .....	12
まとめ .....	18



# スーパーコンピュータ 京×データサイエンス シンポジウム

[ データ meets シミュレーション ] シンポジウム概要

日時: 2017年8月25日(金) 13時30分~17時50分  
 場所: 丸ビルホール(東京都千代田区丸の内2丁目4-1 丸の内ビルディング7階)  
 主催: 理化学研究所 計算科学研究機構、高度情報科学技術研究機構  
 後援: 文部科学省、日本経済団体連合会、経済同友会、日本商工会議所、HPCIコンソーシアム、  
 スーパーコンピューティング技術産業応用協議会、計算科学振興財団、  
 理化学研究所 革新知能統合研究センター

参加費: 無料  
 参加者数: 197名

## || 講演プログラム

時間	内容	講演者/パネリスト
13:30~13:35	開会挨拶	平尾 公彦(理化学研究所 計算科学研究機構 機構長)
13:35~13:40	挨拶	原 克彦(文部科学省 研究振興局 参事官)
13:40~14:20	基調講演:AI	機械学習の新展開 ~シミュレーション科学~ 上田 修功(理化学研究所 革新知能統合研究センター 副センター長)
14:20~14:50	講演:医療・ヘルスケア	予測と予防の個別化医療実現に向けたデータとシミュレーションの融合 桜田 一洋(ソニーコンピュータサイエンス研究所 シニアリサーチャー)
14:50~15:10	コーヒーブレイク(産業利用相談会)	
15:10~15:40	講演:社会経済	「未来観測所」から見える日本企業の現状と未来:「京」をつかって 高安 美佐子(東京工業大学 科学技術創成研究院 准教授)
15:40~16:10	講演:データ同化	シミュレーションと実測データを融合するデータサイエンス 三好 建正(理化学研究所 計算科学研究機構 データ同化研究チーム チームリーダー)
16:10~16:35	コーヒーブレイク(産業利用相談会)	
16:35~17:45	パネルディスカッション	データmeetsシミュレーション モデレータ:佐藤 三久(計算科学研究機構 プログラミング環境研究チーム チームリーダー) パネリスト:上田 修功(講演者:理化学研究所 革新知能統合研究センター) 桜田 一洋(講演者:ソニーコンピュータサイエンス研究所) 高安 美佐子(講演者:東京工業大学 科学技術創成研究院) 三好 建正(講演者:理化学研究所 計算科学研究機構) 青山 聖子(サイテックコミュニケーションズ サイエンス・プロデューサー) 松葉 浩也(理化学研究所 計算科学研究機構 利用高度化研究チーム チームリーダー)
17:45~17:50	閉会挨拶	平山 俊雄(高度情報科学技術研究機構 専務理事・神戸センター長)



主催者あいさつ

理化学研究所  
 計算科学研究機構  
 機構長

平尾 公彦

スーパーコンピュータ「京」はこの9月末で運用開始から5年になり、科学技術のさまざまな分野で素晴らしい成果が出ています。産業界のスパコン利用も大いに進みました。「京」の出現で、我が国の計算科学は一気に花が開いた感があります。

最先端のスパコンは国の競争力を左右します。我が国は2014年に「京」の後継機であるポスト「京」の開発プロジェクトを開始しました。理化学研究所が開発主体となり、世界最高水準の汎用性あるスパコンの、2020年頃の実現を目指しています。

ポスト「京」は最大で「京」の100倍のアプリケーションの実効性能を持ち、「精度」と「サイズ・時間」という従来の2つの軸に、「統計上の精度向上」という新たな軸が加わります。アンサンブルシミュレーションは多次元空間での最適化問題にも大変有効であり、エンジニアリングや設計分野でも力を発揮することが期待されます。

現在、高性能コンピューティングの世界には、従来のシミュレーションに加えてビッグデータ、機械学習、人工知能(AI)が新たな学問領域として出現しています。

そもそもシミュレーションとは、元となる法則あるいは方程式に初期条件と境界条件を与えれば、データなしに独自に計算が進んでいく演繹的推論です。

他方、データサイエンスはデータが与えられてから作業がスタートするものであり、統計科学に代表される帰納論的推論です。

データ解析やデータからのモデル抽出の先にも、スパコンの性能がフルに利用されます。シミュレーションの演繹的手法と、データサイエンスや機械学習の帰納的手法の融合が強く求められているのです。

いくつか取り組みも始まっています。その一つであるデータ同化は、観測、実験データとシミュレーションを結びつける数理科学です。すでに天気予報に飛躍的改善をもたらしており、今後は脳神経科学やタンパク質科学、ものづくり、社会科学などにも展開されることでしょう。

計算科学の方法は、これまでの予測、設計にとどまらず、新しい論理や概念の創出へも向かうべきです。スパコンという強力な道具を用いて、シミュレーションがデータサイエンス、機械学習等々と連携できれば、計算科学はさらに魅力ある学問へと進化することでしょう。

連携を進めるには、それぞれの学問的基盤をしっかりと確立して、お互いに自己主張をする中で相互理解を行うことが必要です。このシンポジウムがその一つのきっかけになれば幸いです。

データサイエンスとの連携により、  
 計算科学はより魅力ある学問へと進化する



基調講演: AI

機械学習の新展開  
～シミュレーション科学～

第5の科学「シミュレーション科学」  
とは

理化学研究所では現在、機械学習をシミュレーションに導入した「シミュレーション科学」を幅広い分野に応用する研究をしています。

近年、IoT時代の到来によりさまざまなものにセンサーが搭載され、ビッグデータの蓄積が加速しています。しかし、ビッグデータをどのように活用すればよいか分からないというのが現状です。

従来ビッグデータ解析には、統計学、人工知能、機械学習、データマイニングという4つの手法が導入されてきました。統計学は不確実性を扱う学問で、人工知能は演繹(論理)をベースに研究されてきました。機械学習は推測統計学と最適化による予測、データマイニングは列挙によるバスケット分析です。これまでこれら4つは棲み分けがされてきましたが、現在は全て人工知能に集約されています。

その中で、国際的に言われてきたのが、人工知能のコア技術は機械学習であるということです。しかし最近、コア技術が、機械学習から深層学習やニューラルネットワークに移ってきています。機械学習の良いところは、正確な物理モデルや第一原理を知らなくても、データさえあればなんとかなるので、幅広い分野に適用できるということです。そのため、サービス科学の土台技術として有望視されています。

次に人工知能と機械学習との関係についてお話しします。1950～80年に人工知能研究が開花し、80年代、ニューラルネットが大流行しました。90年代に入り、データマイニングが台頭し、2000年からは私の専門である統計的機械学習が出てきました。さらに2010年に入り、ビッグデータが注目されるようになり、機械学習の重要性が高まってきています。一方、ニューラルネットは1968年に今の土台となる技術が生まれ、近年のコンピュータの性能向上に伴い、2007年にヒントン教授らの論文によってブームが再燃、現在に至っています。

機械学習は、データが正確かつ大量に与えられた「教師

あり学習」では最強のツールとなります。画像、音声、自然言語処理では必須の要素技術です。しかし、学習データが容易に準備できない場合などには適用できません。そこで、膨大な学習データや稀有な学習データを収集するための手段として、シミュレーション+機械学習が有用であると考えています。これが「シミュレーション科学」です(図1)。科学研究の変遷は、実験科学、理論科学、計算科学と進み、現在、第4の科学としてデータ科学が注目されています。我々は、第5の科学としてシミュレーション科学を考えています。

シミュレーションに求められる  
リアルタイム性と先行性

次に、現在、我々が取り組んでいる人工知能に関するNICT(国立研究開発法人情報通信研究機構)受託研究プロジェクト「SODA(Social big data)」を紹介します(図2)。研究代表は慶應義塾大学の徳田英幸教授(現・NICT理事長)です。これは、さまざまなセンサーを使ってソーシャル・ビッグデータを収集し活用することで、スマートシティの実現を目指すというものです。センサーが収集したデータには、位置情報と時系列情報が含まれているので、「時空間データ」と呼ばれています。

ここで重要なのは、現在、何が起きているのかというリアルタイム情報です。これはシミュレーション科学においても重要です。例えば、東日本大震災が発生した夜、東京は帰宅難民で溢れかえりました。高度情報化社会にもかかわらず、我々はこのような事態を避けることができませんでした。原因は、個々ではなく、集団全体を最適かつリアルタイムに誘導できなかったからです。加えて、渋滞が生じる前に予測に基づいた先行的な誘導を行うことも不可欠です。

