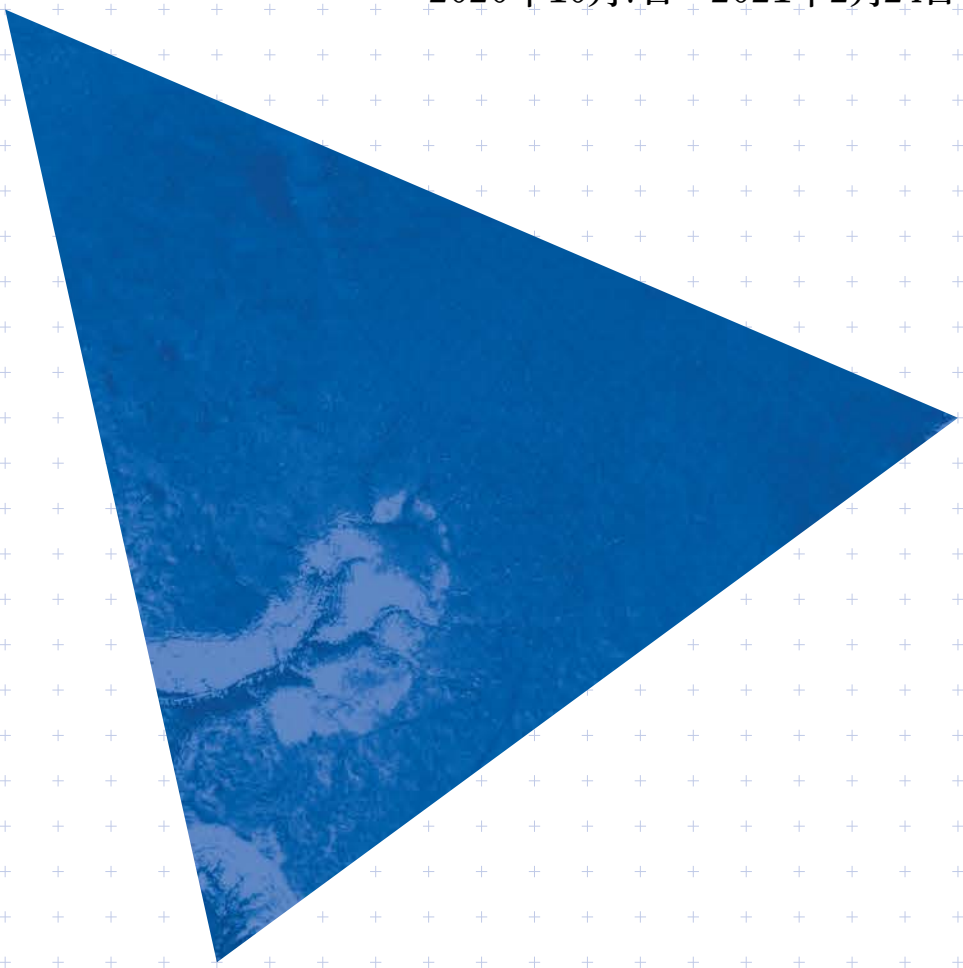
ワークショップ報告書
CRDS-FY2020-WR-09

俯瞰ワークショップ報告書

システム・情報科学技術分野 数学と科学、工学の協働に関する連続セミナー

2020年10月7日～2021年2月24日 全 16 回



国立研究開発法人科学技術振興機構
研究開発戦略センター

エグゼクティブサマリー

本報告書は、国立研究開発法人科学技術振興機構（JST）研究開発戦略センター（CRDS）が2020年10月から2021年2月まで、全16回にわたって開催した『俯瞰ワークショップ　システム・情報科学技術分野「数学と科学、工学に関する連続セミナー」』に関するものである。

社会課題の解決に向けた科学技術にはシステム・情報科学技術、ナノテク・材料技術などさまざまなものがある。それらはロボティクスやAIなどの成果を通じて社会で活用される。この目に見える成果を下支えするのが工学であり、自然科学である。さらにその根底には数学が存在し基礎を支えている。数学と科学、工学が協働し、科学技術を通じて社会の課題解決に貢献しているのである。また科学技術の社会適用から得られる新たな課題を受け取り、科学技術の基礎としての次なる発展を目指す。このような循環を描くことで社会のイノベーションを実現していくのである。

CRDSでは、数学と自然科学、工学の連携についての俯瞰調査、政策提言を目指して、2020年から活動を開始した。まず、現状の数学と自然科学、工学の連携の状況を調査するために連続セミナーを企画した。2020年10月から2021年2月まで、全16回のセミナーを開催した。31名の多彩な講師をお招きし、各回1時間30分にわたり、主に基礎的な数学とその応用を組み合わせて、それぞれアカデミアとインダストリーから1名ずつの講師による講演と質疑応答を行なった。セミナー聴講の申し込みは、各省庁並びに関係機関、アカデミア、インダストリーなどから120名超に及び、各回においても40名を超える参加者を得た。本報告は、講演録を中心として、数学に関する考え方、CRDSの取り組み等について述べたものである。

日本に限らず、世界はビッグデータ・AIが席巻する社会となりつつある。 AIなどの先導もあり、広い意味で企業における“数学化”が進んでいるかのようである。しかし、日本においては AIに関しても工学的利用のみが大きく進んでいるように見受けられる。 AIを利用する研究や開発には、より深い数学・数理科学の裏づけが必要である。いまだ多いAIに存在するブラックボックス、背後に動くアルゴリズムにおけるリスク可能性やAIが導く結果の検証や評価においては、確実性を高め信頼性への配慮を怠るわけにはいかない。また、安定した持続的な力を備えるためにも、将来に続く今後の人材育成に多くの力を注いでいく必要がある。したがって、わが国の将来を考えると、大学等の学術機関はもとより、産業界での真の意味での数学利用、産業数学、数理科学研究の振興が必須である。 CRDSでは、わが国における科学技術政策に関するシンクタンクとして、今後の日本において必要なことはもちろん、世界的に高まる数学・数理科学の役割をどう推進していくべきかの調査・検討を進める予定である。そこでは、国際的な比較や積極的な情報収集に基づき指針をまとめて提言に向けたい。本報告書はその序章・前段階であり、指針作成の第一歩として位置付けている。数学・数理科学とその応用分野の発展に皆様の深い関心が寄せられることを期待する。

目次

1 ｜ はじめに

“万物は数”とは、ピタゴラスによる。現在のAI、ビッグデータの到来くらいは知っていたかのようである。そのころに示された素数が無限に存在することと素因数分解の一意性が、およそ2500年後に公開鍵暗号を生んだ。もっとも技術的には17世紀のフェルマーの小定理[1]を使っている。しかし古代ギリシャ人が、どういう意図から素数たちに思いを馳せたのかは類推するしかない。いずれにしても、応用までに随分と時間がかかっているように感じるかもしれない。ただし、宇宙と言わずとも人類の長い歴史に比べれば瞬き程度である。このように、重要な数学の発見は多少の時間を超えて社会とともにある。ところで、レオナルド・ダ・ビンチは多方面で時代を超えていた天才である。彼はいう。“工学は数学的科学の楽園である。 何となればここでは数学の果実が実るから”[2]。また、ガリレオ・ガリレイの次の言葉はあまりにも有名である。“宇宙（自然）は数学の言葉でかかれている”。しかしながら、物理学者ユージン・ウィグナーによる1960年に公開された講演録、“自然科学における数学の理不尽ともいえる有効性”ほど、多くの天才たちにも共有されて数学の特長が端的な表現は見当たらない。もちろん、数学自体は発見するものであり発明できるものでもない。また、数学にも様々な分野があるが、じつは秀れた数学研究の営みには根底から繋がる一体感がある。このことは数学を生々と発展させてきた理由の一つでもある。

わが国では、先進諸国にあっては上記の感覚があまり共有されてこなかった。たとえば世界ではSTEM、STEAM[3]といわれて久しい。だが荒っぽくいうと、受験数学を除けば、実質的にMが希薄である。計算機性能が貧弱な時代には、実は日本にも超一流の応用数学を実践した工学者が多くいた。それが大学での教育にも生かされた。日本の高度成長の原動力の一因になったと考えられる。したがって、ずっとMが欠けていた訳ではない。数学は、諸科学・技術において、いわば通奏・横断的な役割をする。ただしその特質は、成果物からはほぼ見えないことにもある。しっかりと前提を踏まえれば、それが導く数学の結論は（少なくとも数学としては）正しい。それがアルゴリズムという面からの数学のひとつの有用性である。また、その抽象性は思考の整理と柔軟性、ときにはそれまでの概念を破壊するに見えるほどの飛躍を導き、同時に実問題に対しても有用性を発揮している。20世紀には、その最も親しかった物理学からも乖離したかのように、抽象性を前面におきながら発展した。（セミナーでは、可能であればこのような発展に根差す諸例の紹介もしていただこうと、講師の方に依頼した。）だが近年、計算機性能とそれに依拠した科学と技術が目を見張るまでに進歩したことで、抽象といういわば空中戦で進んだ数学が再び実社会へ降りてきている。米国のGAFA[4]を持ち出すまでもなく、数学とエンジニアとの共同作業は社会を先導しているようにみえる。数学は、なにもそれを大学・大学院で専攻した者のみの学問や武器・道具ではない。多くの研究開発現場では、その解決のために、さらなる数学を待っているはずである。他方、中途半端な数学の利用は危険と背中合わせである。さらに、数学は計算とは似て非なる場合があることにも注意すべきである。実際、与えられた微分方程式などに対して、その解の存在と一意性を知ることは数学における重要テーマである。これは同時に、応用上もきわめて重大・深刻な課題を孕んでいる。事実、解が求まらないときには数値計算で臨む。ここに誤差が伴うことは皆が知っている。しかし本当は、誤差どころか望む性質をもつ解が存在しないかもしれないのだ。どんなに見た目の現

1 フェルマー（1665年没）予想で有名なフェルマーの最終定理（1995年に A. Wiles により証明された）ではない

2 レオナルド・ダ・ヴィンチの手記（下）(杉浦明平訳) 岩波文庫 1958年

3 STEM = Science, Technology, Engineering and Mathematics, STEAM = STEM+A（=Arts）

4 GAFA = Google, Amazon, Facebook, Apple

象と話が合うようでも、時間が進んだときにそれらがずっと正しいという根拠はない。一寸先はわからないのである。たとえば流体の方程式としてきわめて有名なナビエ・ストークス方程式にせよ、解の存在と滑らかさはいまだ未解明である[5]。また、微分方程式などに関する逆問題では、解が与えられるデータの微小変化にも敏感であり安定性に欠けることが多い。同様に、AIの導く答えの有効性を確認するには数学や確率論をベースとする深い統計学が必要である。また、数学の衣を纏ったアルゴリズムに隠れるリスクは見えにくい。

本セミナーシリーズでは、数学的手法の有効性・効率性、数学を用いることでの信頼性の向上や系統的な探索など、その考え方や手法が豊富な例とともに著名な講師により紹介された。毎回のセミナーには、アカデミアおよび産業界（あるいは産業界で活躍されたご経験のある研究者のなか）から、数学そして統計学を主たる道具とされ研究開発現場で活躍されている方に依頼してお引き受け頂いた。たいへん広がりがある、そして奥深く面白いお話しを拝聴できた。この場を借りて深く感謝の意を記しておきたい。成果や応用からはおよそ予見し難いながら、素晴らしい数学的アイデアや理論が私たちの住む現実世界に、いま生かされている。セミナーでは、そのことを初めて耳にされた聴衆の方々もおられたと思う。CRDSにおいては、数学ポテンシャルが高いにも拘らず、先進国としては日本が遅れをとる数学の真の活用を活発化させたいとの考えがある。セミナーで紹介された研究者・技術者個人のなかでの数学利用と実社会での応用、数学者と諸科学や産業界とのチーム連携により生まれた技術を通じて、新しい数学の課題が生み出され、数学の発展にも資することを期待している。本セミナーでの目的とはしなかったが、このような発展の連鎖を通じて、数学が、形を変えた場面で、将来にまた役立っていくという好循環を促すことになるはずだ。

この機会に、上記では日本での数学とその応用研究への認識の薄さを指摘した。限られるが、以下ではすこし世界動向の一端も見てみたい。

米国では、数学で博士号を取得した人たちが、アカデミアのみならず、産業界、国や州政府、たとえば、DOEやDARPA、NIH[6]で多く活躍している。2015-16年（学年暦）での米国数学会の調査資料によると、アメリカにおいて数理科学で博士号を取得した人数は1642名である。日本数学会の同年の調べによると、理学系の数学に限られた数であるが、博士号取得者は137名である。工学系で数学の学位を取得した人の人数を加えるともう少し多くなると思われる。ただし、総人口でいえば米国は日本のおよそ2.6倍である。やはり日本の数学系の博士号取得者は少ない[7]。また米国では、学位取得直後、このうち半分強の900名近くはアカデミアに残り、3分の1が民間企業に活躍の場を得る。その他の進路は政府機関等であるが、軍事・防衛といった文脈から必須のセキュリティー関係部署では1000人以上の数学研究者が働いていると聞く。加えて米国の数学への政府投資額は、年間10億ドル強である。トランプ政権時代、科学・技術予算はフラットに推移した。それ以前の10年間も全体としては停滞気味であったが、数学予算は増加していた。注目すべきは、そのなかで日本のJSPSとJSTを合わせたような、米国の科学・技術振興のために科学研究費等を統括する

5 米国クレイ研究所のミレニアム100万ドル懸賞問題

6 DOE：米国エネルギー省、DARPA：国防高等研究計画局、NIH：国立衛生研究所

7 経済産業省・文部科学省報告書「数理資本主義の時代」(2019.3)：https://www.meti.go.jp/shingikai/economy/risukei_jinzai/20190326_report.html　ただし、該当する日米比較表（本文図7）において、日本側の数字は日本数学会の調査であり、理学系の数学に絞られた範囲になっていることに注意。一方で、ここには日米社会における“数学・数理科学”浸透の差も垣間見える。なお、より新しい日本数学会による2019年度「数学・数理科学分野の博士後期課程修了者の進路調査アンケート」調査結果についても、数学会会報180（in「数学通信」第24巻第4号：2021.2）にある。参考にしていただきたい

NSF[8]予算が占める割合は全体の23%にとどまっていることだ。そのほかの77%は、上記の省庁をはじめ国のセキュリティー分野に関わるなどである。これは数学を用いた応用や出口研究への投資の大きさを示すもの

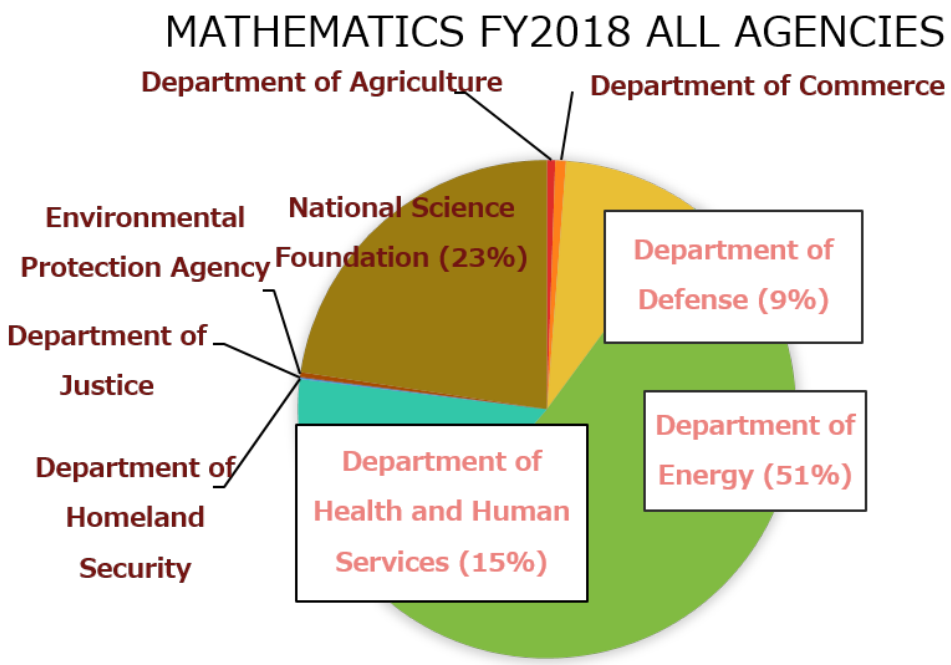


図1-1　　米国における数学への投資

である。

他方、Brexit などもありダイナミックに国が動いているイギリスでは、2020年1月末、ジョンソン首相が、数学への新しい投資として5年間で3億ポンド（約400億円：年間80億円）の追加を宣言した。例年の科研費にあたる数理科学に対する予算は1.8億ポンド余[9]（約255億円）である。また、ミッドランドに新しい数学国際研究所を建てるために一億ポンドを拠出するとした。後者はCOVID-19対策などで、資金繰りが不透明になっているが、前者については昨年秋よりすでに予算が執行されている。かくなる追加予算であるが、その目的は、研究プロジェクトの支援ではない。博士課程学生を倍増以上にすること、海外から秀れた数学者を招聘すること、ポスドクに教育機会を多く与え、同時に大学教員である数学者の時間を確保することとある。

ここでは、米国、英国のみの国家投資について述べた。中国は米国までには公表されておらず不明であるが、この20年間の伸びは著しくみえる。なおこの3国は、数学論文生産数でいえば世界上位を占めている。ここで国家の数学への投資を述べたのは、国や社会が数学をどのように見ているか、ということの一端を示しながら紹介したいという意図からである。 GAFAにおける研究開発費は、企業 IRレポートによると年間9.5兆円に上るが、その大半は人件費であるという。サイモンズ（Simons）財団やカブリ（Kavli）財団をはじめ、米国では数学そのものへの個人の名を冠する財団からの支援も大きく、きわめて特徴的である。これらの投資には、目的あるいは結果として、研究人材の育成にも大きく寄与している。見えない数学の力への期待でも

8　NSF：アメリカ国立科学財団

9　2021年1月末日現在

ある。ところが一方、日本は、その人材育成も含めモノづくりに長けた国である。そして、医療・診療データが丁寧に集取・整理されていることなどは国際的にも評価されている。そうした意味で、たとえば世界一流のセンサー技術などからは、データを丁寧に収集する能力はGAFAにも勝り、一日の長があるように見受けられる。そしてそこには、数理科学の力が発揮できる豊かな場面があると考えられる。大いに期待するところである。

このような状況下、これまでに数学がどのように使われているかの一端に、とは言いつつも幅広く触れていただくために本連続セミナーを企画した。講演をお願いした方々には、聴衆は普段数学に特段の馴染みがある方ばかりではない（むしろ少ない）ことを予めお伝えし、種々の工夫もお願いした。16回のテーマは異なっていたが、そのなかにも、どこかで互いに通じていることを感じ取ってくださった聴衆もおられたと思う。実はそうお感じになることは、数学の特長をひとつ捉えられたに等しい。また、半年間に16回ものセミナーに、すべては難しかったにせよ、お忙しい時間の合間を縫って参加してくださった皆様に深く感謝申し上げたい。

日本に限らず、世界はビッグデータ・AIが席巻する社会となりつつある。 AIなどの先導もあり、広い意味で企業における“数学化”が進んでいるかのようである。だが、日本においては AIに関しても工学的利用のみが大きく進んでいるように見受けられる。気になるところである。実際、AIを利用する研究や開発には、より深い数学・数理科学の裏づけが必要である。いまだ多いAIに存在するブラックボックス、背後に動くアルゴリズムにおけるリスク可能性やAIが導く結果の検証や評価においては、確実性を高め信頼性への配慮を怠るわけにはいかない。少なくとも上記で挙げた米国におけるAIの開発利用・研究では、この点も重視されている。同時に、安定した持続的な力を備えるためにも、将来に続く今後の人材育成にも多くの力を注いでいく必要がある。したがって、わが国の将来を考えた折、大学等の学術機関はもとより、産業界での真の意味での数学利用、産業数学、数理科学研究の振興は待ったなしである。 CRDSでは、わが国における科学技術政策に関するシンクタンクとして、今後の日本において必要なことはもちろん、世界的に高まる数学・数理科学の役割をどう推進していくべきかの調査・検討を進める予定である。そこでは、国際的な比較や積極的な情報収集に基づき指針をまとめて提言に向けたいと思う。本報告書はその序章・前段階であり、指針作成の第一歩として位置付けている。数学・数理科学とその応用分野の発展に皆様の深い関心が寄せられることを期待したい。

2 ｜ 数学と科学、工学の協働

2.1 数学の必要性

国家が持続的に発展するには、短期の国益のみにとらわれず、いかに世界に貢献できるかを深く踏まえたビジョンを持って、ある程度の傍証はありながらも、いまだ計り知れないことへの投資をすることが我が国にとって不可欠である。数学は自然科学・工学全般の基礎力を高める。また、アルゴリズムのみならず、数学の持つ抽象性に基づくことで初めて知ることができる応用による産業上の効果も報告されている。こうした点に鑑み、我が国の数学と自然科学、工学との連携を実現に導くために、国際協力も踏まえた数理科学への投資と人材育成の重要性を明確にしなければならない。

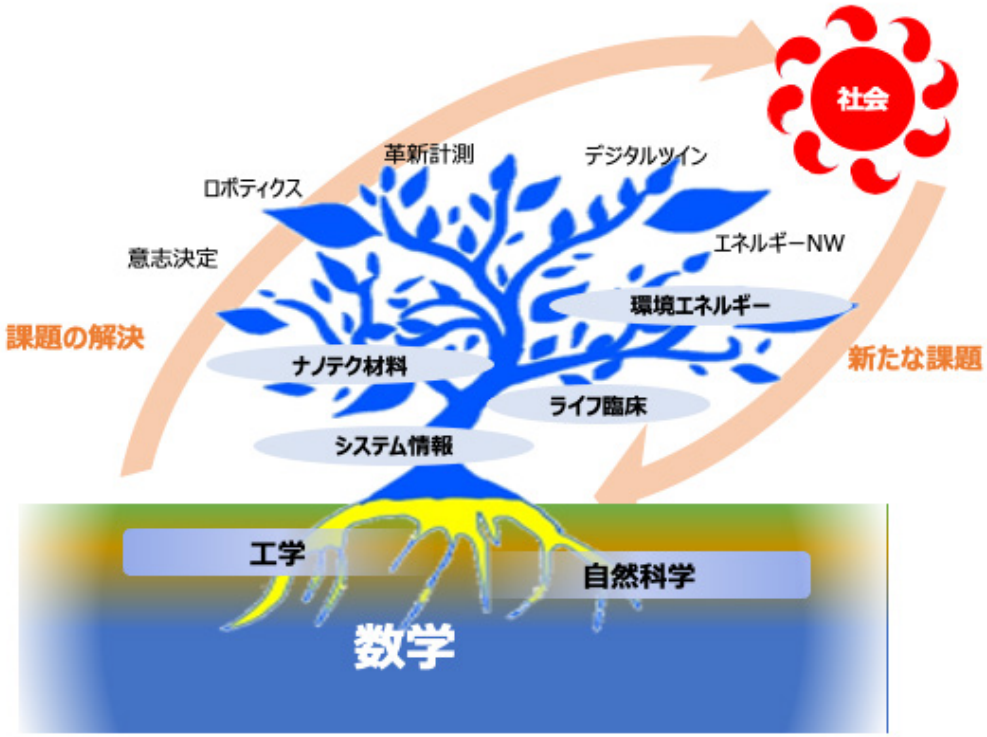


図 2-1　　イノベーションの基礎としての数学

図 2-1 に示すように、社会に向けた科学技術にはシステム・情報科学技術、ナノテク・材料技術などさまざまなものがある。それらはロボティクスや AI などの成果を通じて社会の課題を解決するために活用される。この目に見える成果を下支えするのは直接的には工学、そして医療・医学であり、さらに自然科学である。さらにその根底には数学が存在し基礎を支えている。数学と科学、工学が協働し、科学技術を通じて社会の課題解決に貢献する。また科学技術の社会適用から得られる新たな課題を受け取り、科学技術の基礎としての次なる発展を目指す。このような循環を描くことで社会のイノベーションを実現していくのである。

2.2 動向

数学と自然科学、工学との連携に関する活動について、いくつかの事例を紹介する。

（1）　欧州

European Science Foundationは2010年11月に、“Forward Look on Mathematics and Industry”[1]と題する報告書をEuropean Mathematical Societyと連携して発行した。この中で、2020年以降ヨーロッパのあるべき姿として、

・スマートな成長、知識とイノベーションに基づく経済発展
・持続的な成長、資源効率、グリーンで競争力のある経済を促進
・包摂的な成長、経済的、社会的、そして地球規模の結束をもたらす高雇用経済の育成

の3点を挙げている。これを実現するためにこの報告書では、数学と産業の協力を刺激し強化するための方策を提案している。そして、戦略的な目標として、応用数学のコミュニティ形成、発展、強化を設定した。具体的な進め方として、

・European Institute of Mathematics for Innovationにおける数学に関する活動に資金を投入
・学術組織と産業界は全ヨーロッパで複数の学術領域にまたがったベストプラクティスを共有し、流通

と言うことを勧告している。

このForward Lookに続いて、イノベーションにおける数学の重要性を示すために、”European Success Stories in Industrial Mathematics”[2]を2012年に発行している。ここでは、数学を使うことが成功のためには決定的に重要であり、産業において多くの数学の分野が関連していることを述べ、数々の成功事例、すなわち数学を利用することによって産業的な効果が得られた事例を列挙している。数学と産業は双方向の関係性があり、数学が産業へのソリューションを提供する一方で、産業応用を進めることによって数学への新たな研究テーマと新たな方法論への刺激が得られると指摘している。

（2）　米国

2013年にthe National Academy of Sciencesはthe National Science Foundationの支援のもと、“The Mathematical Sciences in 2025”[3]を発行した。この中で、数理科学はあらゆる科学における不可欠な要素であり、数学者は学際的であると同時に研究活動も学際的であるべきであるとしている。そのためには新たな数学教育が必要であるとともに、ファンディングが学際的活動に対応しなければならない。しかし、現状では数学の役割の増大に比して、ファンディングは不十分である。女性、マイノリティへの数学教育を充実させなければならず、MOOCsなどの活用を考えるべきだとしている。

全米科学・工学・医学アカデミー（NASM：The National Academies of Sciences, Engineering, and Medicine）は、大学、国立研究所、企業の専門家からなる特別委員会を招集し、数理科学の影響を特定し、それを説明するための報告を作成する。この報告は、米国経済、国家安全保障、健康、医学、その他の科学、工学、技術分野における数理科学の基本的な役割を大まかに説明するものである。

この活動では、以下の 3 つの成果が期待される。（1）説明と図、およびレポートの要約を含むパンフレット、

1 ISBN :978-2-918428-28-2
2 ISBN :978-3-642-23847-5
3 National Research Council. 2013. The Mathematical Sciences in 2025. Washington, D.C.: The National Academies Press

（2）それぞれのトピックに関連するWebインタラクティブインフォグラフィック、（3）公開Webinarシリーズである。

Webinarは2020年の3月から開催され、下記のようなテーマが取り上げられた。

・From Solving to Seeing
・Revolutionizing Manufacturing through Mathematics
・Abstract Geometry, Concrete Impact
・Keeping the Internet Safe
・Understanding Uncertainty
・Computed Commutes
・Climate and Weather
・Smart Materials
・Personalized Medicine
・Leveraging Data Through Mathematics

（3）　日本

• FIRST合原最先端数理モデルプロジェクト[4]

最先端研究開発支援プログラム「複雑系数理モデル学の基礎理論構築とその分野横断的科学技術応用」プロジェクト（中心研究者：合原一幸教授、通称：FIRST合原最先端数理モデルプロジェクト）が2010年3月から2014年3月まで実施された。これは、世の中の様々な現象に学びそれを社会に活かす数学である「数理工学」の観点から、様々な複雑な科学技術の問題を解くための基礎理論研究とその応用研究である。国内外の32の大学・研究機関および4社の企業から計約80名の研究者が研究分担者として参画した。「複雑系数理モデル学の基礎研究」、「複雑系数理モデル学の工学応用研究」、「基礎研究と応用研究の融合による複雑系数理モデル学の体系化」という3つのサブグループで研究が実行された。

• 九州大学マス・フォア・インダストリ研究所[5]（MI:Mathematics for Industry）

マス・フォア・インダストリとは、純粋・応用数学を流動性・汎用性をもつ形に融合再編しつつ産業界からの要請に応えようとすることで生まれる、未来技術の創出基盤となる数学の新研究領域である。 MI の研究と人材育成を推進するために２０１１年４月に創設された。

・産業界との共同研究とそれを支える数学研究
・グローバルな場でリーダーとして活躍できる若手研究者、優秀な人材の輩出
・共同利用・共同研究拠点として共同利用研究の企画・運営

などの活動を行なっている。

マス・フォア・イノベーション卓越大学院では、数学力・統計力を基盤に、数学モデリングを通して組織や分野の垣根を越えて各分野で共創し、イノベーションを創発する卓越した数学博士人材を育成する5年間の修士博士一貫プログラムである。令和2年度には文部科学省卓越大学院プログラムに選定された。

• 東北大学材料科学高等研究所[6]（AIMR:Advance Institute for Materials Research）

材料科学における世界拠点となるべく、新たなシステム作り・研究活動に取り組むため、2007年、文部

4 https://jom.jsiam.org/11370/
5 https://www.imi.kyushu-u.ac.jp/pages/about.html
6 https://www.wpi-aimr.tohoku.ac.jp/jp/about/outline/

科学省プロジェクトであるWPI（世界トップレベル研究拠点プログラム）のもとに「原子分子材料科学高等研究機構」として創設された。2012年からは、数学との連携により、予見に基づいて材料を開発できるような新学理を創出するための基礎的研究を進めている。2017年4月には、組織名称を「材料科学高等研究所」へと改称した。

・研究所レベルでの材料科学と数学とのコラボレーション
・研究者の約40%を外国人が占める国際的な研究環境

などがAIMRの特徴である。

- 京都大学ヒト生物学高等研究拠点[7]（ASHBi:Institute for the Advanced Study of Human Biology）

ヒト生物学高等研究拠点（ASHBi）は、ヒトや霊長類を用いた体系的な研究を推進し、進化が付与した多様性 = 種差 の表出原理を解明する、先進的なヒト生物学を創成するため、文部科学省「世界トップレベル研究拠点プログラム（WPI）」の採択を受け、2018年10月に設立された。中でも、ASHBi数学グループでは生命科学グループとの融合研究として、多階層・多スケールにわたる多生物種・多細胞種データ解析のための新規方法論を、トポロジカルデータ解析・力学系理論・機械学習などを用いて開発している。

- 理化学研究所数理創造プログラム[8]（iTHEMS:Interdisciplinary Theoretical and Mathematical Sciences Program）

数理創造プログラムは、理論科学・数学・計算科学の研究者が分野の枠を越えて基礎研究を推進する、理化学研究所の新しい国際研究拠点である。 iTHEMSでは、様々な科学のしくみを支えている「数理」を軸とする分野横断的手法により、宇宙・物質・生命の解明や、社会における基本問題の解決を目指している。

・現代数学の活用で未知の科学に挑む
・柔軟に変化する組織でテーマを深める
・顔の見える交流で異分野連携を促進

といった取り組みが具体的になされている。

- 理化学研究所 革新知能統合研究センター[9]（AIP: Center for Advanced Intelligence Project）

革新知能統合研究センターは、革新的な人工知能基盤技術を開発し、それらを応用することにより、科学研究の進歩や実社会における課題解決に貢献することを目指し、文部科学省AIPプロジェクトの研究拠点として2016年度に設置された。AIPには3つの研究グループ

・汎用基盤技術研究グループ
・目的指向基盤技術研究グループ
・社会における人工知能研究グループ

を設置し、基盤技術の開発、サイエンス研究の加速、社会問題の解決、人工知能の倫理的・法的・社会的課題の分析及び人工知能研究者・データサイエンティストの育成を行っている。

また、JSTでは、戦略的創造研究推進事業において、表1に示すような数学に関連する取り組みを行っている。

7 https://ashbi.kyoto-u.ac.jp/ja/research/interdisciplinary-research/
8 https://ithems.riken.jp/ja/about/outline
9 https://www.riken.jp/research/labs/aip/
https://aip.riken.jp/about-aip/

アメリカ、ヨーロッパの動向は、研究開発活動およびその社会適用を推進する活動であり、個別の研究開発活動や組織はかなりの数が存在する。体系的かつ持続的発展のために、研究および大学レベルでの教育などの人材育成も強く打ち出されている。一方、我が国にはこういった上位の活動があまりなく、個別の研究開発プロジェクトや組織を紹介するに止めた。今後の我が国の数学関連研究開発をより活発化するためには、個別の研究プロジェクトや組織だけではなく、上位の政策的、あるいは社会的、産業的な意識の向上が必須である。

表1　　これまでの数学系の戦略的創造研究推進事業一覧[10]

（研究総括の所属は当時のものである）

種別	研究領域	活動期間	研究総括
ERATO	合原複雑数理モデルプロジェクト	2003～2008	合原 一幸（東京大学 生産技術研究所）
CREST さきがけ	［数学］数学と諸分野の協働によるブレークスルーの探索	2007～2015	西浦 廉政（東北大学）
ERATO	河原林巨大グラフプロジェクト	2012～2017	河原林 健一（国立情報学研究所）
さきがけ	［数学協働］社会的課題の解決に向けた数学と諸分野の協働	2014～2019	國府 寛司（京都大学）
CREST	［数理モデリング］現代の数理化学と連携するモデリング手法の構築	2014～	坪井 俊（武蔵野大学）
ERATO	蓮尾メタ数理システムデザインプロジェクト	2016～2021	蓮尾 一郎（国立情報学研究所）
CREST	［数理的情報活用基盤］数学・数理化学と情報科学の連携・融合による情報活用基盤の創出と社会課題解決に向けた展開	2019～	上田 修功（NTTコミュニケーション化学基礎研究所）
さきがけ	［数理構造活用］数学と情報科学で解き明かす多様な対象の数理構造と活用	2019～	坂上 貴之（京都大学）
ACT-X	［数理・情報］数理・情報のフロンティア	2019～	河原林 健一（国立情報学研究所）

10　https://www.jst.go.jp/kisoken/index.html

2.3 CRDSの取組み

CRDSでは、数学と自然科学、工学の連携についての俯瞰調査、政策提言を目指して、2020年から活動を開始した。まず、現状の数学と自然科学、工学の連携の状況を調査するために連続セミナーを企画した。詳細は次章にゆずるが、主に基礎的な数学とその応用を組み合わせて、それぞれアカデミアとインダストリーから1名ずつの講師を招き、全16回のセミナーを開催した。今後、セミナーの内容を整理し、数学とその応用に関する俯瞰図を作成する。その過程で重要な課題、喫緊のテーマなどを特定する。その後、関連する学界へのインタビューとともに、産業界へもインタビューを実施し、現状を把握するとともに、課題の明確化を行う。海外も含めて、数学界の状況、各国の政策やファンディングあるいは教育、人材育成、産学協同のあり方などについても調査する。また、必要に応じてワークショップを開催する。これらの成果は2023年度に発行する予定の俯瞰報告書にまとめる予定である。

3 ｜ 連続セミナー

２０２０年１０月から２０２１年２月末まで、16回に渡って「数学と科学、工学の協働に関するセミナー」を実施した。 COVID-19の影響により、一堂に会する形での開催は困難であったため、全てオンラインで行なった。直接顔が見えない、臨場感に欠けるという欠点はあるものの、移動を伴わないため時間的な制約が緩和され、次表に示すような多忙を極める著名な講師の方々を招くことができた。聴講者にとっても参加しやすかったのではないかと想像される。また、オンライン会議システムの機能を利用して、セミナーの映像を記録し、参加できなかった聴講申込者には後日ビデオ視聴が可能なように配慮した。講師の発表資料についても、講師の協力によってビデオと同様、聴講申込者に限定公開することができた。

1	2020/10/7	13:00-14:30	数学の広がり	徳山 豪（関学）	
2	2020/10/21	13:00-14:30	社会、産業と最適化	土谷 隆（政策研究大学院大学）	田辺 隆人（NTTデータ数理システム）
3	2020/11/4	13:00-14:30	量子情報処理	富田 章久（北大）	萩原 学（千葉大）
4	2020/11/11	10:30-12:00	数理ファイナンス・金融工学	谷口 説男（九大）	池森 俊文（統数研）
5	2020/11/18	10:30-12:00	力学系と安定性、制御、感染症の数理	國府 寛司（京大）	稲葉 寿（東大）
6	2020/11/25	13:00-14:30	Uncertaintyと数学	竹村 彰通（滋賀大）	恐神 貴行（日本IBM）
7	2020/12/2	10:30-12:00	ネットワーク、グラフとSNS	藤澤 克樹（九大）	山下 長幸（NTTデータ経営研）
8	2020/12/2	13:00-14:30	耐量子計算機暗号	髙木 剛（東大）	高島 克幸（三菱電機）
9	2020/12/9	13:00-14:30	データマイニングと知識発見	山西 健司（東大）	上田 修功（NTT・理研）
10	2020/12/16	13:00-14:30	データドリブン数理モデル構築（因果推論、情報量基準）	二宮 嘉行（統数研）	伊藤 公一朗（シカゴ大）
11	2021/1/6	10:30-12:00	Machine Learningと数理モデル	河原 吉伸（九大）	渡辺 澄夫（東工大）
12	2021/1/13	13:00-14:30	数学と可視化技術	高橋 成雄（会津大）	安生 健一（OLMデジタル）
13	2021/1/27	13:00-14:30	シミュレーションとデータ科学	吉田 亮（統数研）	三好 建正（理研）
14	2021/2/10	10:30-12:00	統計とスパースデータとAI	椿 広計（統数研）	木虎 直樹（HACARUS）
15	2021/2/24	10:30-12:00	科学・工学・医学における数理	水藤 寛（東北大）	中川 淳一（東大）
16	2021/2/24	13:00-14:30	トポロジカルデータ解析：数学的発展と諸科学への応用	平岡 裕章（京大）	平田 秋彦（早大）

文科省、内閣府等関連省庁、JST関連部署、企業等に参加を呼びかけ、124名の申し込みを得た。それぞれの回の出席者は40名前後であった。

3.1 数学の広がり

［概要］

徳山氏から「数学の広がり：行列式と因数分解から幾何、計算理論、量子計算へ」というタイトルで、三角形の面積公式、大学初年級レベルの行列式の計算から統計力学や量子誤り訂正で用いられるアルゴリズムに至る道筋や流れについて講演をいただいた。

また初回セミナーなので、冒頭には若山氏からも、開会の挨拶と学問の菩提となっている数学の意義、諸国の動向を踏まえた話を頂いた。

［Key point］

・行列と行列式には美しい公式達があり、数学的対象だが、計算・情報・AI でも利用される根本概念である。
・随所で“学生への数学教育”にも言及しながら、講演は進められた。

［徳山氏の講演：詳細］

まず、ご自身の略歴紹介があり、東大の数学科で指導教官 岩堀長慶先生のもと、群と表現論と組合せ論で学位を取り、その後、日本IBM東京基礎研究所（1986-99）、東北大学教授（1999-2019）、関西学院大学教授（2019-）という経歴で、今やっている理論は、アルゴリズム理論との紹介であった。尚、徳山氏が博士課程を修了した際、東大の数学科では、博士の学位を取って卒業後企業に行った例はないと言われたとのことであった。

つぎに、本題の行列式の話として、大学生も出来ない問題として三角形の面積や三角錐の体積を求めるための利用される行列式（下記）の紹介があり、

$det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ （3点（-1、5）、（3、7）、（2、15） の作る三角形の面積の説明があった）。

三角錐の体積の計算 4点（0, 0, 0），(1, 1, 0），(0, 1, 1），(1, 1, 0）を頂点とする三角錐の体積の説明もあった。

勿論、三角形の面積は、小学生ならば三角形を囲む長方形の面積を書いて計算する。大学生でも、ほとんどこの方法で解くことが多く、積分で解く人もいるが、大学生ならば知識を使って解いて欲しい（今日のテーマの行列式の利用）。