現在、内閣府のタスクフォースとして、NTTの機械学習・データ科学センターが企業と共同で取り組んでいるプロジェクトの一環に、国立競技場への入退場の人流誘導技術があります。我々が開発した時空間予測技術を使うことでどのゲートが混むかを予測し、機械学習技術を使って混まないルートを見つけ出し、リアルタイムに誘導しようとい

理化学研究所  
革新知能統合研究センター  
副センター長  
上田 修功



うものです。これは、シミュレーターのパラメーターを、取得したデータを基に学習しているという点で、「学習型マルチエージェントシミュレーション」と呼んでいます(図3)。このように、センサーから入ってきた情報をリアルタイムかつ多角的に分析し、実社会にフィードバックすることで、人々のQOLを向上させようというのが、私が目指すIoT社会です。これは、内閣府が掲げる「Society 5.0」にも合致するものです。

スーパーコンピュータに不可欠な  
シミュレーション科学

私が最後に申し上げたいことは、スーパーコンピュータを使って新たなブレイクスルーを起こすことを考えた時に、データを生成する領域やパラメーターを推定する領域にシミュレーション科学を導入することが今後重要になっていくのではないかと思います。

理化学研究所の革新知能統合研究センター(AIP)でも、防災など幅広い分野で、シミュレーション科学をベースとする研究を進めています。パラメーター学習を含めデータ同化をすることにより、高効率化、高精度化を図ろうとしています。

しかし、現在、このようなシミュレーション科学などを扱うデータサイエンティスト人材が不足しており、文部科学省などでも盛んに議論されているところです。

また、サービス産業の時代になり、高度な技術の確立よりも、まずはアイデアを行動に移すことの重要性が高まっています。そのためには、専門分野に閉じた研究開発だけでは不十分で、異分野間の連携・融合が大切です。その点においても、いろいろな領域をシミュレーションという1つの枠組みで連携可能なシミュレーション科学が、重要な役割を担うと考えています。

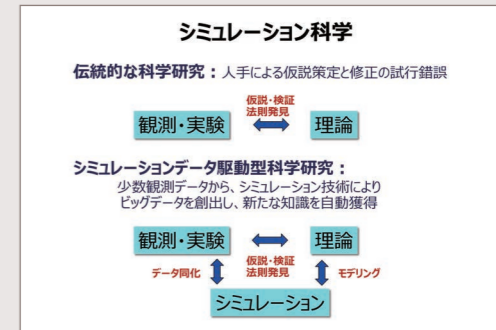


図1 「シミュレーション科学」  
シミュレーションと機械学習を組み合わせることで、膨大なデータや稀有なデータを収集し、そこから新たな知識を自動的に獲得することが可能となる。

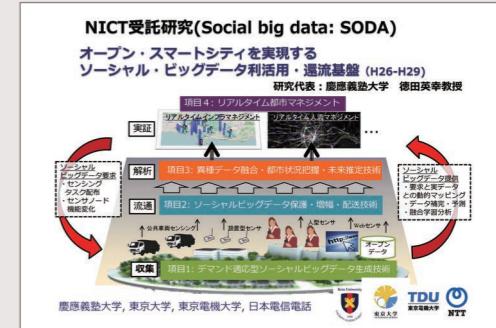


図2 現在取り組んでいる人工知能に関するNICT受託研究プロジェクト「SODA(Social big data)」  
さまざまなセンサーを使ってソーシャル・ビッグデータを収集し活用することで、スマートシティの実現を目指す。



図3 学習型マルチエージェントシミュレーション  
従来のマルチエージェントシミュレーションとは異なり、学習型マルチエージェントシミュレーションでは、シミュレーターのパラメーターを、取得したデータを基に学習することで、リアルタイム性と先行性を実現している。



講演:医療・ヘルスケア

# 予測と予防の個別化医療実現に向けたデータとシミュレーションの融合

ソニーコンピュータサイエンス研究所  
シニアリサーチャー  
桜田 一洋



## 高精度の予測に基づく個別の予防医療と治療の実現に挑戦

私の研究は、人を健康にするための技術やサービスの探究です。その一環として、データとシミュレーションの融合に取り組んでいます。しかし、これには困難が伴います。まず、生命医学分野では、データを体系化して推論する方法が確立していません。また、シミュレーションにおいても、病気を理解するための方法が未確立です。これらの問題の解決には、現在の知識体系だけではうまくいきません。そこで私がこの10年間にわたり取り組んできたのが、ビッグデータや人工知能を用いた推論、そして、それらとシミュレーションを融合できる新しい概念道具の開発です。

ソニー内で実際の医療データを使った研究は難しいため、2016年度からは理化学研究所を拠点に、「医科学イノベーション推進プログラム」の「高精度の予測に基づく予防医療の実現に向けた疾患ビッグデータ主導型イノベーション」を開始しました。高精度の予測に基づく一人ひとりに合った予防医療と治療の実現に取り組んでいます。今日はそのコンセプトを中心に紹介します。

今や世界の死亡原因の60%が生活習慣病です。一方で、遺伝性の難病や感染症もなくなっていません。病気の難しいところは、個別対応が必要な点と、病気になるまでの過程を追っていく必要がある点です。

そもそも健康・病気には個人差があります。個別の問題を、信頼性を担保しながら解いていくのは難しい問題です。さらに、健康・医療は当事者(内部観測者)視点の推論が必要です。そのためには、散逸力学と呼ばれる力学を生命医学分野でも使えるようにしなければなりません。それには、機械学習や次元圧縮を組み合わせた適切なシミュレーションが不可欠です。

20世紀初め、生命科学は生物をメカニズムとしてとらえ、説明しようとしてきました。しかし、生物は高次の非線形のシステムであり、病気はさまざまな要因が相互作用して発症しているので、局所的に治療すると、逆に全体がおかしくなったりします。多様性の問題もあり、標準的な治

療では治る人と治らない人がいる上、健康に対する価値観も人によって異なります。

生命現象の推論を実現するには生命の持つ知覚・応答特性を反映する必要があります。認知の場合それは二種類の認知に整理できます。言語的な認知と非言語的な認知です。言語的な認知に関する情報は、効率性や合理性を重視します。一方、非言語的な認知に関する情報には不確実性という課題があります。現在、人工知能では、「知覚があり、情報を計算する連合野があり、行為がある」というモデルを使っています。しかし近年、ミラー・ニューロンなど連合野を介さず行為が行われるしくみが見つかりました。そのため、人間の持つ身体的認知の特性を考慮した推論が求められます。

さて今、自動運転やソーシャルロボットなどデータと人工知能を主体とするプラットフォームを構築する時代に突入しています。データや人工知能、シミュレーションが現実世界のサービスの中に入ってきており、医療分野では、ウェアラブルセンサーが一例です。医者も患者さんのデータ量が増えていけば、負担が重くなります。

医者の経験や知識を人工知能やシミュレーションに反映させることができれば、個別の患者さんに対して最適な解を提供できるようになります。

## 生命医学分野における異なる3つの枠組みの統合

では、収集したデータからどのように推論すればよいか。最後に知識体系の構築方法について紹介します。

今、生命医学分野では、異なる3つの枠組みが存在します(図1)。1つ目は還元主義モデル(入出力モデル)、2つ目は非線形システムモデル(非平衡開放系)、3つ目は医学モデルです。還元主義モデルとはネオダーウィニズムをベースとした知識体系で、機械論、線形因果を基に作られたものです。一方、非線形システムモデルとは複雑系科学をベースに生命現象を捉えようというもので、機械論との融合がうまくいっていません。医学モデルとは、病気は遺伝子が原因で起こるという考え方に基づくもの

です。しかし、生活習慣病などはこれに該当せず、新たなモデルが必要です。私はこれら3つの枠組みを同化し、統合しなければならないと考えています。

機械論では、階層を分けて、マイクロからマクロに向かって階層をつなげていくことで、マクロな現象である疾患をとらえようとします。それに対し、非線形モデルでは、動力学モデルを用いたシミュレーションの導入が行われています。動力学は、現象をマクロで表現しようとしています。これには、機械学習を使ったアルゴリズムとシミュレーションが不可欠です。なぜなら、機械学習により病気の発症予測ができるようになったとしても、それだけでは価値が創出できないからです。発症の予防などの介入法を発見す

ることが求められます。その解決にも、シミュレーションが大きな威力を発揮します。

最後にまとめとして、イノベーションのスキームを紹介します(図2)。現状の生命医学の知識体系である線形モデルから、ビッグデータ、機械学習、情報幾何学を取り込み、新たな自然科学の枠組みである非線形科学を構築することで、人工知能とシミュレーションがつながり、ソリューションが出来上がっていくという流れです。

今後、企業の皆様には、「ソリューションを使ってさらにデータを収集し、性能を高める」というバリューチェーンを回していくくみを、我々のような基礎研究者とともに構築していただきたいと思います。

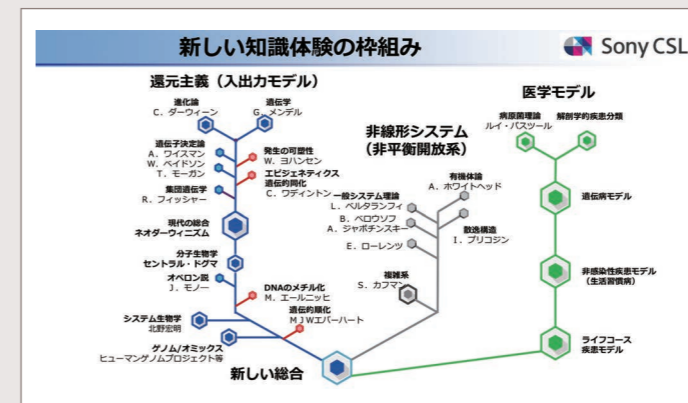


図1 生命医学分野における異なる3つの枠組み  
生命医学分野には現在、還元主義モデル(入出力モデル)、非線形システムモデル(非平衡開放系)、医学モデルという異なる3つの枠組みが存在する。今後、これら3つの枠組みを同化し、統合していくことが大きな課題。

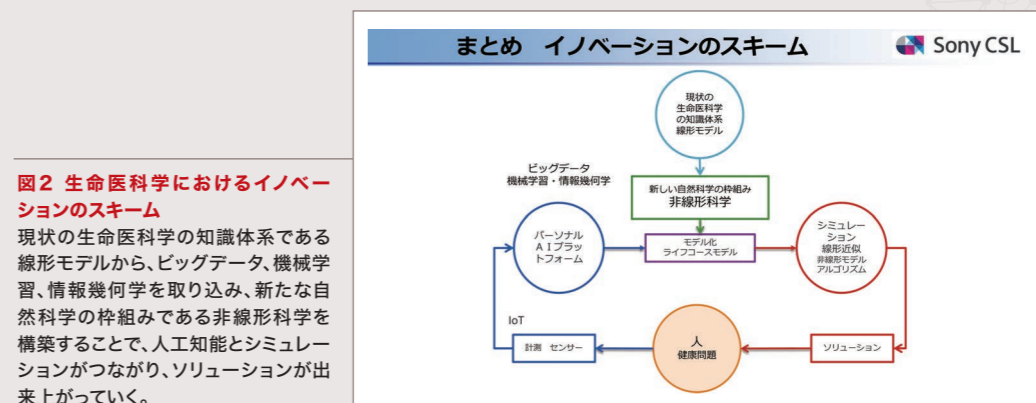


図2 生命医学におけるイノベーションのスキーム  
現状の生命医学の知識体系である線形モデルから、ビッグデータ、機械学習、情報幾何学を取り込み、新たな自然科学の枠組みである非線形科学を構築することで、人工知能とシミュレーションがつながり、ソリューションが出来上がっていく。



講演:社会経済

# 「未来観測所」から見える日本企業の現状と未来: 「京」をつかって

## 超スマート社会の実現を目指す 未来観測所

2017年度に東京工業大学の科学技術創成研究院に設置したビッグデータ数理科学研究ユニットの活動内容について紹介します。本ユニットには多角的に人間社会や経済現象を理解するための研究拠点である「未来観測所」を併設しており、超スマート社会に向けてスーパーコンピュータやシミュレーター、ビッグデータを使った最先端研究と実証実験を行っています。

私が取り組んでいる経済物理学では、人間社会や経済現象に関するビッグデータを解析することで、普遍的な法則性の発見を目指しています。解析方法としては、機械学習が有効で、複数のパラメーター同士の因果関係の発見が得意です。

## 企業ビッグデータを使って 各種リスクを予測

では、現在、未来観測所が帝国データバンクと共同で2014年から進めている研究成果を紹介します。

帝国データバンクは日本中の企業、約100万社の30年間分のデータを保有しており、現在、東工大ではそのデータを解析しています。データには企業の財務諸表、取引関係、詳細な倒産情報などが含まれており、このようなデータを使いながら日本企業を数理モデル化していくことで、各種リスクを予測することができます。

日本は自然災害が多い国ですが、2007年7月16日に発生した中越沖地震は局所的であったにもかかわらず、日本中の自動車工場の生産ラインがストップし、経済的な被災規模が大きかったことで知られています。理由は、その地域のリケンという会社が被災したことで、高いシェアを占めていたある自動車部品が作れなくなったからです。

これを機に、製品のサプライチェーンや資金の流れを把握することの重要性が見直されました。帝国データバンクのデータからは、それがある程度、推測できます。例えば、企業Aと企業Bとの関係性が分かれば、資金や製

品、サービスが流れる方向が分かるというわけです。

そこで、推定モデルを開発し、そのモデルを使ってデータを解析した結果、相手企業の規模に応じて取引額が決まっていることなどが分かりました。

このモデルを使えば、企業間だけでなく、日本中の地域間のお金の流れなどを調べることができます。その例が、内閣府の「RESAS(地域経済分析システム)」というプラットフォームです。ここでは、我々が開発したモデルを使って、どの地域からどの地域へどのくらいお金が流れているかを計算しています。

さて、取引関係が分かれば、それに基づき、方程式を作ることができます。我々は取引量と売上げを推定する方程式を作りました。この方程式を使ったモデルのパラメーターをデータ同化することで、推定を試みた結果、売上げの分布が出てきました。それにより、取引関係のネットワーク構造だけから、高精度で売上げと取引額が推定できることが分かりました(図1)。これは、どの企業とどのような取引をしているかだけで、その企業の売上げが推定できることを意味しています。

これにより、大企業の倒産によってお金の流れがどう変化するか、災害によって特定の地域が機能しなくなった場合、経済的な被災規模はどれくらいかなどをシミュレーションできるようになりました。現在は、災害時の企業の売上げの被災額を推定するモデルとして利活用されています。

実際、2016年4月16日に熊本県で起きた大地震に対して、我々が推定した売上げ被災額について紹介します。まず、震度6以上だった地域にある企業および、避難勧告が出ている地域の企業のノードをネットワークから取り除きました。その上で方程式を解き、売上げの変化を推定した結果、1カ月そのままの状態が続けば、約0.7兆円の被災額となることが推定されました。計算は、実際のデータを入手してからたった2時間で完了しました。非常に短時間で被災状況を把握できることが分かります。のちに、この推定結果が実態とほぼ合致していることが確認されました。

東京工業大学 科学技術創成研究院  
ビッグデータ数理科学研究ユニット  
ユニットリーダー・准教授

高安 美佐子



## 「京」を使い 2050年の日本を予測

さて、ネットワーク構造から推定される現在の日本の企業の安定性についてですが、シミュレーションのパラメーター数は100万以上あり、非線形の方程式になっているため、これを解くのは容易なことではありません。そのため、スーパーコンピュータを使って計算しました。その結果分かったことは、現在のところは、さまざまなサイズの企業が共存・共栄する安定領域にありますが、ある臨界点に達したら一気に不安定化し、大企業の独占状態に陥るといことです。これにより、今後、日本で企

業合併が進み、より大きな企業と合併したいという意識が働けば、メガファームが誕生し、さらにそのメガファームが倒産したとすれば、景気が一気に後退するということが推定されました。

このように、未来観測所では、人間社会や経済現象に関するビッグデータ解析により未来を予測しています。「京」を使った計算では2050年ぐらいまでをシミュレーションしました。それにより、各地域の未来の経済状況が推定されました。今後、推定結果を基に政策提案などを行っていくことも考えています(図2)。このような我々の研究が、よりよい未来の構築に寄与できることを願っています。