3点（a, b），(c, d），(0, 0）のなす三角形の面積は $|\frac{1}{2}det\begin{pmatrix} a & b \\ c & d \end{pmatrix}| = \frac{1}{2}|ad - bc|$（高等学校で既習）。

三角錐ならば、行列式

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - afg - bdi - ceg$$（サラスの方法）

を使えばよい。今日は、これら行列式（determinant（直訳 決定式））と計算理論にまつわる話をしたい。

行列式は、n次の正方行列Aに対し、det A=$\sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$，（ただし、$S_n$ は置換群）と定義され、置換の数（置換群の位数、順列の数）は、n! 個あり、定義通りの計算や線形代数で教わる余因子展開は効率的でない（Exponential time in n ）。数理を考えず教則本の通りプログラムすると、いつまでたっても終わらないアルゴリズムになることに気付く。

その中でも、行列の基本変形（ある列（行）の定数倍を他の列（行）に加えても行列式は不変）は、定義に比べると圧倒的に効率的でミスしにくい。尚、学生は講義では習ったが、「なぜ優れているか（計算量の観点）」を教わらなかったので覚えていないことが多い。数学は「美しく」ないといけない学問で、美しい解き方はキチンと教わらないといけない。ここでは、美しい解法＝優れたアルゴリズム（氏の専門）になることが

強調された。

またさらに、「なぜ優れたアルゴリズムが？」の説明がされた（下記）。

行列式の自然な変形と永久式について、転倒数（転位数）inv(σ)を使えば、

det A= $\sum_{\sigma\in S_n} sgn(\sigma)\prod_{i=1}^{n} a_{i,\sigma(i)} = \sum_{\sigma\in S_n}(-1)^{inv(\sigma)}\prod_{i=1}^{n} a_{i,\sigma(i)}$,　S_n は置換え群

となっている。ここで、数学者の習性として（−1）を 変数（tとかq）に変えたい。つまり、

det [t]A= $\sum_{\sigma\in S_n} t^{inv(\sigma)}\prod_{i=1}^{n} a_{i,\sigma(i)}$ のように行列式の「自然」な一般化をする。とくに、

det [1]A= $\sum_{\sigma\in S_n}\prod_{i=1}^{n} a_{i,\sigma(i)}$ は（永久式）になり、アルゴリズム理論や物理モデルなどで重要な概念として知られている。永久式は計算が難しいと信じられている。おそらく、Det[t]（A）が効率的に（多項式時間で）計算できるのは、前述のt=0,−1 のときだけであろう。

それではなぜ、行列式の計算 det[-1]Aは易しく、一般化した式は難しいのか？

というのも、*n* 次の永久式は*n* の多項式のサイズの行列式で表せないという予想がある。結局、行列式と永久式の探求することは、P vs NP問題とも見える。つまりMillennium 問題、行列式 vs 永久式の計算複雑度や群論と関係がある。

また、基本変形は、基本行列の掛け算と同じであるので、群作用による不変性がある。

GCT：群の表現論を用いた計算理論への挑戦として、代数幾何学と群の表現論を用いて、Valiantの問題を解けないか？についてもここで言及された。

戸田による算術回路アルゴリズムと呼ばれる方法もあるが、今回はDodgson のアルゴリズムの紹介（Charles Dodgson は、作家Lewis Carroll のこと）がなされた。 行列式のアルゴリズムはデータ解析やAI（機械学習）はもとより、コンピューター科学全般で利用されている。さらに補足として、行列式で使われる転倒数、一般化（永久式）の道筋も述べられた。優れたアルゴリズム = 美しい解法であることが分かる。ここで、行列式は万能計算機で、多項式などの計算でも利用可能との説明もあった。 Valiant予想（計算理論の主要問題）、「P vs NP」問題との関係もあり、群論と代数、幾何学との関係：GCT（群の表現論）が深い。

因数分解からGCT、計算理論、物理学でも使われる計算の中でもDodgson の凝集法を徳山氏が応用拡張したことも説明された。Vandermonde の行列式と因数分解、Weyl の分母公式とSchur関数の導入は量子

行列式は万能計算機である

- 任意の多変数多項式は、(アファイン線形行列の)行列式で表せる
  - f(x) = det (A(x))
  - x はn個の入力変数、 A(x)は、変数の一次式を成分に持つ行列
- 大きさTの算術回路で計算できる関数…、ほぼTのサイズの行列式として表現できる （Valiant, 1979)
- Valiantの提唱した予想 (計算理論の主要問題の一つ)

「n次の永久式は、nの多項式のサイズ(次数)の行列式としてあらわせない」

- 同値な命題
  - 「計算論的に難しいと信じられている関数(#P完全計算階層の関数)は、多項式サイズの行列式としてあらわせない」
  - 「計算論的に難しいと信じられている関数は算術回路で効率的計算が不可能」
- 現状： だいたい2^n の次数あれば十分(上界)、 $n^2/2$ 以上は必要(下界)
  - 「雲をつかむ」ほど難しい未解決問題。でも、価値は高く、望みはある。

力学の基本的な粒子を表す公式とも関係がある。

$$\begin{bmatrix}1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\\ 1 & x_3 & x_3^2\end{bmatrix}=((x_3-x_2)(x_3-x_1)(x_2-x_1)$$　（Weyl の分母公式）

Gelfand-Tsetlin による、Weylの指標公式のDeformationやDodgson の行列式計算アルゴリズムでは、小行列式（行列式のi 行目からi+k行目、j列目からj+k列目）を取り出した行列 M([i、i+k]; [j、j+k]）に対して帰納的にDesnnanot - Jacobi adjoint matrix theorem

det M det M([2, n-1]; [2, n-1])=det M([2, n]; [2, n])det M([1, n-1]; [1, n-1]-det M([2, n];[1, n-1])

を繰り返す。つまり、大きい行列の計算では一個ずつ大きくしながら計算すれば、ボトムアップで計算すればよい。 Robbins-Rumsey によるアルゴリズム（Desnnanot-Jacobi adjoint matrix の導入）にも発展している。

講演の最後には、

・上記は、離散数学に類する話で、アメリカ（PROMYS でのイベント）の高校生から、「数学サマーキャンプでのテーマ研究で上記の徳山氏の成果を学び、inspireされて新しい成果を得た」というメールが来たとのこと。その数学サマーキャンプ（PROMYS）を例に、アメリカにおける人材育成、活用事例の紹介もなされた。

・PROMYS：CLAY研究所の開催しているボストン大における数学キャンプの活動も説明された。

また、

・柏原クリスタル・λ行列、ASM、Kuperberg、Square Ice、6vertex model

・量子アニーリング：統計物理における量子計算での利用、Ising model

も簡単に言及された。

まとめ

- 行列式と行列式の美しい公式たち：数学の古典的な対象
  - 計算機科学(CS)の進歩で数学が進み、数学が進むとCSも進む
  - 物理の思想と手法：　量子計算だけでなくビッグデータの時代にはとても重要
  - 数理＋計算機科学＋物理＋…　⇒　新しい時代の情報科学技術
- 数学の高度教育は計算機科学・情報技術・AI人材育成に重要です
  - 自分の体験：　数学科で勉強したことが、CSの研究者としてとても役に立つ
    - アルゴリズム理論やデータ解析では、線形代数を熟知しているだけでも大変な強みです
    - 暗号や量子計算やAIなどの話でも、抵抗なく話に加われます
      - 「量子アニーリング？ああ、Isingモデルを解くのね！」 即座に反応できます
- 自分のアイデアが何十年も読み継がれる：　研究者の幸せ
  - 願わくば100年200年：「行列式の凝集」も120年後に再評価された

【質疑応答】

Q：量子計算、量子コンピューターが早いのはなぜか？説明可能か？

A：素因数分解や物理現象、例えばIsing modelの計算などでは説明されている。基本的に量子entanglementという現象が理由である。一般の実用問題で実際にどこまで効率が良いかは数学理論でははっきり解明していない場合も多い。

Q：Natural Intelligence として発展するのではないか、生命現象ではDNA折りたたみ、物理、生命でははっりと完全には解けないが、実験としては非常に上手く行く場合がある。

A：数学も実験科学となっているのではないか。やってみたら上手く行くことがある。

C：我が国では数学の本が売れているが、ライブで話をする強さもある。諸外国の例でも本当にすごい人（A. Wiles など）が一般向けに話をしている。

C：広中平祐先生などが開催しておられる「数理の翼セミナー」が、PROMYSの活動に近いか？

C：数学科出身者が日本の場合、数学者になる。アメリカでは、Google などの道を目指すことも多い。産業界にもすごい研究所がある。

C：カリフォルニア大では、数学専攻人数が6.5倍などになり、数学へのモティベーションが上がっている。日本でもimprove すべき

C：応用数学（数学とコンピューターサイエンスも）を学ぶ大学が少ないように思う。

3.2 社会、産業と最適化

［概要］

CRDSの連続セミナーシリーズ、今回は「社会、産業と最適化」をテーマとして、土谷氏より「最適化の研究の発展」、田辺氏より「最適化を価値にするためには」というタイトルでご講演をいただいた。

土谷氏の講演は最適化理論が「物理・工学的なモデリング、数理的視点、アルゴリズム論」が絡む学際分野であることの説明から始まり、その歴史的経緯から基礎的な解析手法、実用例の紹介へと展開された。

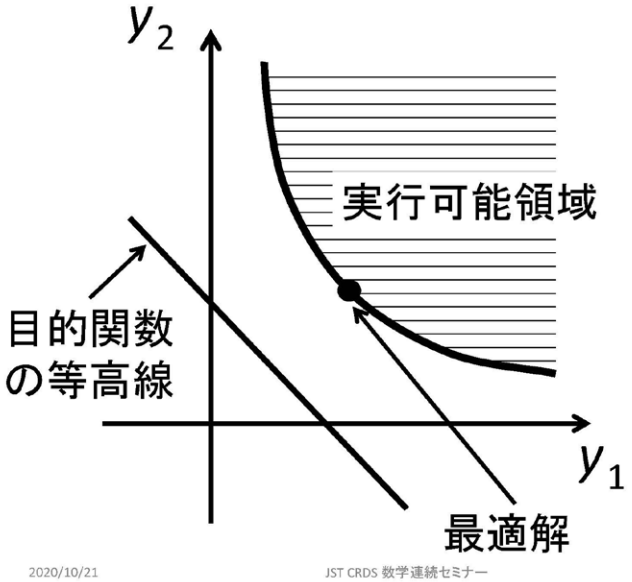
田辺氏の講演はビジネスシーンにおける最適化問題の応用、およびそれを基礎とした技術・ビジネスの発展プロセスに関する話題で構成された。

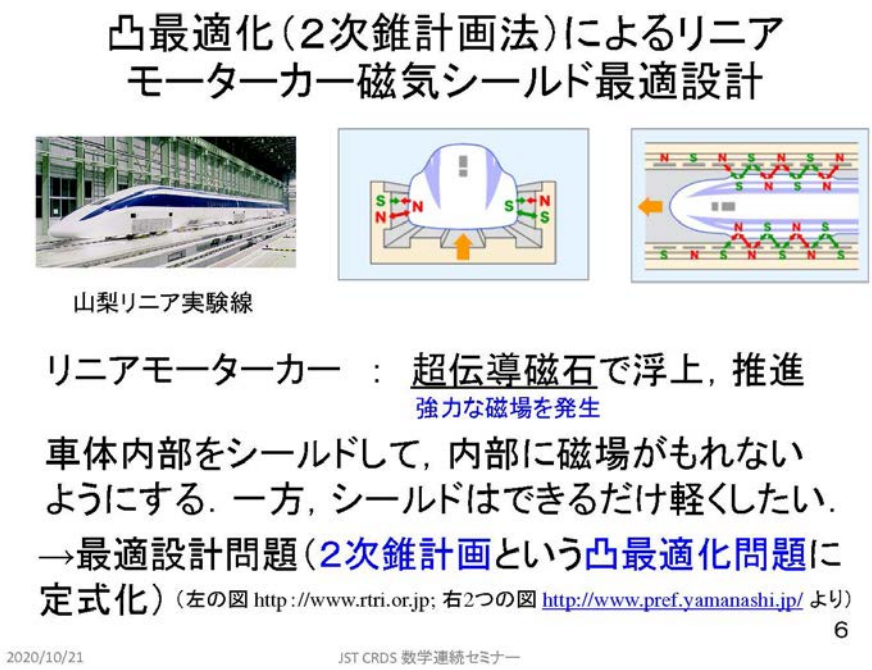
［Key point］

・大きな問いが（ブレイクスルー含め）研究・課題解決の道標となり、原動力となる。（土谷）
・緩やかな人のつながり（たった1日の研究訪問でさえ！）がブレイクスルーを生み出しうる。（土谷）
・実用では、問題の提起からその解法の実装までが求められる。「解けることが知られている」だけでは価値を生まない。ビジネスシーンでは実際に「解いて」初めて価値が生まれる。（田辺）
・「解ける」の意味は様々であり、アカデミアとインダストリでは大きく意味が異なるケースが多い。それを理解した上で課題に取り組むことが大事。（田辺）
・数学では解の「存在」が保証されれば「解ける」と言える向きもある（そもそも「考える意味がある」かどうかに重点を置く）が、実用では「常に実用的時間で具体的データを伴った答えが出る」、またアルゴリズムも「動いて当然」というレベルまできて「解ける」と言える。
・社会人はマラソンランナーのようなもの。方法論やモデルは細分化・進化され、それらを積み上げ続けることで「価値」が出来上がる。これはある意味、「攻略法を積み上げる」プロセス。（田辺）

［土谷氏講演：詳細］

最適化理論の始まりはコンピュータによる計算が本格利用されたのと機を一にするが、1970年代に「計算





複雑度」の概念が提唱されたことがさらなる飛躍を生んだ。特に、現実的な時間、詳細には「多項式時間で解ける問題」の研究が軸となって分野が進展した。このクラスの問題として特に、線形計画問題、凸2次計画問題、半正値計画問題が重要な研究対象となってきた。これらは数学的には簡潔に記述できるものであるが、数理・アルゴリズム分野における深い結果を生み出し、モデリングにおける幅広い応用に直結しているという意味で数理科学の一つの「良い公理」の例となっている。

上記の問題そのものは多くの分野で用いられる基礎的なものではあるが、ブレイクスルーとして圧縮センシングとスパースモデリング（線形計画より派生）、代数幾何学との融合課題などがあり、理論・実用の両面で大きな広がりを見せている。

歴史的には問題解決のアプローチとして線形計画問題の解法として用いられる「内点法」を基礎とし、内点法は、目的関数が2次関数である凸2次計画問題、線形計画問題の行列版とも言える半正値計画問題の解法に拡張された。

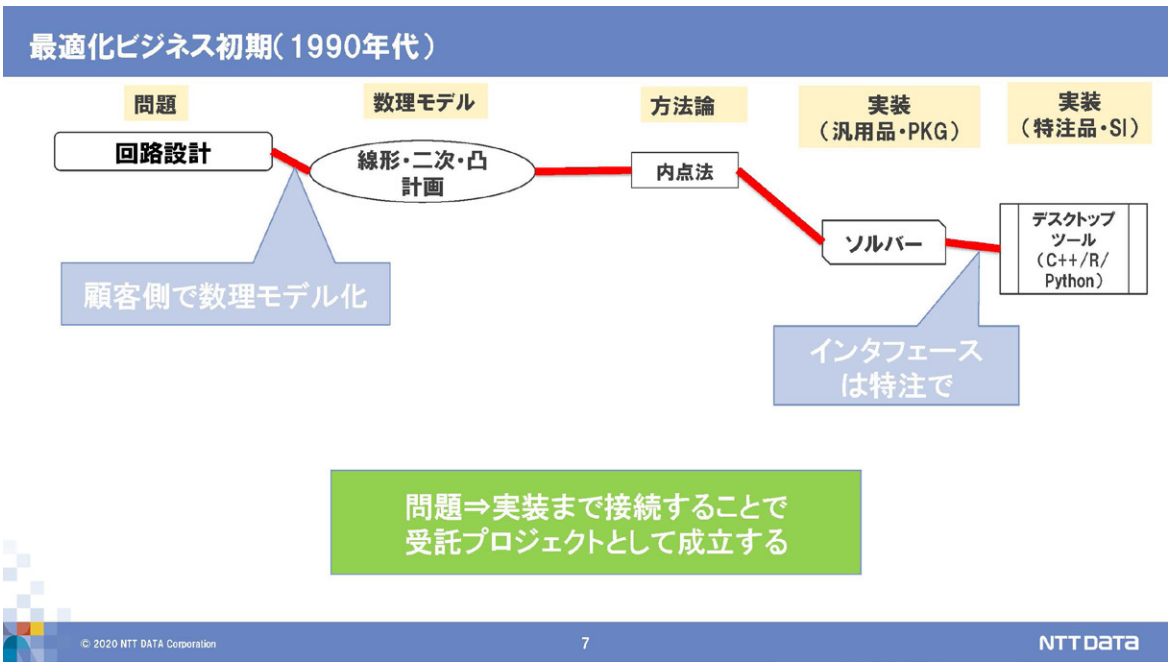
ポートフォリオ最適化、サポートベクターマシン（凸2次計画より派生）の開発などでは、内点法が一つの発展の契機となった。機械学習を始めとしたデータサイエンスの問題でも、計算複雑度の解析に裏打ちされた最適化アルゴリズムは重要な役割を果たしている。

氏自身の研究例として、Jordan代数を援用した「2次錐計画問題」とそのリニアモーターカー磁気シールド設計への応用、内点法におけるアルゴリズムの反復と最適化問題の「曲がり方」を対応させた「最適化の情報幾何」、最適化問題の主問題と双対問題の背後にある悪条件性の構造の「代数幾何学」（Tarski-Seidenbergの定理を基礎とする）による解明、メソポタミア文明における古代都市ヌジ社会の解析など、非常に多岐にわたり紹介された。

実用応用もさることながら、問題自身の数理的意味を問うものも含まれ、最適化問題というトピックの広がりと深みの両方を感じさせるものであった。

［田辺氏講演：詳細］

最適化手法を応用したビジネスの初期段階から「問題から実装までの課題解決フローを接続する」ことを基本姿勢としている。これを実現することで、受託プロジェクトとして成立する。

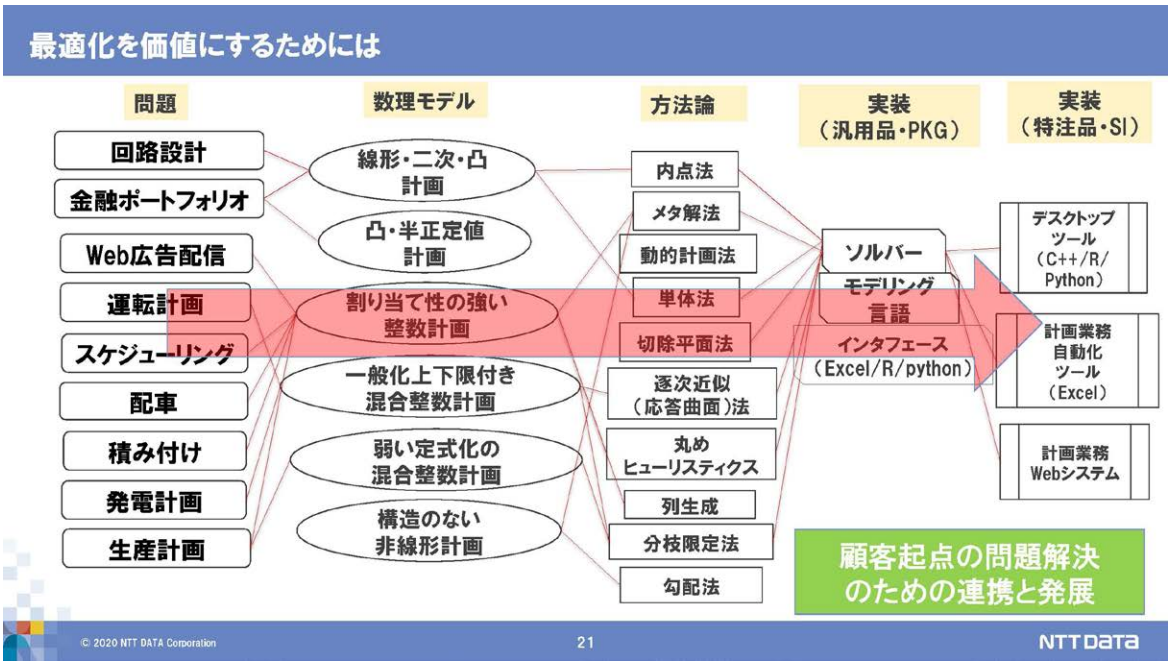


初期段階では顧客側に数理モデリングの素養がある人材が在籍しており、それに手持ちの方法論（非線形最適化向けの内点法を軸とした最適化手法）を組み合わせ、課題解決につなげた。依頼を受けることに伴うノウハウは蓄積される（回路最適化、金融ポートフォリオ最適化の解法の提供など）。

同社では内点法を軸とした最適化ソフトとして「NUOPT（現、Numerical Optimizer）」を提供し始めた。

しかしここで「整数計画問題の壁」にあたる（2000年代）。多種多様な問題が持ち込まれたが、現実規模で「解ける」問題の幅が狭かった。それは「整数計画問題」を攻略するモデリングとアルゴリズムのノウハウが不足していたためである。

そこで内点法の他に単体法、分枝限定法、メタヒューリスティクスなどさらに多くのアルゴリズムを拡充し、よりタイトな（線形緩和で元の問題の構造が保たれやすい）数理モデル作りのノウハウの蓄積など「方法論・



モデルの多様化・深化」を行っていく。価値のアンカーとなる「実装」も、ソフトウェアシステムをC++, R, Python, Excelなど多様なものへ対応させ、「出口」の拡充も行ってきた。

最適化技術はこのような「数理モデル」「方法論」「実装」の接続、相互作用による発展が鍵となる。

特に最適化技術をビジネス上の「価値」にするには、以上の接続を通した顧客起点の問題解決のための連携と発展が重要となる。

さらに最適化の理論は「制約・目的関数の設定が最適解を決めている」という視点を提供していることも見逃せない。最適化モデリングは現実問題に存在するリソースの制約と目的関数の設定を見直し、我々が「最適化（できるだけ良くしたい）」と漠然ととらえていることの意味を問い直す可能性を持っている。最適化技術は「不定形の計画業務」という人間に残された「聖域」における数理科学の価値を生み出すことに貢献し続けることを確信している。

【質疑応答】

Q：研究と実用のタイムスパンの違い（例：問題を「解ける」ように持っていくまでの所用期間）

A：お試し期間が3ヶ月。2週間に1回のペースでブラッシュアップを重ね、これで見込みがなさそうであればご破算。うまくいきそうであれば、半年で完成・運用開始まで持っていく。
これに限らず、「実用思考」は研究活動には向かない（逆も然りか）。時間の使い方も含めて、考え方や取り組む姿勢の大きな違いは意識すべき。

Q：ユーザーはどうやって技術開発・提供を行っている企業（今回はNTT数理データシステム）を知るのか。

A：定期的なセミナーの実施及び宣伝、あとは事例の紹介。これを頻繁にすること。

Q：依頼を受ける側は、依頼する側（ドメイン）の知識をどのように取り入れるのか？

A：経験とパス（様々な話題の間の繋がり）の多さが決め手となる。別の課題でノウハウを貯めて、それを生かせる時に前面に出す。引き出しの多さが勝負で、それは学術的知識の深さよりも顧客の幅が断然有利に働く場合も少なくない。

Q：人との繋がりにつき。JSTのさきがけ・CRESTのアドバイザーはこれに資するものか？

A：さきがけは領域合宿があり、若い人たちが互いに知り合う良い機会としてうまく機能していると思う。CRESTはどちらかというと集中投資で、まだ工夫の余地がある。「分野を超えた繋がり」を作ってほしい。

3.3 量子情報処理

［概要］

CRDSの連続セミナーシリーズ、第3回は「量子情報処理」をテーマとして、富田氏より「光量子技術 - 数学を具現化する技術と実装を救う数学」、萩原氏より「量子情報処理におけるイノベーション符号化 - とくに数学の視点から」というタイトルでご講演いただいた。

富田氏の講演は「光量子技術」で想定される応用分野の紹介に始まり、量子計算は萌芽の段階にあるとの話があった。量子力学の導入から、量子状態と量子ビットの説明があり、観測の結果起こる量子力学の重ね合わせ状態の収縮や連続的に起こるデータ誤りの訂正の意味が説明された。 Google による、大規模な量子計算の試みも紹介された。光による量子計算の実現によるメリット、デメリットや複素振幅を利用した計算が説明され、最後に、光コンピューターの展望について、光情報通信、量子情報伝達やネットワークに組み込む応用も紹介された。

萩原氏の講演は「量子情報処理」に必要な符号化の理論の紹介に始まり、産業や情報通信社会の基盤技術としての誤り訂正の紹介があった。量子情報、量子計算、量子暗号を見据えた量子符号、量子誤り訂正のイノベーションが望まれている現状が紹介された。

［Key point］

・大規模量子計算では、量子状態に生じる誤りが伝搬しない仕組みが必要。（富田）
・光コンピューターにはメリットも多いが、克服すべきデメリットもある。（富田）
・データのコピーや伝送では、量子力学的な効果によって量子誤りが発生する。（萩原）
・量子情報に対応した符号理論、つまり量子誤りに対処するための量子符号が必要 。（萩原）

［富田氏講演：詳細］

量子情報技術は実用化を目指し盛んに研究されている。ただし現状では、実用化には遠いことも強調された。

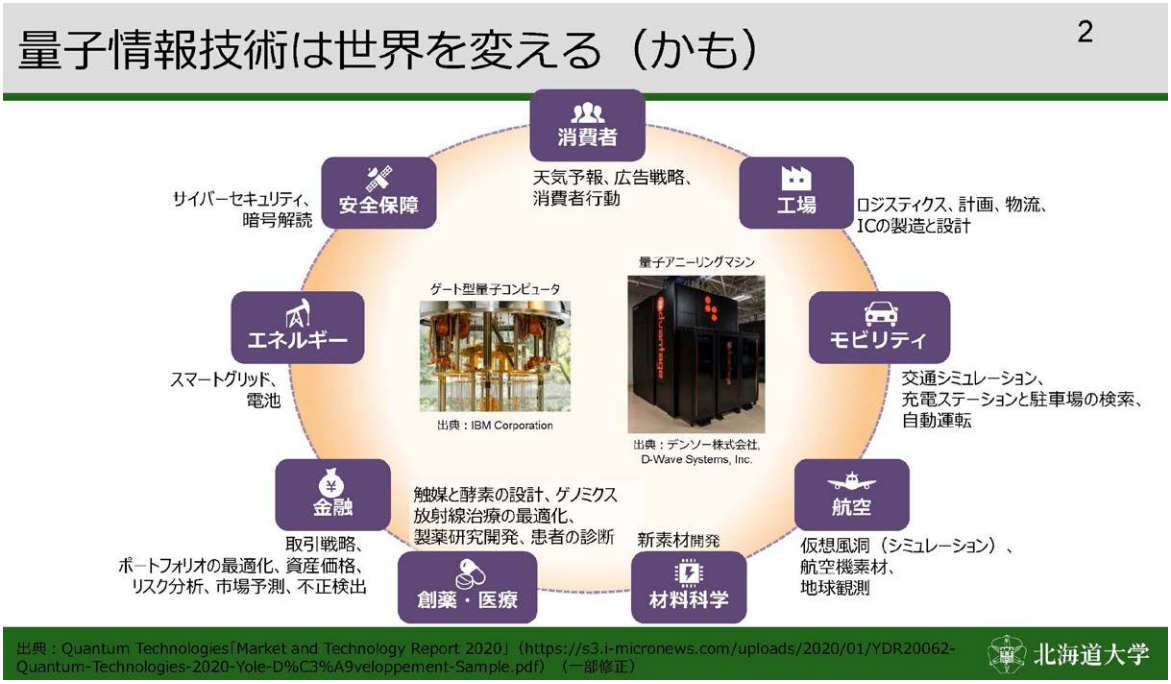
次に、数学と物理学の関係から始めて、量子状態の基礎概念と原理が紹介された。とくに量子状態はケットベクトル $|\psi>$ として表される。測定が行われると量子状態はエルミート作用素の固有空間 $|a_i>$ に射影されて、さらに射影が起こる確率は、固有状態のベクトルとの内積の絶対値で表されている。つまり、

$$|<\psi|\,a_i>|^2$$

となりこれを確率解釈と呼ぶ。結局、測定値が得られたときに、 $|\psi>$ という状態が固有空間 $|a_i>$ に変化する。この変化を状態の収縮と呼ぶ。また測定を行わないときは、状態の変化はユニタリー変換に限る。量子状態はベクトルであるから、量子状態には重ね合わせの原理が成り立ち、$|0>$ と $|1>$ の和からなる二次元の状態は、複素数値 α、βを用いて $\alpha|0>+\beta|1>$ と表せる。これをqubit（量子ビットの単位）という。さらにn粒子の状態は、テンソル積 $|0>\otimes|0>\otimes\cdots\otimes|0>$ から $|1>\otimes|1>\otimes\cdots\otimes|1>$ の2^n個のベクトルの線形和

$$a_1|0\cdots0>+a_{2^n}|1\cdots1>$$

で表される。このn個の粒子で表される状態がn qubit（n量子ビット）状態になる。このベクトルの表す状態が2のn乗個あることが重要である。結局、量子ビットを1ビット増やすごとに指数的に扱える計算量が増え、計算処理が早くなることが期待される。なお、量子ビットはテンソル積の空間のベクトルの線形和で表されており、量子もつれと言われる状態が発生する。重要なことは、Aの状態が0でBの状態も0、さらにAの状態



が1でBの状態が1になるようにお互いが相関を持った状態を作れることにある。つまり、Aの状態と、Bの状態に関連付けられる量子状態

$$\frac{1}{\sqrt{2}}|0>_A\otimes|0>_B+\frac{1}{\sqrt{2}}|1>_A\otimes|1>_B$$

も作ることが出来る。この量子状態は2量子ビットゲートで実現される。

なお歴史的には、量子ビットを利用する量子状態の利用は、1990年代Feynmanなどが提唱し、IBMの科学者Rolf Landauerが原理的な問題を論じている。つまり、重ね合わせの線形結合の係数が連続的に変化するため、利用したシステムはアナログコンピューターになっており、無限精度であれば良いが、精度が悪かったら上手く働かないのではないか？ 量子状態が環境と相互作用することによって重ね合わせ状態が壊れてしまうのではないか？さらに、アナログなデータなので、デジタルだったら誤り訂正が出来るけども、アナログでは出来ないのではないか？といった問題が提起されている。

上記、誤り訂正について、具体的には、観測測定をすると射影されて状態が収縮してしまう、量子計算の量子情報処理で誤りの検出はどうしたらいいか？さらに量子状態の係数は連続的に変化してしまう。こういった誤りの訂正は、そもそもどうやって行えるか？という疑問が説明された。

このような疑問に対して、すでにGottesmanの博士論文（1997）で理論的な概念が提唱されている。つまり、測定によって重ね合わせが壊れなければ良い。具体的には、想定される量子ビットの状態を固有状態にもつ演算子で測定すればよい。さらに、誤りがない状態と誤りがある状態を区別出来るように、正しい状態は固有値1エラーがある時は固有値-1を持つような作用素を作ればよい。測定によって有限個の誤りパターンに対応する状態のどれかに収縮させる。結局、連続的な誤りは離散的な誤りに変えられる。

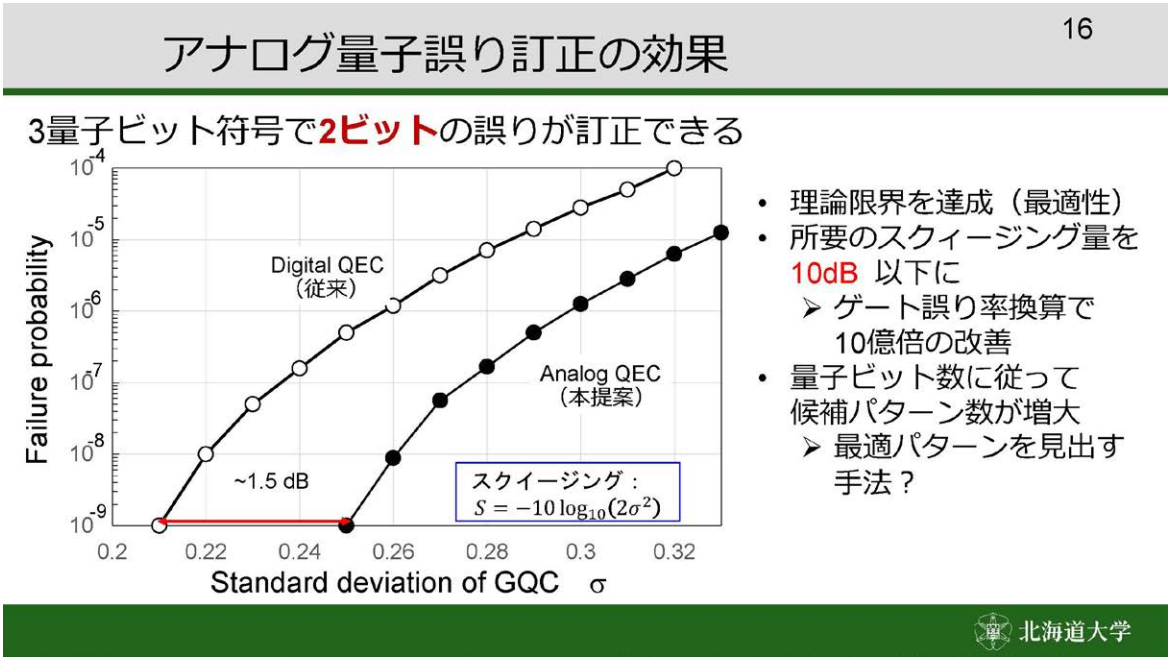
さらに、大規模量子計算に向けて、ゲート操作の誤り率が一定率以下なら、どんなに長い量子計算も可能になる閾値定理があることも紹介された。1996年に発表された結果では、ゲートの誤り率が100万分の1以下と非常に小さければ、どんなに長い量子計算も可能になる。その後の理論家の工夫の結果、ゲート誤り率は1-2%くらいまで許容範囲が上がっている。

一方実験では、Googleが2014年に9 量子ビット集積した量子回路で1量子ゲート0.1%の誤り率で実行できるということを示し、さらに2019年には53 量子ビット集積回路において、2量子ゲートで0.36%のエラーまでできている。

すでに上述のエラー限界を超えているので、理論上はGoogleの量子ゲートを使えばどんな長い量子計算もできるコンピューターを作れるはずというところまで来ている。 ただし、大規模量子コンピューターを実現するのは非常に難しく、実用上意味のある計算をしようと思うと100万ぐらいの量子ビットを使わないといけないため、この53量子ビットでは全然足りていない。

富田氏は光学を研究しており、できれば光で量子計算を行いたい旨の説明があった。光の良い点として、光は光通信のように情報伝送に向いており、伝送の損失が小さい。 また、環境との相互作用が小さく、偏光を使うと量子ビットの実装も容易で、1ビットゲートの実装も簡単にできる。良いところも多いが、大規模情報処理のための、2量子ビットゲートを作るのは難しく、確率的にしか実現出来ない。また、損失を補償するためのアンプは使えない。普通の光通信ならば弱った信号を増幅し、強くする技術があるが 量子計算には使えない事情がある。また、波長が1ミクロン程度なので、素子はそれよりも大きくならざるを得ない。さらに光は飛んでいってしまうので、メモリーも作りにくい。共振器を使ったメモリーを用いて、2016年に実験では10 量子ビットの光量子コンピューターが実現されている現状が紹介された。

さらに偏光を利用するのではなく、光電場の複素振幅（位相、振幅）を用いた光を使った量子計算も考えられている。複素振幅の二つの自由度が位置と運動量に対応した関係になり、位置演算子、運動量演算子の類似となっている。複素振幅を用いると透過率と反射率が50%の半透鏡で単に光を混ぜるだけで足し算と引き算ができ、実際、東大のグループなどで量子ゲートが実現されている。スクィーズド光源を用いて、鏡で光を混ぜ、ディレイをおいて、また混ぜていくことをすると、100万モードを使った量子もつれまでできている。Google では 量子ビットが53とか50、100のレベルだが、すでに100万モード（量子ビットではないが…）になっている。問題点もあり、複素振幅はアナログ量なので、アナログ的な誤りが蓄積されてしまい、量子誤り訂正ができず、大規模計算が出来ない。量子誤り訂正を行うためには、アナログ量をデジタル化すればよいことになる。そのために、例えばGKP量子ビットのように、とびとびの振幅の値だけが0でないような状態を作る。 GKP量子ビットは半透鏡を使った量子ゲートで演算を行うことができる。量子ビットになっているため量子誤り訂正することもできる。このため、大規模量子計算ができることになる。 GKP量子ビットを用いた量子コンピューターで問題になっているのは、実際にはとびとびの値とした振幅には幅があり、それがエラーを引き起こすことである。また、とびとびの値の状態を無限に重ね合わせることはエネルギー的にできない。重ね合わせの数が有限であるということによってもエラーがおきる。これらのエラーをなくすためには、頑張っ



てスクィージングを強めて、光振幅の取りうる幅を狭くすることになる。

富田氏のグループではGKP量子ビットのアナログ性を逆に活用することで、量子誤り訂正の性能が向上できることを示した。実際、3ビットの符号で2ビットの誤りを訂正することができる。また、許容されるゲート誤り率を10億倍にできるという改善が見られた。この方法は物理的な考察に基づくものであるが、誤り訂正の性能をちゃんと最適化して求める手法には、数学の助けが必要との説明であった。

最後に、光量子コンピューターは自然に複素振幅を利用出来、室温動作、ネットワークとも親和性といった利点があることが指摘された。実際の実験機材の画像も提示された。さらに海外では実際、光チップを使って、ベンチャー企業が、光を使った量子コンピューターを作ろうとしているとの現状の紹介もあった。

［萩原氏講演：詳細］

経歴紹介から始まり、群論や表現論を学んだあとに符号理論、離散数学を専門にしているとの話があった。自身の現在の活動では、日本数学会、アメリカ数学会、IEEE IT Soc.や電子情報通信学会の活動とInternational Symposium on Information Theory and its Applicationの実行委員長や電子情報通信学会（電子広報）、同学会SITA（幹事）としての活動も紹介されていた。本講演では数学を重点とした話を紹介する旨の導入があった。

氏の研究では、量子誤り訂正符号、符号の濃度への数論的手法の利用、ガウス二項係数の符号理論への応用がある。加えて、量子符号の導入として（古典）符号の説明があった。導入として、広義の符号としての暗号、データ圧縮、署名を挙げた。また、狭義の符号としての、聞き取りミスや書き間違いなどのミスが生じたときに、元のデータを推測するために利用する誤り訂正符号の説明を行った。

符号訂正は情報通信社会の基盤技術として、既に一般的に使われており、携帯電話、DVD、ブルーレイ、ハードディスク、ＵＳＢメモリーやテレビのデータ受信の例もある。 すでに実装されている符号理論を使うことで、快適な通信や記録が出来ることが強調された。

さらに、誤りの起こる要因の説明としCDの500倍の拡大写真が紹介された。拡大写真ではCDの溝の顕微鏡写真の様子が表され、溝のあるなしに対応して、それぞれ1,0に対応している様子の説明があった。 溝には小さな傷があったり、CDに髪の毛や埃が載ったりするとほぼデータが読めなく、データが変化する様子も推測され、さらに1秒間にCDの場合だと溝を430万個ぐらい読み続けており、さすがに読み間違いが発生