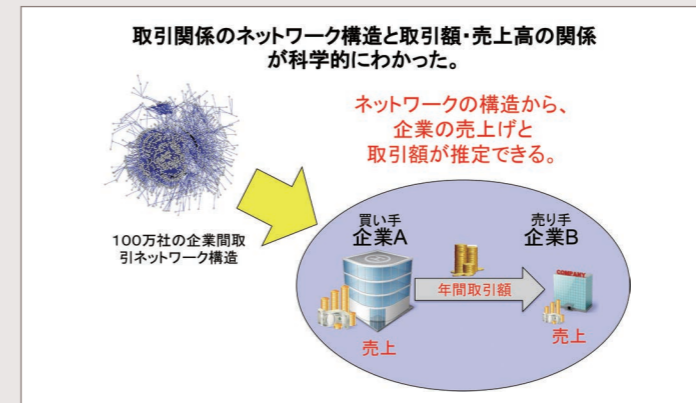
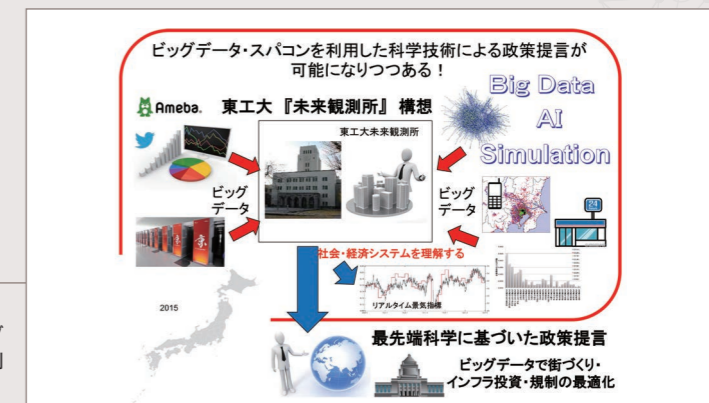


図1 企業同士の取引関係のネットワーク構造だけから、高精度で売上げと取引額が推定できることが明らかに。

図2 人間社会や経済現象に関するビッグデータ解析により推定した未来予測を基に、政策提案も可能に。





講演: データ同化

# シミュレーションと実測データを融合する データサイエンス

## シミュレーションと現実世界を 結びつけるデータ同化

私の研究テーマはデータ同化です。きっかけは大学卒業後に気象庁に入り、そこで、データ同化を担当したことです。大学時代にはカオス力学を学んでいましたが、気象庁時代に留学したメリーランド大学では、気象学で博士号を取りましたので、気象学者でもあります。5年前に理研に移り、現在は幅広い分野のデータ同化に取り組んでいます。

そもそもデータ同化とは、シミュレーションと観測・実験データ(現実世界)を結びつけることです。これにより、新たな価値や相乗効果を生み出そうというのが、データ同化の目的です(図1)。

データ同化のワークフローを説明します。まず、シミュレーションを走らせるため、初期状態と境界条件を与えます。そして、シミュレーションを実行します。その結果に観測データを同化すると、最適な初期状態が算出されて出てきます。

しかし、これを実行する際には、観測データとシミュレーションのデータを比較するというプロセスが必要となります。観測データとシミュレーションのデータの差を求めるには、単位を同じにしなければなりません。これが簡単そうに見えて、必ずしも簡単ではありません。

例えば、気象衛星ひまわりが観測している赤外線放射輝度はケルビンという単位です。一方、気象をシミュレーションする場合、大気を格子状のメッシュに区切り、各格子点に温度、湿度、雲の量などを与えます。ひまわりが観測する放射輝度をシミュレーションでは直接扱っていないのです。そのため、両者を関連づける必要があります。これは、気象に限ったことではありません。それがデータ同化の難しい点です。しかし、それにより、さまざまなデータを統合することが可能となります。

データ同化は、逆問題の解法としても有用です。逆問題とは、結果から原因を探る数理的アプローチで、結果からそこに至るプロセスを推定しようというものです。

例えば、大気中の二酸化炭素量を観測し、そこから、地面や海からどれくらい排出されたり吸収されたりしているかを推定するといった問題があります。このような時に、データ同化を使うとうまく推定することができます。

## データ同化で ゲリラ豪雨を30分前に予測

次に、「京」を使ったデータ同化の最先端研究を紹介します。2016年8月に、ゲリラ豪雨を30分前に予測できる新技術を発表し、国内外の数多くのメディアで取り上げていただきました。

ゲリラ豪雨は、わずか10分間程度で急変化します。現在、気象庁ではシミュレーションモデルを使って天気を予報していますが、新たなデータを使った更新頻度は1時間に1回です。したがって、10分間で急変化するゲリラ豪雨を予測することはできません。そこで、我々は「京」上で、データ同化を使ってゲリラ豪雨を予測しようと考えました。

「京」による全球シミュレーションは870メートルメッシュで、世界最高の解像度ですが、ゲリラ豪雨は数キロメートルの範囲で起こるため、100メートルメッシュの領域シミュレーションを設定しました。

また現在、ゲリラ豪雨はPhased Array Weather Radar(フェイズド・アレイ・ウェザー・レーダー)という気象レーダーを使って観測しています。これは、発射した電波が跳ね返ってくる割合を観測するもので、単位はデシベルです。一方、シミュレーションでは、例えば、1立方メートルあたりに雲が何グラムあるかといったことをベースに計算しています。そのため、これらをデータ同化させなければなりません。

さらに、気象レーダーからは30秒毎に観測データが送られてくるため、30秒以内にデータ同化を行う必要がありました。大量のデータを高速に処理する必要があったのです。現在では、このようなビッグデータの同化を「ビッグデータ同化」と呼んでいます。

完成させたシステムを「京」を使って走らせたところ、実際に2014年9月11日に神戸で発生したゲリラ豪雨を高

理化学研究所 計算科学研究機構  
データ同化研究チーム  
チームリーダー  
三好 建正



精度で再現できました(図2)。

この成功を機に、今後、気象予測に関しては、これまで活かされてこなかったビッグデータを積極的に活かしていけるだろうと考えています。例えば、人工衛星の観測データなどが考えられます。新世代の気象衛星ひまわり8号は従来のひまわり7号の約50倍の観測データ量です。

さて、最近、集中豪雨による川の氾濫が各地で頻繁に起こっています。防災情報というのはリアルタイム性が非常に重要です。高い頻度で観測データを取り込み、高い

頻度で予報を更新できるシステムへのニーズが高まっています。ビッグデータ同化は、そうした社会のニーズに応えることができると考えています。

今後、ビッグデータ同化は気象に限らず、幅広い分野に広がっていくことが予想されます。そのため、現在、理研では、「データ同化をハブとした数理科学、実験・観測科学、シミュレーション科学の融合研究イノベーションハブの形成」に取り組んでいます。今後もデータ同化の研究を通してよりよい社会の実現に努めていきます。

### データ同化



データ同化は、シミュレーションと現実世界を結びつけ、相乗効果を生み出す。

双方の情報を最大限に抽出

図1  
データ同化とは、シミュレーションと観測・実験データ(現実世界)を結びつけること。これにより、新たな価値や相乗効果を生み出そうというのが目的である。

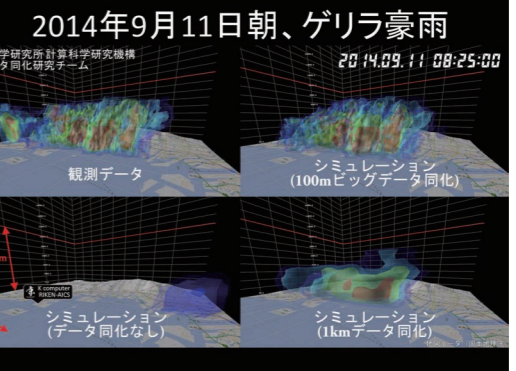


図2  
「京」を使って、2014年9月11日に神戸で発生したゲリラ豪雨をデータ同化により高精度で再現。データ同化した場合としない場合とでは、再現性に大きな差が出ていることが分かる。



## パネルディスカッション データmeets シミュレーション

### モデレータ

理化学研究所 計算科学研究機構  
プログラミング環境研究チーム  
チームリーダー  
佐藤 三久



### パネリスト

理化学研究所  
革新知能統合研究センター  
副センター長  
上田 修功

ソニーコンピュータ  
サイエンス研究所  
シニアリサーチャー  
桜田 一洋

東京工業大学 科学技術創成研究院  
ビッグデータ数理科学研究ユニット  
ユニットリーダー・准教授  
高安 美佐子

理化学研究所 計算科学研究機構  
データ同化研究チーム  
チームリーダー  
三好 建正

サイテック・  
コミュニケーションズ  
サイエンス・プロデューサー  
青山 聖子

理化学研究所 計算科学研究機構  
利用高度化研究チーム  
チームリーダー  
松葉 浩也



### データサイエンスの重要性と 社会還元

**佐藤** 本日のパネルディスカッションのテーマは「データmeets シミュレーション」です。モデレータを務めます私は、理化学研究所 計算科学研究機構 プログラミング環境研究チームのチームリーダーの佐藤三久です。これまでいろいろな計算機の開発をしてきました。現在、ポスト「京」の開発ではアーキテクチャの開発チームリーダーを務めています。

ではまず、私同様パネルディスカッションからご参加いただきますサイテック・コミュニケーションズの青山聖子様と理化学研究所 計算科学研究機構 利用高度化研究チームの松葉浩也チームリーダーに自己紹介をお願いします。

**青山** 私は科学技術と人を結ぶ仕事をしております。研究機関などの広報媒体の作成を手掛けており、計算科学研究機構の「京」やポスト「京」を紹介するお手伝いをしています。本日は一般人代表として登壇しました。

**松葉** 利用高度化研究チームでは、多くの方々にスパコンをより簡単に使っていただくための研究をしています。アプリケーション側ではなく、計算機側の研究者の立場でお話しさせていただきます。

**佐藤** さて、平尾公彦機構長の話にもあった通り、シミュレーションは極めて演繹的な手法です。特に「京」の開発で大きなテーマとなったのが、このような大規模計算で何ができるのかということです。その背景として、「パラメーターと基礎方程式さえ与えられれば、自然は解明できる」という演繹的

な考え方があります。

一方、ポスト「京」では、計算機システムの中でのアンサンブルサーチによる大規模なアプリケーションも1つのターゲットになっています。これは、データサイエンスにも通ずる帰納的な考え方で、今日のテーマとも合致しています。

本日は、なぜ、近年データサイエンスに注目が集まっているのか。それに対するシミュレーションの役割とは何か。そして、データサイエンスとシミュレーションの融合による成果をいかにして社会に還元するか。これらについて、講演を行っていただきました。今回のシンポジウムを振り返って、改めて、それぞれのお考えを自由にお聞かせください。

**上田** これまでいろいろな領域で第一原理が研究されてきました。これを使ってシミュレーションし、現象やしくみを再現することの重要性が高まっています。しかし、逆問題を解くことは困難です。それに対し、機械学習や人工知能の登場により、逆問題が解けるようになりました。これが、データサイエンスが注目されている1つの理由だと思います。今後、シミュレーションとデータサイエンスの2つが両輪となって進展することで、相乗効果が発揮され、イノベーションが起こるのではないかと期待しています。今はその転換期にあると思います。

そのため、昨今、スパコンと人工知能との距離が縮まってきていると思います。ポスト「京」の開発に向けては、スパコンと人工知能の連携強化を図っているところです。

**桜田** 私はまず、生命医科学の面で直面している課題についてお話しします。それは、人に関する各種データの収集が難しいことです。個人情報保護などの観点からセキュアなデー

タマイニングのしくみ作りが不可欠です。一方で、健康・医療は個性が高いため、個人個人のデータが収集できなければ最適なソリューションは提供できません。そのために、どのようなシステムを構築していくかが大きな課題となっています。また、健康・医療の場合、気象のように再現性の高いデータを得ることができないため、生命現象をどのように表現すべきかについて、研究者間で考えを共有する必要もあります。

データの収集に関しては、患者さんや市民が費用を負担しながら、自分のバイタルデータを計測、蓄積していくシステムがなければ、ブレイクスルーは起こりにくいと思っています。本人が問題解決につなげられる形で、共通データベースにデータを上げていき、それを高いセキュリティシステムの中で解析し、問題を解決していくというエコシステムを作っていく必要があるでしょう。データを提供するメリットを我々が実証できれば、社会に受け入れられると思います。

生命現象の原理的なことを理解するには、非常に多くの変数を使って、生命現象を再構築しなければなりません。その場としてスパコンによるシミュレーションが不可欠です。しかし、最終目標は社会実装ですから、そのためには、一般的なコンピュータでも利用できるなど、よりシンプルな形に落とし込んでいくことが重要です。データ主導型によるイノベーションを期待しています。

**高安** 私の現在の問題意識として、例外的な事象というのは全く予測できないものなのか、科学者は経験したことがないものは予測できないと言い切ってよいのかということがあります。現象に関するメカニズムをきちんと理解できれ

ば、予測は可能になるのではないかと。その点で、今後、ビッグデータとスパコンを使ったシミュレーションにより未来を予測し、国の制度設計や防災に役立てていくことが重要になっていきます。

しかしながら、現在、私が扱っている企業のデータに関しては、我々の研究グループしか扱うことができないものばかりです。今後、科学を深めていくには、皆で共有財産としてのデータを利用していくことが重要で、そのためには、より多くの企業のデータが必要となります。企業の皆様には、自社のデータを社会の共有財産として活かすという視点を持ってほしいと願っています。

**三好** 私は「京」とデータサイエンスという観点で意見を述べさせていただきます。私は「京」でなければできないことを研究していますが、これは同時に、桜田さんも指摘されたように、現時点では普及が難しいことを意味しています。今日紹介したゲリラ豪雨の予測手法も、そのまま実用化することはできません。とはいえ、これから10年すれば普及機(通常のスパコン)のスペックは、今の「京」並みになると思います。ですので、社会への還元という点では、10年後に普及機でできることを我々が今、試していると考えています。ここで得られた知識や技術は必ず将来活かされるでしょう。

社会還元に関しては、今後、人材育成に取り組んでいくことも重要だと考えています。今や気象については知識が浅くても、用意されているシミュレーションソフトを動かせば、それなりのアウトプットが出せる世の中です。データ同化の研究も同様です。しかし、このままでは中身をきちんと理解して



II パネルディスカッション

いない「専門家」が増えていくばかりです。今後は、中身をきちんと理解しているデータサイエンティストをしっかりと育成していく必要があります。そういった人材の育成も我々の責務であり、社会還元の一つだと思っています。

**佐藤** 青山さんと松葉さんは、講演者の方々の講演内容や今の発言を聞いてどのように思われましたでしょうか。



サイテック・コミュニケーションズ  
サイエンス・プロデューサー  
青山 聖子

**青山** 私は一般人代表として、素朴な疑問をいくつか投げかけたいと思います。まず、ビッグデータ解析についてですが、そもそもデータの信頼性は担保されているのでしょうか。私にとって一番分かりやすいビッグデータは国勢調査です。あれは、自分も参加しているのどのような種類のデータか、どういうバイアスを含む可能性があるか分かります。しかし、インターネットなどを通じて自動的に集められたデータに関しては、仮にどこかでバイアスがかかっていたとしても、そのことが外からは分からないのではないかと不安に感じてしまいます。

加えて、今、皆さん、「インスタ映え」などと言って写真うつりの良いものを食べたり着たりして、それをSNSにアップしていますよね。私などは、皆がSNSにアップするために消費行動をしているのではないかと疑ってしまいます。同様に、ビッグデータを活かすことに集中しすぎると、データを取るための社会になってしまうのではないかと心配になります。どこにでもセンサーが設置されていて、勝手にデータが集められ、使われる世の中というのは、決して居心地の良いものではないのではないかと思われますが、いかがでしょう。また、それに伴い、データストレージへの投資などがどんどん増えていくのではないかと懸念しています。