する様子が推察される。

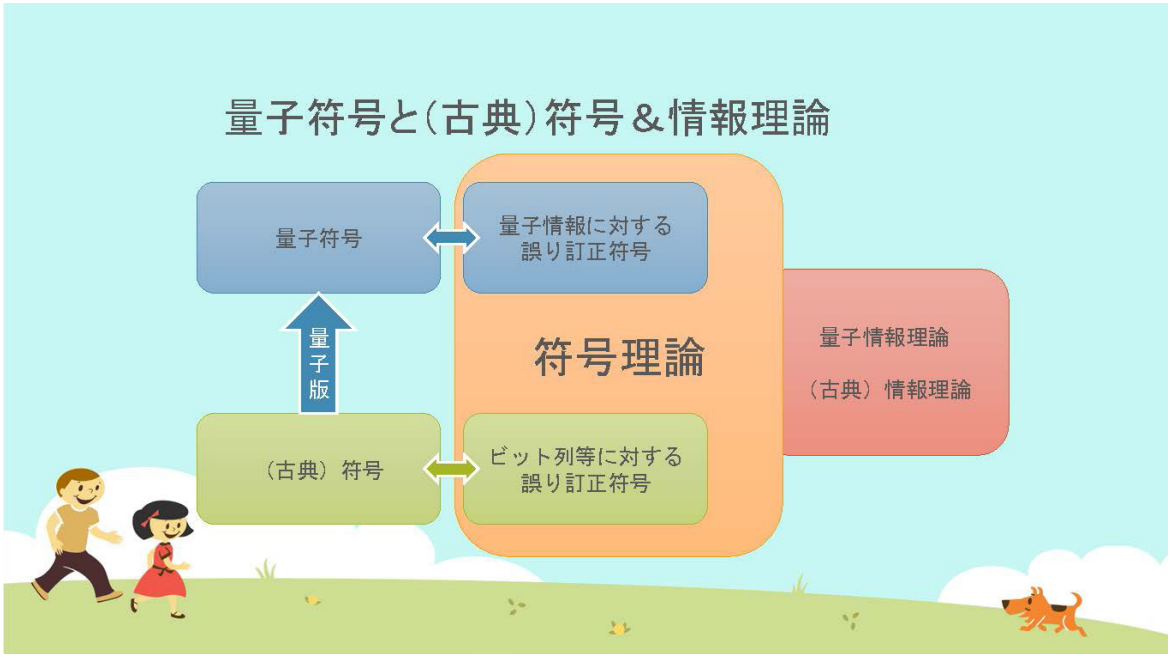
勿論、既に符号訂正は実用化出来てはいるものの、新しい技術が必要とされている。例えばハードディスクの容量は増えたが、それは、溝が狭くなりもっとたくさん書けるようにしている。溝が小さくなって幅が狭くなったらエラーも乗りやすくなるので、そこで強烈な誤り訂正符号を発明することでバランスが取れており、信頼性が高くて質の高い符号によるデータ処理ができるようになっている。 結局、エラー発生率が一段と高くなっている実情も見て取れた。

例として、画像とともに誤り訂正符号に使うデータも載せ、元データの白を0、黒を1で画像を表したものを挙げた。画像に描かれているデータがあったときに、符号化という処理で、冗長なデータを付随する。 この冗長なデータが、元画像のヒントになる。エラーが載ってしまって、白が黒に黒が白になることを考える。誤り訂正ではうまく処理を使うと、最後は元のデータまで戻せる。 色々な技術を使って、つまり、元の画像に対して余計なデータをつけることで、それがヒントになって正しいデータが読めるように工夫をする。

学術的な価値としては、日本国際賞、京都賞などの受賞者もおり、アカデミックでもかなり高い評価がある。1990年には、ウエスレイ・ピーターソン博士が符号理論の研究について日本国際賞を受けており、30年後の2020年にもロバート・ギャラガー博士が発明した符号が世界を変えたとも言われていて、賞が与えられた。他にも京都賞では、情報理論とか符号理論の発明の父シャノンが京都賞の第1回の受賞者であり、かなりの高い評価を受けている。シャノンの業績は、数理科学と純粋数学を含む分野の賞になっている。

量子符号は、古典符号が古典情報の誤り訂正符号に対応するように、量子情報の誤り訂正符号である。つまり情報理論では、古典符号の量子版を考えている。量子版は、富田氏が紹介した量子力学を考えるときに、古典符号を使ったデータを扱い直すものになっている。つまり量子情報に対する誤り訂正符号であって、それは古典のビット列に対する誤り訂正符号の量子版になる。これ全体を使う理論として符号理論があって、一部が量子符号と言える。

どんな数学を利用するかということについては、最先端の数学よりも、既存の結果を利用しやすい。例えば、距離空間、有限体の線形代数、計算量の理論、エントロピーやグラフ理論を理解することが大切である。日本国際賞のギャラガーはそのグラフ理論ベースに、とても計算量が低く、扱いやすくて実用的な符号をつくれることを発見しており、2000年から2010年くらいまでは、符号理論の論文にはグラフの言葉が沢山出てくる時代もあった。 数学的なこと、それ以上にいろんな特別な符号、ハミング符号、リードソロモン符号、LDPC



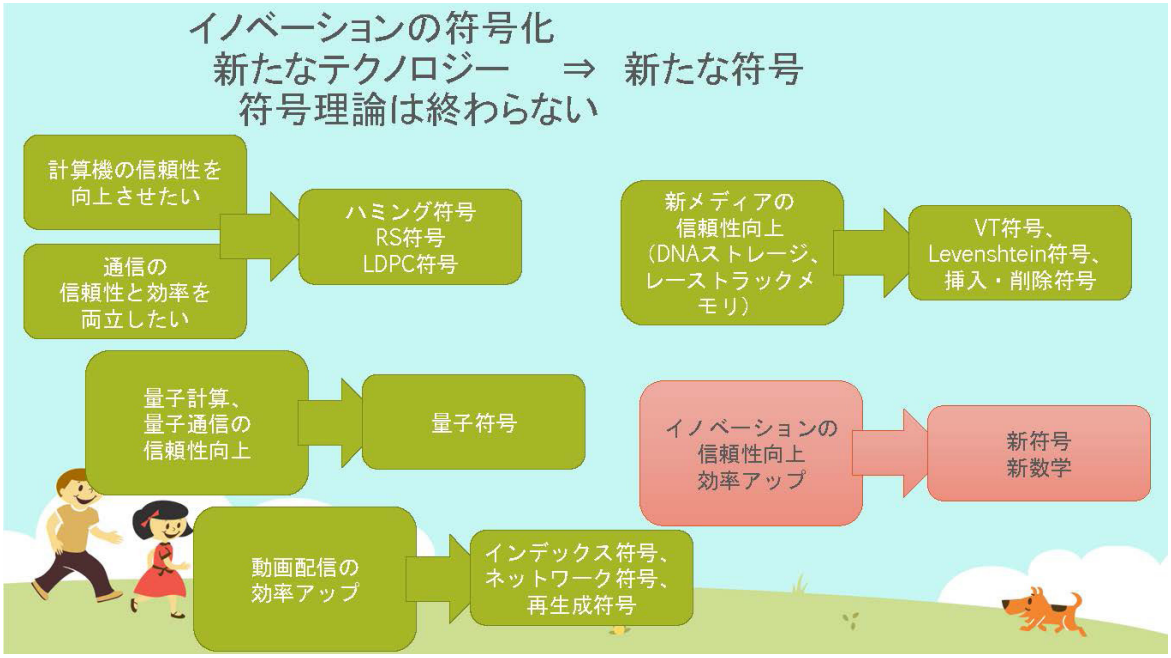
符号など、個別な理論がいろいろたくさん出てくるので、幅広く色々理解するということも大切になる。

符号理論では、一般の符号の誤り訂正の計算量が指数関数的に大きくなることが知られている。小さいデータだったら易しいが、大きいデータに対して、効率の良い符号を作っていこうとすると計算量的に破綻することが知られている。 そこで、計算量をいかに理解できるかというのも、符号理論を研究するために必要となる。確率の視点で、研究されることも多い。符号を研究するときには、それが何種類あるか？という問題もあり、そこに組み合わせのテクニックを使う機会もある。

代数幾何学を使った符号ではリーマンロッホの定理とかも大切である。これらのことは一般的な符号理論の教科書に書かれている。実際、符号理論の本は定義・定理・証明のスタイルをとっており、数学の教科書と変わらない。一瞬、実用的な話なので、工学的な実験の話とか、たくさん書かれているような錯覚を受けるかもしれないが、実際には、ほぼ数学の本を読んでいるのと変わらない。量子版になるとどうなるかというと、富田氏の話では複素ベクトル空間であったり、テンソル空間であったり、ユニタリー変換などの知識が必須となる。古典の符号を前提とし、それの量子版を考えるとかなり色々なヒントが出てくる。古典符号をやった上で、数学の素養も身に付いていると、量子誤り訂正を研究しやすい。

量子符号を含め、古典符号理論、情報理論などが持続可能な研究領域になるためには、電子情報通信系と数学が総合発展していくのは良い筋道である。 そのうち電子情報通信学会で数学の発表が、数学会の方で情報の発表を堂々とできる時代が来ると、総合発展がどんどんできるのではないかと思ってたる。こういった、境界研究的なことをやろうとすると、どうしても自分の専門の話に話を持っていきたくなる。実際、符号理論では有限単純群の分類理論であったりは、様々な符号の構成に使われている。ただし、自分の分野に立ち止まらず、新数学、新情報の領域を目指すと、持続可能な発展が見込めるのではないかと思っている。

持続可能な領域にするためには、特定の技術の符号にとどめず、イノベーションの符号化ととらえる視点が大切である。新たなテクノロジーは新たな符号を生み出すという視点が良い。計算機の信頼性、通信の効率性などを向上させる視点からハミング符号、RS（リードソロモン）符号、LDPC符号などの有名な符号が生まれている。 同じように量子計算の信頼性向上を動機として量子符号が生まれた。 最近ではDNAにデータを書き込むDNAストレージだったり、次世代メモリーと呼ばれているものに対して、非常に面白い符号がこの4、5年から10年くらいで誕生し、研究が進んでいたりする。動画の配信の効率アップのためにインデックス符号、ネットワーク符号や再生成符号という新しい分野も生まれている。イノベーションは、社会の発展に



伴い、常に生まれ続けてくるものだと思われる。それが新しい符号、新しい数学の発見につながる、というふうに捉えれば持続可能な研究となる。

氏の研究では、LDPC符号を応用した量子QC-LDPC符号を作った。インパクトの高い仕事として評価されている。QC-LDPC符号の枠組みで初の量子版を氏がつくり2007年に発表したとの旨であった。量子削除符号についても言及された。富田氏の話には出てこなかったデータ長が短くなってしまうエラーで、ここ数年、情報系で注目されている削除誤りは、通信上の情報のロス、データが届かないときのモデルである。そういうエラーは、DNAストレージと次世代メモリーの自然なモデルと言われていて、さらなる応用が期待されている。非常に理論が複雑であるが、上手く行けば富田氏も話されたような誤りも含む一般的な誤りを全て訂正できるような凄い符号が作れるようになる。これの初の量子版が誕生したのは最近の萩原氏の業績である。

まとめとして、数学と符号理論は互いに支えあい強化しあう存在である。符号理論はイノベーションの信頼と効率を向上するための理論であり、イノベーションとして量子情報を考えれば、量子符号が出てくるのがひとつの自然な見方になる。量子計算通信のバージョンアップにも非常に重要な役割を持っていて、またイノベーションのええ視点でいろいろ観ていくことで、数学と符号理論の発展に寄与できると考えている。

【質疑応答】

Q：理論と実験がうまくかみ合っているように見えるが、物理か数学どちらの問題と考えれば良いか。

A：両方あると思う。物理の問題としては、大きなノイズを小さくできるかが問題にされる。普通の半導体よりもはるかに改造が必要である。 100万ビットともなると、かなり巨大な運動場ぐらいのスペースを必要としたり、かなり巨大なあの冷凍庫を使わないといけない。

一方、非常に大規模なものができなくても、うまく分割することや肝心なところだけを利用し、うまく組み合わせることによって、意味のある計算をするとか、そういう計算テクニックの話もある。さらに、もっといいゲートの誤り率向上を考えて、量子ビットを大きくして頂けると良い。これらは数学的研究だと思う。（富田）

Q：国外の動向は？

A：surfaceコードというのが出てきて、かなり良く実装もしやすく、いいものだということになっている。

より良いフォールトトレラント計算は、国内でも藤井先生、東芝の後藤氏も最近、研究されている。（富田）

Q：光を使った量子コンピューターの話で量子誤りなんかも含めて、古澤 氏、NTT山本氏と富田氏の研究の位置関係は、古澤氏との研究の関係はどのようなものか。

A：今回の話で１００万ノードの生成の結果は古澤氏の結果。光と超電導では研究の状況が違う。（富田）

Q：surfaceコードのことについて、半導体の場合は２次元に実装する。光の場合にも、誤り訂正の話をされたが、surfaceコードと違う何か新しい数学的なモデルあるのか？また違ったものが展開されているのか？

A：大きい物を作って誤り訂正をのっけていく方向にはなっている。大きいものを作って誤り訂正を実装し、誤り訂正も同時にやっていこうというゲートを実現しようという方向になっている。 誤り訂正やるためのハードルは１部成功しているが、さらに洗練させて、量子ゲートを乗せて制御回路付とかをやろうとはしている。光の場合は、一応出来る。 超電導の量子ビットの場合などでは、量子回路ではゲートを順番に並べて行くタイプで実装しているものが多い。光の場合はそれほど２次元ゲートが得意ではなく、むしろ１量子ビットのゲートや測定の方が得意なので、まず最初にたくさんのものを作っておき、それに合わせて測定と１ビットの変換をする。測定を使った量子計算。（富田）

C：超伝導の量子ビットの展望は、大きくしていくに伴って制御系が重要。ノイズが増えていったり、現在は、本当に量子ビット１個１個非常に細かい調整をしている。

Q：今日は光のお話だったと思うんですけれども、他にもやり方があるという事ですけども、方式が違う場合でも、そのプログラムの仕方ソフトウェアは同じようにできるのか？

A：もちろんアルゴリズムの段階は一緒です。機械語レベルでは、つまりゲート操作するところは、それぞれの実装の仕方によって全く違って、量子ビットを物理的に動かすレイヤーがあって、抽象化して、その論理的なビットができるようになる。（富田）

Q：つまり、どこかで区切って、そこから下の物理的なレイヤーの制御は実装によって違ったことをもちろんしていて、その上の抽象化したレベルでは同じになる。

A：どこかの段階で共通化した形で万能プログラムができるはず。（富田）

Q：誤り訂正では古典版と量子版の差異がどれくらいあるのか。 量子情報には双方向性や古典ではない誤りがある、量子計算であればとくに計算中も誤りが発生する。レジスターの状態を戻すために、少し条件が違う話は、どのように拡張されて量子版となっているのか。

A：繰り返し符号の例ならば、0というデータを書くのを繰り返し、例えば1というのを111と書く。 この時に、もしエラーが起きて、例えば101というのを受診した場合に、元は0だったか？1か？と聞かれると、おそらくなんとなく雰囲気で1に似ているので、元は１個だけずれて1だった、と推測するのが一般的だと思います。量子誤り訂正の恐ろしいのは、受け取ったものは分からないが、元は0か1かデータが間違っているかもしれないというような、とんでもない問題がまず最初に出ている。というわけで、受け取ったものがなんだか知る由もないのに、元のデータを探ってくださいということです。エラーの種類が普通に0と１が変わるだけでなく、さらに恐ろしいのは、マイナスの111という状態もあり得る。結局、複素数係数だったら、いろいろ変化しうる。そういったところが、量子情報量、誤り訂正と古典符号の大きな違いになっていると思います。ですので、観測というのを使って、データを破壊しない限り、破壊しない範囲で許される量子力学的に許される。その操作で0と１のヒントになるものというところから元のデータを探すことにする。ここが、大きな違いになる。（萩原）

Q：量子特有の情報もうまく使って対応するという感じですか。測定もそうでしょうし。 ある種特殊の量子特有の事情をうまく使って、符号を作ることになる。

A：データを何回もコピーしても良いなら誤り訂正は簡単だが、それが最初から許されない量子では無理であるところを意識するということが大切だと思う。（萩原）

Q：本来、無限次元の現象のような気がするのですが、Hilbert空間を有限次元的に全部扱っているようにみえる。有限次元の代数的構造で追えばよいという ことでしょうか

A：物理系の方と理論の方、うまくバランス合わせてるという。それであえてちゃんと有限の意味がある世界に落とし込んでいる。（萩原）

Q：量子群の表現という形で符号というのを研究するということはありえる方向でしょうか？

A：特別な非常にシンプルな表現をうまく考察して行くと、そのデータが短くなるような誤りとか、逆にデータが増える誤りというのを、最近ちょっと行ってみました、といった感じのものでしたらある。もっと複雑な数学的な構造の方から、何かが出てくる可能性は充分にある。そこで、問題はそこに固執するのか、その実用的な意味があるから役に立つものだというところまで落とし込めるのか、いろいろその後の戦いがたくさんあると思うので、そこのバランスで、いろいろ考えることになるのではないかと思う。（萩原）

C：情報理論と関係の深い話と、数学で言うと、コンヌの埋め込み予想というのがあって、それが否定的に解決されているということが非常に深く関係しているということが話題になったと聞きました。 群そのものを考えるというより、作用素環論的なアプローチであの符号ができていくというのもありかなと思い、質問しました。

Q：例えば量子の符号で良い符号というときに指標というか、何を基準に良いって言っているのか。 いろんな限界もあると思うが、それに対して量子コンピューターのところだと、もっと実装性とか、ゲート操作だけで実装できる、物理ビット数が少ないに越したことはないが、どういう評価をされているっていうのを聞きたい。

A：鋭い質問だと思う。車のハンドルで言うと、その遊びの部分であるような感じがある。良い方法というのは、まあそうなったら面白い数学程度の言葉にとどめるのが、かえってその理論を発展させているところなのではないかと考えている。シャノンの業績は、通信効率と信頼性というのが、どこかでトレードオフになると信じられていたのが、実はそうではなくてthresholdがあって、突然ある所まですごく良い符号ができて、ある瞬間からとてつもなく悪いものしか作れないというthresholdの存在が真実であって、理論ができているがそれと同じ。 量子版でも起きるか起きないかというのは1つの問題になっている。ただし、確実にゴールだというふうに今は言えない状況で、そのthresholdの量子版は多分まだ私の知識だと発見されていなかったと思う。冗長をどこまで減らせるかなどという細かいところで見たり、実装するときに回路がどれだけ易しいか、みたいなものであったりとかは、遊びの部分とか、むしろ符号理論で懐の深い理論だという言い方が、良いのではないかと思っている。（萩原）

Q：萩原先生から最後の方に、情報と数学の出会いのようなところをもっとやるべきだというお話があったと思うが、富田先生のところでも、光の物性のようなところと情報理論的なところとか、そういう繋がりっていうのはやっていると思う。具体的にはどうやれば広がるんでしょうか。

A：面白い所を見せないといけないと思っている。数学でも自分の領域に持って帰って得になる所が見えるようになっていれば、3つくらいそういう活動していても良いと思う。今回のセミナーをやっているのも、そういうのが見つかればという願望もあるので、こういうことを続けていきたいと思う。違う分野は敷居が高い。とくに量子情報では、物理の言葉、数学の言葉と情報の言葉があり、相手の言っていることが理解できるようになるところで障壁があるので、そこをどう広げていくかが大事かと思う。（富田）

3.4 数理ファイナンス・金融工学

［概要］

CRDSの連続セミナーシリーズ、第4回は「数理ファイナンス・金融工学」をテーマとして、谷口氏より「数理ファイナンスに現れる確率モデルと数学」、池森氏より「なぜ数学が金融に不可欠か」というタイトルでご講演いただいた。

谷口氏の講演「数理ファイナンスに現れる確率モデルと数学」では、確率の取り扱いを公理的集合論と期待値の計算から始め、ブラック・ショールズ・モデルの導入があった。数理ファイナンスの原理原則やブラック・ショールズ・モデルから現れる幾つかの量の紹介に加え、一般化した理論が研究途上にあることが強調された。

池森氏の講演は、「なぜ数学が金融に不可欠か」として、金融における技術史の流れの紹介からはじまり、数学を利用した効用、経営管理まで言及された。さらに、数学と金融の協働について、提案を交えた紹介もあった。

［Key point］

・ブラック・ショールズ・モデルが一般化モデルの元になっている。（谷口）
・金融市場をモデル化する際に、いくつかの仮定がある。（谷口）
・金融における数学の効用は多岐にわたる。（池森）
・数学と金融の協働には、現場の人間、数学者と数学教育にも意識変革が必要である。（池森）

［谷口氏講演：詳細］

谷口氏からは、数理ファイナンスに現れる確率モデルとして、最初に確率を操る意味から始め、基本的なモデルとしてブラック・ショールズ・モデルが主に紹介された。さらに一般化である確率ボラティリティ・モデルの話と別の形で使った金利変動モデルの話が紹介された。ここでは、参考文献として、

1° 技術に生きる現代数学、若山正人 編、第3章「顕微鏡をのぞくと株価が！」
2° 確率微分方程式、谷口説男、第6章「ブラック・ショールズ・モデル」、共立出版

が紹介された。

この講演での詳しいことは、1° を参照して欲しいとの紹介があった。ブラック・ショールズ・モデルの数学は、2° の方を参照して欲しいとの説明もあった。

確率を操る公理的確率論と呼ばれるものは、1933年にKolmogorovが最初に導入した。考え方は、ランダムな現象を表すパラメーターをωとし，実際の観測値はX（ω）で表せるというものである。パラメーターの全体をΩで書いておく。未知変数ωの関数としてX（ω）を取り扱うことは難しいが、Xがaより小さい、つまり、観測値がaより小さい出来事を図ることができる確率

$$P(X \leq a)$$

だけで、数学が展開できるというのがKolmogorov流の考え方である。実際には、ωや事象の集まりΩには数学的構造がいるが、単に確率が与えられている状況だと考えよう。さらに確率が分かった後、大事なものは期待値である。確率を与えておくと期待値が計算できる。計算には積分が必要になるので、まずは簡単のために、重みをつけた和とだけ思っておくことにする。例えば、サイコロを投げると、目1に1/6という確率をかけて，各々の目に1/6をかけて加えることで期待値は、7/2になる。重みをかけて計算することで期待値を計算するのは、高校の教科書に記載されている。和の一般化されたものが積分であるから、一般に積分を用いて期待値E[X]（Eは、expectationの略）を定義できる。

大事なことは、統計でも多く取り上げられていると思うが、Xの独立なコピーと言われ、全く無関係なコピー

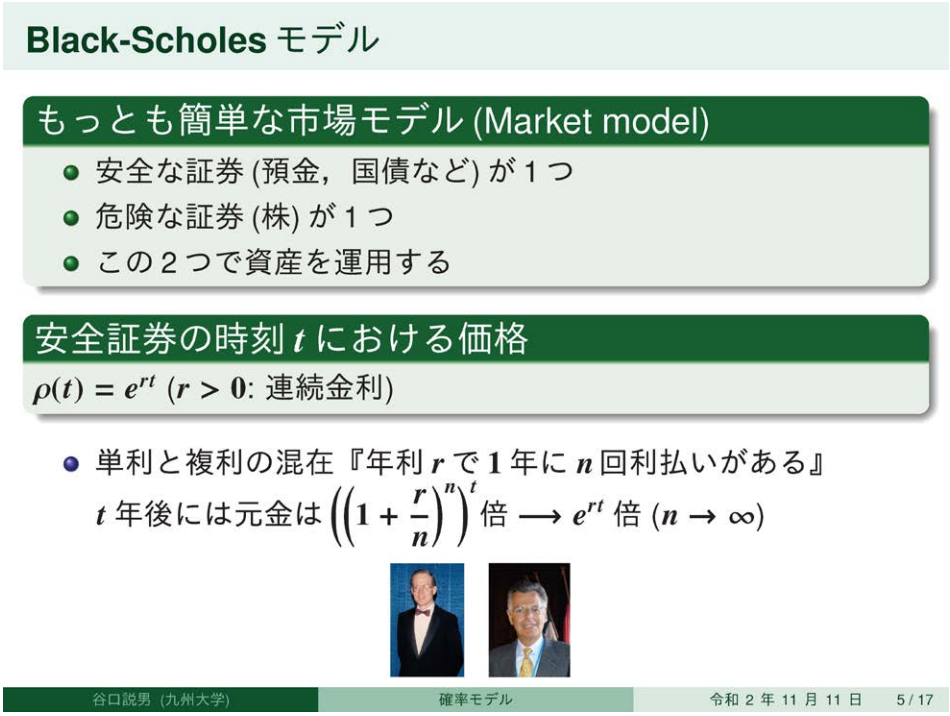
をｎ回観測することである。それら観測値を加えてｎで割った標本平均と呼ばれる値（算術平均とも言われる値）を計算する。ｎを大きくするとランダムネスが消えて、E[X]が見える。結局、理論的にはE[X]は積分として期待値が計算できるが、現場では、たくさん集めて平均を取ればE[X]と思って良い。確率は1回観測するだけではわからないので、時間に合わせてどのように変化するかを集めると、tという時間のパラメーターが1つ増えて、X（t, ω）という形になる確率過程と呼ばれるものになってくる。

確率を使うと話が面白い例としては、待ち行列の話の最初の簡単な例として紹介されるものがある。バス停にA系統とB系統2種類の系統のバスが来ているとする。どちらのバスに乗っても目的地には着く状況とする。その際、Aは5分おきに来て、Bは10分おきにくる。どちらのバスに乗っても良い場合、バスは何分おきに来るか、何分待ったらバスに乗るかを計算したい。Aは5分おきに来ているということは、1分ごとに1/5台バスが来ていると思える。Bの方は10分におきに来るから1分ごとに1/10台来ていると思える。1/5と1/10を合わせた3/10台が、一分ごとに来るバスの台数なので、逆数にした10/3分おきにバスは来る。直前にバスが到着していれば、3分20秒待てばバスに乗れる。

問題は、バスは遅れたり早かったりするような交通事情があることである。きっちり1/5分で来るわけではない。ランダムであるから、実際の問題としては上の考え方で十分かには疑問が残る。ランダムネスを確率と思うと、期待値だと思って足し算すれば、A系統のバスがランダムに来ていると思っても、その期待値は1/5台で、B系統のバスが何台来るかということの期待値はやはり1/10台になる。AかBのバスが来る台数の期待値はやはり3/10であるという計算になる。

このようなことを、数理ファイナンスの方でもやってみるのが、今日紹介するブラック・ショールズ・モデルになる。実はブラック・ショールズ・モデルは非常に簡単なモデルであり、応用上充分であるかどうかは別にして紹介する。ブラック・ショールズ・モデルは二つの証券からなっている。1つは安全な証券と呼ばれているもので、預金とか国債とかいうものになる。もう一つの危険な証券と呼ばれるものは、株だと思うことにする。株が1種類だけあるとする。この2つを使って、預金と株購入を上手に組み合わせることで資産を運用していく市場を考える。お金を減らさない努力をするには、預金を出し入れするか、株を売買するか、もしくは両方をミックスして使うかしかない状況のモデルになる。

安全な証券の単位当たりの価格は e^{rt} つまり指数関数（rは連続金利）で与えられる時間tの連続関数にな



り、確実に増えていく。rは非常に小さいかもしれないが、増えていくという状況にはなっている。実は単利と複利の考え方が混在している形で、金利のモデルを使うということで、指数関数を用いることは正当化される。

危険証券も株価だと思って、株価時間tにおける株価はランダムに時間発展するので確率過程S（t, ω）と書いておく。そのS（t, ω）がどんな形で与えればいいのかが問題になっている。確率過程としての表現は、1900年バシェリエが確率過程としてブラウン運動を覚えていたのがきっかけになる。パリの株式市場を見ながら、株価としてブラウン運動を用い、バシェリエは一定時間の株価の最大値を計算することを行った。ブラウン運動（微粒子がジグザグに動いていく様子を1次元に落としたもの）は正の値も負の値も取りうる。株価が負の値になるのはありえない話なので、モデルとしてはいささか問題がある。そこで数学としてはあまり興味をひかなかったが、1960年頃に経済学者サミュエルソン（ノーベル経済学賞受賞者）が、ブラウン運動を指数関数の指数に入れた確率過程を扱った。MITの雑誌にいくつかの論文を書いて、いろんな先駆的な仕事をした。指数関数の指数に入っているので、時間発展のことを記述するならば、ニュートン方程式のようなものがあるとよい。指数関数の指数を見ることになるので、tで微分したら方程式が導出される。

$$\mathrm{S}(\mathrm{t},\omega)=\exp\left(\beta+\sigma\frac{dB(t,\omega)}{dt}\right)$$

と考えて、

$$\mathrm{dS}(\mathrm{t},\omega)=\mathrm{S}(\mathrm{t},\omega)\left(\beta+\sigma\frac{dB(t,\omega)}{dt}\right)\mathrm{dt}$$

になる筈である。厳密には、ブラウン運動は微小時間dtの間に $\sqrt{dt}$ のようにふるまうので、このままでは定式化できない。マートンは、1944年に伊藤清が提唱した伊藤積分を用いればよいことに気づいた。伊藤積分とは、

$$\int_0^T f(t,\omega)dB(t,\omega)=\lim_{n\to\infty}\sum_{k=0}^{2^n-1}f(\frac{kT}{2^n},\omega)\left\{B\left(\frac{(k+1)T}{2^n},\omega\right)-B(\frac{kT}{2^n},\omega)\right\}$$

とリーマン和として積分を定める。伊藤積分の定義にのっとって計算すると、

$$\boxed{\mathrm{dS}(\mathrm{t},\omega)=\mathrm{S}(\mathrm{t},\omega)\{\mu dt+\sigma dB(t,\ \omega)\}\qquad\left(\mu=\beta+\frac{\sigma^2}{2}\right)}$$

となり、$\left(\mu=\beta+\frac{\sigma^2}{2}\right)$ には、$\frac{\sigma^2}{2}$ が付加されている。このことは、ブラウン運動が $\sqrt{dt}$ に応じた動きをしているからとなる。

このことを使って資産運用してみる。例えばT年後に小麦2400kg（80袋）を12,000円で買うと契約をしたとする。T年後に買うためには、当然、12,000円を用意しないといけない。つまり、国債と株を売り買いして、12,000円の利益を生み出さないといけないことになる。少し技術的ではあるが、$\frac{kT}{2^n}$ のように時間Tまでの区間を細分しておいて、そのそれぞれの時間に株と国債の様子を見ながら、持ち替えて行く。そうすると、$\frac{kT}{2^n}$ のときに国債を $\theta_0(\frac{kT}{2^n},\omega)$, 株を $\theta_1(\frac{kT}{2^n},\omega)$ 買うことをすると、$\frac{T}{2^n}$ 秒ごとに利益は差分の入った、

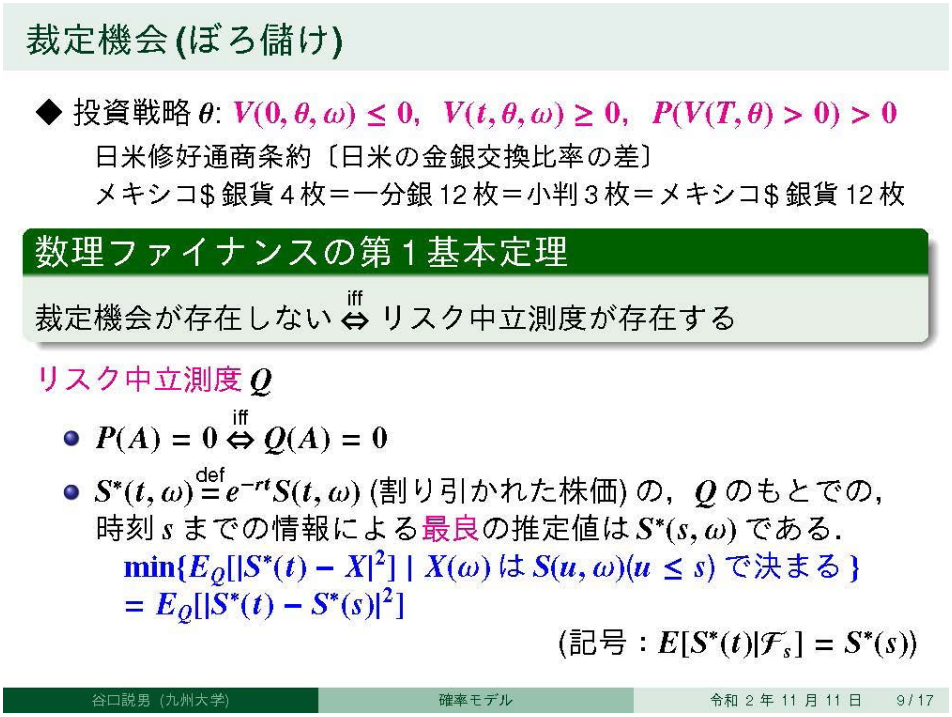
$$\theta_0\left(\frac{kT}{2^n},\omega\right)\left\{\rho\left(\frac{(k+1)T}{2^n}\right)-\rho\left(\frac{kT}{2^n}\right)\right\}+\theta_1\left(\frac{kT}{2^n},\omega\right)\left\{S\left(\frac{(k+1)T}{2^n}\right)-S\left(\frac{kT}{2^n}\right)\right\}$$

となる。よって、時刻Tにおける利益は和 $\sum_{k=0}^{2^n-1}\cdots$ を取り、$\mathrm{n}\to\infty$ の極限を計算して、

$$\mathrm{V}(\mathrm{T},\theta,\omega)=\mathrm{V}(0,\theta,\omega)+\int_0^T\theta_0(t,\omega)d\rho(t)+\int_0^T\theta_1(t,\omega)dS(t,\omega)$$

となる。式に書いてある、差分が入ったような利益が得られている場合、投資戦略 θ に対する富過程もしくは価値過程と呼ぶ。

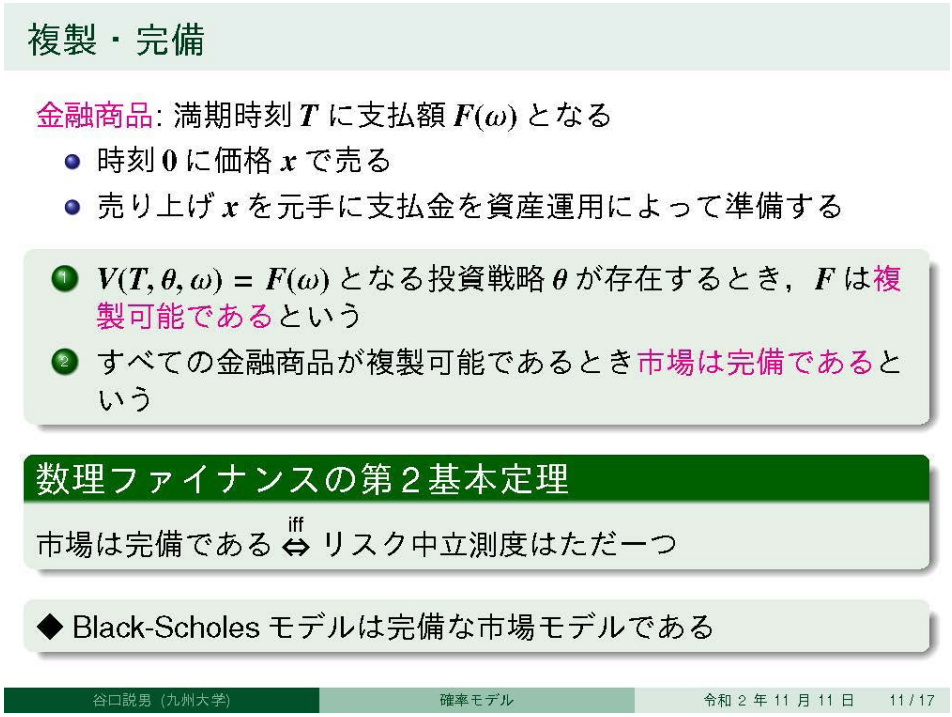
最初の問題はぼろ儲けができない公正な市場を数学的に特徴づけることである。ぼろ儲けとは、初期資産



0でどの時刻tに対しても資産は負にならなく、かつ支払いが行われる時期には資産が正になっている確率が正であることをいう。日本でも日米修好通商条約の際に、金銀交換比率のレートが日本とアメリカで違った為、ぼろ儲けが発生した。ぼろ儲けを許さないのはどんなときかは、数理ファイナンスの第１基本定理で保証することになる。ぼろ儲けできる状態を裁定機会と呼ぶが、リスク中立測度というものが存在することと同値になる。リスク中立測度というのは、“未来の割引価格を現状で持っている情報で推定するとそれは今の割引価格に他ならない”ことが起きる現象のことである。リスク中立であれば、未来に関与する情報は何も持っていない。未来は、ある意味で独立に起きているという状況である。例えば、インサイダー取引のようなことが起きれば、リスク中立測度の存在は壊れる。未来の情報が現在に流れ込むので、その株を買えば得するということで買うことになる。リスク中立ということは要するにインサイダー取引のようなものが市場に存在していないことを意味し、ぼろ儲けができないということは、未来と現在との独立性を要求する。ブラック・ショールズ・モデルでは、リスク中立測度が構成できる。その意味で、ブラック・ショールズ・モデルは、良いモデルの一つである。

次に問題になるのは、市場でへんてこな商品が売られることである。例えばリーマンショックの元凶であるサブプライムローンをもとにつくられた複雑怪奇なオプションが挙げられる。オプションのような商品が、どのような価格になるか誰も判定できない状態で、皆が危険を感じて売り出したら買い手がいなくなり価格破壊が起きた。オプションに市場で正当な価格が付けられるのかということが問題になる。これは、言い換えると、金融商品と呼ばれているものの価格が投資戦略で再現できるかということである。先ほどの例でいえば、12,000円を預金と株式の売り買いで確保できるという条件が市場で満たされねばならない。このような金融派生商品の価格が投資戦略で再現できることを「市場が完備である」という。市場が完備であることと、先ほどのリスク中立測度はただ1つになることが対応する。ブラック・ショールズ・モデルは完備な市場モデルである。結局、簡単なToyモデルだが、理想的な市場に要求されていることをすべて満たすのが、ブラック・ショールズ・モデルになっている。

一般に大事なことは、その金融商品に価格が付けられることで、実はリスク中立測度に関する期待値という形で価格をつける。ブラック・ショールズ・モデルのときはリスク中立測度が1つしかないので、それを使って期待値 $E_Q[e^{-rT}F]$ を計算すれば価格になる。一般に、池森氏の話にも出てくるが、f（S（T））という形をした支払額は



$$e^{-rT}\iint_{-\infty}^{\infty} f\left(S(0)e^{\left(r-\frac{\sigma^2}{2}\right)}\right)\frac{1}{\sqrt{2\pi\sigma^2 T}}e^{\frac{x^2}{2\sigma^2 T}}dx$$

の形になる。金融商品の価格というのは、ブラック・ショールズ・モデルであれば、正規分布に従った計算で出来るので、とくに大事なコールオプションも計算できる。コールオプションというのは、満期時に、ある株を１株Ｋという価格で買うと言う権利のことで、価値は

$$\mathrm{C}(\omega) = \max\{S(T,\omega) - K, 0\}$$

である。これは絶対損をしない。Ｋより高かったらそれをすぐ市場で売れば $\mathrm{S}(\mathrm{T},\omega) - \mathrm{K}$ の利益が出て、Ｋより低い価格だったらこのコールオプションは権利を放棄してしまえば何も支払わなくて良いからである。つまりコールオプションには何らかの価格をつけることになる。さらにコールオプションからは、ボラティリティという値が計算できて、例えば日系225オプション価格からは日経VI（ボラティリティー・インデックス）などで使われている。

さらにブラック・ショールズ・モデルを一般化したボラティリティー・モデルのスライドが紹介された。加えて、ゼロクーポン債に関する価格を計算する方式に、今のブラック・ショールズと同じような確率微分方程式と呼ばれているものが使える金利変動モデルがあり、そういうモデルにも使われている。

現在は、ブラウン運動で駆動することが正しいかどうかが良く分からないので、様々なモデルが使われ、レビ過程と呼ばれているジャンプがある確率過程を使ったり、フラクショナルブラウン運動と呼ばれる、さらにジグザグしたガウス過程を使うとこともモデルとして取り込まれ、色々計算がさらに研究されている。

［池森氏講演：詳細］

池森氏の講演では、自己紹介と簡単な経歴の紹介から始まり、なぜ数学が金融に不可欠か、についての所見が順次述べられた。

講演の導入として、銀行で数学が利用されてきた経緯が示された。前史として金融技術革新以前、状況が一変し、金融技術革新と金融自由化、金融危機を経て状況が変化した流れを紹介した。数学によって金融を組み立てることの効用、数学と金融の協働の強化に向けての提言も述べられた。