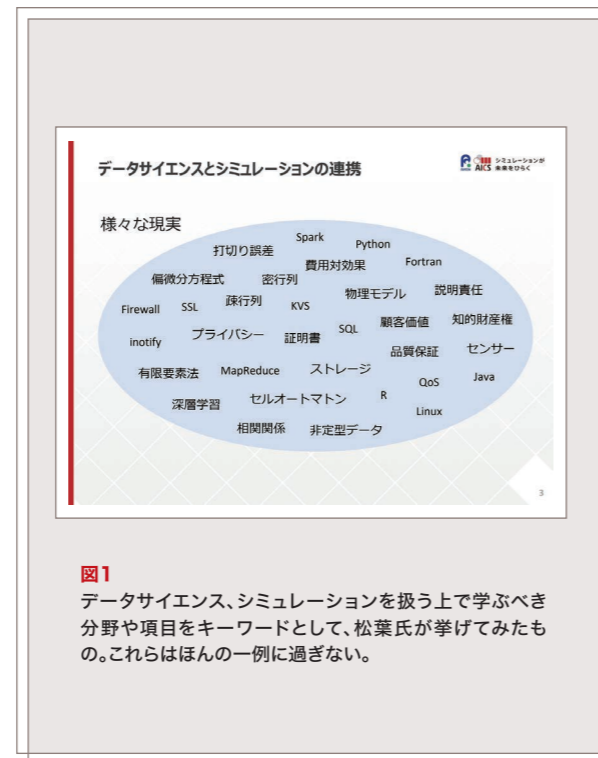
次に、機械学習や人工知能についてですが、上田先生のお話にもあったように、今の機械学習はデータを入れると勝手に自分で学習するといいます。ブラックボックス化しているわけですが、出てきた結果は本当に正しいのでしょうか。シミュ

レーションによってデータを生成して、それを機械学習すると、出てくるのは、元々のシミュレーションで作ったモデルではないのでは、という素朴な疑問もあります。

**松葉** 私は、データサイエンスとシミュレーションの連携について意見を述べさせていただきます。

連携にはいろいろな可能性があり、絶対に進めていかなければならないと、今回の講演を聞いて改めて強く感じました。

とはいえ、データサイエンスだけでも、シミュレーションだけでも習得するのは大変難しく、1人で全部を扱えるようになるのは並大抵のことではありません。深層学習やセルオートマトン、偏微分方程式、有限要素法、打ち切り誤差など思いつくままにちょっとキーワードを挙げただけでも、数十はあります(図1)。



**図1** データサイエンス、シミュレーションを扱う上で学ぶべき分野や項目をキーワードとして、松葉氏が挙げてみたもの。これらはほんの一例に過ぎない。

したがって、広く実社会で応用されるためには、シミュレーションやデータ解析が簡単にいける環境を整えていくことが必須です。具体的には、まず、分かりやすいガイドラインやマニュアルを整備する必要があります。次に、これはコンピュータサイエンス的な仕事になりますが、新たなコンピュータ

語の開発なども含めて、使いやすい最適なツールの開発、整備が不可欠です。使いやすいサービスの立ち上げ、事例の共有も大切です。それに対し、我々のところでは、シミュレーションやデータ同化、データ解析が簡単にできるツールの開発と提供を実施していきたいと思っています。サポート体制も整え、ワンストップサービスの提供を目指します。我々のこういった取り組みに関しては、企業の皆様も含め、各方面からのご意見や情報提供が不可欠です。この場をお借りして広くお願いしたいと思います。

機械学習をうまく組み合わせることがシミュレーション科学のカギ

**佐藤** ここからは、会場の方も含めて議論したいと思います。3つくらい課題があるかと思います。1点目は、データを収集し使うための基盤についてです。これは、桜田先生がおっしゃったように、生命医学の分野ではまだ標準化もされていない上、セキュリティの問題もあります。2点目は、社会実装についてです。上田先生や桜田先生、三好先生のお話にもあったように、個人個人のサービスに展開するためには、より簡単化しなければなりません。3点目は、人材育成についてです。シミュレーションだけでなく、データサイエンスにも精通した人材の育成が急務です。では、青山さんのご指摘なども含めて、会場からのご意見をお願いします。

**来場者** 一般市民として2点、質問があります。1点目ですが、「京」のアーキテクチャは1970年代からあまり変わっていないように思います。その「京」を使って人工知能をどうやって実現させるのでしょうか。2点目ですが、先程の青山様の意見に非常に共感しています。そこで、上田先生に伺いたいのですが、ビッグデータを処理する汎用データ解析技術を使えば、解析する現象に対する知識などなくても、数理モデルができてしまうということに非常に不安を感じます。数理モデルの検証はどのようにして行われているのでしょうか。

**佐藤** 1つ目の質問ですが、人工知能はプログラムですので、どんなコンピュータでも原理的には動きます。「京」でも、汎用的に使われている人工知能のパッケージソフトが動いています。「京」にはたくさんCPUが計算リソースとしてありますので、一個一個は速くなくても、それなりの数を使えば、速く計算できます。ポスト「京」では、人工知能に使われるような

演算を高速に処理するアーキテクチャの構築にまさに今、取り組んでいるところです。

**上田** 2つ目のご質問についてです。確かに今のニューラルネットワークは、原理を知らなくても使うことができます。機械学習が進化すればするほど不安になるとおっしゃっているのだと思いますが、だからこそ、シミュレーション科学が重要です。つまり、シミュレーターは、その分野の専門家がきちんとした理論に基づき構築しています。ただし、モデルにはパラメーターがあるので、かなりの自由度があります。観測データからパラメーターを推定する必要があります。

一方、機械学習技術は、説明機構を無視しても入出力データさえあれば実行できます。だからこそ、逆問題を解けるのです。つまり、観測データが得られている時に、どのようなパラメーターが正しいのかについては、説明は不要なのです。機械学習をうまく組み合わせることが、シミュレーション科学のカギを握ります。

ですので、例えば、地震に関しては、地震のモデルがあります。そのモデルパラメーター同士は相互作用を持つので、その組み合わせは百万通りにもなります。それを試行錯誤でやっているのが現状です。しかし、シミュレーターでデータを作れば、逆問題を解くことができますので、機械学習をうまく使えるということです。

データの扱い方とデータ同化を学部生のうちから教育することが重要

**来場者** データサイエンスという点では、人材育成が避けられません。その観点からは、今、現場にいる人間は、今後、何を学んでいく必要があるのか聞かせてください。また、大学1年生や2年生といった新世代に向けては、こういったカリキュラムの提供が望まれるでしょうか。

**三好** 今、私は京都大学理学部で教えているのですが、学部生の段階からデータ同化を教えることは、非常に効果が高いと考えています。理学部には、物理学や化学、生物学、地球科学などを勉強している学生がいて、どの分野にも基本的に実験データやシミュレーションデータがあります。ただ、データの扱い方に関する基礎知識がすっぽりと抜け落ちているように感じられます。まずはこの部分をしっかりと教育する必要があります。特に今、データのソースが多岐にわたっ



## II パネルディスカッション

てきています。先程、青山さんからもお話がありましたが、データの信頼性はまさに誤差の問題です。したがって、データの信頼指標の作成もデータ同化をやる上では必須となります。実験データにどのような誤差があるのかを知って初めてデータ同化ができるということを、データ同化の教育で伝える必要があると感じています。

**高安** 私は、データサイエンティストを育成するためには、基礎科学として、統計科学は絶対に学ばせなければいけないと思っています。なぜなら、解析ツールは使い方が限定されているものが多いからです。しかしそれを知らずに、あらゆるデータを入れてしまっている学生が多いように思います。解析方法の根源を理解している人材に育ててから、解析ツールを使わせるべきだと思います。

**来場者** 今後、社会実装を考えた場合、システム全体のスピードについては、どなたかが考えているのでしょうか。例えば、将来、人の命にかかわる深刻な状況の中で、医療データを使うといった場合、たとえデータ解析が高速でも、データ転送にすごく時間がかかっていたら、意味がありません。その点に関する研究はされているのでしょうか。

**桜田** 極めて重要なポイントです。例えば、今、私はウェアラブルセンサーをつけています。Bluetoothでデータ伝送していますが、Bluetoothで送れる情報は限られます。ですから、例えば、「心拍の細かい変動のデータまではとても送れない」といったことが起こります。そのためには、大きな情報量のデータを送ることができる機器をセンサーに格納しなければなりません。しかし、費用の問題もありますので、そのあたりをしっかりと検討する必要があります。



理化学研究所  
計算科学研究機構  
利用高度化研究チーム  
チームリーダー  
松葉 浩也

**松葉** つまらない答えで申し訳ありませんが、現在のスパコンは全てベストエフォートになっています。リアルタイム性の確保は、個々の研究レベルで取り組まれてはいますが、それにはやはりリソースの確保が重要で、コストとのトレードオ

フになるのが現状です。

**三好** ゲリラ豪雨の予測の場合、30秒毎に予報を出しますので、データ送信に何分もかかっていたら話になりません。その点に関してプロジェクト内で取り組んでいます。「それぞれプロジェクト毎にやっている」というのがご質問の答えになるかと思います。