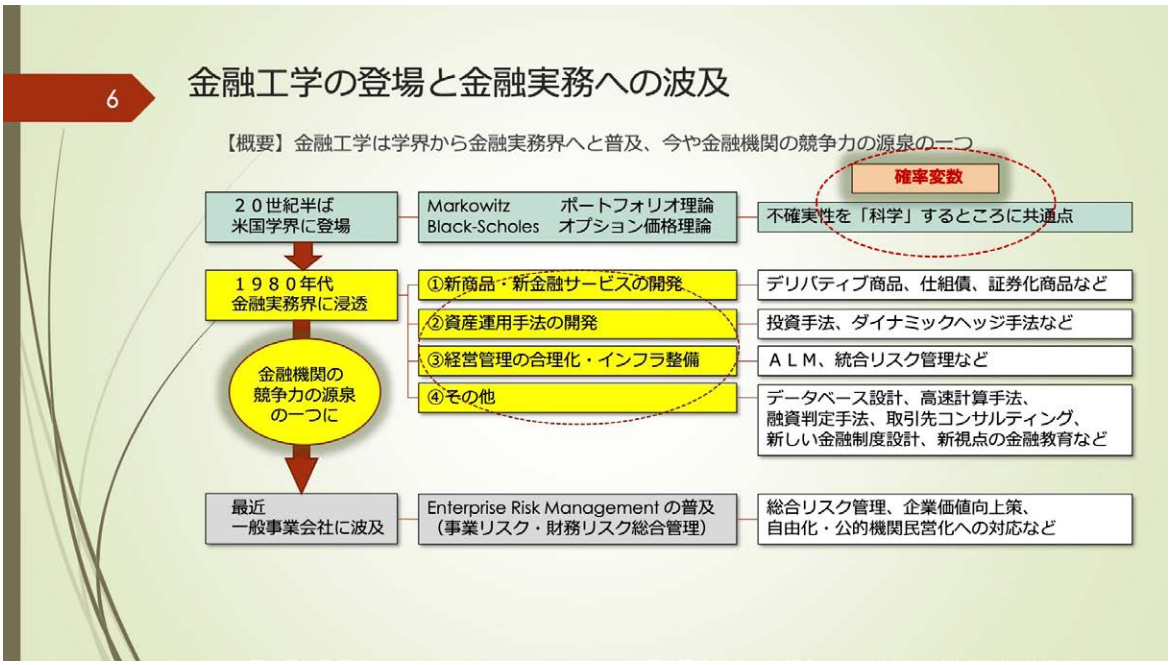
まず前史では、銀行はそろばんが大活躍の職場であり、池森氏が銀行に就職した経緯として、数学と遠い職場の銀行に就職したという説明もなされた。しかし、四則演算を超える数学もあり、例えばキャッシュフローを現在価値に割り引いて投資プロジェクトを評価したり、多変量解析を使って財務分析をする、あるいはマクロ経済予測を多変数の連立方程式系を立てて推定することが行われていた。

細々と数学が使われてはいたが、金融技術革新と、金融自由化で状況は一変した。 1970 年代に状況が変わった。71 年の 8 月にニクソンショックがあり、それに引き続いてブレトン・ウッズ体制の崩壊、変動相場制へと為替制度が変わり、73 年の 10 月には第 1 次オイルショックが起こって物価が急上昇した。 先進国は国債を大量に発行して景気後退を支えたが、その国債が流通市場を形成し、金利が大きく変動する、いわゆるガルブレイスのいうところの不確実性の時代が到来した。 この不確実性に対処する手段として登場したのが、金融工学であった。その金融工学をベースにした金融技術革新が各銀行に波及したのが、1980 年代の初め半ばぐらいで、この頃から金融工学が大きく進展した。

金融工学は、新商品開発、資産運用手法の開発、銀行の経営管理などの分野に波及していったが、順番に述べていく。金融工学に特有な新商品は、デリバティブと証券化商品で、まずデリバティブについて紹介する。

従来型の古典的な金融商品には受信取引と与信取引がある。まずお金を受け取って、満期にはお金を返済するのが受信取引で、具体的には預金である。逆に、まずお金を渡して満期に回収するのが与信取引で、具体的には貸出である。このように時間を隔ててお金の受け払いがペアになったような構造を持っているのが、従来型の商品だったが、金融工学で登場したのは、例えば将来の満期時点において固定金利と変動金利を交換する、つまり、現時点でキャッシュフローが起こらないが、将来の交換をベースにして取引が成立するものが登場してきた。これがデリバティブで、この特性がレバレッジ効果の原因となった。もう一つの証券化商品は、銀行のバランスシートから、既に実行した貸出を SPV（special purpose vehicle＝特別目的会社）に譲渡し、その貸出のプールから発生するキャッシュフローを一括し、別の基準で纏め直して新しい商品を作るという操作を行う。それが証券化商品である。投資家のニーズに合った商品を合成する方法ということで登場した。

商品開発のポイントというのは、谷口氏の話にもあったが、商品構造をどのように設計するかがテーマになる。この際、将来のペイオフ（＝キャッシュフロー）を、関数形で表すことになる。2 つ目の重要なテーマは、



商品の値段をいくらにするのかである。これも谷口氏の金融工学の話にも登場したが、具体的に書きおろすことが、実務では非常に重要になる。さらに、高速で計算しないといけないので、そのための技術が発達してきた。その他に、新商品の持つリスク構造を分析してどうやって制御するかということも、重要なテーマになっている。 以上が新商品開発である。

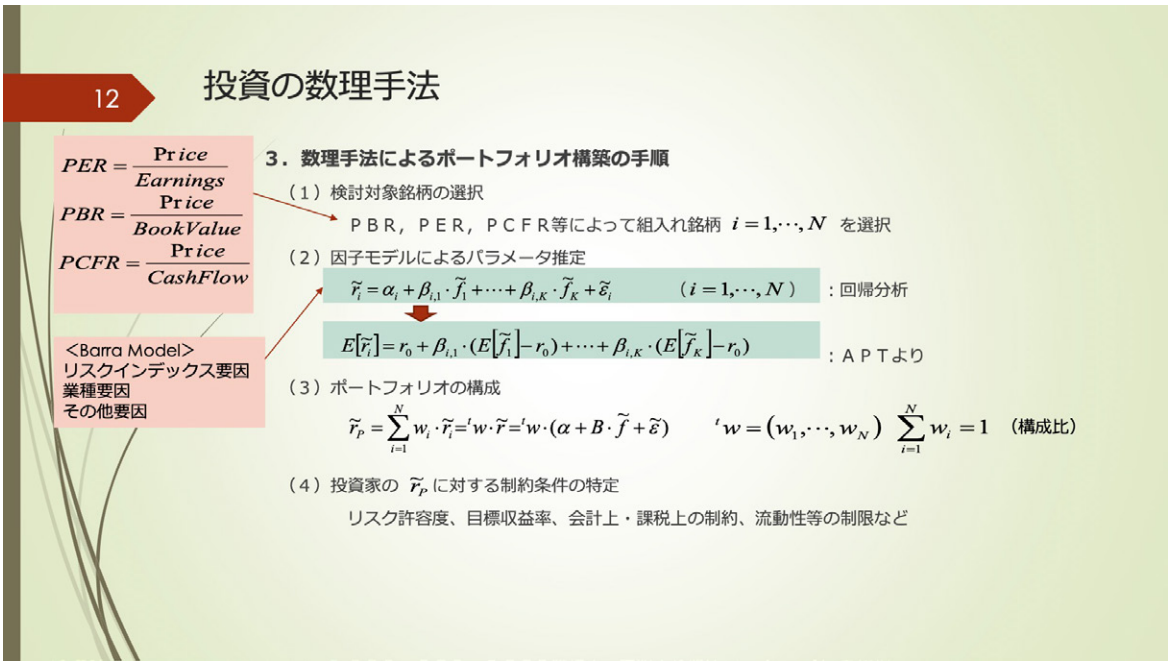
次のテーマは、投資などの資産運用手法に金融工学がどう使われているかということである。 従来の投資手法というのは、儲かる銘柄をどうやって発見するかという、収益性のみに着目した手法だったが、金融工学の登場に伴い、そこにリスク評価という要素も加わり、収益性とリスクを投資目的に合致した形で、どのようにバランスさせて具体的な投資内容を決定するかが問題となっている。

次に運用をもっと短期間で行うのがトレーディングである。１日の中で売ったり買ったり、せいぜい１週間くらいの期間に売買する。安く買って高く売る。高く先売りして安く買い戻す。そのようにして、売買益を得るというのがトレーディングである。その際には、売買の対象になっている金融商品の価格をいかに判定するか、あるいはリスクを定式化できるかが重要なテーマになる。具体的な投資戦略としては、相場を張って値上がり・値下がりすることにかけてポジションを作るdirectionalトレーニングが典型だが、その他にも Carry Trading, Arbitrage Trading、Gamma Tradingなど、様々なトレーディング手法があるが、これも数学を使うと上手く整理することができる。

このように、トレーディングに対する金融工学の適用が行われてきた。

３つ目の金融工学の応用分野として、自主的な経営管理がある。自主的というのは、それまでの金融行政が規制・保護という観点で行われていたが、1980年代になってからいわゆる金融自由化が実施された。自由化になると、色々なことが出来ることになったが、裏を返せば、今まで規制によって間接的に外側から防御がかけられていた状態から、銀行は無防備になった。無防備な銀行に対して、これからは自主的に経営管理をすることが要請された。大事なことは、収益性で経営破綻をしないように、資金繰りで経営破綻をしないように、という２つの観点から経営状況を注視することである。

個々の取引から始めて、それを銀行のポートフォリオとして合算していく。経営破綻という観点からは、銀行全体でどのようなことが起こっているか見ないといけない、つまり統合管理になる。これが第一義的だが、その上で、収益性やリスク構造を制御・改善しようとすると、全体を見ていたのでは構造が複雑すぎる。そこで部門別に分けて、部門別管理をするような仕組みもつくっていかないといけない。これらが経営管理の重要



なテーマである。具体的には、銀行全体のバランスシートから、ALM部門を通じて各部門が必要にとなる資金を内部資金として供給することによって、業務部門別バランスシート分解を行い、各業務部門の収益性とリスクが議論できるようにする。その上で各部門に資本を配布し、その資本に対する収益性が確保できる構造、損失が生じた場合にその損失を配布資本でカバーできるような構造を順番に考えていくという枠組みであった。

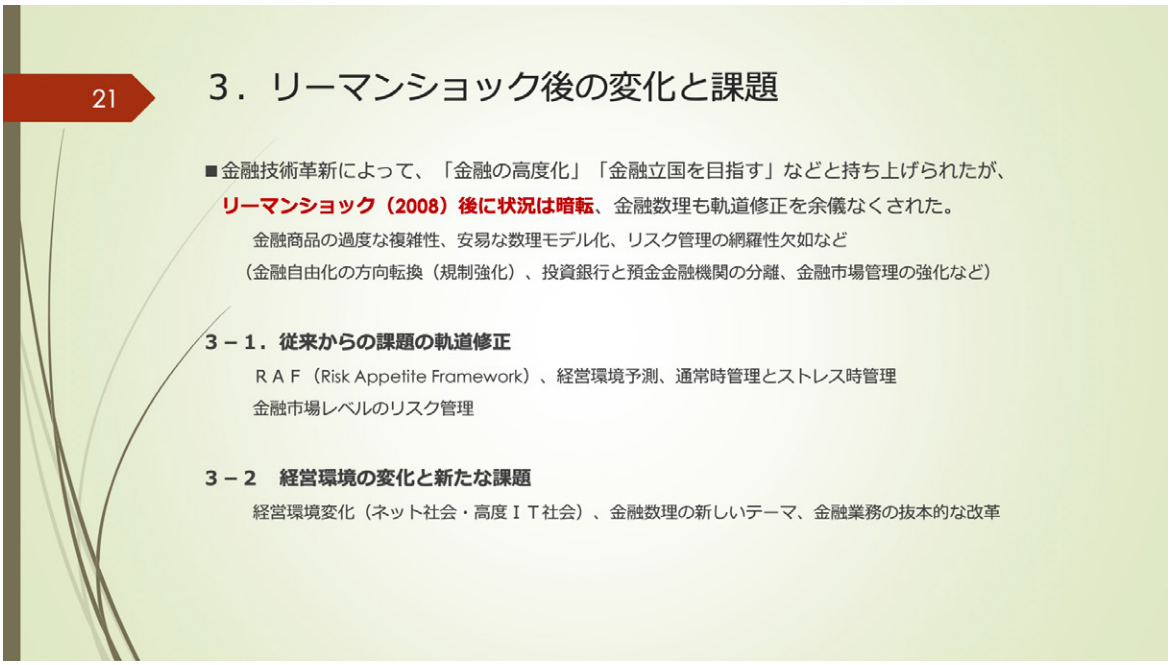
このように、金融工学が3つの分野、つまり、新商品開発、投資手法、それから銀行の経営管理、これらの分野で展開してきた。これらは、金融技術革新によって行われた銀行経営の大きな改革である。

このような変化によって、一時は「金融高度化の時代」、「日本も金融立国を目指す」などと持ち上げられたが、2008年のリーマンショックを機会にこれが暗転することになった。金融商品の過度の複雑化や安易な数理モデル、リスク管理の網羅性の欠如などが指摘されて、再考を余儀なくされることになった。これをターニングポイントにして、従来からの軌道修正と、経営環境の変化に伴う新たな課題に取り組まないといけないような状況になった。

まず、従来からの軌道修正では、RAF（risk appetite framework）、具体的に言うと、従来の経営管理をよりフォワードルッキングに、網羅的に、かつ経営管理に直結していかないといけないという方向づけが要請された。

次に、金利と為替と株価の変動を相互に関連付けて議論すべき状況下でも、それまではそれぞれ個別にモデル化して組み立てていた。 谷口氏の話も個々の金融商品について、その原資産を個別にモデル化することであった。しかし、特に金融危機のような時には、金利と為替は同時に動くことがよく観測されている。これらを相互に関連付けてモデル化するということである。

3番目に、平常時とストレス時では見るべき視点が違うということである。平常時では、収益性の確保が重要である。従来、監督当局主導の経営管理では、収益性の議論は余り扱われていなかった。 ストレス時という厳しい状況下では、今まで以上に、もっと厳しい状態を想定しないといけない。それに加えて更にもっと重要なテーマは、金融システム自体を如何に保全するかというテーマである。これは個別の金融機関だけでは対応できないような重要なテーマになっている。さらに4番目のテーマは、金融工学のテーマの階層アップということである。谷口氏の話でも紹介のあった、金融商品の値段をどのようにつけるか、というのはキャッシュフローの集合としての金融取引がテーマである。その金融取引の集合としての金融機関の経営管理が次の階層、さらに、その金融機関、投資家がたくさん集まった金融システム・金融市場が次の階層である。この金



融システム・金融市場がどのように変化するかというメカニズムを議論しようという階層が、新たなテーマとして加わって来ている。これらが、従来の延長としての新たなテーマになっている。

次に経営環境の変化に伴う新たな課題では、ネット社会、高度IT社会になり取引先や顧客の行動様式が変化し、新しいタイプの大量データが発生してきている。それに加えて、周辺業界からの攻勢、新たな技術集団の台頭、これらは金融機関をサポートしてくれるという可能性もあるが、これらに対してどのように対処していくかが新たなテーマとして起こってきている。このような状況の変化を受けて、金融数理の新しいテーマは、得られるビッグデータに対して、どのように集めて、どのように業務に活用するのか、具体的にどういう業務に繋いでいくのかが非常に重要なテーマになる。金融市場から得られる高頻度の取引データをどうやって解析するかもテーマになっている。加えて、取引をネットワークで行うようになると、取引のプロテクション、ブロックチェーンによる業務の効率化をどのように進めていくかということがテーマになってきている。従来の銀行では、ホストコンピューターを用意して、それに専用回線で専用端末をぶら下げてデータが外に漏れないようにして管理してきた。インターネットの世界では、それぞれのパソコンの中にブロックチェーンを置き、それぞれ分散管理して効率化していくことを考えることになる。

暗号理論では、現在、公開鍵暗号が実用化されており、金融業務の中にも組み込まれているが、量子コンピューターの実用化とShorのアルゴリズムによって脅威にさらされつつある。これらによって常識的な時間で暗号が破られる状態が生じることになると、ネットワークを通じた金融取引が安全に行えなくなる。そのような事態に対処するために、量子暗号や耐量子計算機暗号もテーマとなる。

以上のような新しい業務形態への取り組みとともに、今重要なのは金融業務でどうやって収益性を確保するかということが重要なテーマになっている。顧客に対して、納得して料金を払ってもらえる仕組みをどのように構築するかが問題になる。今、あらためて「金融とは何か」ということについて考える必要があるのではないかと思っている。

以上が、これまでの金融に数学が使われた経緯と、これから取り組まないといけないと思われるテーマである。

次に 数学で金融を考えることの効用について、2点だけ述べたい。1つ目は数学によって問題を見通しよく考えられるということ、2つ目は主要なコンセプトの存在性の議論ができるということである。具体的に例をあげると、まず見通しよく考えられるということについては、金融商品開発をファイバー・バンドル空間で考えると考えやすい。次に、数理ファイナンスの基本定理は谷口氏の講演の中にも出てきたが、金融商品の価格を与える一般式を示したこの定理は、キャッシュフローが構成する線形空間から実数空間へのプライシング関数を考えることになり、これが線形構造（空間）を（価格）に写像する線形汎関数になっていることから、リースの表現定理を用いて内積表示できることを考えれば、数理ファイナンスの基本定理が納得できる。このように数学によって見通しよく議論することができるという効用がある。

さらに、主要なコンセプトの存在性の議論ができる。Coherent Measure of Riskと呼ばれる理想的なリスク指標は、4つの公理によって特徴づけることができる。金融実務で様々な形で使われている標準偏差やValue at Risk やExpected Shortfall などのリスク指標が、このコヒーレントな公理を満たすかどうかというチェックをするとともに、実は、有限個の将来の不確実事象の下では、コヒーレントなリスク指標（ρ）は

$$\rho = \rho(\widetilde{L}) = sup\{E^{Q}(\tilde{L}) : Q \in P\}, \qquad \text{P: 将来の不確実性に対する確率測度の集合}$$

という形状に限る、ということが示される。 このような例によって、数学が金融になぜ不可欠かと言う、本日のテーマにも、納得できるのではないかと思う。

まず金融のテーマを数学者に相談する場合、課題を数理の問題として定式化する人材が必要になる。それは多くの場合、現場の金融の専門家になるが、この定式化能力を持った現場の人間が必要になってくる。これが出発点になる。

そうした上で問題解決にあたっては、双方の当事者がもっと課題に取り組むための意識を高める必要がある。銀行の現場の担当者は文科系が多いが、彼らがもっと数学を理解しないといけない。そのためには、例えば文系の数学教育の方法を変えないといけないと思う。

それから、相談される数学者も、もう一歩踏み込んで、自らが中心になって解決しなければならない問題として考えないといけない。「役に立つことがあれば」程度の意識では駄目である。

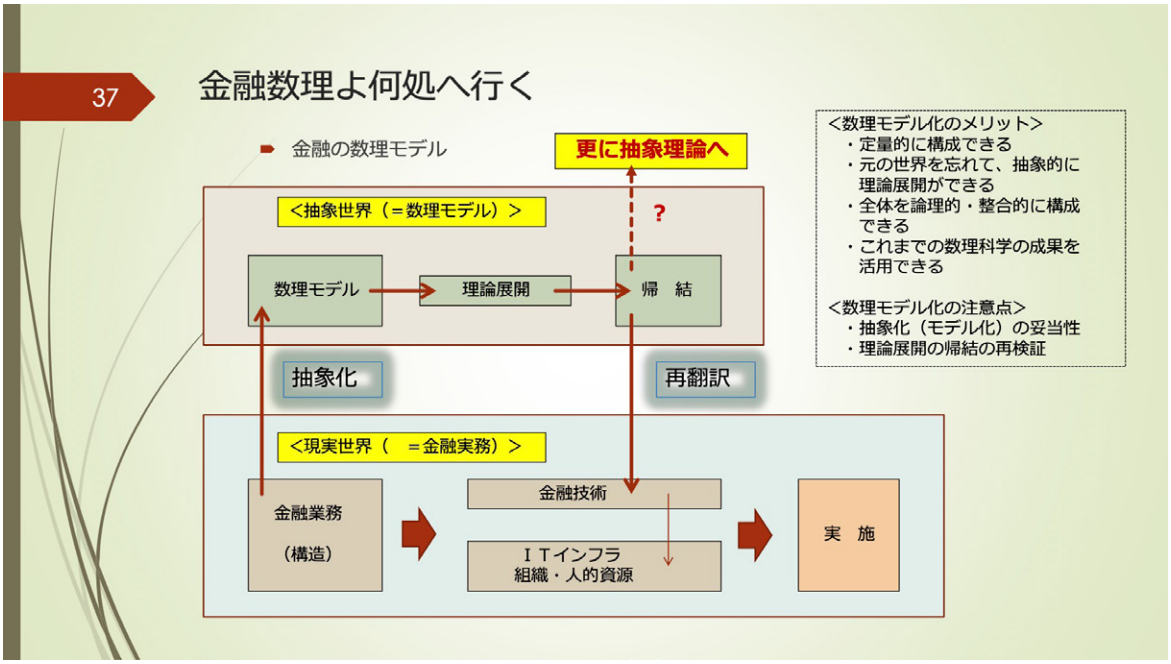
それから、金融界にとって、金融に向かう数学人材の確保が、これからは重要になってくる。現在は残念ながら、金融は優秀な学生が最初に選ぶ職場ではなくなっている。博士の積極採用や、一旦就職した後の、企業と大学の行き来を柔軟にすることも考えられる。

最後に、金融業界（金融機関とその監督当局）では、金融技術は個々の金融機関の競争力として議論する問題から、業界全体として、例えば、金融システムの保全の問題等として考えないといけないような共通問題がずいぶん多くなっている。このような共通問題に対する仕組みを作らないといけない。

想妄莫。芝浦工業大学の学長であった柘植綾夫の座右の銘と言われている。工学をする人たちは、現実をしっかり見据え妄想してはいけない。実用を追うように心がけることが肝要であるという、日本経済新聞の夕刊に載った記事の紹介があった。池森氏は、学生や会社の社員には、とくに新しいことを構想する段階では、あれこれと夢を見て「妄想する」ことも必要だと諭していた。

金融が理論として進展して非常に実のあるものになってきているが、元は、実務の世界を抽象化してモデルにして、理論展開をして帰結が出てきたら、それを改めて現実世界に再翻訳し、更に具体的に実施して行くための要件を整える。

再翻訳のところで、本当に使えるかもう一度チェックする必要がある。往々にしてモデルにして帰結が出るとそれが必ず正しいものと考えがちであるが、もう一度、現実への適合性のチェックを行う必要がある。また、現実世界に持って行くよりも、もっと抽象的な議論を作り上げて、綺麗な数学の世界に昇華させていくことがずいぶん行われている感じがする。実学としての金融工学を見失わないようにしなければならない。



【質疑応答】

Q：ブラック・ショールズがとても単純なモデルで、中立測度が1個だという話だが、今の市場に拡張して使われているのか。

A：当然1個でできたら、有限個の株でも出来る。ただし、ブラック・ショールズのときの完備の条件は少し書き換えてやらないといけない。リスク中立測度があるかという問題も少し違う形で記述する。少し記述を変えていかないといけないが、ブラック・ショールズのシグマにあたる部分を行列に変えることが出来て、Sは多次元のモデルに変わる。なお、国債の方は基本財と呼ばれるので、基本的には1個だけ指定しておいてモデルを考えるのが普通だと思う。国債を変えて、違うものに変えてもいいと思うが、要するに社会が経済成長していて、どんどん増えていくというのが、ある意味で基本財が増えていく意味になる。例えば預金高が、1年で2倍になった時に株価が毎年1.5倍にしかならないのなら、株価は全然上がっているとは思えない。本当に株価が上がっているかというのは、基本財が2倍になったら株価は2倍以上のものが上がっているというふうに考えるべき。（谷口）

Q： 今のようなマイナス金利でも適用可能なモデルになるのか。

A：基本的にブラック・ショールズ・モデルはポジティブで取っている。（谷口）

Q：ブラック・ショールズはポジティブだということだが、資産運用をやるときに、時刻を等分しているが、必ずしも等分しないとか、社会状況というのを加味してやらないといけないことがあると思う。どのように実際使われているのか？

A：確率ボラティリティ・モデルでは、ブラック・ショールズはシグマしか入ってないが、そのシグマをランダムに変えて、ミューもランダムに変える。一般だと、ファイナンスで実際に計算できないといけない。確率論屋がこれだったら解けるという存在定理だけではダメで、具体系が必要になる。もしくは具体的に近いものが何らかの形でいるという形で、WigginsとかHestonが言っている形にしてくれば解ける。それがいいのかどうかは別にして、計算して使われている。 Vasicekの言っている式は方程式としては非常に易しく、ランダムネスが加味されていないように見えるが、破産のモデルに使われている。 CIRではベッセル関数が出てきたりして解けるので、良い形になる。結局、具体的に書けないと話がなかなか動かないので、皆困っているが、色々なモデルを使いながら、現場では、実際に計算している方も沢山いると思う。（谷口）

Q：数学的かもしれないが、力学系の概念、アトラクターや埋め込みの話などはモデルで取り入れられたりしているか。

A：1番最初に紹介した大数の法則は、たくさん観測して算術平均取ると、実は本質的なパラメーターの期待値が出てくる。それは、金融でも、保険関係と銀行関係を数学的に分けているのは何かと言われると、中心極限定理と大数の法則の2つがあり、銀行とかは株の取引になる。そのような金融市場は多くの買い手がいて株が動くので、中心極限定理の正規分布が最終的に出現する分布までわかる確率が見えてくるような極限定理になる。一方、損保の方は一方に保険会社があるので、損害額とかには大数の法則までは効くが、中心極限定理までは効かない。モデルとして扱うべきものは違うが、アトラクターとかそこまでのレベルまでは扱っていない。ただし、漸近展開の理論というは統計の方にあるので、漸近統計の方は少なくともそのアトラクターに近いものを探すような行為はあると思う。（谷口）

C：決定論的なものと、そうでない話をしているが、そのあたりがどう組み合わされて考えられているのかに興味があったので、質問した。

Q：裁定取引、つまりアービトラージは一般的にあると思うが、それとブラック・ショールズ・モデルの関係は何か。一般に現物と先物の価格差を見て取引をしている投資家とか、金利差とかで見ている投資家とかもいると思うが、そういうのが増えると結果的に裁定状態がない方に落ち着いて行くというのが今回の話なのか、その関係が分かればと思う。

A：ノーベル経済学賞を取ったとき、ショールズとマートンはそれを受けて投資会社を作っている。彼らの

理論を背景に展開すると言ったが、実際は質問の中にあった、鞘取りをコールとプットオプションを2つ構えて、どちらでも損しないような形で、微妙な鞘取りをやることで稼ぎを出そうとしていたが、うまくいかなかった。理想的な経済モデルをあの数学として定式化しているというだけで、現実に使うときは、例えば、どういう株を買ったらいいかというのを決めるために非常に簡単なシミュレーションをやる。50人とか100人とかいう人間がみんなブラック・ショールズに基づいて株の売り買いをやったシミュレーションをかけてどれが上がるかをシミュレーションで出してみる。それに従い買ってみるということをある投資会社がやっていたということは聞いたことはある。理想論なので、ここで書いてある話から、本当に儲けを出そうと思ったら、何らかの形で裁定状態を壊してアービトラージのない状態で、自分だけが知っている情報で稼がないと稼げないと思う。（谷口）

Q：コンプレックスタイムというかファインマン積分について、フラクショナルなブラウン運動の話もあったが、物理対応版というか、コンプレックスタイムにしたらどういうことになっているか。

A：フラクショナルブラウン運動が捕まるようになった理由は、ブラウン運動はルートdtを使うと言ったが、確率論としてもなかなか悩ましい。テイラー展開すると2次の項まで出てくるということが邪魔になっている。そこで一次項だけでなく、2次近似も込みでペアにして計算すると上手くいくのではないかということを1900年代の後半ぐらいから、完成されてきた。2次近似までやるとフラクショナルブラウン運動がきっちり捕まえられる。ある意味で、それに付随する確率微分方程式とかかけると言うところまでベースにして、数理ファイナンスなんかも出来るということを考える人たちがいる。数学的にフラクショナルのブラウン運動がうまく捕まっていることになる。物理のモデルでどのように対応するかは良く分からない。（谷口）

Q：金融というより、経営全体に対して適用できる考え方なのか。

A：まず、商品開発は、どの業種の企業にとっても重要なテーマだと思う。会社全体がどういう経営状態になるのか、同じように収益性と資金繰りはどの会社でも当然考えないといけない。毎期の決算を、具体的に制御していこうとすると、部門別の会計を行ったり、部門別のリスク管理をやったり、戦略を立てたりと言う事が必要になって来る。銀行も金融の自由化とともに本来やるべきだったことをあらためてやるようになった。一般事業会社の経営でも、同様の枠組みが使えるのではないかと思う。（池森）

Q：このような考え方が共通に使えるなら、なぜファイナンスや金融に絞られているのか。むしろ、他の業界ではそういう考え方がすでに適用されていて、金融ではまだそういう考え方がされていなかったと理解するのが正しいか。

A：以前に勤めていたみずほfinancial technologyという会社では、電力会社などからリスク管理の相談があった。つまり、部門別の経営管理で収益性は計量しているが、リスク構造については見ていなかった。銀行で展開している手法を適用するためのコンサルティングをしてくれという案件が何件かあった。池森は東大大学院の経済学研究科で講座を1つ持っているが、その経営財務IIでやっていることは、銀行の収益リスク管理の仕組みを紹介した後に、それを一般事業会社に展開するとしたらどうしたら良いのかというような内容も含んでいる。一般的な会社の経営管理、いわゆる経営学まで話が広がったら、もっとこのような考えが浸透すると思われる。金融危機以降の新たな展開の中で、銀行はもっとフォワードルッキングに経営管理・リスク管理をしていこうということだったが、これは何かというと、将来どういうことが起こりうるかということを経営者として予測し、予測からずれる部分をできるだけ小さくなるようにする。従来の銀行の管理では、今のポジションはそのまま継続されるとか、過去の変動がそのまま将来にも伸びるとかを前提にしてきたが、一般事業会社の場合、なかなかそうはいかない。この1年間の事業計画をたてて、それがどうぶれるのかを議論し、対処することが既に普通に行われている。それに合わせるような数理的な枠組みの組み立て方を、もっとしっかりやっていかないといけない。（池森）

Q：谷口氏の話でもあったが、金融理論は理想的なところでの考察になるのか。しかし市場というのは、実際には完全なことはないし、仮定している完備性も崩れていると思う。理想的なところで考えていても、もちろんリスクは考えてやっていると思うが、その市場が不完全であるとか完備性が欠如しているところにも大きなリスクがあると思う。その辺りはどのような考え方で、業務を進めているのか。例えばALM部門ではどうか。全体を考えている部門があると思うが、例えば不完全性を前提にした場合に、どのような数理的な方法でリスクは捉えられているのかお伺いしたい。

A：金融理論では市場がアービトラージフリーである、裁定取引が不可能であるという仮定のもとに理論が組み立てられているが、実はそうではないところが沢山ある。それが崩れている場合にはアービトラージ機会の問題になる。理論的な価格はこうだけれども、そうでないとすれば、安い方を買って、高い方を売ることによって、リスクを取らないで利益が得られる。ノーアービトラージじゃないところを発見して収益を上げていくことをやっていたりする。それから市場が完全であるという仮定も現実では成り立たない。ここでいう完全市場というのは手数料がかからない、無限に短い期間で取引が可能である、言い換えれば、売り買いが自由にできるということであるが、例えば有限の間隔でないと取引ができないとか、取引はそれぞれコストがかかってしまうことをモデルの中で修正した学者がいる。コスト込みの離散時間の取引を想定するとすれば、プライシングモデルをどのように修正したらいいのか、理想的な世界に対して、現実はそうではないというところを如何に修正して、実務にどのように繋げていくかという研究は、いくつかすでにある。実際には、現実の世界がスタートポイントになるので、理想化してモデル化した上で、現実に合わせるためにはどのようにモデルを修正するかとか、あるいは人間の知恵をそこにどう組み入れて実務で運用するかを別途考えないといけない。理想と現実をどのように近づけるかは、実務では非常に重要なテーマになっている。（池森）

C：結局、数学的なことはリスクヘッジというか、リスクの評価をどうしていくかということか。

A：ベースは、理論的な評価モデルをどう修正するかということだと思う。それがそのままリスク管理につながってくる。価格変動をリスク管理としてどうしたら良いか。それは投資の問題にもつながってくる。現実を加味して投資をどのように組み立てていくかが重要になってくると思う。（池森）

　ノーアービトラージだけど、完備ではないという市場なので、数学モデルとしてどう扱うかというのは、やっぱり大事な話になる。中立測度で期待値を取れば価格というように、数学のモデルを紹介したが、現実は完備ではないし、数学モデルも沢山ある。その時どうするかというと、ひとつやられた研究は、エントロピーを計算する。エントロピー最小のものを良いものだとすることを提唱した人がいた。現場で役に立つのかどうか全然わからないが、たくさん中立測度がある中で、どの中立測度で価格を期待値として計算するかというのは、数学屋としてはそれぐらいの答えをとりあえず思いつく。（谷口）

　エントロピーの話をしたら、現場のトレーダー達は悲鳴を上げるかもしれない。（池森）

　そう思う。あまり何か役に立つ話じゃないと思うが、池森氏が紹介した、さらに抽象理論へ向かっている間違った方向の進展だと思う。（谷口）

Q：最後の図で、理論から現実に戻るというのがあるが、他の技術でも大事で、理論的にどういう限界があるかを見た上で、システムに実装するということになる。ところが、その実装技術が必ずしも理論どおり動いてくれるかというと、そうはいかないので、現実がついてこない事があると思う。そうすると、もう一度モデルに戻してやらないと、うまく動かない。つまりこれは、巡回のように、金融業務から数理モデルへ行って、その帰結を実装して、その実装をもう一度数理モデルに回すような話も入っていた方が良いと思う。

A：ご指摘の通りだと思う。例えば典型的な例で言うと、金融商品やデリバティブのプライシングモデルに色々な要素を考慮して、谷口氏の話にもあったような、複雑なモデルを組むと、1日に1回必ず再評価をするうえで問題になる。つまり、価格を再計算して、その日の評価損益を計算する上で問題になる。これをrevalueと呼んでいるが、それに耐えられる高速の計算ができるかが非常に重要なテーマになる。

モデルができても、ひとつの商品を計算するのに20分も30分もかかってしまうとすると、2000もポジションを持っていたら、1日では終わらなくなる。計算を早くする方法を別途見つけたり、もう1回戻って改めて条件をいくつか緩やかにして、モデルを組み立て直さないといけないと思う。スライドの説明図では、抽象化するところでは現実をよく見ながら、モデル化するというストーリーであるが、ここから数理理論を展開して非常にきれいに展開されたとする。そうすると、そのまま無条件に世の中が間違っていて、モデルが正しいと感じてしまう。ご指摘の点も含めて、もう一度この再翻訳のところでのチェックが必要だと思う。

Q：数理の問題として、協働の強化に向けて、数理の問題として定式化できる人材という話があったと思うが、具体的な取り組みはあるか。実際の現象を数理的な問題に落とし込むための議論ができるような状況や場など、何か取り組みはあるか。

A：実務から離れて10年ぐらいになるが、当時、筑波大学の先生に、我々の抱えている問題を最適化の問題として依頼したことがあった。口頭で一生懸命、やりたい内容を伝えたが、結局、あまり理解してもらえなかった。視点の違った回答が返ってきた。その時、痛感したのは、言葉ではなく数式に書き下して問題を提示することが大切だったということである。現場では、習慣で、なんとなくの言葉で指示していることが多かった。池森自身も「最適化モデルを作ってほしい」という依頼を受けたことがあるが、「最適化というのは、制約条件が何で、どのような関数をmaxにしたいと考えているか」と尋ねると、「要するに最適化です」ということになり、話が進まなかった経験がある。思っていることを数式にできる人を現場に育てないといけない。そうすると、現場には数理ができる人を入れていくことが必要になる。業務として直感としてわかっているようなことも、数式として表現しようとすると、わからなかったりすることもある。そのためには、数理が判る人が現場に入ることが重要なことだと思う。

Q：金融業界は企業なので、全部はオープンには出来ないと思うが、どういう取り組みがあれば嬉しいか。例えば、こういう問題があるが、頭をつき合わせるとか。例えば、筑波大学に持って行った話があったが、そういう場が設定されると嬉しいというのはあるか。

A：人材を育てるのももちろん必要になる。今後の課題となると思うが、例えば今の時点で実際に取り組める事として、数理学者と金融実務の担当者との交流みたいなものが考えられる。実際にそういう場があると、課題が促進できそうだと思う。実際、池森は統計数理研究所の統計思考院というところに週に1回ほど行っているが、そこに「共同研究スタートアップ」という仕組みがあって、現場でも大学でも良いが、もしくは研究所でも良いが、ちょっと困った統計的な問題があったときに、無料で1時間から1時間半位、相談に乗れる場がある。そういう場があると良いかと思う。相談を持って行くにも、現場でも相談を持っていけるような人が必要になる。そうでないと逆に相談を受けた数学の専門家も困ってしまう。（池森）

Q：ちょっと間に入れるような人が欲しいということか。

A：定式化まではいかないとしても、上手く数理の言葉である程度、こういうことがしたいと言える人材が育つと嬉しい。（池森）

3.5 力学系、安定性、制御と感染症の数理

［概要］

CRDSの連続セミナーシリーズ、今回は「力学系、安定性、制御と感染症の数理」をテーマとして、國府氏より「力学系理論：ダイナミクスの数学」、稲葉氏より「感染症の数理」というタイトルでご講演いただいた。

國府氏の講演は「決定論的法則に従って時間変化する系」という数学としての力学系の話から始まった。高校時代に我々は既に漸化式やNewton力学という形で力学系の一端に触れている。様々な自然・社会現象が力学系による記述で理解され、逆に自然・工学的現象の深い理解が力学系理論を発展させてきた。実例を挙げ、随所に力学系理論と社会の発展の相互関連性を強調された。

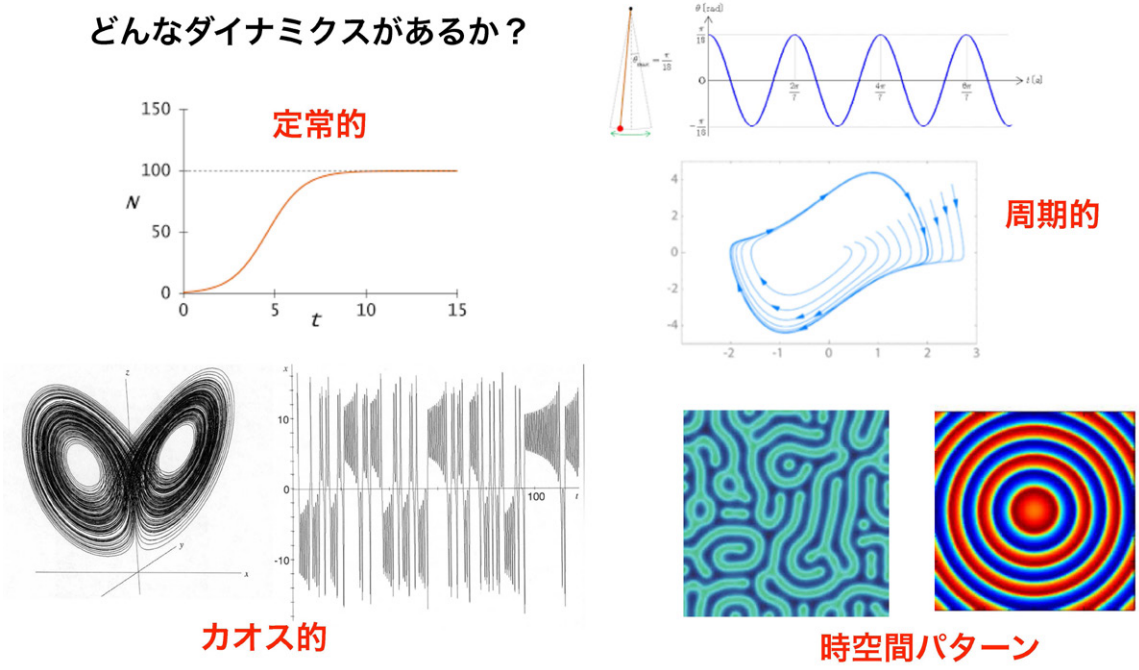
稲葉氏の講演は「感染症」の研究の歴史から始まり、感染症を数理的に扱う際の重要な因子である「基本再生産数」、SIRモデルに基づく力学系としての感染症ダイナミクスの話につながり、スペイン風邪、COVID-19への適用例なども含めてその有効性：流行動態の因果的理解における有用性が展開された。

［Key point］

・ダイナミクスに関係する現象は世の中にたくさんある。その基礎的考え方は高校時代に（漸化式や物理という形で）既に習っている。力学系理論の基礎を学ぶことは理系に必須の素養の1つといえると思う。（國府）
・数理モデル、特に力学系理論は「動態の因果関係」を理解するのに有用である。（稲葉）
・数理に基づく社会研究を実施できる体制を作っておかないといけない。何かが起こったときに現象そのものやその原因、予測をできるようにし、損害を最小限に調査できる体制を作るべき。それには教育が重要。（稲葉）

［國府氏講演：詳細］

Poincareの発見（1887年のスウェーデン国王オスカーII世の懸賞問題の一つであるNewton力学に基づく多体質点系の振る舞いの解明に対する解答として提唱）により力学系に潜む「複雑さ」が数学的に本質であることが認知され、力学系理論の原点となる。その後（1950年代）太陽系の安定性の解明を目指す中で



発見された KAM 理論として知られる複雑なダイナミクスの「閉じ込め（安定性）」が宇宙探査における衛星軌道の計算に応用されたり、電気工学（van der Pol、1927 年）に端を発する複雑な振動を伴う振る舞い（Cartright-Littlewood による力学系としての研究、1945 年）が、複雑でありながら「構造安定」なダイナミクスの幾何モデル、すなわち　Smale による「馬蹄」力学系の発見につながるという歴史的展開があった。他には Lorenz による気象予測の研究から発見された非線形系の原理的予測不可能性、電気工学における上田睨亮や数理生物学における May によるそれぞれの分野の数理モデルにおける「カオス」的振る舞いの発見、それらを包括する Yorke によるカオスの特徴付けの研究など、自然科学や工学の多様な現象と力学系理論の相互作用による発展は随所に見られる。

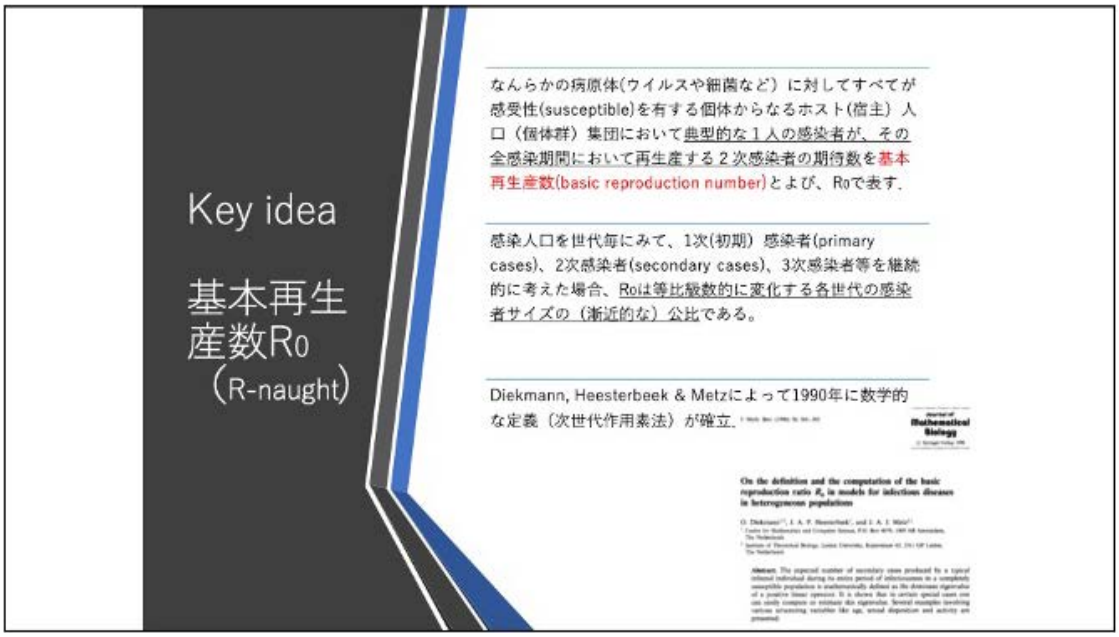
力学系理論そのものは「決定論的」、すなわち初期時刻での状態で未来の状態が一意に決まる系を対象としているが、力学系理論は決定論的システムに潜む「予測不可能性」を明らかにして、今日の複雑系科学の発展の基盤となった。他にも生物学・生命科学におけるパターン形成、結合振動子系における同期現象、遺伝子制御ネットワークなど、現象解明研究における力学系理論の貢献は大きい。

「どこにでも現れる力学系」をキーワードに、その研究の重要性を説いて氏の講演は終了した。

［稲葉氏講演：詳細］

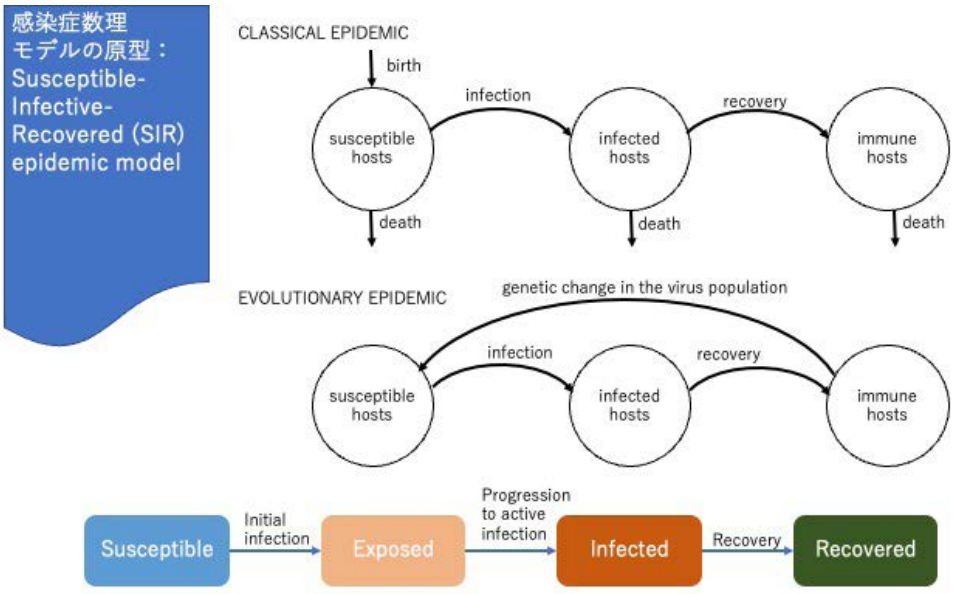
感染症は社会最大のリスクと言われており、近年の COVID-19 が特別なわけではなく、古くは 1918 年ごろからヨーロッパを中心に大流行したスペイン風邪、1980 年代より流行している HIV があり、マラリアなど多くの死者を出す感染症は現在も多数ある。「新興型」として SARS や BSE、「再興型」として結核も感染症に含まれる。

感染症の力学系としての理解は 1920 年代から 1930 年代に、Kermack と McKendrick により提唱された。直前に大流行したスペイン風邪において、そもそも何が起こっているか、すなわち流行動態が不明であった。これが動機の一つとなり、提唱された内容は微分方程式モデルを用いて流行ダイナミクスを因果的に理解できるという革新的アイディアであった。力学系としての感染症を理解するための重要なパラメーター：「基本再生産数 R_0」は感染者 1 人が何人 2 次感染を生み出すかを記述し、典型的にはこれを閾値として流行動態は大きく変化し、その様子は力学系理論で説明できる。系の外から感受性を持つ（感染の可能性がある）対象を取り入れることで、非自明なエンデミック定常解を生じうる事も説明できる。



R_0の（数学的）厳密な定義は1990年、Diekmann, Heesterbeek, Metzにより与えられた。しかし各感染症に対するR_0の値そのものは生物学的因子だけでなく社会的因子も絡むため、その値の導出は自明でなく、ばらつきも多い。2000年代からはデータに基づいたR_0の推定が世界的に行われているが、日本のデータサイエンス、疫学分野はR_0を始め、「データから重要な特徴量を抽出する、またそれを活用する」動きについて非常に遅れをとっている。

感染症の基本モデルであるSIRモデルにより、流行初期の感染動態（指数増大的振る舞い）や最終的な感染者数とその推移を数学的に記述できる。 COVID-19にも修正モデルにてその様子を提唱しているが、モデルにおける総人口、PCRをはじめとした検査や隔離政策についての応用など、社会的側面によって結果が大きく変わり得るため、議論の余地は多く残されている。対して体内感染に関しては社会的側面を極力排除できる実験レベルでの系となるため、SIRモデルで信頼できる結果を得られる。



【質疑応答】

Q：國府氏の最近の研究について。

A：力学系を構成する数理モデル、すなわち微分方程式や写像の反復（漸化式）などが陽に与えられているとは限らないシステムの振る舞いを力学系の観点から説明する「モデルフリー・ダイナミクス」の研究。対象を観測して得られる時系列データがある力学系で支配されると仮定し、力学系の解空間構造の要素分解の一種である「Morse分解」を求めて、系の大まかな振る舞いを記述する。生命現象や気象データに適用して、一定の成果を得られている。

Q：「系が決定論的に振る舞う」ことがわかる指針はあるのか？

A：実際の現象は多かれ少なかれ確率的因子（ノイズなど）を含むと考えられるので、決定論的とするのは1つの理想化である。むしろ系が決定論的に振る舞う、すなわち微分方程式などで記述できると「仮定」し、そこから何が見えるかを考察、比較するという現象理解のための一つの切り口として力学系理論を適用するという考え方が妥当であろう。

Q：「観測がgeneric」であることの目安はどう確認するのか？

A：Takensの時系列埋め込み法は内在する漸近挙動を仮想的空間に実現して記述する方法であるが、観測時系列が特殊でないという仮定の確認は必要であり、それがgenericという仮定の意味である。数学的に厳密に定式化されている状況であれば、その仮定の確認が可能な場合もあるが、現実のシステムから得られる時系列からのダイナミクスの解析の場合には、得られた解析結果と現象の比較などで判断するしかないと思われる。

　関連して、Takensの方法も含めて従来の時系列からのダイナミクスの解析法では、単一の長い時系列を用いることが多いが、國府氏の時系列データからMorse分解を得る方法では多くの初期データに対する多数の「短い時系列」を用いており、得られる結果も不安定なダイナミクスの情報も含むなど、従来の方法と補完的なものとなっている。

Q：SIRモデルの解析によると、感染者はある地点でピークを迎える。日本でのCOVID-19流行第1波ではピークに達するまでに緊急事態宣言で抑えたとも解釈できる。その際流行初期は感染者数が指数増大していて、第2波、第3波でも同じような（初期）動態となっているのか？

A：おそらくそう。

Q：GoogleのCOVID-19感染予測はどのように？

A：SEIRモデル（SIRモデルの拡張版）を基礎としている。各種パラメーター推定は機械学習。非常にマンパワーとリソースが必要となる試み。

Q：社会経済の相互作用のモデル化はあるのか？

A：経済との相互作用モデルにおいては多く仕事がある。感染症対策において、これらを考慮していくことは重要であるが、もともと研究者が少ない。さらに社会現象のモデル化は「過去とのマッチング」は多くなされているが、「未来予測」はさっぱりである。

政策などの「非薬剤的効果」の定量化も、社会現象のモデル化としてCOVID-19がらみでは重要となる。

C：「情報を受け止める側のリテラシー（数理的素養）」が重要。出てきた情報が不完全でも、それを吟味し解釈できる素養を一人一人身につけるべき（稲葉）

C：（Key Point 2つ目について補足）例えば感染症数理に携わる研究者でも、目の前、短期間でローカルに起こることにフォーカスしがち。国レベルなどグローバル（マクロレベル？）な視点を持った議論をできる人材を増やさなければならない。感染症は近い将来克服できるという楽観が支配した時代もあったが、そうした見方は過去のものになった。今後も感染症が社会の最大のリスクの一つであり続けるだろう。（稲葉）

3.6 Uncertaintyと数学

［概要］

CRDSの連続セミナーシリーズ第6回は「Uncertainlyと数学」をテーマとして、竹村氏より「統計学や確率論における不確実性のとらえ方」、恐神 氏より「強化学習の数理と応用」というタイトルでご講演いただいた。

竹村氏の講演「統計学や確率論における不確実性のとらえ方」では、前置きから昨今のコロナウイルス関連で2つの過誤の説明があり、さらに不確実性とその有用性からゲーム理論的確率論への言及も行われた。

恐神氏の講演は、「強化学習の数理と応用」として、強化学習の紹介をゲームからはじめ、現実社会での利用、解決すべき課題にも言及された。強化学習の原理については、多くの例（ゲーム、実用例）、これからの応用を見据えた例も交えて紹介された。

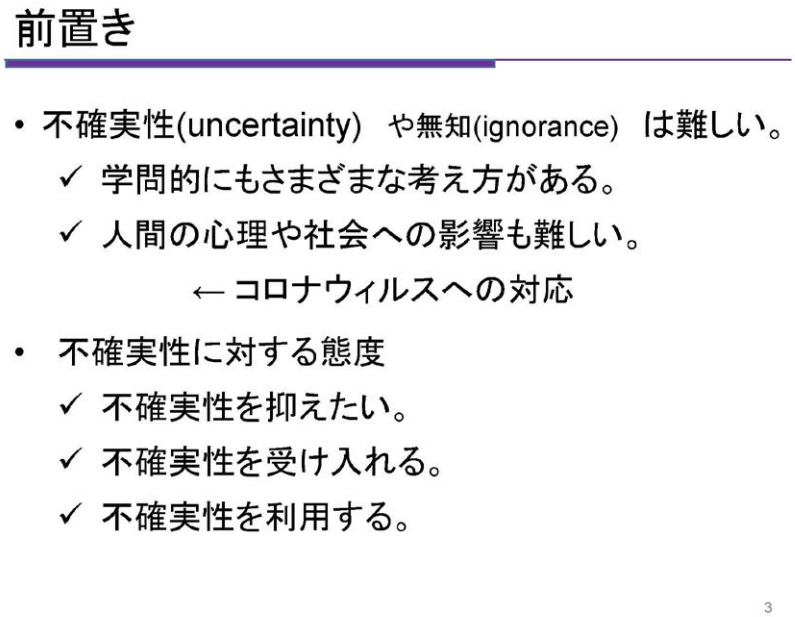
［Key point］

・コロナウイルス関連は、統計データとしても注視している。（竹村）
・不確実性には、有用性もある。（竹村）
・強化学習の技術は、ゲームから現実社会も見据えた研究開発が肝要である。（恐神）
・強化学習の数理には課題も残る。（恐神）

［竹村氏講演：詳細］

竹村氏からは前置きから始め、コロナウイルス関連、不確実性、不確実性やランダムネスの有用性の紹介があり、最後にゲーム理論的確率論にも簡単に触れられた。

はじめに前置きから、コロナウイルス関連、不確実性に関する論点について述べるとのことであった。加えて一般的なこととして、不確実性にはさまざまな側面があり、有用性もあったりするので、それらの側面に興味を持っているとの旨であった。最後にゲーム論的確率論は簡単に紹介するとの話であった。今日の講演を準備している時には、竹村氏自身でもわからないことが沢山あると感じており、色々な文献を見てみたが、様々な議論や流儀があるのが実情であろうとのことであった。学問的に明確に捉えられていないところもあるので、



今日の話は参考として聞いて欲しいとのことであった。学問的にはまだまだ未熟だが、一方で社会的、とくに昨今のコロナウイルスや不確実性に対する人間の反応など、社会への影響は大きくなっている。

不確実性は、基本的には制御したい、抑えたいということになる。抑えられないときは、諦めるかもしくは受け入れる態度もあるし、一方で、不確実性を積極的に利用するという面もあるので、色々なトピックについて、触れてみたいと思う。

最初、コロナウイルス関連について、日経の経済教室に記事を載せたが、その時に、3つのポイントを書いた。

1つは数理モデルである。感染症のモデルで、標準的にSIRモデルと呼ばれているが、現象を数理モデルで捉えていくことで、具体的な計算ができて、予測ができるという一般に理解されたことが非常に良かったと思う。一方で、数理モデルに対する批判めいた論調もあるが、竹村氏は有用性が理解されたことが良かったと思っていた。気になったのはPCR検査の精度で、当初は非常に不確実性が高く、議論もちょっと極端に振れているところがあった。状況が徐々に分かってくれば、人々が「正しく恐れる」というか、合理的に行動できると思うが、とくに初期的には非常に混乱を起こす。トイレットペーパーの買占めなどの不合理な行動も起きたので、「不確実性」に対する人々の反応も気になる。

ここでは、PCR検査の精度について述べたい。統計では統計的検定論の枠組みになるが、2種類の誤りがある。PCR検査でも、本当は感染していない人を感染していると判定する誤り（偽陽性、誤検知）。逆に、感染しているのに陽性にならない誤り、つまり見逃しがある（偽陰性）。これらがどのくらいの率で起こるかが問題になる。実は、この辺の用語も沢山あり、感度や特異度など疫学の方で使われる用語があり、分かりにくい面がある。 PCR検査では偽陽性率は低い。しかし、偽陰性率、つまり見逃しは3割ぐらいあり、陽性なのに陰性と判定されることが結構ある。一方で、ウイルスがいないと検知できないので、検知されれば殆ど本当に感染している。つまり誤検知は少ない。

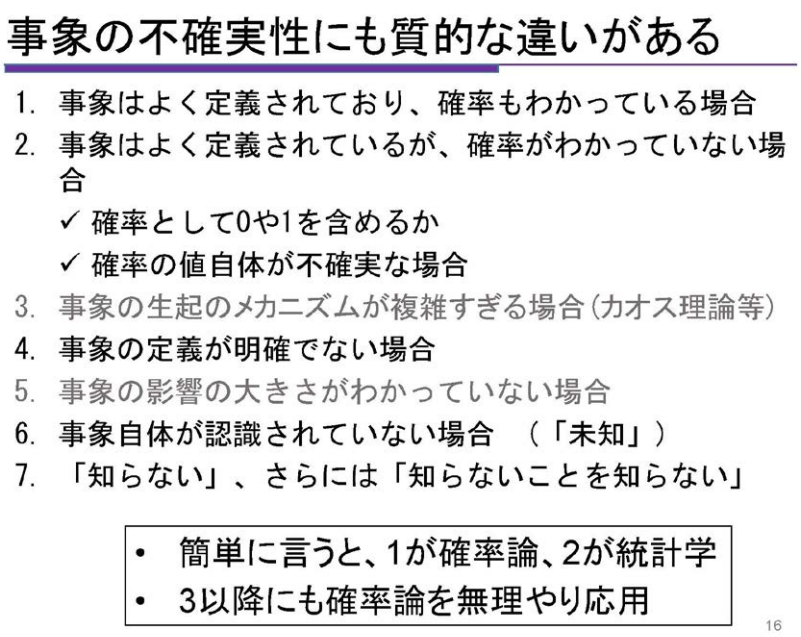
偽陽性率と偽陰性率の数値は1%や30%などとされており、スライドには政府の数値が紹介してあるが、実は、7月6日の政府の第1回新型コロナウイルス感染症対策分科会資料では、数値を「仮定する」と記載してあった。これを見た時に、政府の文書で「仮定」と言うのはまずいと少し感じた。そこで、文献を調べてみた。偽陽性率は実は非常に小さいと考えられるので、報道を見てみると、例えば香港では10万人以上調べて陽性と判定されたのは6人だったということで、誤って陽性としてしまう偽陽性率は非常に低く、恐らく0.01%くらいである。偽陽性率については率よりも、大勢を検査して何人ぐらい陽性になるかの絶対数が問題になる。このことをなぜ気にするかは、1%も偽陽性が出ると、本当は陰性の人も無駄に隔離することになり良くないという議論があるからである。実は日本ではPCR検査を拡大しない理由として、1%という数値が使われたことがあった。ここが非常に問題だったと考えている。

偽陰性は見逃しが結構多く、ウイルス量が少ない場合は検知できないことがあり、何回か繰り返さないと本当に陰性かどうかわからない。残念ながら、仕方ない側面があると思う。偽陰性率や偽陽性率を、コロナ時代で人々が考えるようになったというのは良いことだと思う。そこで少し紹介した。

次は、不確実性とは何かということで、まずは確率論で扱うことになるけれども、実は不確実性というのは、確率よりは広い概念だと思う。スライドでは、通常の確率論の範疇にない本を6冊ほど提示されていた。竹内啓氏の著書や話題になった「ブラックスワン」という本、さらに確率論の哲学的議論も含む本、70年代や古くは経済学では「ナイト不確実性」などが紹介された。ここでは、通常の確率論の範疇外にあるような文献として取り上げられた。

確率論で想定している不確実性と、より広い不確実性があり、スライドでは様々な例が紹介された。教科書的なコインの表裏が同様に確からしく現れる例から、宇宙人は存在するかまで、そもそも確率を考えることができるかが紹介されていた。他にも自分が病気になるかのような主観確率も提示された。一概に確率論が適用しづらい局面があるとの説明があった。

確率論の前提では事象が定義されて、確率も分かっている場合は扱いやすいが、確率が分からない場合な

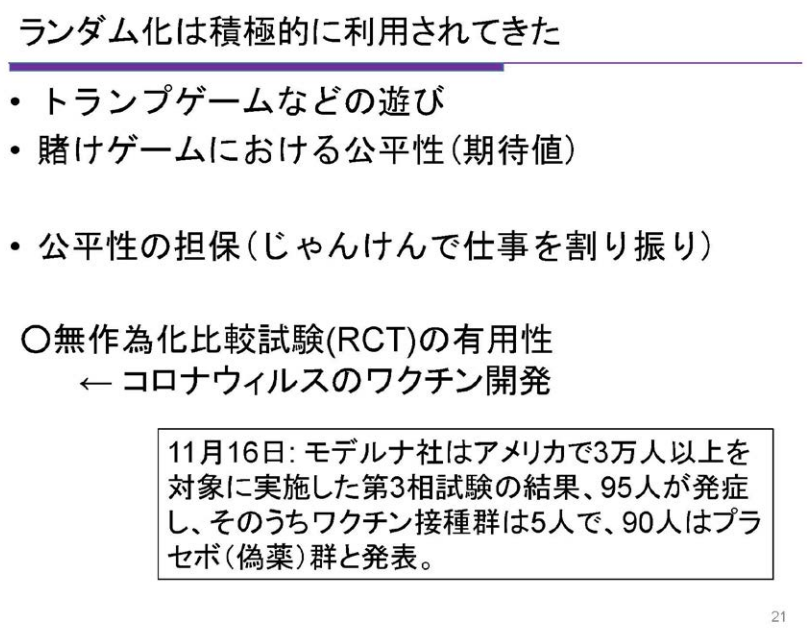


ど、様々であるし、事象の定義が明確でないことも多い。例えば、「汎用人工知能は実現するか」というと、汎用人工知能の定義がなければ、実現するかどうかは言えない。定義が明確でない場合など、不確実性に加え未知の側面も大きくなる。確率というのは、ある意味では定義がはっきりしている場合で、例えばサイコロを投げて１から６みたいに起こり得る事象のリストがあり、通常は確率が割り当てられている。ただし割り当ては古典的な頻度論であったり、対称性を用いている。中高等学校では、組合せ論と対応して一緒に教えるが、同様に確からしいという設定することが基本で、そこから外れると主観確率が出てきたり、賭けの公平性から見たりもする。また頻度論と主観確率（ベイズ）の違いも問題となる。頻度論と主観確率（ベイズ）で哲学的にはかなり違うが、実際の計算はあまり違わないことが多い。古典的な確率の想定があって、かつ頻度論が成り立つような局面では、大数法則、中心極限定理が標準的な枠組みになる。一方、外れた状況ではべき則とか“heavy tail”がある。 Dempster-Shafer理論は確率から外れてしまうが、例えばコインを投げたとき表が出る確率の1/2の値にさらに不確率性を考えるようなこともする。

不確実性は通常は避けたいものであって、例えば明日の天気予報の精度が上がれば、傘を持って行かなくて済む。あるいは典型的な保険などでは、リスクは避けられないけれども、沢山の人をプールすることでリスクを制御しようとする。逆に、不確実性とランダムネスの違いを考えていくと、ランダムネスはより積極的な有用なものとして扱われる。

もともと確率論はカードゲームやサイコロを使った賭けのゲームで現れたものである。例えば、じゃんけんで仕事の割り振りするなどは、公平性を担保する一つの方法である。無作為化比較試験（RCT）は、非常に有効だと思う。最近の１週間ぐらいの報道では、コロナウイルスのワクチンについて、モデルナ社の発表では３万人以上を対象にワクチンをわりつけて、そのうち95人が発症したが、その中でワクチンを接種した人はわずか５人だった。要するに、ランダムに割り付ければ、ワクチンの効果だけが残ってくるので、非常に有効な無作為化比較試験になる。ゲームから確率論が出てきたが、ゼロ和２人ゲームのところで、ミニマックス戦略という形で現れるのも一つの例である。そこには哲学的な何かimplicationもあると思われるが、遺伝法則が確率的であるというのが、メンデルの法則や進化論として出てきた。なぜ生物の進化にランダムネスが現れているのかは、戦略としてランダム化が有利に働いているからと思われる。ミニマックス戦略では、相手が非常に強く賢い場合に、自分としてはランダムに手を選んだ方が相手から読まれないという利点がある。

環境の変化があるときにランダムネスによって、多様性を保持することが種の生き残りにつながる。つまり、



ランダムネスは生存戦略と考えられる。一方で、別の観点では、最近ではベイズ法が典型的だと思うが、工学的な応用で確率計算を積極的に用いる。１番いい例は、かな漢字変換だと思うが、機械が意味を理解する必要がなく、確率の計算によって、高い確率の変換を出力する。それが有用であれば良い。その時の確率の計算というのは、エンジニアリングの意味であり、哲学的な意味などは不要と割り切ったところが良いと感じる。役に立つ場面が増えたので、注目されているのかと思う。

最後に、ゲーム論的確率論に関して、2001年に本（Shafer-Vovkの本）が出版されて、改訂版も出た。背景を述べると、コルモゴロフの公理的確率論（1933年）があり、これは非常に成功した数学理論と思う。なぜ成功したかというと、確率に関する哲学的な議論を公理系に置き換えて意味を問わないことにした。確率の意味を問わないところが、逆説的だが、非常に有用だったと思う。主観論や頻度論には無関係で、確率の操作を公理化した。ただし、いい面と悪い面があって、悪い面としては、確率の意味を全く考えないということに繋がったかと思う。コルモゴロフは自身の作った理論に対して満足できなくて、コルモゴロフ複雑度を提案することになる。

典型的な例だが、コイン投げで

10101010101010101010

10101001100111001101

上の列は表裏が規則的に表れている列。下の列は擬似乱数の列。確率論的に言うと、どちらの列の出現確率も$1/2^{20}$で区別がつかないが、一方で上の方が規則性があって、同じ確率を割り当てるというだけではランダムネスを捉えられない背景がある。それに対してランダムネスを捉える議論というのは、１つの列に対して賭け事をするときに儲けられるかが出自になっている。賭けゲームからゲーム論的確率論が出てくるというので、測度論が不要で面白いと思う。ゲームから確率で出てくるのは、ゼロ和２人ゲームにも似ているが、極限定理が現れる所が違っていたりすることも補足して講演を終わりにしたい。

［恐神氏講演：詳細］

恐神氏の講演は自己紹介から始まり、博士号はコンピューターサイエンスで取得しているとの紹介があった。数字を武器として研究に取り組んでいるとの導入もあった。2013年からビッグデータCRESTの主たる共同研究者を務め、現在は、社内で2つの研究プロジェクトをリードをするとともに、数理科学分野の基礎研究を推

進する役割も担っている。 IBMの研究部門は、世界で3千人ぐらいからなる組織で、5年先や直近のビジネスに貢献するというミッションも当然あるが、10年先のビジネスを作っていくことが期待されている。これまでも、新しいビジネスを作ってきた組織である。基礎研究は、コンピューターサイエンス、physical science、Math. Science、最近はCOVIDの関係でLife Scienceも重要な領域になっており、systematicにプロジェクトを募集して基礎研究を推進している。このような枠組みの中で、恐神氏がリードしているプロジェクトの一つは基礎研究のプロジェクトで、もう一つはAIになる。

今日は、強化学習に関する話題で、特にゲーム理論との融合というところを講演する。竹村氏もゲーム理論による確率の定義の話をしているが、この講演ではゲームが2つの異なる意味で出てくるので、気をつけながら紹介する。今日のテーマの強化学習が、どのようなところに使われているのか、どのように発展してきたのか、実社会や産業のどのような所に応用できる可能性があるか、どういう課題があるか、その辺りを中心に話をしたいと思う。その中で数式は書いていないが、役に立っていくのではないかという場面をちりばめて紹介する。

まず記憶に新しいところでは、囲碁のチャンピオンにAIのプレーヤーが勝利した。これには強化学習や関連技術が使われている。もう少し遡ると、ビデオゲームであったり、少し単純なバックギャモンというゲームであったり、そのような所で強化学習の技術が大きく成功し、人間を超えるようなパフォーマンスを出すことができている。スライドには、同じようなタイプのゲームである、ポーカーとチェッカーも載せてある。これらのゲームで使われたのは探索ベースの手法だが、関連技術として掲載した。強化学習や探索の技術はゲームを解くということで発展してきた。

強化学習技術は、産業でも大きな成功を収めた例がいくつかある。カタログ送付や徴税支援は後でまた詳しく紹介したいと思う。カタログ送付は1960年頃の事例だが、強化学習の最初の手法の一つである方策反復法は、この実問題からの要請に応じるために出来た技術である。これらの強化学習の応用例に共通するのは、本日のテーマに合わせて言うと、不確実な環境の中で、逐次的に意思決定をしていく必要があるということである。

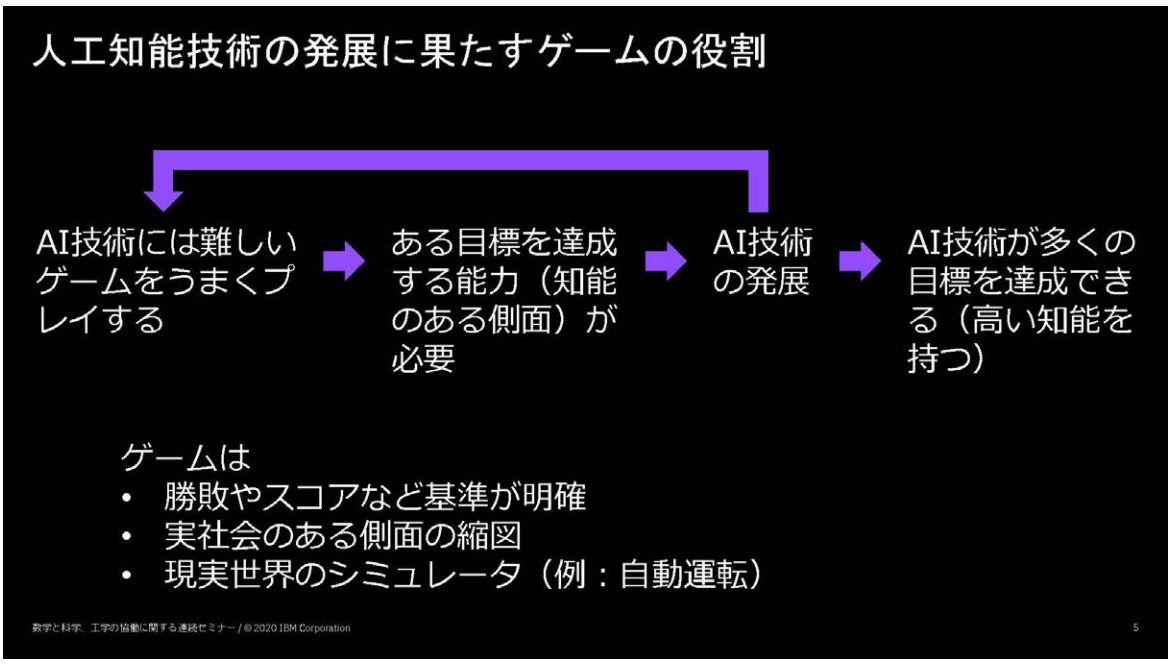
スライドのスーパーボンバーマンRというゲームは、爆弾を置いて壁を壊し、出てきたアイテムを取得して新しいスキルを身に付け、そのスキルを使って敵を追いつめ、最終的に敵を倒して勝利する流れになる。最終的に目的を達成するために今何をすればいいのかを、逐次的に各ステップで決めていくゲームになっている。カードゲームであったり、ボードゲームやVideo Gameには、勝利するとか高いスコアを達成するとかの目標があり、そういう目標を達成できるようにすることで、強化学習や人工知能が発展してきている。ある目標を達成するには、知能のある側面が必要である。つまり、そのような知能を機械が獲得していくことを意味している。

AI技術が多くの目標を達成できるということは、知能の多くの側面を持つということで、高い知能を持つということもできる。ゲームがＡＩ技術の研究に利用されているのは、スコアなどの基準が明確で、どの程度技術が発展したのかを正確に把握しやすいからである。加えてゲームは実社会のある側面の縮図になっていて、ゲームで良い性能を出せれば、実世界でも何かの役に立つと考えられる。これに関して、ゲームの専門家からコメントがあったので、専門家の生の声を紹介したいと思う。

ただの縮図というだけでなく、最近のゲームというのはすごく高度な技術を使って、本当に現実の世界をゲームの中に丸ごと再現する。そういった技術が結構AIの研究開発に役に立っていて、最近だと自動運転のAIを学習させるために、そのゲームの世界を使ったりするという事例がありました。あと、ゲームというのはこう現実世界だけではなく、非現実的なSF的だったり、ファンタジーだったり、未来の世界とか、そういったものを再現することもできるが、そういったところ、今存在しない、でもすごくリアリティのある世界を作れるということは、そこでAIがまだ現実には起きてない状況に対処できるようになる。

このような形で、ゲームと実世界は密接に結びつくことができる。

強化学習技術は、データに基づいて意思決定を最適化する。強化学習のためのデータは、どの状態で、ど



の行動をとり、その結果どの状態に遷移して、どれだけの利得が得られたかの情報からなる。そのようなデータに基づいて、それぞれの状態でどういう行動をとると、目標を早く達成できるのか、利得の積算値を大きくするにはどうしたらいいのか、そのような行動の方策を導いていくのが強化学習技術である。

ここでは論文の紹介も行われた。

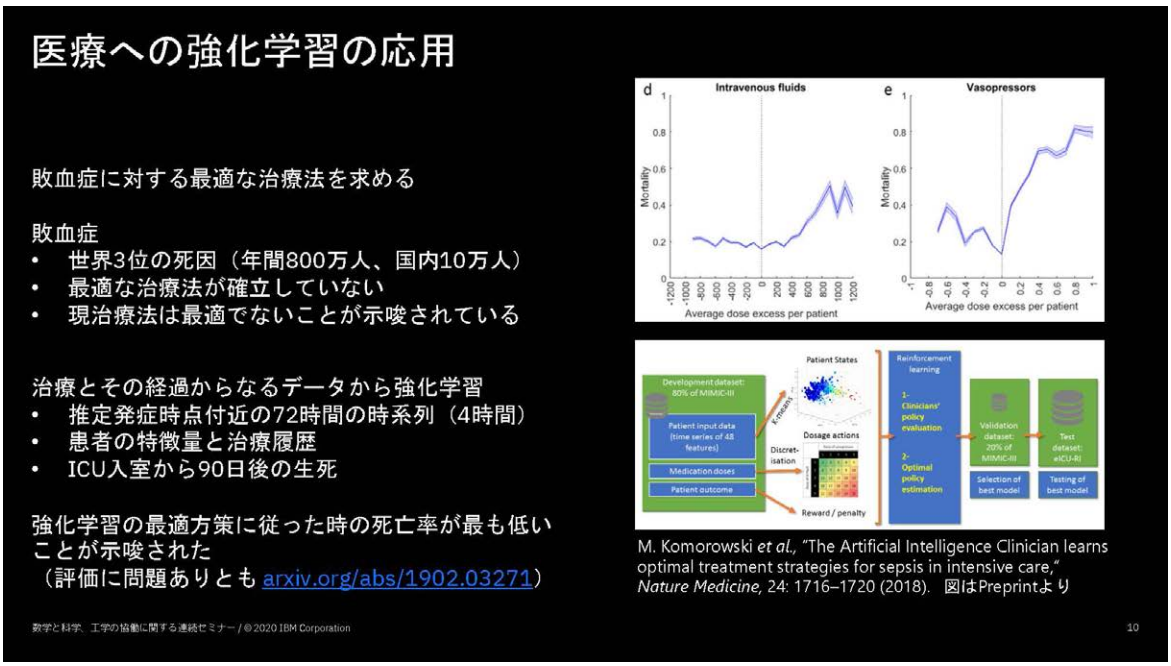
Y. Li, Reinforcement Learning Applications, arXiv:1908.06973

強化学習の応用先としては、さまざまな産業が挙げられていて、いくつか紹介したいと思う。まず1960年に強化学習の技術ができるきっかけとなった事例として、アメリカのデパート、シアーズの例が紹介された。カタログを顧客にカタログを送付するとき、カタログを送付すること自体はコストがかかるだけで、利益が上がるわけではないが、顧客のデパートに対する興味を惹きつけたり、状態を変えて、将来的に顧客がデパートで買い物するようになることを期待してカタログを送付する。1960年以前は、購買履歴を記録していて、直後の利益だけを考慮していたが、強化学習的なアプローチでやったことは、将来に利益をもたらすようになる効果を考慮して最適な方策を導出することであった。

2010年頃には、ニューヨーク州で行われた、延滞税を徴収する事例がある。税金を徴収したいが、突然訪問したりは出来ないので、あらかじめ手紙を送ったり、電話したりなど、税金徴収を実施するために、予めやっておかなければならないことがある。そのようなことを考慮しつつ、訪問したり電話したりするのに必要なリソース（人）に関する制約も考慮して、将来的に延滞税の徴収額を最大化したい。そのような目標を達成するように、延滞している人や未納の人に対してのアクションを最適化した事例になる。2010年の延滞税の徴収額の増加分が83 million dollarsということで、100億円近い増収になった。

さらに最近注目されているアプリケーションに医療が挙げられる。敗血症と呼ばれる病気があるが、最適な治療はまだ確立していないし、今行われている治療は最適ではないとも言われている。病気が発症してから、どのように治療をしていくと、90日後に命が助かったか、または助からなかったか、というデータから、生存確率を最大にするように、患者の状態に応じて、治療の最適な戦略を強化学習で導くという研究がある。 強化学習によって導かれた方策が実際に良いかどうかを評価することが実は非常に難しい。評価に問題があるのではないかと言われていて、恐神氏もそのように思っているとのことであった。

さらに、新しい材料を見つけたり、金融ではリスクを考慮した投資などにも、強化学習の応用が検討されている。先日のファイナンスの講演でも似たような話があったが、AI的なアプローチの特徴として、ニュース記



M. Komorowski *et al.*, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," *Nature Medicine*, 24: 1716–1720 (2018). 図はPreprintより