**佐藤** 社会実装という観点からは、今、三好さんが神戸限定でゲリラ豪雨の予測をしていますが、今後、実用化という面では、シミュレーション結果の信頼性に対して、社会的な理解が必要になってくるかと思っています。それを踏まえた上でインフラ投資ということになるかと思っています。

**来場者** 2点質問です。1点目ですが、高安先生の社会経済を予測する研究に興味深く伺いました。企業が自社のデータをなかなか出してくれないという話もありますが、「取引データを全て出してくれば、御社の1年後、3年後、5年後を予測します」ということになれば、企業にとっては大きなメリットになるかと思っています。現在は、人間が経営判断をしていますが、今後、コンピュータが経営判断を行う、もしくは、経営コンサルタントの役割を担えるようになるのでしょうか。

2点目ですが、株式市場でもコンピュータが1秒間に千回単位で受発注を繰り返していますが、人工知能が発展していけば、例えば取締役会でも人工知能の判断を反映させるといった話になるかと思っています。社会全体が人工知能によって動くことになり、その動きを人工知能が解析して未来を予測する、ということになっていくのでしょうか。

**高安** まず、経営コンサルタントのお話ですが、それはすでに展開しております。そもそも帝国データバンク自身がコンサルティングに使っています。データ自体、帝国データバンクのものなのですが、我々の成果を使ってビジネスをしているのも帝国データバンクなのです。ですので、今のところ、私の周りでは、我々の研究を信頼し貢献してくれている企業に貢献しているという関係になっています。今後は、「未来を推定するので、代わりにデータを提供してください」といったこともあり得ると思います。

2点目の市場予測についてですが、これについても今、研究を進めています。現在、端末の8割以上に人工知能が搭載されていると言われていています。今、金融市場で一番問題になっているのが、取引時間の速度です。光は1000分の1秒でも東

京から大阪くらいまでしか行けないので、世界中で取引をする場合、通信速度は大きな課題です。最近、このような取引では、複雑な人工知能は入っていないことが分かっています。人工知能といえども人間の模倣なので、すごく複雑なことをやっているわけではないということも段々見えてきました。

**佐藤** 終了の時間になりました。最後に、本日のディスカッションの論点を簡単に振り返って終わりにしたいと思います。まず、データをどういう形で集めて公開、共有するか、あるいは

はセキュリティを確保するかということが、課題にありました。次に、シミュレーションをどういう形で社会実装に持っていくかがありました。最後に、人材教育をどうするかという問題が提起されました。今後、理化学研究所ではポスト「京」の開発も進めていますので、それと合わせて、こうした課題に取り組んでいきたいと考えております。今日はありがとうございました。

## Summary of Key Points

## ▶なぜ、近年データサイエンスに注目が集まっているのか

- ① Webや携帯端末、各種センサーによりデジタルデータ量が指数関数的に増大するビッグデータ時代を迎えたため
- ② 機械学習、深層学習などの発展とコンピュータの飛躍的な性能向上によりビッグデータを利活用できる環境が整ってきたため
- ③ データサイエンスは帰納的な推論であり、演繹的な推論であるシミュレーションと融合させることで相乗効果や付加価値が見込めるため

## ▶それに対するシミュレーションの役割とは何か

- ① 機械学習と組み合わせることで、膨大なデータや稀有なデータを収集するための手段となること
- ② 取得したリアルタイムのデータを基にシミュレーションすることで少し先を予測でき、さらにそれを実社会にフィードバックできること
- ③ データ同化させることで逆問題(結果から原因を探ること)を解くことができること

## ▶データサイエンスとシミュレーションの融合による成果をいかにして社会に還元するか

- ① 生命科学現象、気象変化、大規模災害発生時の対応など、自然現象や社会・経済現象に関して、データサイエンスとシミュレーションの融合により“未来”を予測することで、人々のQOLの改善やそのための政策に役立てる
- ② まず多数の変数や高精細なデータを使って「京」でシミュレーションを行い、そこから変数を減らしシンプル化したアルゴリズムを開発することにより、個人や一般組織が簡単に使えるものにしていく
- ③ 教育により、データ解析及びシミュレーションを個々の領域に則して行える人材を育成、またそのためのツールを提供



まとめ

「京」×データサイエンス

現在、「京」に代表されるようなHPC(ハイパフォーマンスコンピューティング)では、従来のシミュレーションに加え、新たな学問領域として、ビッグデータ、機械学習、人工知能を含むデータサイエンス(データ科学)が台頭しています。シミュレーションが、物理法則に基づいた方程式に初期条件と境界条件を与えれば実行できる演繹的な推論であるのに対し、データサイエンスは、データが与えられてから実行される帰納的な推論です。

現在、このシミュレーションとデータサイエンスとの融合が強く求められています。データを解析しデータからモデルを抽出することで、シミュレーションの精度が大幅に向上したり、リアルタイム性が実現できたりするなど相乗効果が発揮されるからです。また、シミュレーションとデータを結びつける手法として、「データ同化」の重要性が増しています。

データサイエンスは、実験科学、理論科学、計算科学に次ぐ、第4の科学として位置づけられています。さらに、シミュレーションとデータサイエンスの融合を、理化学研究所 革新知能統合研究センターの上田修功副センター長は今回の

基調講演「機械学習の新展開～シミュレーション科学～」の中で、「シミュレーション科学」と命名。今後、シミュレーション科学は第5の科学として位置づけられていけると述べています(図1)。シミュレーション科学ではさらに、従来の計算科学で行われてきた予測や推定に留まらず、新たな概念や論理の創出が期待されています。

期待が高まる社会への還元

今回のシンポジウムでは、基調講演を含めて4人の研究者に登壇いただき、シミュレーションとデータサイエンスを融合することで初めて実現された研究成果を報告していただきました。その分野は、医療・ヘルスケアから社会経済、気象(防災)に至るまで多岐にわたります。しかもそのすべてが、我々の安心・安全で豊かな社会の実現に貢献する内容です。

例えば、上田修功副センター長による前述の基調講演では、NICT(国立研究開発法人情報通信研究機構)受託研究プロジェクト「SODA(Social big data)」などが紹介されました。これは、さまざまなセンサーを使ってソーシャル・ビッグデータを収集し、利活用することで、スマートシティの実現を目指すというものです。センサーが収集したデータには、位置

情報と時系列情報が含まれているため、この情報をリアルタイムかつ多角的に分析し、シミュレーションに反映させることで、例えば、災害時などに混まないルートを予測し、リアルタイム的に的確に誘導するといったことが可能となります(図2)。上田副センター長は、東京の国立競技場の8万人の観客の誘導に関するシミュレーション研究についても発表しました。2020年に開催される東京オリンピック・パラリンピックを視野に、こうした研究がますます加速することが予想されます。

また、ソニーコンピュータサイエンス研究所の桜田一洋シニアリサーチャーによる講演「予測と予防の個別化医療実現に向けたデータとシミュレーションの融合」では、理化学研究所をハブとする「高精度の予測に基づく予防医療の実現に向けた疾患ビッグデータ主導型イノベーションハブ」が紹介されました。これは、さまざまなバイタルデータを個人個人から収集することで、病気の発症などを高精度に予測し、一人ひとりに合った予防医療と治療を実現しようというものです。

東京工業大学 科学技術創成研究院の高安美佐子准教授による『「未来観測所」から見える日本企業の現状と未来:『京』をつかって』では、現在、高安准教授が帝国データバンクと共同で進めている研究成果が紹介されました。これは、帝国データバンクが保有する約100万社の30年間分の企

業データを解析することで、例えば、被災などにより、ある企業が機能しなくなった場合に、社会に与える経済的な影響などを推定できるというものです(図3)。この研究成果は、現在、内閣府の「RESAS(地域経済分析システム)」というプラットフォームにも採用されています。これは、人間社会や経済現象に関するビッグデータとシミュレーションとの融合により、未来社会がある程度、予測できることを示した事例です。

そして、理化学研究所 計算科学研究機構 データ同化研究チームの三好建正チームリーダーによる「シミュレーションと実測データを融合するデータサイエンス」では、「京」を使ったデータ同化の最先端研究が紹介されました。これは、気象レーダーから30秒毎に送られてくる観測データをデータ同化することで、短時間かつ局所的に発生するゲリラ豪雨を30分前に予測することに成功したという事例です(図4)。近年、ゲリラ豪雨による土砂災害などが世界的に大きな課題となっている中、1日も早い社会への還元が期待されています。

このように、スパコンという強力な道具を用いて、シミュレーションとデータサイエンスが融合することで、さまざまな社会課題を解決することができます。これは我々のQOLの向上に直結します。その他、あらゆる分野への応用と社会実装が期待されています。

科学研究の変遷

1. 実験科学  
アリストテレスの天動説(自然哲学)
2. 理論科学  
ニュートン(プリンキピア)
3. 計算科学  
数値計算, シミュレーション
4. データ集約型科学(第4の科学)  
データからの仮説発見

シミュレーション科学!?