事なども入力データとして活用して、アクションの最適化に活かしていくことも行なわれ始めている。

色々なアプリケーションが考えられているが、現実世界における成功は、ゲームにおける大きな成功と比べると、まだ非常に限定的だと思われる。何が難しいのかというと、まず試行錯誤が現実世界においては難しい。学習データを能動的に収集する、特に失敗しながらデータをためていくことが難しい。データの量も限られている。 実際に応用しようと思うと、特に医療やファイナンスでは、高い信頼性がないとなかなか使えない。つまり、リスクを考慮する必要がある。また、ゲームは閉じた世界で、現実世界はオープンになっている。加えて非定常でどんどん変化していく。例えばファイナンスであれば、規制が変わると思う。さらに、使う側の立場からすると、失敗したときになぜ失敗したかとか、説明可能性が重要な要素になっていると思う。

まとめでは、強化学習は難しい問題を解こうとしているという説明がなされた。

マルコフ決定過程を背後にあるモデルとして考えているが、強化学習は、このモデルが未知である状況で、最適方策を求めようとしている。マルコフ決定過程の最適化方策を求める問題自体が難しい問題のクラスになっている。特に、無限期間の部分観測マルコフ決定過程において、ある一定以上の累積期待報酬を達成する方策が存在するかを判定する問題は決定不能な問題として知られている。解くことが出来ないことがわかっているような問題のさらに難しいものを解こうとしている。

また、使われている手法に関しても、例えばDQNというビデオゲームで成功した手法は、収束しないことがわかっている。現状うまくいっている手法がなぜうまくいったのかわからない上、実はうまくいかないことがあることもわかっているのに、使われているのが現状になっている。

現実世界で応用して行くためには、この状況を改善して行く必要があると思う。数学は、このような課題を部分的に解決してきている。収束の保証や、ほぼ最適なものの最適性の評価など、これらのことに、数字的なアプローチをとることがさらに必要とされている。

最後に、ゲーム強化学習の技術はゲームによって発展してきた。現実世界でも、いくつか大きな成功を収めているが、ゲームでの成功に比べると限定的ではある。幅広い応用が期待されているので、ここに数理科学によるさらなる貢献を期待している。

【質疑応答】

Q：前回の講演で、経済学者は感染防止と経済効果を考えているが、両者の総合作用を入れたモデルみたいなものは、研究できているか。

A：そこまではちょっとわからない。数理モデル化が、限定的なものでも研究が進むと良いと思う。現象の本質を単純な問題で捉えることがやはり有効だと思う。単純であるが故に、数値的な計算ができて、8割削減すれば感染が抑えられるみたいなこととも言えると思う。一方で、経済的な効果では経済データが充分取れてない。 例えばGoToの効果などが数量的に計算できるかというと、なかなかそこまで具体的なものができていないと思う。そういう意味では、世の中のデジタル化が進んで、具体的に経済的な効果も金額で計算できるようになると良い。人命を守る側面は金銭的に測るには価値観として抵抗もあるかもしれないが、経済効果と合わせて考えるためには、やっぱり割り切って金額で計算するのが良いと思う。そのような議論をするにはまだまだデータが不足していて、具体的な議論がしづらいと思っている。（竹村）

Q：dynamicsというか、データが揃わないとなかなか難しいと思うが、試みている人もいるか。

A：しっかりしたデータが取れるかというのが難しいところかなと思う。色々な効果があるということで、定性的な議論だけでなく、定量的な議論も必要かとは思う。（竹村）

C：SIRモデルで、感染者の削減の意味を人々が知ることが非常に重要だったと思う。定量的なことが、少しでも捉えられれば、皆の考えも新鮮になって良いかと思っている。

A：8割削減することに関して政治的な反発がある中で、そこをきちんと主張していかないと、モデルの良さを主張していかないといけないと思っている。批判ももちろんあるとは思う。（竹村）

Q：大学では、ベイズ検定を教えることになる。学生には感度や特異度も説明する。例えば、診断で陽性だから怖いなど、癌検診でも良くあるが、感度や特異度のデータっていうのは、コロナウイルスに関しても公開されているか？

A：報道で少し調べる限りは、例数が少ない場合もあるし、外国の文献をそのまま利用しているような感じもする。まともに調べていない感じがしており、統計的な調査をして、少しきちんとした数字を得てから議論した方が良いと思う。現場の医者はきちんと調査するより、まずは早く患者を治したほう良いと言うかもしれないけれども。 PCR検査はまだ良いが、抗原検査はさらにデータ少ない感じがする。簡

便にたくさん使えるので利用されているが、偽陽性や偽陰性についてはやや疑問がある。
調べたものを紹介したが、政府の委員会の文章には「仮定する」とあるので、少し驚いた。（竹村）

C：偽陽性なのか偽陰性なのかを正確に判定する手段がそもそもないと思う（恐神）
PCR検査が1番スタンダードで、その他の検査についてはPCR検査のデータが再現できるという議論をしている感じ。 PCR検査は偽陽性は少ないのでその点は良いが、。偽陰性についてはPCRを基準にしてもわからないので、このような議論だけではまずいと感じている。（竹村）

Q：敗血症の話もあって、強化学習というのは、価値を最大化するような行動を学習していることになっている。 AIの成功というのは、教師あり学習が中心で進んできたと思うが、人間の判断によるバイアスもあるので、教師なし学習がとっても重要になると思う。例えば教師なし学習で得られて、その後、数理的な評価もなされるのはどのようなところになるのか？

A：強化学習の扱っている課題では、教師データは直接的には与えられていない。教師あり学習とは違う問題になっている。例えば、勝利したかどうかは教師信号になるが、どのアクションが勝利につながる良いアクションだったのかは直接与えられない。なので、最後に与えられる勝ったか負けたという信号をもとに、この盤面で何をすればよかったのかを導く、もしくは推定しなければならない課題になっている。各盤面で、こういうアクションが良いという情報を与えて、エキスパートの行動を真似るように学習する技術として模倣学習がある。強化学習では、それがなくて、報酬をもとに何が良かったのか、何をすれば良かったのかを導くことになる。（恐神）

Q：強化学習の中に最適化理論のようなものは、色々組み込まれているという風に理解すればよいか？

A：基本的に、価値関数というものを最大化している。そのような問題設定になるので、そこでは最適化の技術が使われている。（恐神）

Q：ボールを蹴っている動画の例で、GAN、日本語で敵対的生成ネットワークとは、どんな関係になっているか？

A：密接な関係がある。 GANでは、Generatorがリアルな例を生成するように学習を進め、Adversaryがリアルな例と生成された例とを判別できるように学習していく。動画のサッカーの例は、強化学習によってうまくサッカーをできるようになった。これに対して敵対的に意地悪な例をつくると、大きく失敗する。そのような例を作って、学習に使うことで、いろんな環境で上手くいくポリシーを作っていくこともできる。そのようなアプローチもあるが、動画の例のように閉じた環境で学習したエージェントは、少し環境が変わると全然うまくいかない。（恐神）

Q：話があったのは最終的に勝ったところで報酬が決まる。そうすると途中の段階で、最初にどの壁を壊すというのはさっき言われた模倣的な学習とか、そういうことを使うのか。その時にどの壁を壊せばいいかというのは、どうなるのか。

A：この問題は難しい。簡単でないが、原理的には勝ったか負けたかという情報だけから、今この壁を壊すべきかどうかを意思決定できるようになる。そのためには、この壁を壊した後、こうなって、結果的に負けたとか、結果的に勝ったとか、非常にたくさんの試行錯誤が必要になる。そういうデータがたくさん集められたとすると、今この壁を壊すべきなのか、それとも別の壁を壊すべきなのか、判断できるようになる。それをできるだけ少ないデータから効率的に学習するのが、強化学習が行おうとしている課題になる。（恐神）

Q：単純に考えると、ものすごくいっぱいやらないと結果が出てこない。

A：単純に考えると、ものすごくいっぱいやらなきゃいけない。そこをいかに工夫するかという問題になる。（恐神）

Q：途中の段階で報酬を与えるようなやり方もあるか。

A：いくつか考えられている。各ステップごとに壁を壊したら少し得点を与えるとか、そのような形で中間報酬を入れていることによって、問題を解きやすくするアプローチもよくやられている。（恐神）

Q：自動的に任せておいて、とんでもない結果になることもあるかと思う。聞いた話では、自動運転でぶつからないように、ぶつかったらマイナスにして上手く走ったら何点とやっていくと良いという話もあった。実際の車でやるとぶつかると大変なのでシミュレーションでやると思うが、それでも結果として上手く行かないこともあると思う。もっとトップダウンで何かルールを与えておくやり方もあるか。

A：強化学習だけで自動運転をするのは、少し危険だと思う。これまでの学習の中で出てこなかった場面に遭遇したとき、何をするかがわからない。わからない時には、わからないというようにして、アクションを選ぶことを放棄して、人に任せることも必要かと思う。（恐神）

Q：ニューヨーク州の税金はどのように学習しているのか。

A：まずは取りうるアクションがいくつかある。手紙を送るとか電話をかけるとか。データとしては、住民のデータがある。税金を延滞している住民がいて、どれ位の額をどれくらい納めていないのかとか、加えて過去にどういうアクションをこの人に対して取ってきたかがわかっている。ある住民に対して、あるアクションをとると別の状態になって、その時に、税金を納めたとか納めなかったとか、電話はこれまでかけたことがなかったのに、電話をかけた状態にかわったとかがデータになる。さらに、いくつかの制約、特にリソースの制約を考慮することが必要になる。電話をかけるオペレーターが限らているので、何人かの人にしか電話をかけることができないなどが、リソースの制約になる。基本的には、この人にどういうアクションを取ると、将来的に少ないコストで多くの税金を取り立てることができるかという課題を解くことになる。（恐神）

C：結局、督促の方法が色々あるということになる。

Q：ATARI2600というのがあったが、一体、何をやっているのか。

A：ATARI2600は、2015年に深層強化学習が流行りだしたきっかけになったビデオゲームである。昔のビデオゲームなので、画面の解像度は粗っぽいものだが、ゲームの種類は30とか50ぐらいある。それらのゲームを、画面の情報と、勝ったとか負けたとかスコアの情報だけを使って、試行錯誤でゲームを繰り返しプレイして学習することで、人を超えるようなパフォーマンスを持たせられるようになった。それも1つの手法で、多くのゲームにおいて、人を超えるようなパフォーマンスを出せるようになったという成果になる。（恐神）

C：ゲーム攻略を一つの手法で行ったということになる。

Q：強化学習では、想定外とか、間違ったことをしでかすこともあるというところを工学的にカバーする方法はあるか。理論的にはまだ難しいにしても工学的というか、危ないことがわかるとか、そういった手立てがないと、なかなか実用は難しいと思うが？

A：どれくらいの確信度でアクションを出しているのかを、一つの参考値として使えると思う。強化学習の手法によっては、アクションを確率的に出したり、確信度を出したりすることができるので、それらの値を参考に、自信がなければアクションはやめておこうなど、自動運転ならば人間に運転してもらった方が良いなどとすることも考えられる。（恐神）

C：確信度に関しては、ある程度信用できるという保証も必要になる。

Q：確信度の話は大事だと思うが、竹村氏の話でも不確実性と確率より広い考え方が、ある程度の行動の指針を与えてくれるとは思う。そういうものがない時にどのようにすれば良いか。やっぱり、シミュレーションなりができないと何もしようがないのではないか。わからない状況というのは、自分が知らないことを知っていれば、ほかの人に判断を任せようということが出来ると思うが、なかなか難しいと思う。確信することがわかるのか。

A：ある程度やろうとはしている。全く仮定を置かないのは難しいが、ある程度もっともらしい仮定の下で、意思決定方策の信頼度を評価するという研究も進んではいる。（恐神）

Q：その安全性の評価、信頼という問題では、竹村氏のいる滋賀大学で、データサイエンス学部のほうでの教育で、この手のことはどのように学生が学べるようにしているのか？

A：そこまで全然いっていない。少し補足になるかもしれないが、先ほど紹介した、カナ漢字変換の例があり、大抵上手くいく。役に立てば良いということになる。かな漢字変換が間違っても、深刻にならないので、一定の精度が出れば良い。そういうところで工学的なアプローチがすごく役に立っているが、失敗したときの損害が大きいようなところでは使いづらいなというところがある。確率ベースの技術が役に立つ場面が注目されていて、そのような場面を学生に教えてしまう。今日のような話はあまり学生に教えていないのが現状である。こういうのも少し教育の中に取り込むべきかなとは感じている。どうしても、役に立つ場面が増えているので、そのような場面に注目されているとは思う。（竹村）

Q：同じ質問だが、今の論点は、社会に実装したときに重要になってくるが、IBMのなかではそういう議論というのは、どのような感じになるか？

A：ゲーム論的なアプローチを、こういう強化学習の枠組みにしっかり入れて行く。環境だとか、車だったらほかの車だったりとかが、戦略的に、自分に対してAdversarialにふるまうと考える。その中で自分がベストに振る舞うようにする。そのためにゲーム理論的にしっかり解析をすると言うか、ゲーム論的な意味の保証を与えるのが1つのアプローチなのかなと思っている。（恐神）

Q：ゲーム理論的に確率を定義して行くと話があったが、それによって、例えばこのKolmogorov的な定義からは言えないようなことというか、できないようなことというのが、何かできるようになるのか？

A：ゲーム論的確率の話題は数学的には面白いが、測度論的確率論と、それほど違う結論が出てくる感じではない。数学的なところは、見方が広がる所があったりするが、そもそもゲームと言っても、結果の集合は与えられているし、プレーヤーのとり得る手の集合も先に与えられているので、それより広い不確実性というところには全然行かない。ゲーム論的確率論で測度論的確率の見方が広がるとは思うが、、不確実性の程度というのは更に広いので、その辺はなかなか学問的に議論しづらいところがあるので課題だと思う。（竹村）

C：ゲーム理論的確率論と言うと、互いに他に影響を与えるという。まあ、そういうところがある。少し違うのかもしれないが、量子力学的な香りもする。不確実性というのは、例えばそれこそ量子力学ならば、非可換性として記述できる。いい加減なこと言ってるだけかもしれないが、ゲーム理論の人が、量子力学の話を持ち出したり。 という議論というのはあるのだろうか？
そこまでの知識がないので、一般論として面白そうなので、機会があれば勉強してみたい。（竹村）

3.7 ネットワーク、グラフとSNS

［概要］

CRDSの連続セミナーシリーズ、第7回は「ネットワーク、グラフとSNS」をテーマとして、藤澤氏より「大規模・複雑データに対するグラフ解析と超スマート社会実現のための産業応用」、山下氏より「機械学習を利用した併買品予測モデル構築・検証プロジェクトからの示唆」というタイトルでご講演いただいた。

藤澤氏の講演は「グラフ理論」を基礎とした応用分野の広がりの紹介に始まり、サイバー空間における共通クラウド基盤を中心とする「サイバーフィジカルシステム」を用いた実世界の問題解決や技術革新の動向、Graph500などのベンチマークテストを通した計算機の評価と、計算の数理的基礎をなすグラフ解析アルゴリズムの重要性が展開された。最後に藤澤氏が新たに携わる「データ格付け」に関する産学共同プロジェクトの紹介を通して、本分野の新展開と数理基盤の重要性が提唱された。

山下氏の講演は所属されている研究所における事業の実例を通して、機械学習を用いたデータの解析がいかに企業の経営や事業戦略を始めとした社会活動に資するか、またその適用において、いかに数理的・社会的背景が重要となるかが展開された。そこには現在の大きな流れである「機械学習・AIの利用」に対する様々なメッセージが含まれる。

［Key point］

・グラフ解析、特に実社会の問題に関連するものは「理論構築→即応用」というスピード感を持った動きができる事が前提となる。（藤澤）
・想定していない動きに対する予測、開発ができるようにするには、「余裕を持たせた投資」が必要。あらかじめ余裕を持った投資しておく事で、想定外の事態にも柔軟に対応でき、また予想外の成果を得られる可能性もある。（藤澤）
・ツールを使用するにも「人の手」、すなわちデータとアルゴリズムの整合性が判断できる人材は必要である。機械学習やAIを使うにも、意味が曖昧なデータを分析、意味のある関連性を抽出・提唱できる人材の手を加えながら扱う事が望ましい。数学・数理科学はこの点で社会に大きく貢献し得る。（山下）

［藤澤氏講演：詳細］

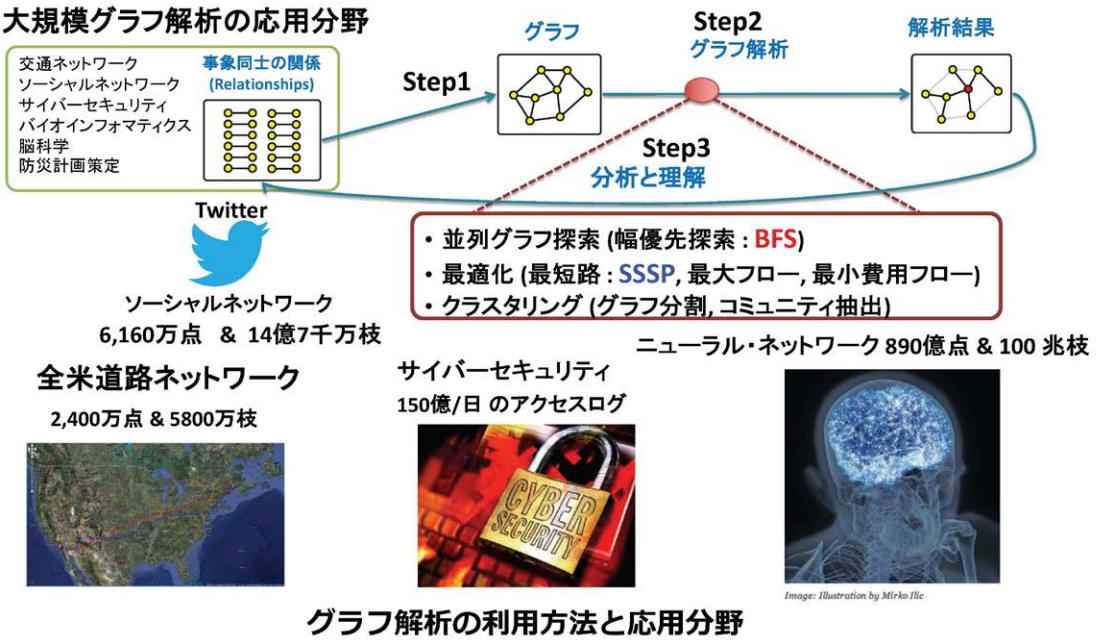
点とそれらを繋ぐ枝からなる集合をグラフと呼ぶが、個々の事象を「点」、事象間の関係を「枝」で表現することで、SNSや交通、サイバーセキュリティ、脳などのネットワークをグラフで表現でき、その解析を用いることで科学的、社会的ネットワークの様々な側面を理解する事ができる。グラフの特徴を示す指標はいくつかあるが、代表的なものの一つとして「中心性」があり、交通ネットワークや「PageRank」などの検索エンジンの肝として用いられる。

グラフの解析手法としては幅優先探索（BFS）、最適化手法とも親和性の高い最短路（SSSP）の探索などがあり、前者はTwitterのフォロワー間のネットワークなど、Graph500（後述）のベンチマーク問題に、後者は中心性指標などと合わせて道路ネットワークにおける各地点間の最短距離の計測などに用いられる。

解析するグラフのサイズは、道路交通ネットワークの場合点や枝の数が多くても10億程度であるが、SNSや脳の場合は点や枝のいずれか、あるいは両方のサイズが1兆程度あるいはそれ以上になり、近年ではより巨大なサイズのグラフを扱う必要もあって、計算機や計算アルゴリズムの必要性はさらに大きくなっている。

実応用では共通クラウド基盤を中心とした「サイバーフィジカルシステム（CPS）」の実現を通して、実世界とサイバー空間を繋ぐことによる都市OS、スマートシティーの推進などの動きがある。

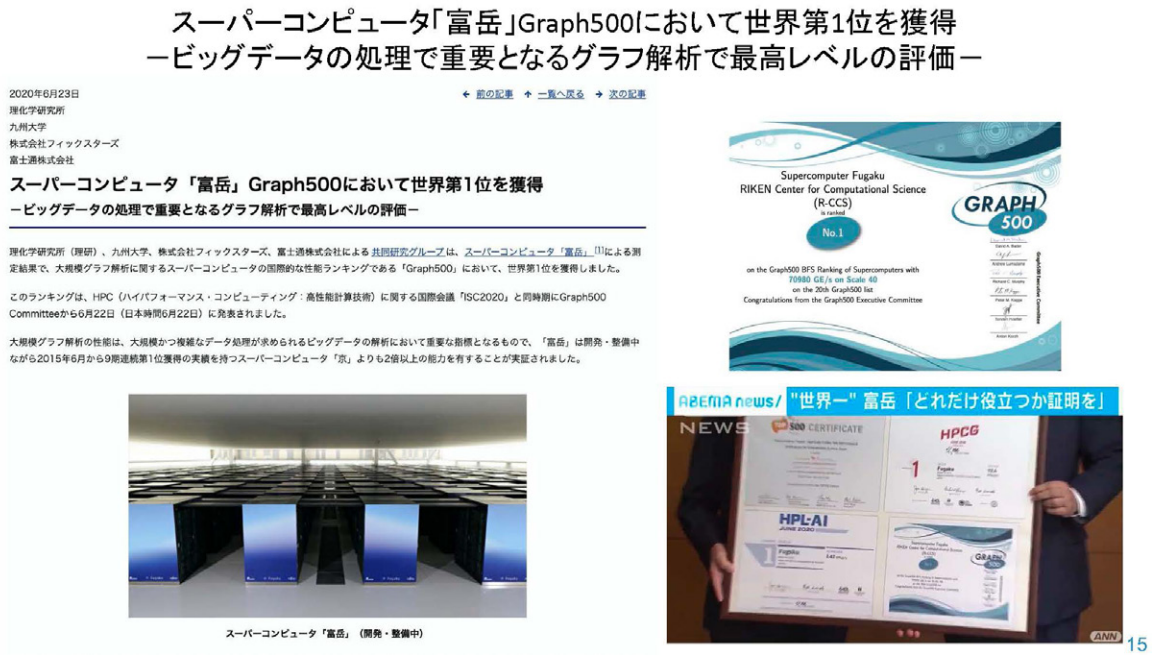
サイバー空間において最適化計算やシミュレーション、機械学習を通してグラフの解析が用いられ、これらを情報・ヒト・モノ・交通の「モビリティー」のデータ群に適用することでより良い社会の創生が期待される。



日本ではデジタル庁発足の動きなどの DX の動きも進んでいる。

モビリティーの解析の応用例として、物体検知、骨格や人物、人流の推定や追跡、最適レイアウトの策定などがあり、この解析を実現・適用するスマート工場建設などの構想もある。

グラフサイズの巨大化、解析の高速化要請に伴い常に高性能の新しい計算機が求められるが、その性能を評価する世界的なベンチマークテストとして「Graph500」がある。これは人工的に生成した巨大グラフに対してBFSを行うことによって、その計算性能を評価するものである。藤澤氏のプロジェクトチームは2014年から2020年までにスーパーコンピューター「京」「富岳」などを用いて通算１２期世界１位を獲得している。これは単純にコンピューターの計算速度だけでなく、グラフの生成や枝探索のアルゴリズムの良さにも依存する。事実、同一のコンピューターでもアルゴリズム次第で性能が大きく異なり、設計時には想定されていな



15

かった性能（ビッグデータの取り扱いによる計算性能）が、相性の良いアルゴリズムの適用で引き出されたりする。上記の結果は「余裕のある資源投資」と「数理的なアルゴリズムの弛まぬ発展」の両方でなされたものであり、どちらかが欠けても計算機の性能向上、計算技術革新が実現され得ない事が氏のチームの結果において示唆される。

なお、現在はデータの品質保証の観点から「データの格付け」サービスに関する共同研究が産学プロジェクトとして発足しており、AI利用における信頼性の向上などに資するものとして期待される。

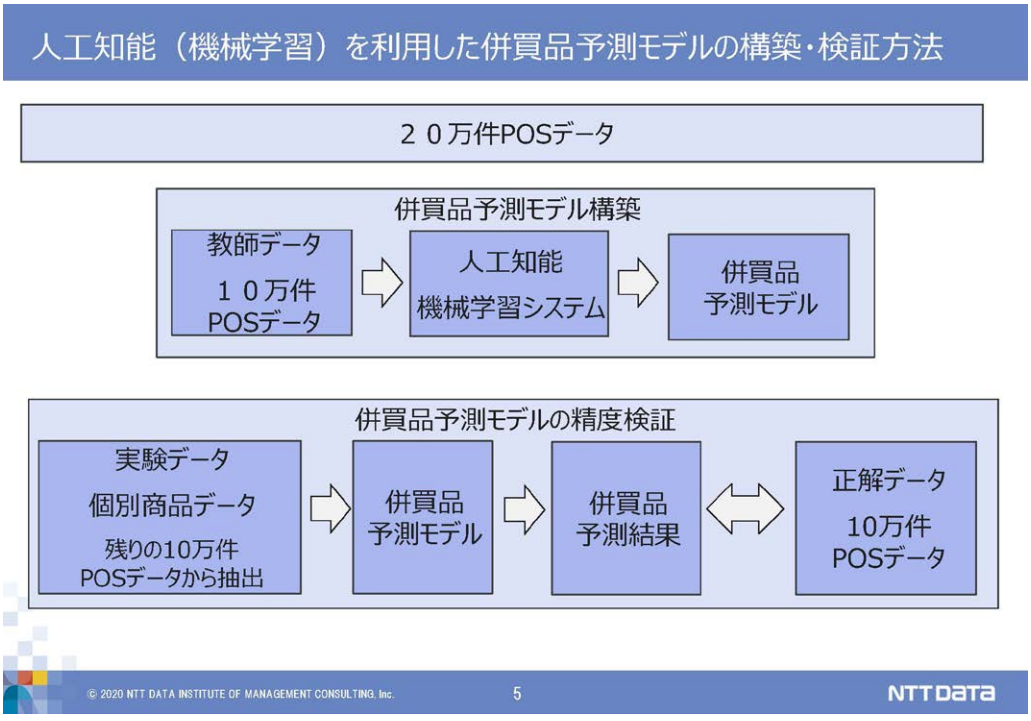
上記の研究開発は産業・実社会への応用と直接結びついており、「理論・手法開発→即応用」が実現できるレベルの理論の確かさと汎用性、スピード感の全てが要求される。

［山下氏講演：詳細］

NTTデータ経営研究所は「デジタル・トランスフォーメーション」（DX）の考え方に添い、企業経営や事業戦略、地域マネジメントやエネルギー・医療・金融、IT戦略など幅広い分野のコンサルティングを行っており、コンピューターを使うシステムづくりを行う企業、「システムインテグレーター」であるＮＴＴデータのグループ会社である。

実例の一つとして、最初に「Twitterのデータを用いた販売戦略」の紹介があった。ある季節性食品を含むツイートを解析して、その食品の販売をいつから始めるのが良いかを分析するというものである。１週間の違いが売り上げに劇的な違いを生み出すようで、データ解析により得られた結果は食品会社の顔色が変わるほど衝撃的なものだったそうである。この解析に基づく予測は大きな成功をもたらした。手元にあるデータはほぼゴミ同然であるが、「目利き」により非常に高い価値をもたらし得る事は留意すべき点として強調された。

講演の主なトピックとして、スーパーマーケットのPOS（Point Of Sales）データを用いた「より効果的な併買案の策定」が紹介された。これはスーパーの売り上げ向上に直接関わってくる課題である。データは「ポイントカード」に含まれる顧客情報と購買履歴が約20万件用意され、AI・機械学習によりデータを分類、併買品の予測モデルが組み立てられた。しかし、男女、年齢、職業、店舗特性など様々な特徴で分類を試みても、学習により予測される結果と実測結果の一致がほぼ見られず、予測精度の向上に非常に多くの考察や検証を要した（実例では店舗特性によるデータ分類・モデル構築が実用性に足る結果として得られている）。

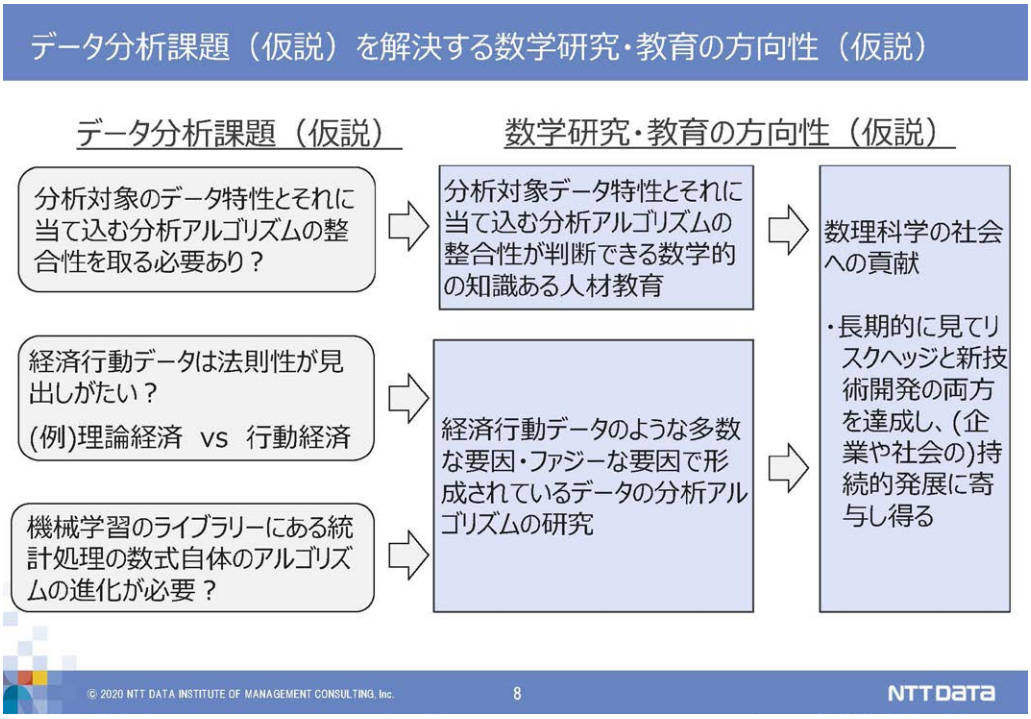


実際、用いられたデータは「経済的データ」で、法則を見出しづらいものである（一例として、理論経済と行動経済の違いが挙げられている）。

上記の実例において、機械学習の既存のライブラリを適用することの是非が非自明な問題として生じている。これは「データ特性とアルゴリズムの整合性」の問題であり、機械学習の闇雲な適用に対する一つのアンチテーゼとなっている。近年、AIの大きな成功例として囲碁やチェスが挙げられているが、これらはルールがカッチリ決まっており、与えられたデータと適用するアルゴリズムの整合性がきちんと取られた上で運用されていると考えられる。対してこれらの整合性が不確かな状態、あるいは機械学習のライブラリにある統計処理アルゴリズムが要求される結果に不充分な回答しか得られない場合、見当外れの予測を出してしまい、かえってリスクを増大させる要因となり得る。

以上の実例より、数学研究・教育の方向性の一つが提唱される。闇雲なAI、機械学習の適用の汎用を防ぐため、「分析対象データの特性と適用するアルゴリズムの整合性」を判断できる数学的見識を持つ人材、また経済行動データに代表される「多数・曖昧な因子で形成されるデータの分析アルゴリズム」ができる人材の育成と登用が、長期的にはリスクヘッジと新技術開発の両方を達成し、企業や社会の持続発展に寄与し得る。

山下氏は上記のメッセージを「データ解析に（常に）人の手を加える」と表現し、数学・数理科学の社会的重要性を強調された。



【質疑応答】

Q：最適化問題におけるデータの「冗長性」は、グラフ解析において考慮に入れる？

A：「データは綺麗だろうか？」という問いと関連。入力データにノイズは入ってくるので、「ロバスト性」評価は応用では必須である。事実、深層学習において「ユーザーを騙す」技術が存在する。並行して「騙しを防ぐ」技術が考察・開発されており（これがデータのノイズに対するロバスト性と関連している）、いたちごっこ状態となっている。

理論構築の際、入力データについて、ロバスト性などの「評価の議論ができるか」が第一に、実際に応用できるかが第二に来る。これらが保証・適用されなければ実サービス等に「使えない」というこ

とになる。

Q：AIそのものにグラフ解析は貢献している？

A：グラフを用いる大きなメリットの一つとして「計算量が少ない」事があるので、深層学習などの性能を上げるためのデータの前処理、具体的にはクラスタリングやデータ補正などに用いることができる。

Q：「データ格付け」の詳細を…

A：次を検証し、データを分類する。

『意味のある結果が出る可能性のある網羅性や正確性などの数理的特徴が担保されたデータたり得るか、それはどの程度信頼性を確保できるか。』

網羅性やアルゴリズムによる品質保証に、グラフ解析は貢献し得る。網羅性は、必要なデータ量や分散、エントロピーなどを評価して判定する。

品質保証はロバストネス、最適化手法による速い収束、複数の手法を当てはめた正しい判定法の選定により実施する。

Q：医療の診断データを機械学習にかけて結果を予測すると、実際の結果と随分異なる（紹介事例と関連する点でもある）。「機械学習にかける」と言っても、適用するアルゴリズムをあらかじめ決め込んでしまうとまずいのでは？

A：何より「合理的に使える結果」を出すのが第一。上層部などで「とにかくAIなどを導入して使え」という声が多いが、そもそもきちんと結果を出すやり方であるのかは、事前に評価が必要である（指摘はごもっとも）。

Q：併買予測に、「前処理」はどの程度？

A：「POSデータを使う」事を前提としていて、あとは詳細な処理は施していない。さらにデータについての知見を加えると、より早く正確に結果を出せたかもしれない。

（関連コメント：因果関係を間に挟むやり方などが一つの解となりうるか?）

Q：機械学習を用いた予測アルゴリズムはどのように用意した？

A：基本的に、機械学習側のアルゴリズムには手を加えていない（同一のものを用いている）。

Q：POSデータの問題に「データ格付け」はつかえる？

A：機械学習の適用前、「データそのもの」が信頼に足るものかを評価するのに使えるが、データ収集という最初のステップに戻る必要があるかもしれない。

Q：与えられたデータがどのような特徴を持っているかを知ることはできる？

A：これがまさに「データ格付け」の考え方。クライアントに「信頼できるデータ」をあらかじめ要求するのは無理があるので、「データ保証サービス」など、格付けを用いた信頼できるデータのアレンジを行うビジネスが展開できるかもしれない。

Q：データ収集、適用、提示のプロセスにおいて、先導は誰が行う？

A：数学側か、エンジニア側か、ユーザー側か、しっかりとした仕組み作りは必要。（藤澤）

　全体のワークフローを一つのセクションが通しで見るのは厳しいので、「誰が」というより「one team」で行うことになるだろう。（山下）

3.8 耐量子計算機暗号

［概要］

CRDSの連続セミナーシリーズ、第8回は「耐量子計算機暗号」をテーマとして、高木氏より「新しい数学：ポスト量子暗号に向けて」、高島 氏より「企業における暗号数理の研究」というタイトルでご講演いただいた。

高木氏の講演は現在使われている暗号技術の紹介に始まり、「量子計算機」出現後に想定される暗号の危殆化と暗号開発の状況の紹介があった。さらに、数学は暗号開発において必要不可欠な技術であることも説明された。

高島氏の講演は、氏自身の企業における暗号開発の経緯と企業における暗号開発の意義、数学の有効性、さらに、耐量子計算機暗号を見据えた暗号開発が望まれている現状も紹介された。数学と暗号の異分野交流についても、さらに活動を広げていきたいとの話もされた。

［Key point］

・暗号開発は、安全性を見据えて開発する必要がある。（高木）
・２０３０年までに、耐量子計算機暗号の具現化に指針を与える。（高木）
・企業における暗号開発では、ユーザーの利便性にも配慮する。（高島）
・安全かつ高機能な暗号が有用になる 。（高島）

［高木氏講演：詳細］

高木氏からは「新しい数学：ポスト量子暗号に向けて」として、自己紹介と略歴紹介からはじまり、現在の暗号の置かれた状況と量子コンピューターの出現による耐量子暗号の必要性について、説明があった。さらに、耐量子計算機暗号の候補として議論されている暗号の現状紹介もあった。

ポスト量子暗号は、昨今注目を集めており、2017年にはNHKクローズアップ現代、2018年にはNHKサイエンスゼロという番組にも出演し説明したとの事であった。最近は、様々な質問を受けるようになり、一般向けの入門書として“暗号と量子コンピュータ”も昨年にオーム社から出版している。本に書かれている内容を、今回は主に講演する予定とのことであった。

暗号と言うと、30年くらい前は軍事・外交で利用されるというイメージにあったが、95年のインターネットの普及に伴い、電子政府やWEBブラウザなどでも利用されている。また、DVDの著作権保護などは、コピーできないのは暗号化されているためである。最近になって有名なものとしては、仮想通貨もある。ビットコインは暗号技術の塊で、暗号無しにデジタルマネーは作れない状況にある。電気自動車やスマートグリッドでも暗号は利用され始めていて、現代社会にはなくてはならない基本的な技術になっている。

暗号技術で現在もっとも使われているものは、RSA暗号と楕円曲線暗号の2つがある。これらは既に広く普及している。なお、RSA暗号の場合は素因数分解問題が難しいことと、楕円曲線暗号は離散対数問題が難しいことに安全性の根拠がある。さらにペアリング暗号ある。

ただし、これらの暗号は、量子計算機を用いると安全ではなくなる。後ほど詳しく説明するが、このことを危殆化と呼ぶ。つまり、今、我々が利用している暗号は、量子計算機の時代には使えなくなり危殆化する。実際に、94年に現れたShorのアルゴリズムによって、理論的にはソフトウェアとしては実現可能であったが、それを動作させるハードウェアはなかったものの、2001年になって量子計算機が現れて、計算規模が拡大している状況にもなっている。

そこで、量子計算機を用いても解読できない暗号を作ろうという動きが暗号の業界で広まっている。これらの研究を総称して、つまり量子計算機の時代になっても安全な暗号を総称して、ポスト量子暗号と呼ぶ。耐量子暗号などとも呼ばれる。

なお、量子計算機は、現在、20量子ビットと呼ばれるものが、実用的に使える状況になってきている。冷却装置が高価なために、現在、家庭では量子計算機を設置はできないが、クラウド経由でプログラムを動かすことは出来て、実行結果も得られる状況にはなっている。結局、量子コンピューターは、現在、誰でも使える状況になっている。ただし、規模がまだ小さく20量子ビットである。

近い将来には、50量子ビットや72量子ビット規模まで拡張されると言われており、50量子ビット以上になると、スーパーコンピューターを上回る計算能力を持つと言われている。このことを量子超越性と呼び、昨年度大きく報道がなされた。さらに大規模な量子コンピューターが出てきた場合、暗号にも脅威となるため、現時点でも量子計算機においても安全な暗号を作る話が盛んに研究されている。