図1 今後、実験科学、理論科学、計算科学、データサイエンスに続き、シミュレーションとデータサイエンスを融合させたシミュレーション科学が、第5の科学として位置づけられていくと思われる。

多種多様なセンサ情報がNWに接続され 人間社会で活用されるICT社会



図2 今後、さまざまなセンサーを使ってソーシャル・ビッグデータを収集し、利活用することで、スマートシティの実現を目指す。データには、位置情報や時系列情報が含まれるため、この情報をリアルタイムかつ多角的に分析し、シミュレーションに反映させることで、スマートな誘導や制御、さらには都市計画のデザイン、実行が可能に。

モデルの応用: 災害による企業の売り上げ被災額の推定

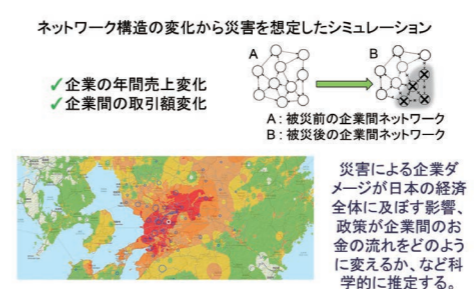


図3 人間社会や経済現象に関するビッグデータとシミュレーションとの融合により、例えば、被災などによってある企業が機能しなくなった場合に、社会に与える経済的な影響などを推定できる。

BDA ビッグデータ同化

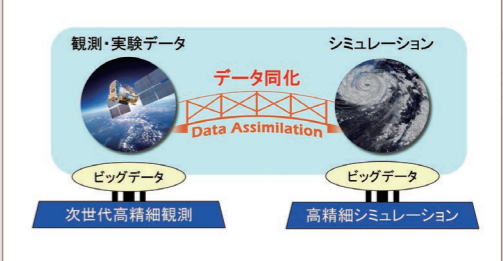


図4 観測データや実験データとシミュレーションを結びつけるデータ同化。特に今後は、次世代高精細観測によるビッグデータと高精細シミュレーションのビッグデータを融合させたビッグデータ同化が求められていく。



## まとめ

### 「京」の産業利用と 本シンポジウムの意義

2012年9月に本格運用を開始した「京」は現在、日本全国のさまざまな分野の方々に利用していただいております。年間150を超える課題を実行する共用施設として運営されています。「京」の産業利用も順調に拡大しており、現在では、延べ170社以上の企業が利用しています。産業利用は競争力強化のカギとなるため、文部科学省と理化学研究所と高度情報科学技術研究機構が強く連携し、利用環境の整備に努めています。

特に最近では、内閣府が「Society 5.0」を掲げ、その実現に向けて取り組んでいます。Society 5.0とは、あらゆるものがデータによって結びつくことで、新しい価値が生み出されていく社会です。このような社会を実現していく上では、質の高いビッグデータの利活用が不可欠です。

しかし一方で、直面している課題もあります。例えば、医療や気象、防災の分野では、実験では作り出すことができない事象、データの収集が困難な稀有な事象、データ量が多量かつ膨大かつ複雑で、仮説を作り出すこと自体が困難な事象などがあります。こうしたデータの収集がままならない事象に対しても、解を導いていくことが求められています。

このような中、その解決策の1つとして、質の高いシミュレーションと、人工知能を中心とするデータサイエンスの融合に期待が高まっているのです。

そしてこのことを、より多くの方々と共有するとともに、議論を進めていくことが重要です。このような思いから、今回のシンポジウムは開催されました。

近い将来、今はまだ全く想像もつかないような新しい学問領域や、社会の実現が想定されます。より多くの方々、異分野の方々をつながることで化学反応が起こり、イノベーションの創出につながるものと思われます。今回のシンポジウムが、そうした未来へのさらなる前進のための一歩となることを望まれます。

### 最後に

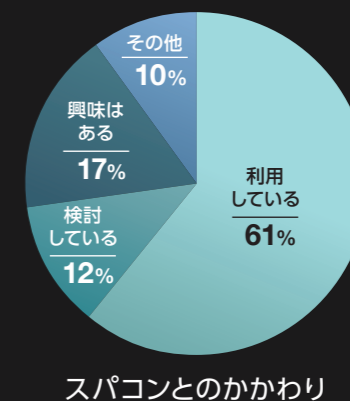
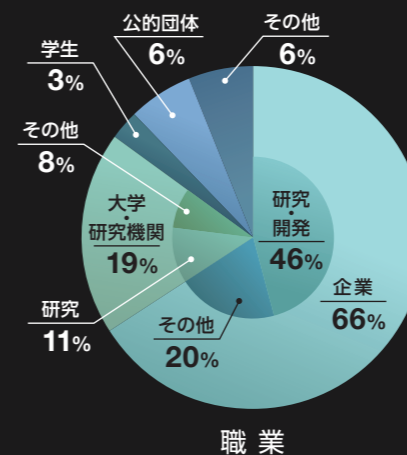
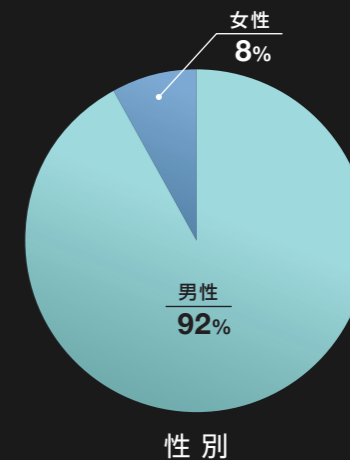
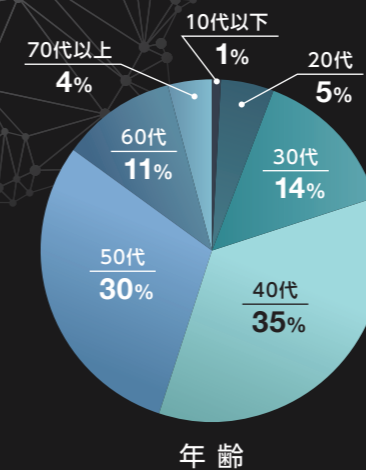
従来、「京」のシンポジウムでは、「京」が誇る高い計算能力を用いたシミュレーションの成果を紹介してきました。一方、今回はそのようなシミュレーションとデータサイエンスとの融合によって、初めて実現した最先端の研究成果を紹介しました。

今後、2020年頃の運用開始を目標に、最大で「京」の100倍速くシミュレーションが実行できるポスト「京」が登場します。ポスト「京」とデータサイエンスが融合することで、社会現象、生命科学、天気予報、防災・減災といった幅広い分野で、研究のさらなる進展が見込まれます。

理化学研究所 計算科学研究機構と高度情報科学技術研究機構は今後も、新たな分野を柔軟かつ迅速に取り込みながら、「京」、そしてポスト「京」を利用するための支援を行っていく計画です。

## 京×データサイエンス シンポジウム 参加者について

### 当日来場者数 197名



多くの方々にご参加いただき、誠にありがとうございました。

### 京×データサイエンスシンポジウム報告書

2018年1月発行

編集協力：株式会社アテナ・ブレインズ

国立研究開発法人理化学研究所 計算科学研究機構  
〒650-0047 兵庫県神戸市中央区港島南町7-1-26  
電話(078)940-5555 FAX(078)304-4956  
<http://www.aics.riken.jp/>

一般財団法人高度情報科学技術研究機構  
〒650-0047 兵庫県神戸市中央区港島南町1-5-2  
電話(078)599-9511 FAX(078)599-9512  
<http://www.hpci-office.jp/>