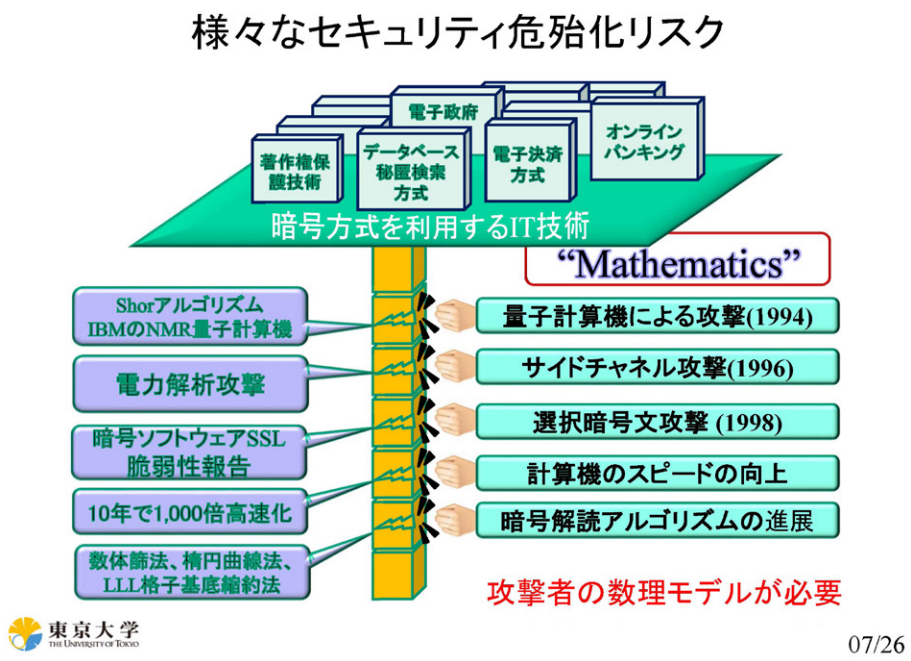
暗号の安全性というのは、実は量子計算機に限らず、様々な脅威が存在する。例えば単独のデバイスの周りの電力を計測して、電力から何らかの情報を抜き出す、サイドチャネル攻撃と呼ばれるものがある。さらに、サーバーに問い合わせをたくさん投げて、中の情報を抜き取るという選択暗号文攻撃もある。加えて、量子コンピューターに限らず、スーパーコンピューターも高速化している状況である。このような状況下で攻撃者はますます強くなっている。その他にも、数学自体が進歩して、その結果、暗号の解読に使われるようになっている。これらの数理モデルを解析することが非常に重要なテーマになっている。

JSTのプロジェクトとしても、高木氏が2014年（約6年前）から暗号数理に関して実施しているCRESTの紹介があった。高木氏が研究代表者で、高島氏や若山氏も参加しているプロジェクトになっている。

現在実用化されている暗号はRSA暗号や楕円曲線暗号だが、純粋数学の問題を扱うことになる。暗号を作るために整数論を使ったが、実際、安全な暗号を作ろうとすると、大きな素数が必要になる。どうやって大きい素数をつくるかということで、新しい問題として暗号から数学にフィードバックがかかった。これは歴史的な成功例として、応用と純粋数学の交流が盛んに行われて、新しい数学も発展するという状況にもなった。さらに現在、IT技術も進歩して、量子コンピューターに対する安全性なども考慮すると、必要となっている数学も格段に増えている。従来の枠組みを超えた数学を使う必要性が出てきた。

そこでCRESTのプロジェクトでは、さらに色々な数学理論を使って、想定される攻撃者の理論モデルを作ろうということもやっている。総勢30名程度の数学者と暗号の研究者が参加して、年に4回程度講演会行ったり、ワークショップを開催したりもしている。若山氏や高島氏も参加している。

さらにこのプロジェクトの一環として、2016年には国際会議も主催した。当時、九州大学で主催したが、



量子計算機に対しても安全性を持たせる暗号を議論する専門の国際会議になる。スライドには、PQCrypto2016（The Seventh International Conference on Post-Quantum Cryptography）の紹介がされていた。このConferenceは様々な所で開催しているが、2016年には日本に誘致して九州大学で開催した。

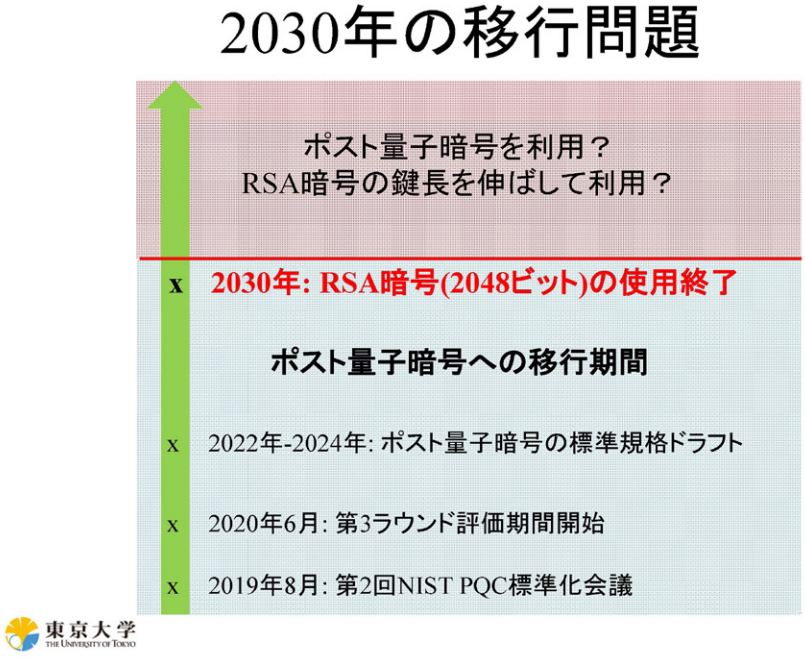
丁度この時期は、量子計算機に対して安全な暗号を作る雰囲気が高まっていた。 NISTと呼ばれるアメリカの標準技術研究所が、耐量子計算機暗号の標準化計画を発表したいということもあり、通常は100名も出席のない国際会議が、このときは240名が北米とか欧州、アジアから集まって議論を行なった。

そこで、NISTは2017年に公募を開始して、3年から5年かけて安全な暗号を選ぶプロジェクトを発表した。基本的な鍵交換、暗号化とデジタル署名と言われるものを作ろうということや、いくつかの数学の問題が使えるのではないかということも、そこの会議では発表された。1番本命と言われているのは格子暗号がある。その他にも符号暗号、多変数多項式暗号などもある。

公募が締め切られて、今は議論が進んでいる状況にある。今のところ7候補まで残っており、もう少しで最終的な候補が、あと1、2 年で決まるという段階まで議論が進んでいる。日本の状況はCRYPTRECと呼ばれている電子政府向けの暗号のリストを作っている体制がある。総務省と経済産業省が主体となって運営している。ここでも、量子コンピューターの時代に向けた安全な暗号や、量子コンピューターでどれくらい暗号が安全なのか、量子コンピューターに対する安全な暗号はどういうものがあるかの調査やワーキンググループなどがある。高木氏は、この暗号技術評価委員会の委員長をしており、日本で使える暗号の安全性の監視や調査を行っている状況にある。

今使っている暗号は、使用期限が2030年までとなる。このことはアメリカ政府の取り決めで決まっている。つまり31年から、現在、標準化しようとしているポスト量子暗号に移行するのか、今使っているRSA暗号の鍵を伸ばして使うのかという、大きなターニングポイントに来ている。つまり、今から10年後に向けて準備をしているという状況にある。10年間の短い期間で、どのように決断するかが非常に重要になっている。現在2020年の12月では、第3ラウンド評価ということをやっている。

暗号の安全性を評価は非自明で、暗号が提案された後に、例えばNISTなどで提案されたときに、国際会議で様々な議論をする。様々な攻撃や、新しい解読アルゴリズムがあるかなど、何ビットの鍵長なら安全かなどを議論する。暗号の業界ではストレステストと呼ばれるが、公開の場で議論する。得られた知見はすべて論



文として発表して、学術的な知見として発表する。その上で、多くの数学者や暗号学者が考えた結果、破ることが出来なかったならば安全ということで、業界のコンセンサスが取れて、最終的に実用化フェーズに入る。鍵の長さなどのパラメーターなどを決めた暗号規格を作成する必要がある。しかし、未来永劫安全ではない。時間がたつと計算機のスピードも上がって、いずれ寿命が来るため、寿命がきたらまたupdateする。今後、2030年に寿命が来るので、新しい暗号にしたい。

なお、現在の暗号は、例えばブラウザをクリックすると暗号の鍵が出てくるが、実際何ビットかを見ることができる。 RSA暗号は今2048ビットを使っている。10進数では約660桁だが、スライドには公開鍵の様子も表示されていた。

結局、この公開鍵が素因数分解されると暗号が破られることになる。2048ビットは素因数分解できないというのが、業界での知見になる。実際、チャレンジ問題というもので、どのくらいの大きさなら解読できるかが、歴史的に実験されてきた。このことを一つの目安としている。2009年には768ビットが一台のパソコンで約1500年になっていた。これらの数字を見て、さらにスーパーコンピューターでどれくらい速くなるかも勘案している。今の富嶽では10^{17}くらいになり、これらのスピードを見ながら、暗号の安全性を評価していく。

ここで数学が重要になってくる。素因数分解を最も高速に計算するアルゴリズムに数体篩法という代数体を使ったものがある。それによると、理論式

$$O\left(e^{(c+o(1))(\log n)^{\frac{1}{3}}(\log(\log n))^{\frac{2}{3}}}\right)$$

があり、実験結果も勘案すると、富嶽のスーパーコンピューターでは1024ビットの素因数分解を1年で解けると言われている。ところが2048ビットでは、10の10乗倍ぐらい時間を掛けないといけない。ムーアの法則に従って計算機のスピードが上がることを想定すると、あと30年以上は安全であろうと言われているので、2030年までは使うように規定をしている状況になっている。このように暗号の安全性は議論されている。

量子コンピューターによる暗号の危殆化について説明する。暗号は鍵を伸ばせば難しくなっていく。ところが、量子コンピューターを使うと、鍵を延ばしても指数関数的なスピードアップが行われる。 正確には$\log n$の3乗だが、例えば1024ビットから2018ビットに伸ばしても、かかる時間は8倍しかならない。8倍にしかならないと、今までは10の10乗倍 難しくなっていたものが8倍にしかならない。結局、鍵はどれだけ伸ばし

ても無駄になる状況になる。この状況を多項式時間の解法と呼ぶ。

そのため安全ではなくなり、代わりの数学として提案されているのがこの格子暗号になる。格子暗号は、ノイズ付きの連立方程式と呼ばれる物に対応する。

具体的に普通の連立方程式では、例えば下記２元連立一次方程式は簡単に解ける。

$$\begin{cases} 2x - 3y = 0 \\ -x + 2y = 1 \end{cases}$$

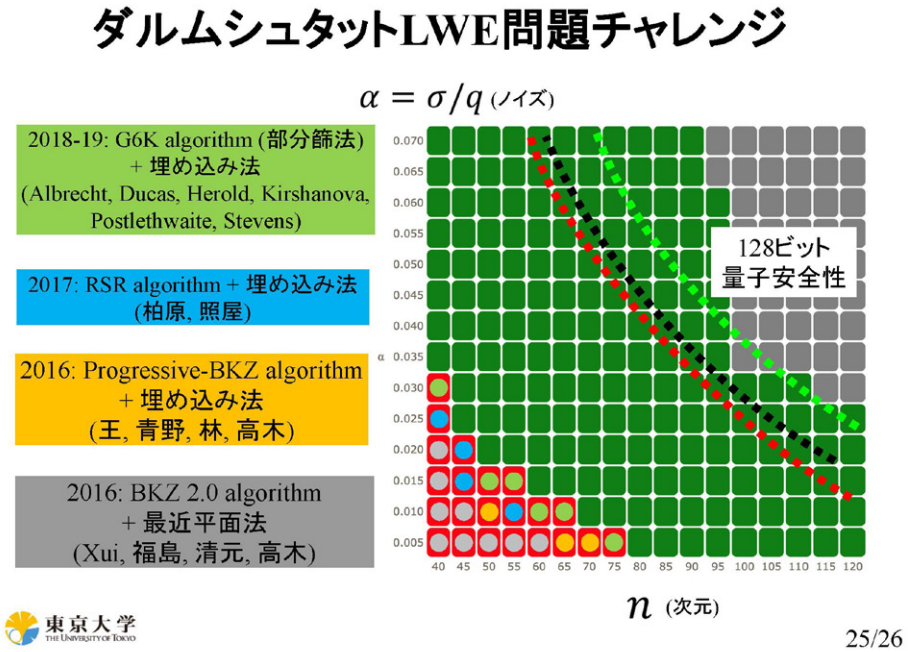
この方程式にノイズを加えて、ノイズ付き連立一次方程式（LWE問題）：

$$\begin{cases} 2x - 3y + e_1 = 0 \\ -x + 2y + e_2 = 1 \end{cases}$$

を考えると解を求めることが困難になる。つまり、（次元orノイズ:増加させると）LWE問題は計算が困難となる。変数の次元、つまり線形空間の次元を上げて、ノイズも上げていく。そのようにすると、LWE問題は計算が困難であるということが理論的に証明されている。結局、NP困難問題になることが証明されているので、この事実を利用して安全な暗号が作れることになる。NISTのプロジェクトで、第２ラウンドで残っている方式にはいくつかあるが、結局、最終的に第３ラウンドまで残っているもので、CRYSTALS-KYBER、SABERとNTRU、これら３つくらいが格子暗号として残っている。現在、この中から最終的な候補が選ばれるだろうと言われている。

LWE暗号に関しても解読問題が存在し、スライドの図で描いてあるように、解読が進んでいる。これはKDDI研究所と高木氏の研究室と一緒に行った回答結果に加え、右のほうに行くと、次元では40次元から75次元ぐらいまで解けており、縦軸がノイズ$\alpha = \sigma/q$を表している図になっており、ノイズが大きいほど解読が難しい様子が示されていた。加えてNICTの結果、総務省関係の研究所の結果、東大の他の研究室で得られた結果が提示されていた。世界記録は、フランスとヨーロッパのチームが得ている状況になる。

これらの解読のデータを見ながら、最終的にどれくらいの難しさになるかを計算し、ここ数年で解読可能性の評価の閾値を決める線を決定しないといけないという状況にある。ただし評価方法が決まっておらず、評価の線は乱れている状況にある。この線が決まれば、128ビット量子安全性と呼ばれる鍵長が決定する。



最後にまとめとして、量子暗号時代に安全な暗号について最新動向は、
CRYPTRECのホームページ　https://www.cryptrec.go.jp/
に詳細な技術報告書があるので、参照して貰えればよいと思う。

［高島克幸氏講演：詳細］

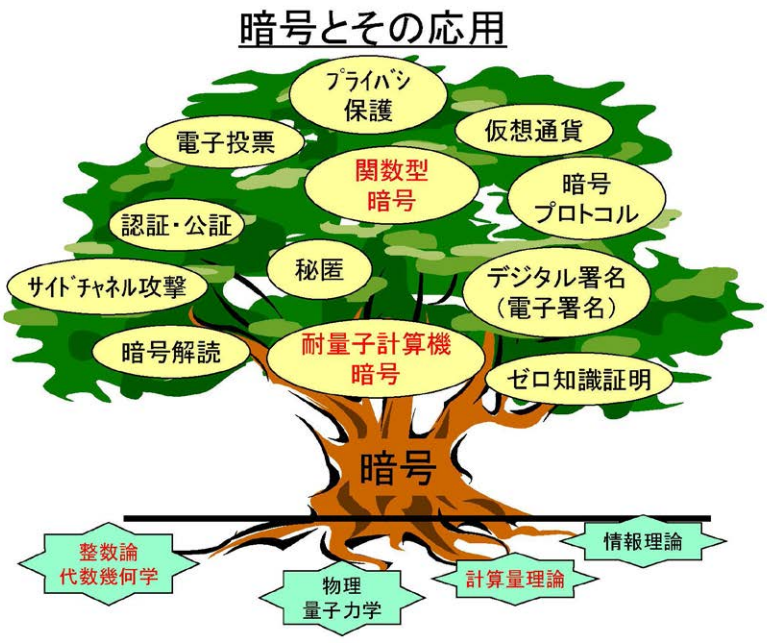
高島氏の講演は、氏自身の企業における暗号開発の経緯と企業における暗号開発の意義、数学の有効性、さらに、耐量子計算機暗号を見据えた暗号開発が望まれている現状も紹介された。数学と暗号の異分野交流についても、さらに活動を広げていきたいとの話もされた。

高島氏は三菱電機で暗号の研究開発をしている。全体のテーマにもある耐量子計算機暗号については高木氏の紹介にも少しあったが、その中でも同種写像暗号と呼ばれる暗号を研究している。この機会に、企業研究者として数学を企業の中で武器として扱うために、どのように考えながらやってきたかについて、個人的な経験も織り交ぜて話をしていく。最初に略歴を紹介し、次に企業で暗号研究をするときに何が求められて、何を考えて研究開発したかを話したいとの旨であった。さらに暗号の中でも、関数型暗号も紹介するとのことであった。

耐量子計算機暗号を紹介する前に略歴の紹介があった。最初、代数幾何を専攻して数学科の修士課程を修了し、NTT 岡本龍明氏（当時 京都大学連携教授も兼任）の下、社会人ドクターを取得したとのことであった。その後、九州大学のマス・フォア・インダストリ研究所 客員教授、日本応用数理学会副会長の経歴の紹介があった。企業で暗号研究をやってきており、企業で１番求められることは、先ほど高木氏の話であったように、安全性を確保することにある。ただし、使い勝手がよく、効率のよい実装ができる暗号アルゴリズムを作りださないと利用されない。

高効率と安全性に加え、2004 年ぐらいからは、暗号の利便性の面についても積極的に考えられるようになってきた。以前からもそのような研究はあったが、例えば電子投票などの研究では、暗号を積極的に使うことで便利になる。現在はブロックチェーンのような技術もあり、一般的に認識されつつある。

多くの人にも、暗号の利便性については共有されつつあると思う。しかし、暗号を使うことは安全性を高めることでもあるので、通常は暗号を使うと不便にはなる。このことはトレードオフで、安全性を高めても出来



るだけ不便にならないようにする。言い換えれば、セキュリィティを高めたとしても、使いやすいものにすることが暗号開発の基本コンセプトになる。

暗号による新しい利便性を生み出すことが、企業で研究する中で大きな意義を持つ。企業内でも、ユーザー目線に立って必要とされる技術や、ユーザーが積極的に使いたいと思う技術でないと、社内でも説明しにくい。これが企業研究者として重要な観点になる。

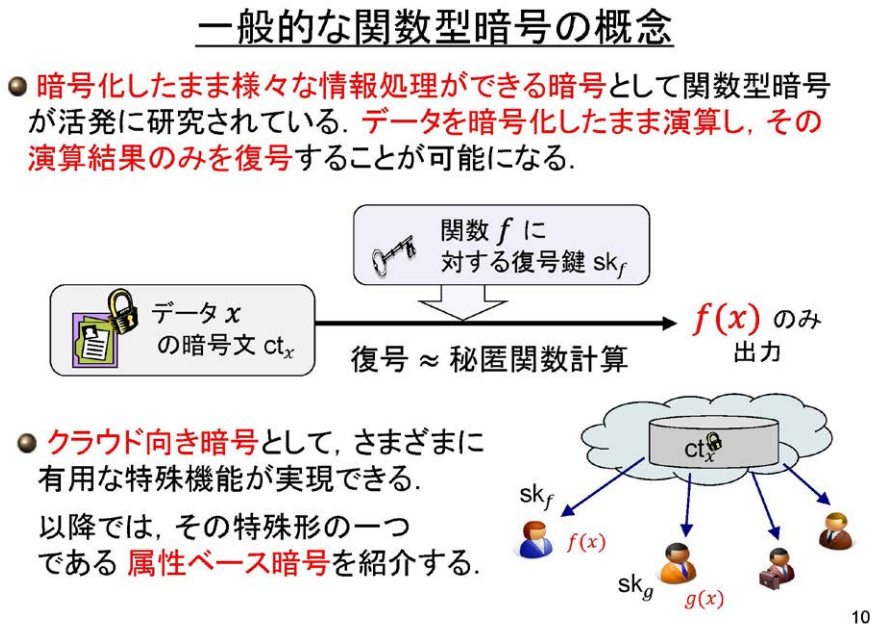
具体的には、独自の方式で特許を取ったり、標準化の取り組みを積極的に行ったり、さらに暗号は学術的な成果でもあるので学会発表を行う。同時にプレスリリースの形で、企業の価値を高めるための報道発表をすることも、強く求められる。学術的な研究成果に加え、ソフトウェアという形で製品化することもでき、情報セキュリティソフトウェアとして、FE（Functional encryption）関数型暗号と呼ばれる暗号方式のソフトウェアで、電子ファイルの保護とアクセス制限を、同時に暗号化機能の中に組み込む暗号方式を実際に製品化もできている。

関連会社からは機密ファイル交換サービスも製品化できた。暗号を設計して、さらに製品まで具現化したことが大変良かったと感じている。

ここまで、企業研究者の一端を紹介したが、関数型暗号についても少し紹介したいと思う。高木氏の話にもあったが、入社した97年頃にはインターネットが普及期に入り、暗号開発が活発化していたので、楕円曲線暗号の実装やパラメーター設計を進めるために会社に入った。更に、2000年代に入ってからは、クラウドを用いて情報利用の形態が変わり、情報処理の進展とともに暗号も進展できたことが、暗号分野での注目すべき点だと思う。また最近は量子計算機が現実味を帯びてきて、耐量子計算機暗号というものが強調されてきた。同種写像暗号という暗号が楕円曲線を使った暗号としてあり、その研究にも携わることができた。

まず、関数型暗号は、暗号を使って便利なサービスを作り出す事にもなる。関数型暗号の概念を一言で言うと、暗号化したまま、様々な情報処理ができる暗号になっている。他にこのような機能をもつ暗号として、準同型暗号と呼ばれる暗号もあるが、関数型暗号はデータを暗号化したまま演算して、最後に演算結果のみを復号結果として出力する暗号になる。

さらに言い換えると、データxがあり、xを暗号化したものをct_xと書くことにする。ctはCipher textの意で、ct_xと表す。通常は復号するとxが復号結果として出てくるが、復号鍵としてsk_fを使うと、復号した際にf（x）



だけを出力するようにできる。例えばxが1000ビットの文字列のときに、上位半分のビットだけ取り出すこともできる。結局、xに対する部分情報だけを取り出すこともできる。つまり、鍵によって関数値という形での部分情報の取り出し方がコントロールできる暗号が関数型暗号になる。社会基盤インフラとしてクラウドがあり、様々な情報が図のように暗号化して保存されている。暗号化することはプライバシー上重要で、例えばAliceさんという人がfという関数に基づいた鍵しか持っていなければ f（x）という部分的な情報しか取り出せないし、Bobさんがgという関数に紐づいた鍵しか持っていなければ g（x）しか取り出せない。情報インフラの中から取り出せる情報の取り出し方が、個々人の鍵によってカスタマイズされる仕組みを関数型暗号によって作り出せる。暗号研究者は長らくこの仕組みを一般的な形で作り出したいと思って研究してきた。

これが本当に一般的な形で実現できることが暗号業界の中で認識されたのは、ごく最近（ここ数ヶ月）になる。一般的に実現されたことが、理論的な話題になっている状況にある。

氏自身は関数型暗号の理論的な側面と同時に、特殊形の関数型暗号も研究して来た。その1つとして、属性ベース暗号をごく簡単に紹介された。属性ベース暗号というのは、例えばコンテンツを持っている人が、コンテンツの中身に応じて暗号化する。例えば、映画コンテンツであれば、洋画のアニメなどという映画の属性をつけて暗号化する。そして、利用者に応じて、例えば、邦画のアニメしか復号できない復号鍵が配布されれば、その範囲内で映画を視聴できるようになる。その際、コンテンツ所有者は、誰が複号するかを考えずに、コンテンツの属性のみを使って暗号化してクラウドに置いておくことができる。アクセスコントロールというのは、人が介在したり、あるいはソフトウェアによって組み込むこともできるが、属性ベース暗号の場合は、その機能の正当性を暗号の安全性として保証することが出来る。

また、素因数分解問題のような数学の問題に帰着することができるようになると、その数学の体系の中に、暗号の安全性が組み込まれるということにもなってくる。高木氏の講演では、素因数分解や離散対数問題の困難性を使って、公開鍵暗号を作る話を紹介してあり、さらに複雑なペアリング暗号や格子暗号の話題もあったが、関数型暗号でも、暗号化は複雑になってくる。そして、関数型暗号では、有用な関数で高効率な暗号を実現する方法を求めてやってきた。複雑になりがちな関数型暗号・属性ベース暗号をどのように効率的かつ安全に構成するかも、暗号分野の1つの大きい目標になっている。

岡本龍明氏と一緒に研究して、楕円曲線に基づく関数型暗号で、独自性のある良い仕事ができたと思っている。楕円曲線の話をしたと思うが、高島氏自身は以前、代数幾何を研究していたので、代数で何か新しいこととして暗号に応用しようとしていた。楕円曲線というのは

$$y^2 = \left(x\text{の}3\text{次式}\right)$$

の形をしているが、楕円曲線上の点には自然な加群構造があり、定数倍ができるようになり、例えば1億倍ができるので、この演算を使うことによって、現在標準的に使われているTLS1.3内の鍵共有、あるいはビットコインでも使われているECDSA署名も実用化されている。αを定数倍とし、Pを楕円曲線上の点としたときに

$$(\mathrm{P}, \alpha) \rightarrow (P, \alpha P)$$

が一方向性関数（つまり、順方向の演算は効率的だが、逆方向の演算は困難）という事実を利用している。

現在でも楕円曲線暗号は、ビットコインはじめ様々な場面で利用されている。楕円曲線は、他にも、ペアリングや同種写像暗号と呼ばれるような演算を有している。とくに、同種写像という演算は、量子計算機が大規模実現されても一方向性が担保される。他の演算は量子計算機が実現すると、逆方向性の演算が容易になって、暗号には使えないという状況になっている。

以上の3つの異なる特性を持つ演算を持つことが、楕円曲線の非常なメリットになる。主にこれら一方向性関数が原点になって、さまざまな特長をもつ暗号が対応して楕円曲線の上で構成できるようになっている。

さらに、楕円曲線よりも高次の曲線

$$y^2 = \left(x\text{の}2g+1\text{次式}\right)$$

でも暗号が構成出来る。このようにして始めたものが、2010年の関数型暗号のはじまりで、次元分だけ線形

空間も大きくなり、高次元空間ペアリング暗号ができるということが研究の始まりとなった。

そのようにして、独自の視点で研究を進めたことによって、最終的には初期の構成法とは大きく形を変える構成となったが、最初のアプローチが独自であったので、他の研究者がやっていないアプローチの仕方で暗号設計に取り組むことができたことが良い成果に結びついたのでないかと思っている。

上記は、数学的な視点での一般化によって始められたことが、数学と暗号のひとつのつながり方になってきた。数学は可能性を秘めており、実際に暗号でも何か新しい研究方向性を見つけるのに有効な場合が多い。

ここで、同種写像暗号という耐量子計算機暗号について説明したい。再度、楕円曲線の説明がスライドでも行われた。楕円曲線間の代数的な変換は曲線から曲線への変換写像になり、曲線2つを補間するような、同種写像と呼ばれる、代数式で定義される写像を計算することが、根本的な考え方になる。暗号では、この計算問題の困難性を考えることになる。つまり、2つの曲線を与えて、その間を補間する写像の計算問題の困難性に基づいた暗号になる。

実は上記の写像は基本的な写像の繰り返しで表されており、繰り返しの仕方が、同種写像がなすグラフの分岐に対応する。グラフ上のWalkにおける分岐の繰り返しパターンを秘密鍵に使うことによって暗号が構成されている。それによって、耐量子計算機暗号が実現できるので、耐量子計算機暗号として活発に研究されている。

SIKE暗号化方式はNIST PQC標準化コンペでも、3rd roundの（補欠）候補になっている。活発にいろいろ研究がされている状況にある。

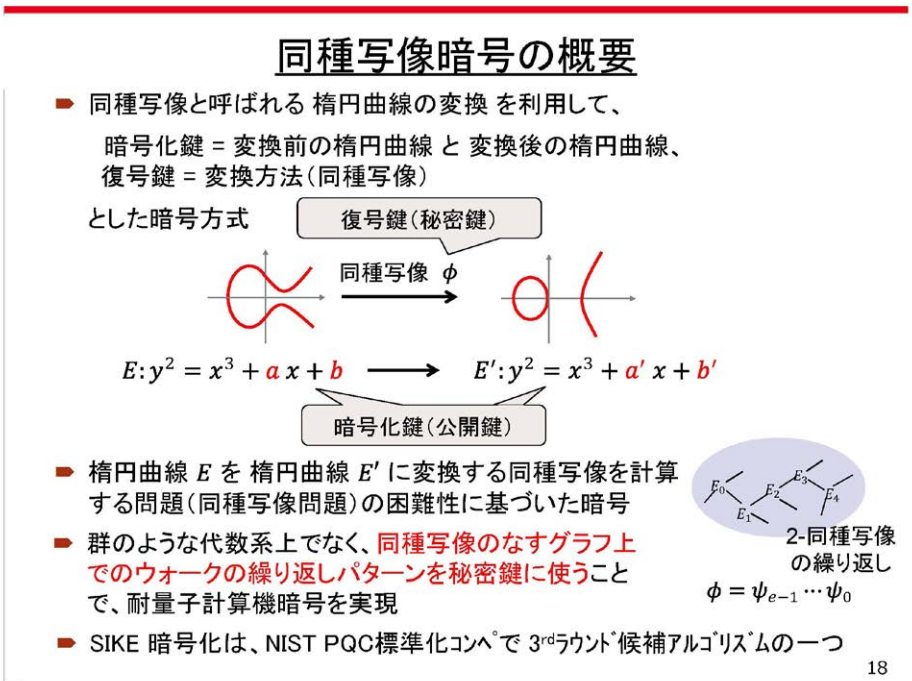
高島氏自身は、楕円曲線を数学的に一般化した

$$y^2 = (x\text{の}5\ \text{次式}), y^2 = (x\text{の}6\ \text{次式}),$$

についても、2017年に同種写像暗号に対応する同種写像グラフ上のランダムウォークを利用した種数2の新しい同種写像暗号をCRESTの論文集で提案している。

種数2の曲線については、他の研究者も参入している。東大数理で代数幾何がご専門の桂 利行 氏と高島氏は共著論文も出している。

数学は素材が抽象的で、しかも素朴なものであるから、後になっても古びないということで、上手く研究出来ていると思っている。耐量子計算機暗号は数学と暗号を関連づける力があり、高木氏のやっているCREST



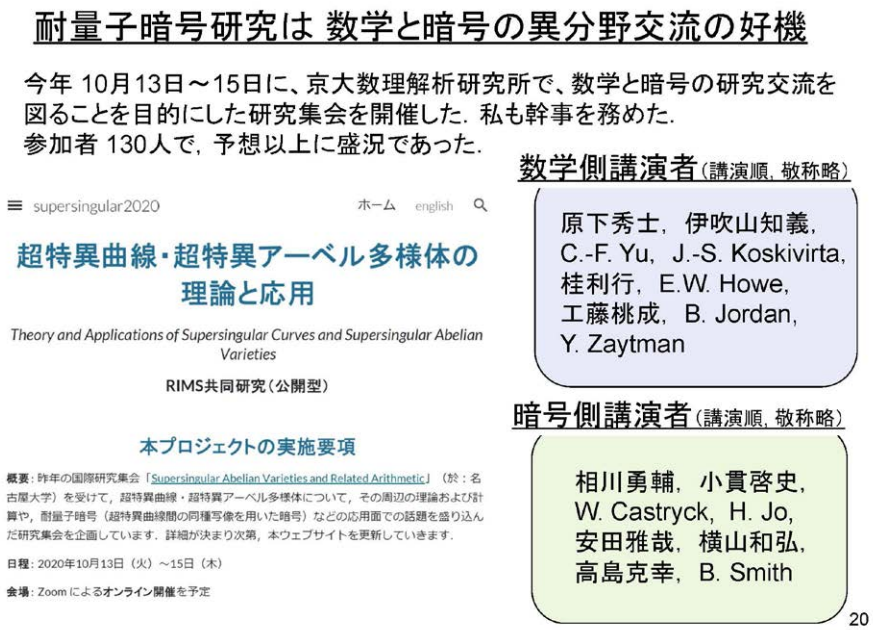
は機を捉えているプロジェクトだと思う。

さらに今年、桂氏との研究の延長で、京大の数理解析研究所における研究集会でも数学と暗号の研究交流を図ることができた。理論が数学で応用が暗号に対応する。参加者132名はZoomによる参加で盛況な会を開くことができた。数学からも、整数論や代数幾何の大学教員も参加して、暗号側は、海外での研究者も含めて比較的若手との交流を深めることができたと思う。

耐量子計算機暗号は、数理の人材をとても惹きつけるので、昨年度は優秀な若手（相川 勇輔 氏）が会社にも入社し、今年のACT-Xで河原林 氏が総括をしている「数理・情報のフロンティア」にも産業界から唯一採択された。高島氏も相川氏には採択されて欲しいと思っていたので、相談しながらやっていた。当初、会社としてはすぐに役立つようには思えないので懐疑的な意見もあった。しかし、採択後には会社の見方も随分変わって、研究水準の高さというプレゼンスを、社会に対して確実に示すことができたと思う。素材としても世の中で注目されている耐量子計算機暗号で、会社の若手が頑張っているということで、会社側の見る目も変わってきた。相川氏のモチベーションも高くなり、良い異分野交流が始まろうとしていると思っている。再度になるが、耐量子計算機暗号は、数理人材を応用に惹きつけると思っている。

まとめとして、企業での暗号研究ということから始まったが、やはり企業研究者である以上、ユーザーが使いたくなるような、ユーザー目線からの動機づけが求められている。広く応用数学を研究している人も、時には、少しユーザー目線を意識しながら研究できると良いと思う。数学を研究している人は、ポテンシャルが高い人が多いと思うが、社会で数学の素養を活かして活躍するためには、“もう少し広い視野から、数学と暗号研究の交わり”を捉えることが重要だと思っている。

講演の中の例では、関数型暗号の話に加え、最後には耐量子計算機暗号としての同種写像暗号を紹介した。これらは数学と暗号を結びつける力を持っているので、若手の人をこの2つの領域で結びつけるような活動が出来ればよいとも思っている。



【質疑応答】

Q：格子暗号など、利用する際の暗号化復号化の計算量については大幅に増えないのか。

A：現在、研究が進んでいて、第2ラウンドで述べた候補で残っている候補は、現在の暗号と処理速度は

ほぼ一緒ぐらいであろうと言われている。（高木）

C：現在あるRSA暗号に比べて極端に遅くなったり、データが非常に巨大化することはない。（高木）

Q：基本的に今までの暗号は計算困難性に依存している。計算困難性以外に、実用化に向けて研究は進んでいるのか。アイデアはあるのか。

A：情報理論的安全性という暗号がある。もう1つ有名なものに量子暗号もある。量子状態を使って鍵を配送する。両方に関して実用化に向けた動きは、もちろんある。（高木）

情報理論的暗号というのは、とくにセキュリティーが非常に高いケースで使われる。実用化は既になされていて、例えば外務省の領事館との通信は、すでに1回使い捨ての暗号ということで成されている。ところが、商用的に我々が使うような暗号では、安価でなるべく高速で、なるべく電力も小さいものというとなると、情報理論的暗号は使い勝手が悪いので使われていない。

量子通信も、光ファイバーとか端末が必要で、一般家庭まで普及するには至っていない。

結局、安価かつ安全性の高いものとして、計算量的な困難性が有望視されていると思う。（高木）

Q：その情報理論的な暗号と他の暗号形態との組み合わせなどはあるか。量子コンピューターを作るときに問題となっている量子誤り訂正符号の問題など、符号の問題もあると思う。それらはどのような関係や将来の見込みになっているのか。

A：情報理論的安全な暗号の1つの側面に秘密分散という概念があり、例えば2箇所に分けて、1箇所だけ攻撃しても情報を盗めないが、2箇所以上に秘密分散する。その分散技術と暗号を組み合わせて、3人が秘密鍵を持っている場合ならば、一人だけなら入られても安全な仕組みはある。分散と呼ばれているこの技術は、実用化に近いレベルで研究が進んでいる。もう1つの量子誤り訂正符号については、業界で心配しているのは量子誤り訂正が進むと、小さな規模で暗号解読ができるので、これは脅威になる。（高木）

Q：最初にNTTに入社するときに、耐量子計算機暗号のことを考えていたか。

A：当時は考えていない。当時はどちらかというと、暗号では１０００ビットという巨大な数字を使うため、暗号をいかに高速にするかという研究テーマであった。 ICチップにどうやって実装するかなどを研究していた。（高木）

Q：学生には、どのような問題が今後やるべき面白い問題だというふうに指導しているのか？

A：やはり、暗号の安全性になる。未来永劫、安全な暗号はないので、数学の理論の進歩に伴って、新しい暗号もいずれ安全ではなくなる。 それらを注視して、どのようなことが起きたり、どの程度、数学が進展するかということもある。本当に純粋な数学の問題でもあり、工学応用にも安全性が要求されるという工学応用をもった純粋な数学問題として、それを学生に指導している。（高木）

Q：超楕円曲線とか、色々あると思う。安全性に加えて利便性を強調していたと思うが、ユーザーの目線というか、安全性と効率は暗号として必須だと思う。その機能を加えるには関数型暗号が好都合であるという話をされたと思う。そうすると楕円曲線の次数を上げるというのもあるとは思うが、例えば、楕円曲線の拡張としてK3曲面とかもあると思う。つまり変数を増やすことを考える。暗号の構成自身は安全かもしれないが、効率的でないとか、利便性の追求は考えられないとかあるのか。

A：一般化をどうとるかという話でK3 曲面の話が出たが、以前、K3 曲面の自己同型群を暗号演算に使えないか考えたことがある。“自己同型がたくさんあるのなら、それを使って暗号できないか”と問うてみたことがあるが、“計算できないから”ということで、それ以降議論していない。実際には出来るかも知れないが、伝統的な純粋数学や代数幾何学では、ある種の存在証明に終始することが多いのではないか。具体的な形として計算できるかとか、効率がどうかとか、そういうことは捉えにくい。例えば同種写像暗号の場合でも、種数2までなら、超楕円曲線で1つのクラスの中で閉じているが、高種数で、1つの標準的な曲線のモデルが取れなくなる状況になると、つまり、複雑になってくると、計算ができるところに納めにくい。数学として計算できて、様々な方向に繋がるものを見つけてくることができれ

ば、とても素晴らしいと思う。そのためには、ある種のアルゴリズムがある程度豊かでスッキリしてないと出来ないとは思う。（高島）

Q：同種写像と言ったときに楕円曲線だと思うが、高次にして耐量子暗号として成立させているということか。

A：楕円曲線のときにすでに、耐量子計算機暗号になっている。その可能性を広げるということで種数も上げて考えてみたということになる。効率的な演算を持つ暗号が構成できたり、安全性の高いものが望ましいので、研究としてやってみたという感じになる。楕円曲線に基づく研究が主流で、楕円曲線の素朴な定義式だけで、ある種の豊かな構造があるので、色々なバリエーションで活発に研究されている。種数が3とか4になると、恐らく同種写像の具体的な計算は暗号では使いづらい。全然計算できないわけでないが、種数1のときとは比べられないし、計算において障害が出ることは代数幾何的な観点からも分かる。数学として深く調べることは色々あると思っており、暗号として、安全性や効率に還元すれば綺麗な話になるが、なかなか一筋縄ではいかず難しい場合もある。（高島）

Q：NISTの第3ラウンドで、同種写像が補欠で残っているという話があったと思う。標準化のまえにコンペで様々な暗号が復活して、本命になるかもしれないし、ならないこともあると思う。そうすると、こういう場合に、標準化から外れた暗号はどのように扱われていくのか。

A：標準化の話は、影響力が大きい。ただし、それを一つの標準化として見ることも可能だが、一つの標準化コンペに過ぎないから、まだこの後に再チャレンジの機会は必ずあると思う。そのエントリーをめざして、研究者は新しい暗号を次々に作って、10年後ぐらいまでに色々な攻撃にさらされることによって、ようやく皆がコンセンサスをもって破れない確信に至る。一つの標準化が決まったからといって、それで終わる訳ではなく、次の機会には別の標準化が必ずあるので、次々挑戦して行くためにも、暗号研究者は活発に研究していると考えている。（高島）

NISTはどちらかというと、一般的に使える暗号の標準化を目指している。そうではなく、暗号は様々なところで使われているので、例えばビットコイン（仮想通貨）で使う暗号とか、あと衛星通信で使う暗号とか、目的によって扱い方も全然違う。同種写像暗号は鍵のサイズが非常に小さい。暗号も小さいので、そういうコンパクトな環境に向いているので、例えばIOTセキュリティに向いているとかになる。暗号は用途を絞って考えるべき局面もある。（高木）

Q：標準化もアメリカの機関であれば、アメリカの企業が有利ということも考えられるか。

A：考えられると思う。（高木）

Q：暗号は流行している話題であることは間違いないと思うが、どのような取り組みがあったら、暗号分野に人材が来るだろうか。

A：講演の中でも、数学に近いところでは、例えば企業でも“一体、これを研究して何が嬉しいのか”は当然聞かれるが、その場面で明確な答えがないと承諾が出ない。企業の場合には、社内でうまく研究成果をアピールできると、継続して暗号研究に取り組めるので、継続して成功事例を作ることが大事になる。これからも、社内で認めてもらえるような成功事例を積み重ねることが非常に重要だと思う。（高島）

Q：暗号は元々、軍事というのがあって、謎めいた文章とかだったと思う。もともとcryptoは秘密とか謎めいたとか不可解の意味だと思う。日本語は見事なものだと思うが、今になると暗号というより、違う日本語を当てた方がしっくりくるような気もする。そういうことは考えたことがあるか。

A：CRESTを立ち上げるときに、暗号の先生方を集めて議論した。それは新しい日本語を作ろうということで、カタカナでクリプトはどうかというのもあった。それくらいしかわからずに良い案がなかった。仮想通貨などもある。なかなか適当な案がなく、困っている状況だと思う。日本語では暗号という言葉に慣れてしまって、なかなかポジティブな明るい言葉がないかとも思う。（高木）

3.9 データマイニングと知識発見

［概要］

CRDSの連続セミナーシリーズ第９回は「データマイニングと知識発見」をテーマとして、山西氏より「記述長最小原理・変化検知・予兆情報学」、上田氏より「ノンパラメトリックベイズ法に基づく関係データマイニング」というタイトルでご講演いただいた。

山西氏の講演「記述長最小原理・変化検知・予兆情報学」では、予兆検知の理論に必要な統計量の紹介があった。COVID-19や工場の検査における実例の紹介も行われた。

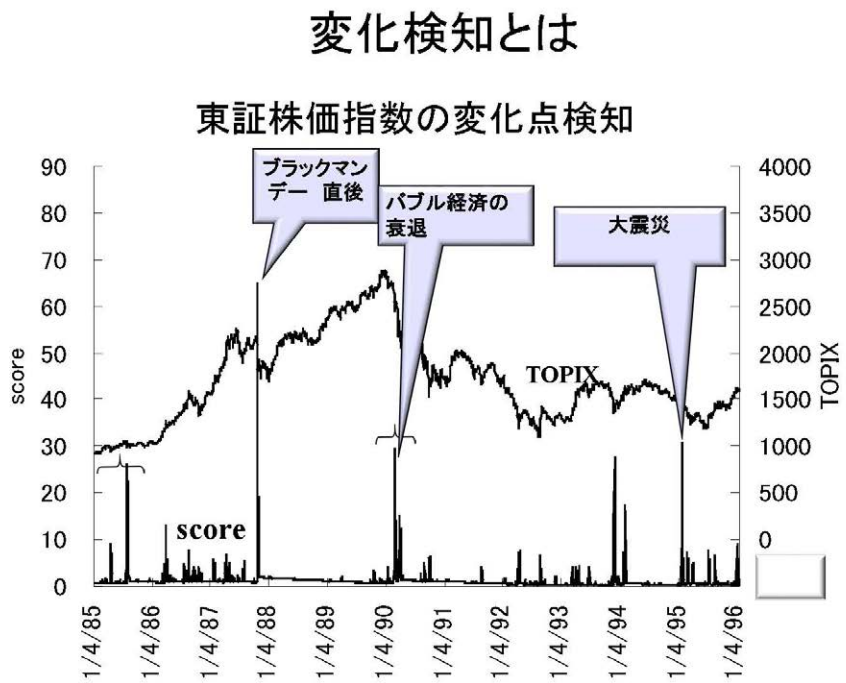
上田氏の講演は、「ノンパラメトリックベイズ法に基づく関係データマイニング」として、関係データの説明からはじめ、応用を見据えた例も交えて紹介された。

［Key point］

・予兆検知に有効な量としてMDL変化量がある。（山西）
・予兆アラートを決める閾値の決定には、統計的検定の技術を利用することも出来る。（山西）
・関係データの分析研究にノンパラメトリックベイズを用いることが出来る。（上田）
・関係データの分析に、Baxter permutationプロセスの利用を提案した。（上田）

［山西氏講演：詳細］

山西氏からは最初に講演の目次として、講演の流れの説明があった。以下、氏からの講演という形で記載する。まず、変化検知とMDL変化統計量について説明する。問題とするのは変化検知、予兆検知というタスクになる。これはデータマイニングの古典的な問題の１つとして知られている。問題を説明するためにスライドでは東証株価指数（TOPIX）の変化検知を例にグラフが提示された。ギザギザに現れるスライドのグラフがTOPIXの時系列を表している。各時刻に変化点のスコアをつけるアルゴリズムがあったとして、スコアが立ち上がったところが変化点に相当する。ブラックマンデー直後、バブル経済の衰退や大地震などの非常に重



要なイベントに対応することがある。

このような変化点検知の意義は、重大なイベント出現の早期検知を促すことにある。この意味で変化点検知は知識発見に直結する技術になる。 時系列データには様々あり、例えば通信ログでは、変化点はマルウェアの発生に対応することがある。また、計算機の利用記録では変化点は犯罪の発生に、そしてシステム動作記録では変化点は障害の発生に、そしてツイートなどでは新規話題の出現に関係することがある。

変化点を求めるために従来から行われている方法としては時系列の背後に確率分布を想定し、各時刻前後の分布のKullback-Leibler divergence（KL情報量）と呼ばれる量:

$$\mathrm{D}(p_2||p_1) = \sum_x p_2(x) log \frac{p_{2(x)}}{p_1(x)}$$

を求め、この値が大きくなるときに突発的な変化が現れたと考えるのが一般的になる。実際にガウス分布の平均が変化する場合や分散が変化する場合、それぞれに対してKullback-Leibler divergenceを計算すると、適切に変化をスコアリングしていることがわかる。スライドにも具体的な図が示された。 しかしながら、実際には分布は未知であって、データから推定しなくてはならない。推定に伴うコストも加味したKullback-Leibler divergenceに変わる理論的な裏付けのある量を定義しなければならない。そこで分布が未知の場合に変化指標として氏のグループが提案している量がMDL変化統計量と呼ばれるものになる。これはデータ圧縮の原理に基づいて変化の度合いを定式化することになる。

$$\Psi_t \stackrel{\text{def}}{=} \frac{1}{n}\left\{\mathcal{L}(x_1^n) - \left(\mathcal{L}(x_1^n) + \mathcal{L}(x_{t+1}^n)\right)\right\} \qquad \text{MDL 変化統計量}$$

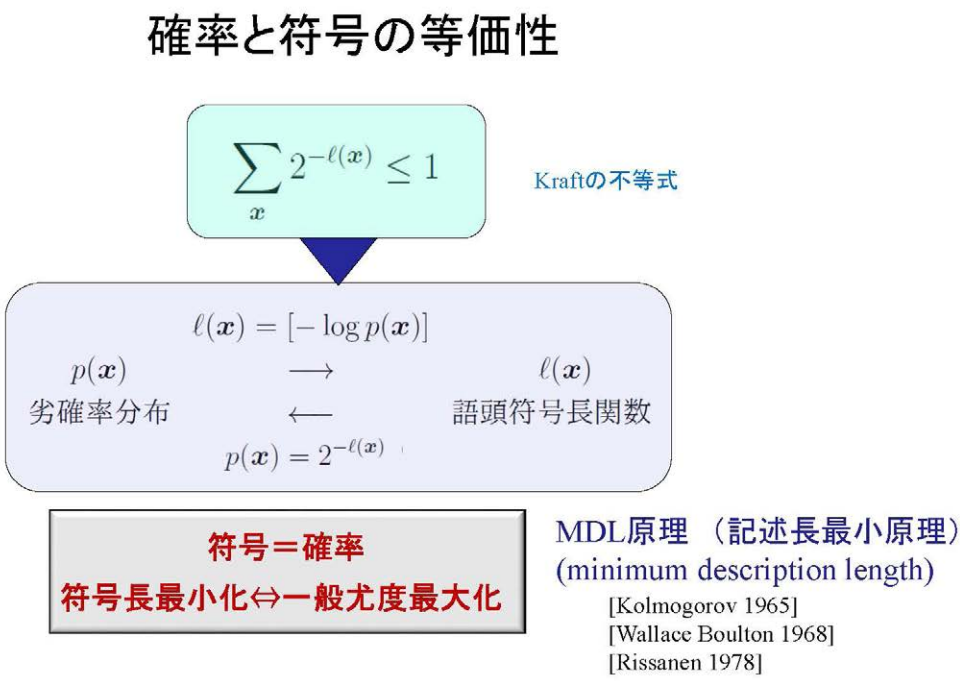
与えられた時刻tに対し、その前後で同じモデルを使ってデータを圧縮したときの記述長に対し、その前後で異なるモデルを使ってデータを圧縮したときの記号長の総和、これがどれだけ短くなっているのかというデータ圧縮の度合いをもって変化の度合いを測るというものになる。 MDLとはMinimum Description Lengthである。さらに記述（符号長）符号化を使ったので、簡単に説明しておく。仮想的な通信問題を考えて、情報をビット系列に変換することに対応する。その場合、符号化としては異なる情報に対しては一意的に誤りなく正しく異なる情報として復元できるように符号を設計する必要がある。 そのために、1つの符号は他の符号語の語頭であってはいけないということ、つまり語頭符号化の条件を満たさないといけない。 実は、情報理論ではそのような語頭符号化が存在するための必要十分条件が知られている。結局、符号長関数 l(x) と書いたときに、

$$\sum_{x \in \mathbb{N}^n} 2^{-l(x)} \leq 1 \quad \text{（Kraft の不等式）}$$

が成り立てばよい。このKraftの不等式を通じて、すでに確率分布が得られていれば、データに対してこの確率のマイナスログを取ることにより、語頭符号超関数 $\mathrm{l}(\mathrm{x}) = [-\log p(x)]$ を得ることが出来る。一方で確率分布がわからない場合に、最初に符号長が与えられていれば、この2の指数として符号長のマイナスを取ることで確率分布を構成することができる。これは符号と確率が等価であるということを示している。

結局、符号長を最小化することは、一般的な意味での尤度を最大化してモデルを推定することと等価になる。 これがMDL原理（記述長最小原理）と呼ばれるもので、情報と統計を繋ぐ一大原理になっている。

さて、そこで記述長を実際にどのように計算するかは、データの正規確率分布pが既知の場合では、与えられたデータベースに対して確率の負の対数をとれば、記述長を計算することができる。これをShannon情報量と呼ぶ。しかし、生起確率分布が未知である場合には、未知パラメーターシートで指定された確率分布のモデルクラスと呼ばれるものを導入する。スライドで提示されたθには最尤推定量、すなわち、データに対して確率を最大にするものとするが、このθはデータごとに違うので、最大尤度では確率分布を成しえない。そこで正規化して、確率分布化したものを正規化最尤分布とする。データの分布が未知の場合でクラスだけが与えられたときにはデータに対して正規化最尤分布に対する符号長を計算して、これをデータの記述長とする。これを正規化された符号長と呼び、実はパラメーターが未知の場合、平均符号長の下限を達成すること



が示されており、これをもって分布が未知の場合の符号長として採用することができるようになった。 MDL変化統計量の定義の中の符号長の計算部分の中に正規化符号長を用いて書き下ろすことができる。スライドには、計算式も提示された。

実はこれはパラメーターが既知の場合には漸近的にKullback-Leibler divergenceを近似することがわかっている。

以上のMDL変化統計量というのは、指定された時刻で 変化があったのかなかったのかを検定する際の統計的検定量としても機能する。 つまり、MDL変化統計量が閾値 ϵ より大きな値をとれば変化があったとし、それ以下であればなかったと判定する統計的検定になる。これに対して変化の過誤、すなわち誤報を出す確率と、第２種の過誤、すなわち変化を見逃す確率を、それぞれ抑えることができる。このことは正規化符号長を使うことによって初めて可能な解析になっている。

これにより第一種の過誤の確率がパラメーターデルタで抑えられるような閾値を

$$\text{Type I error prob.} < \delta \quad \text{つまり、}$$

$$\varepsilon > \left\{ logC_n + log\frac{1}{\delta} \right\} \approx \frac{1}{n}\left(d\log n + \log\left(\frac{1}{\delta}\right)\right)$$

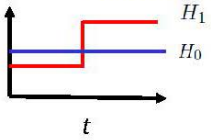
のように決定することができて、スコアがこの閾値以上になるとアラートを上げるという設定をすることなどもできる。

以上の理論に基づき、data streamが与えられる場合に逐次的に変化スコアを計算するアルゴリズムを設計することができるようになる。それは、固定幅のウィンドウを与えて、その中心点のMDL変化統計量をスコアとして算出し、そしてウィンドウをずらす。これらのスコアをプロットした時系列に対し、閾値を越えたところでアラートを出す。

このアルゴリズムを現実の問題に適用した例を２つ示す。 １つはセキュリティーの事例である。アクセスのログのデータを同一IP-URLの通数の時系列に変換し、先のアルゴリズムに適用したものになる。その結果、SQLインジェクションと呼ばれるマルウェアの発生と、その予兆に相当するスキャニング行為に対してアラートを上げることが出来ていることがわかった。もう一つの例はセンサーデータからの障害検知の事例になる。工場の40次元のセンサーデータになる。スライドにも１部が提示された。工場には踏破事故というものがあり、

MDL変化検定

Hypothesis testing: t is a change point or not



$$\begin{cases} H_0: & x_1^n \sim p(X^n;\theta_0), \\ H_1: & x_1^t \sim p(X^t;\theta_1),\ \ x_{t+1}^n \sim p(X^{n-t};\theta_2). \end{cases}$$

$\theta_0, \theta_1, \theta_2\ (\theta_1 \neq \theta_2)$ are all unknown.

Test statistics:

$$h_0(x^n;t,\epsilon) = \Psi_t^{(0)} - \epsilon$$

H_1 is accepted if $h_0(x^n;t,\epsilon) > 0$, otherwise H_0 is accepted.

Theorem　Error probabilities: [Yamanishi Miyaguchi IEEE BigData2016]

変化点誤報確率

$$\text{Type I error prob.} \ \leq\ \exp\left[-n\left(\epsilon - \frac{\log C_n}{n}\right)\right],$$

変化点見逃し確率

$$\text{Type II error prob.} \ \leq\ \exp\left[-n\left(d(p_{\mathrm{NML}}\cdot p_{\theta_1*\theta_2}) - \frac{\log C_t C_{n-t}}{n} - \frac{\epsilon}{2}\right)\right]$$

$$p_{\mathrm{NML}}(x^n) = \frac{\max_\theta p(x^n;\theta)}{\sum_{y^n}\max_\theta p(y^n;\theta)},\quad p_{\theta_1*\theta_2}(x^n) = p(x_1^t;\theta_1)p(x_{t+1}^n;\theta_2).$$

その予兆としての原因を知りたいということで、過去に遡って変化検知を行ったものになる。モデルとしてはGaussianモデルとlinear regressionをもちいて調べた。事故に1週間先立ってアラートが出ていることがわかる。事故の原因を調べたところ、実際に事故の予兆となるボイラーの目詰まりに相当する部分が発見されたとのことであった。これらの事実は変化検知が知識発見を促す事例を示していると言える。

次に、変化予兆検知と微分的MDL変化統計量について紹介する。これまで考えてきた変化は突発的な変化であった。徐々に変化が起こっていて、いつの間にか気が付いたら、大きな変化に成長していることもある。この場合のほうがむしろ多く、このような変化を漸進的変化と呼ぶ。このような漸進的変化の開始点を変化予兆点と呼ぶ。 これはなかなか気がつくのは難しいが、ここでは変化予兆点を早期に検知することを目標とする。 そのためには、まず各時刻の 前後の分布のKullback-Leibler divergenceは計算できるとすると、スライドの2番目のグラフのようになり、時間で微分すれば、微分値が立ち上がるところと変化予兆点が対応することがわかる。つまり、変化予兆点を検知する1つの鍵は、Kullback-Leibler divergenceの微分値を捉えることであると言うことができる。ただし、ここでも分布は未知であるから、データから推定しなければならないという問題は残る。そこで分布が未知の場合にKullback-Leibler divergenceの微分

$$\Phi_t^{(0)} \stackrel{\text{def}}{=} D(p_{\theta_{t+1}} || p_{\theta_t}), \qquad \Phi_t^{(1)} \stackrel{\text{def}}{=} \Phi_t^{(0)} - \Phi_{t-1}^{(0)}, \qquad \Phi_t^{(2)} \stackrel{\text{def}}{=} \Phi_t^{(1)} - \Phi_{t-1}^{(1)}$$

を考える代わりに、微分的MDL変化統計量を考える。これはdifferential MDL change statistics（D-MDL）と呼び、次のように定義する：

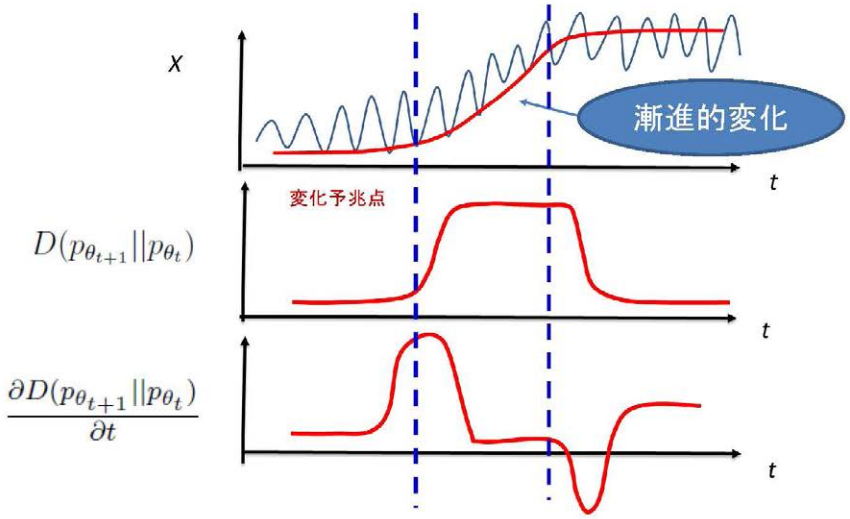
$$\Psi_t^{(1)} \stackrel{\text{def}}{=} \Psi_{t+1} - \Psi_t, \quad \Psi_t^{(2)} \stackrel{\text{def}}{=} \Psi_t^{(1)} - \Psi_{t-1}^{(1)}$$

0次のD-MDLがもともとのMDL変化統計量に相当し、一次のD-MDLと二次のD-MDLがMDL変化統計量のvelocityおよびacceleration速度加速度に対応する。

実は、一次のD-MDLは統計的検定量としての意味を持つ。つまり、変化が時刻tであったのか、それより進んだ先の時刻t + 1であったのか、このいずれかの場合を検定するときに一次D-MDLが閾値 ϵ より大きければ時刻があとの方に変化があると判定する。このことは時間がたつにつれて、大きい変化があったかどうかを検定することに相当する。閾値 ϵ に対して第1種の過誤と第2種の過誤を理論的に導出することができる。

変化予兆検知

漸進的変化の開始点＝変化予兆点　をKL情報量の微分で捉える



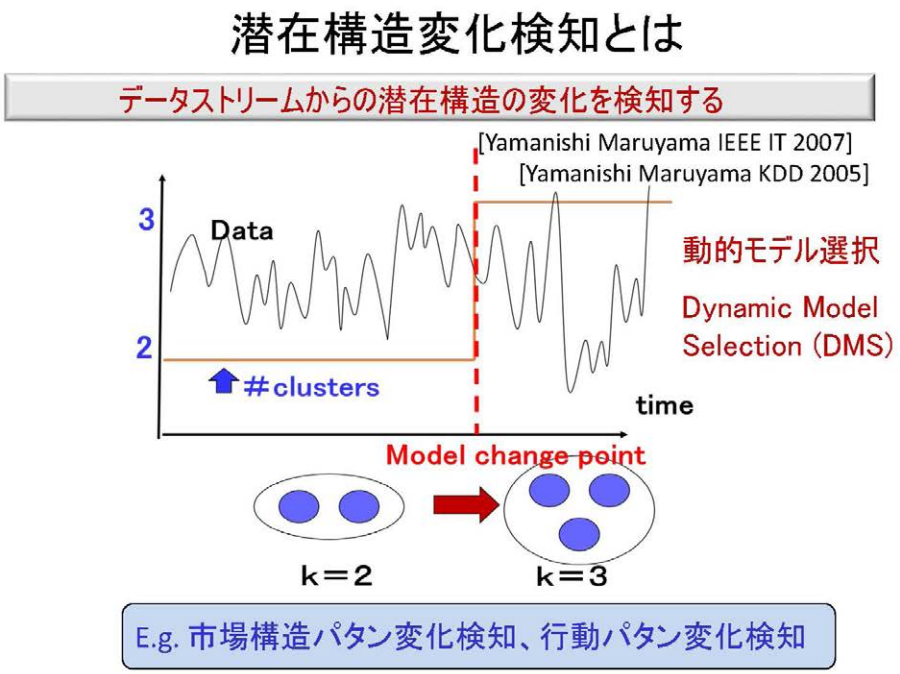
二次のD-MDLについても、これは時刻tで変化があったのか、それともその前後で2回に分けて曲率をもつ変化があったかを検定する検定統計量として機能する。その際の誤り確率も同様に評価し、閾値を決定することができる。

これらを総合して、変化検知と変化予兆検知を階層的に行うアルゴリズムを設計することができることになる。まず、ここでは可変ウィンドウとして第１種の誤りが有限個で止まるための閾値が設定できることになり、ウィンドウをこの値を超えるまで伸ばして、ひとたびこの値を超えるとウィンドウをシュリンクして、またリスタートする可変ウィンドウの構成を行なう。 そしてそこに至るまでに、１次と２次の変化予兆スコアが値を超えると、１次と２次の変化予兆アラートを上げるという仕組みにする。

このような変化予兆検知手法をCOVID-19のパンデミック解析に応用した。 データはECDCという機関が発行している各国の感染者数時系列データになる。 目的は、COVID-19の感染爆発及びその予兆点を検知することになる。モデルとしてはGaussian Modelと Exponential Growth Modelを用いている。Exponential Growth Modelでは累積感染者数が指数的に増大する際の指数が基本再生産者数―１に比例するということが知られているので、基本再生産者数の変化検知を行うことになる。

このような解析を３７か国で行ったところ、１０６個の感染爆発に相当する変化点も検知された。うち６４％については平均６日前に予兆アラートがあった。さらに、Social Distancing の前に変化予兆アラートを高々５個に抑えられた国は、感染抑制に対応する変化点までの期間が平均３０日になり、有意に短いことがわかった。このような分析は完全にデータドリブンの方法で、SIRなどとは本質的に違うものになる。

次に、構造変化検知と動的モデル選択について紹介する。 ここまで扱ってきたのは連続的なパラメーターの変化であった。次に離散的な潜在構造の変化の検知を対象とする。例えば、時系列として入ってくる多次元データをクラスタリング、つまり似たようなデータをまとめてみる。その際のクラスターが潜在構造になる。そのクラスターの数が時間とともに変化する状況を考える。このクラスター数の変化の検知を、潜在構造変化検知とする。 例えばマーケットの行動パターンの変化の検知などに応用できる。 この潜在構造変化検知は再びMDL原理を用いて実現できる。 つまり クラスター構造の時系列を求めるために、モデルとデータはともに未知の確率モデルに従って時間的に推移していると仮定してモデル系列の記述長、さらにモデル系列が与えられたときのデータの記述長それぞれを予測的に符号化して考えて、これらのトータルを最小化するようなモデ



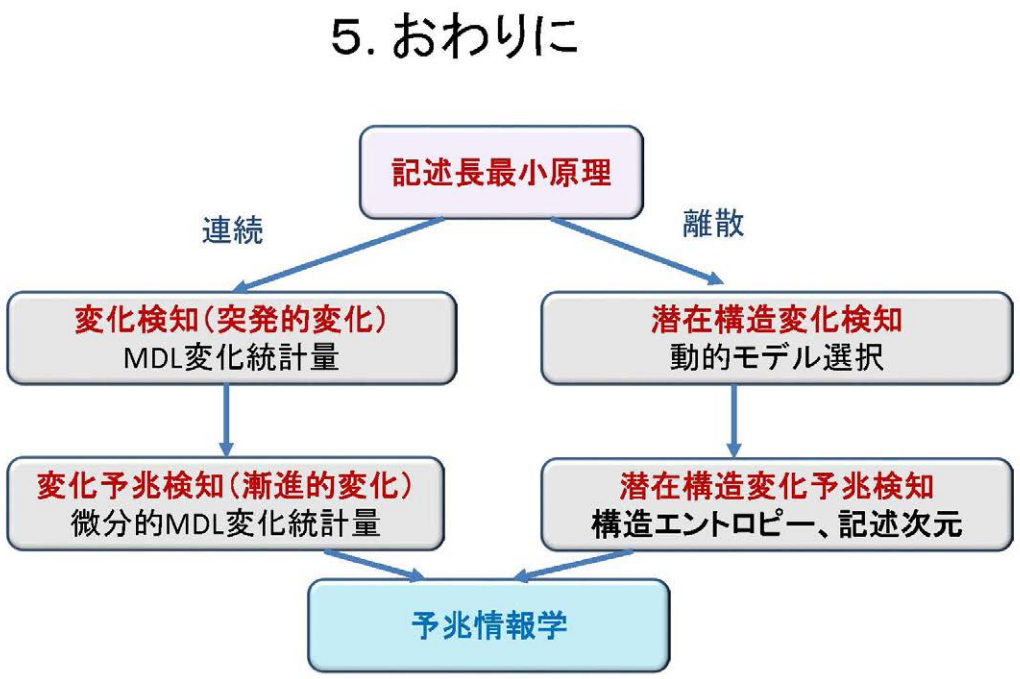
ル系列を選ぶということになる。一般にクラスター数を求めるというのは、モデル選択の問題と呼ばれているが、ここではモデル選択と変化検知の両方の問題を考えていることになる。

実際に動的モデル選択を現実のマーケティングデータに適用した事例を示す。データは、博報堂とマクロミル社による、ビールの購買データになる。300人余りに対して、14種類のビールの購買履歴を時系列データから似たような購買パターンを持つユーザーのクラスターの推移をトラッキングした。その結果、ある時刻でクラスター数の変化が検知できた。スライドの図では年末商戦に喚起された通常とは異なる重要に対する購買パターンのクラスターが発生したことが示された。特定のプレミアムビールを好んで購買するクラスターが発生したことが確認された。これは市場動向に関する知識発見の実例になっている。

最後に、潜在構造変化予兆検知と構造的エントロピーを紹介する。クラスターなどの構造は離散的なものなので、その変化は突発的に起こるものと考えられがちである。しかし、例えばクラスター数が3から4に変化する際に、どちらともつかない状況が現れる。このようなモデル選択に伴う不確実性を定量化して、これが最大になった所で潜在構造変化の予兆が生まれたと考えることができる。このような、モデル選択の不確実性を定量化したものを構造的エントロピーとよんでいる。モデルの事後確率に関するエントロピーとして計算され、このモデルの事後確率は正規化最尤符号長をexponential の指数に入れて分配関数を作ることがポイントになる。温度パラメーターβを適切に取ることにより、構造エントロピーからalarm rateが信頼性のあるものになる量になることを明らかにした。構造変化の3、4日前に、構造的エントロピーが高くなっており、新しい購買パターンが出現する予兆をとらえていることがわかる。

データマイニングと知識発見の側面として変化検知を紹介した。連続的な変化の検知の問題と突発的な変化の検知および漸進的変化、予兆検知の問題もある。離散的な潜在構造の変化の検知の問題としては、その潜在構造変化検知およびその予兆検知を紹介した。いずれも記述長最小原理に基づいて統一的な方法論が与えられること、それらの有効性があることについて論じてきた。

この方法論は広く医療や経済にも展開している。予兆を捉えるための手法は数多く存在するが、これらを統合し、現在、予兆情報学という体系を構築している。今回はその極一端を説明した。

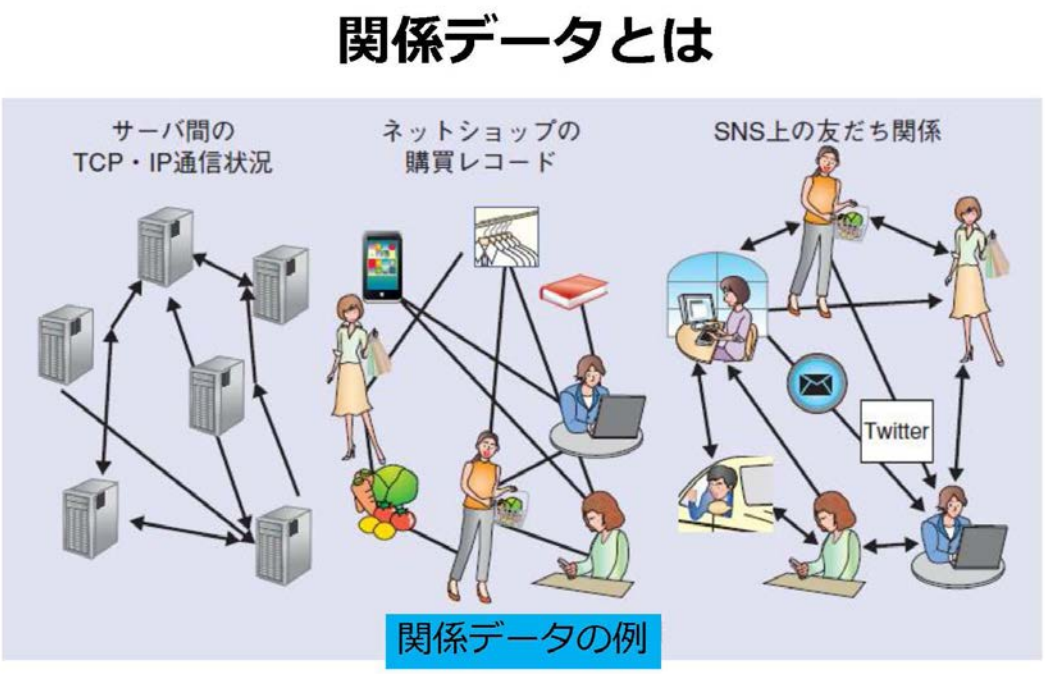


［上田氏講演：詳細］

今回は、NTT基礎研における研究を主に紹介したいと思う。

まず関係データということについて説明したい。関係データというのは、リンク関係で繋がっているようなデータである。例えば、サーバーのTCP/IPの通信状態、購買レコードの情報、やSNSでの友人関係、このような情報は一般に関係データと呼ばれているが、関係データ解析とは、この関係データから隠れた関係を抽出するための技術である。

具体的に解析するための観測データは、例えばオンラインショップの購買履歴などでは、ある人が何かを



何個購入したというような情報になる。SNSならば、例えばAさんはBさんにメールを送った、あるいは送っていない、などの情報が観測される。これらの情報を行列表現で表わす。

重要なことは、個人情報は一切なしに、つまり、顧客あるいは個人情報が一切使えない状況で解析することにある。具体的な解析では、関係データを表す行列の行と列をうまく並び替えて、特徴が見えるようにする。例えば、“この辺の人達はこういうものをよく買いますね”などとわかるようにする。

我々は、このようなことを自動的な計算で実現した上で、関係データとして得たい。山西氏の話は情報理論にベースを置いているが、上田氏は統計的学習を基に考える。統計的学習は何かということを説明すると、まず観測データを観測変数 x_n で表し、観測に現れない変数として潜在変数 z_n 、さらにパラメーターΘによって統計モデルとして

$$\mathrm{p}(x_n, z_n;\ \Theta)$$

確率モデルからデータが生成されていると仮定する。生成モデルというのは、この確率モデルを提示しないといけないが、潜在変数はモデルを柔軟にするために観測されないような構造を入れないといけない。ここがこのモデルの重要な要素になる。モデルのパラメーター（確率分布を特徴づける変数）もある。ここの議論はベイズ理論に基づくので、パラメーターそのものも確率変数として見なすことにする。こうやって観測データが得られたら、学習はベイズ推定に基づいて事後分布を推定することになる。

$$\mathrm{P}(\mathrm{Z}, \Theta|\mathrm{D}) \propto \mathrm{P}(\mathrm{D}|\mathrm{Z}, \Theta)\mathrm{P}(\mathrm{Z})\mathrm{p}(\Theta)$$

講演では、上式の説明 として、ベイズ推定の紹介もされた。クラスタリングをしたい積分消去可能な場合や事後予測分布つまり事後分布による期待値についても補足された。

今回の講演のノンパラメトリックベイズでは、潜在変数の事前分布をどう定義するか、クラスタリングをしたい場合は、クラスター数を何個にするのかというようなことになる。 AICやBICなどの情報量規準もあるが、このような組み合わせ問題の場合は、すべてにわたって 計算して、どれが一番モデルとして良いかと見るのは非常に非効率になる上、ある仮定のもとでの計算をしいるので、任意のデータに使えるわけでもない。ノンパラメトリックベイズは、その１つの解法を与える理論となる。

なおこの理論は70年代に、統計の大家のT. S. FergusonがDirichlet 過程（DPM）を論文としてまとめた。機械学習の分野で、氏が初めて知ったのは90年半ばのR. Nealの研究で、当時、無限の中間層のニューラルネットの論文がレクチャーノートに出た。ただでさえ学習が大変なのに、中間層を無限大にすることは殆ど無謀なことで誰も相手にしなかった。氏自身もあまり注目していなかった。 このとき注目すれば良かったという後悔もあるとのことだが、2000年頃に、機械学習の大家であるM. JordanがNIPS業界でDPMを流行らせた。2010年ぐらいまでに多くの論文が出たが、これは後で紹介する、新たな展開もある。

Dirichlet分布とDirichlet過程について説明する。まずDirichlet分布は非負で和が1になるような3つ以上の変数の同時分布と思えばよい。例えばサイコロの場合ならば、1から6の目が出る確率を$\theta_{1,}, \theta_{2,}, \ldots, \theta_{6,}$とすると、これらの同時分布がDhirichlet分布に従うというように思えばよい。Θ の事前分布をどうするかについては、例えば θ_1 は正であるので、ガンマ分布にしておいても、和が1にならないといけないので、不適切となる。ところで、Dirichlet分布というのはkは固定されている。 Dirichlet過程では、kが可算無限になる。γ なるハイパーパラメーターと G_0: 基底測度が定義されていて、数学的には Θ は連続の空間だが、$\theta_{1,}, \theta_{2,}, \ldots$ は離散分布になる。このような興味深い確率過程になる。

ハイパーパラメーター γ が小さいときは、G_0 の分布の包絡線に非常に近い形になる。実際に、Dirichlet過程を実現する確率過程として、Chinese Restaurant Process（CRP）を、Aldousが提唱している。これは比喩で、中国人というのはおしゃべりが好きなので、レストランに行くと人がたくさん座っているテーブルに好んで座るという比喩になっている。潜在変数として、クラスタリングの例を考えてみる。第k番目の人はどのテーブルに座るかというのを、確率過程で規定する。ハイパーパラメーターをいろいろ変えることで、第k番目の人が、多くの客が座っているテーブルに座るか、新たなテーブルに座ることを規定する。重要なことは、CRPは交換可能なことである。つまり、座る順番によらない。この交換可能性がCRPの重要な性質であり、ベイズ推論（ギブスサンプリング）が適用可能になる。

次に、IRM（Infinite Relational Model）というMITとの共同研究を紹介する。

例えば9人の人がいる。1の人が3の人にメールを出して、5の人にメールを出した。2の人は7の人にメールを出したというデータがあったとする。この観測データがどういうふうに生成されるかというと、前述のDirichlet過程でクラスタリングができたとする。それぞれクラス内の関係を規定する。関係の有無は2値なのでベルヌーイ分布で考える。生成したデータのインデックスを並び替えることで観測データが得られていると仮定する。逆に言うと、学習のときは観測データから分割とパラメーターを学習することになる。

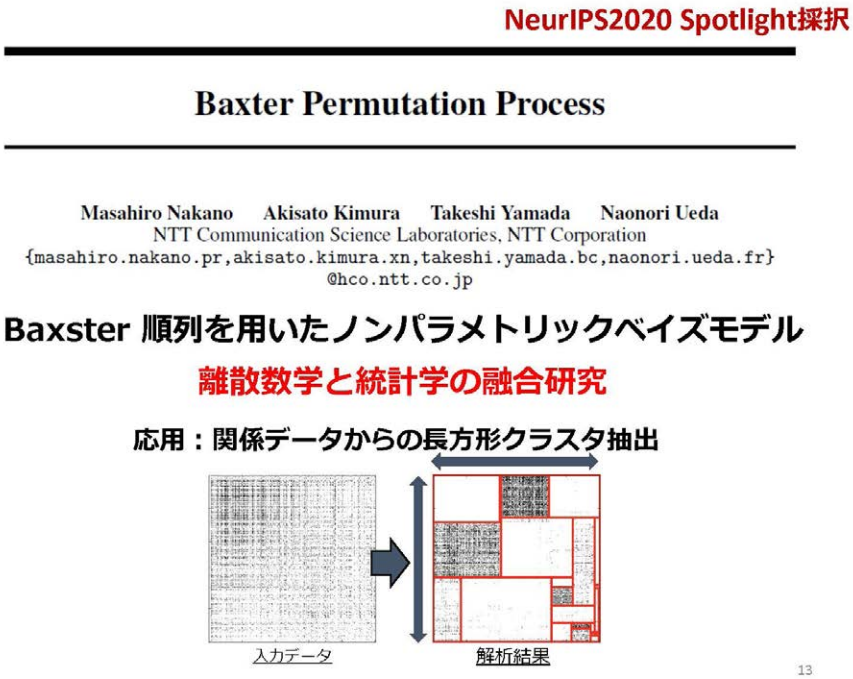
具体的には、$P(R|Z^1, Z^2)$ が Z^1, Z^2 が既知の下での関係Rの生成モデルになる。$P(Z^1)$ と $P(Z^2)$ をそれぞれ行列Rの行方向と列方向のCRPとすると、結局、事後確率

$$\mathrm{P}(Z^1, Z^2|R) \propto P(R|Z^1, Z^2)P(Z^1)P(Z^2)$$

を最大にするような、Z^1, Z^2 のassignmentが事後確率最大の観点で最良解となる。 ただし、Z^1, Z^2 の全ての組み合わせを計算するのは非現実的なので、CRPの交換可能性を生かして、モンテカルロサンプリング（ギブスサンプリング）により効率的に解く。以上がこの分野の非常に基礎的な話になる。

具体例としてアニメサイトの推薦の例を紹介された。ある人達が、あるアニメをたくさん見るということがわかれば、その人たちにrecommendしたりするサービスに実際に使っている例になっている。このような応用がある。

今日は数学がメインのトピックなのでNeurIPS2020 でSpotlightとして採択（採択論文のトップ5%）された、前述のノンパラメトリックベイズ理論をさらに拡張した話の紹介をしたい。歴史的な経緯としては、前述のMITとの共同研究であるIRMは直積的な分割を与えている。それに対してMITのDan Roy氏がモンドリアンプロセスと呼ぶ階層的な分割を2008年に提案した。氏のグループの中野允裕氏は、2014年にRectangular tiling processを提案して、これで任意の分割ができることを示した。しかし、モデルが複雑で、推論が非常に大変で、大規模データに適用困難という理由から、ノンパラメトリックベイズ理論に基づく関係



データ解析の研究も止まっていた。

それに対してNeurIPS論文では、Baxter順列、つまり離散数学の話を確率統計の分野に持ち込んだという点で独創的で高く評価されたと言える。 Baxter順列の定義に加え、例として{1, 2, 3, 4}の順列のうちで2413と3142はBaxterにならないという説明がなされた。

当初、中野允裕氏がBaxter順列と長方形分割との関係を発見した際、関係データ解析に新たな発展の可能性を感じた。離散数学の研究では、Baxter順列の研究が進んでいることも後から知った。例えばtwin binary treeと呼ばれるデータ構造など多数の結果が知られている。例えばLSI設計などに使われているmosaic floor planというfloor plan（部屋割り）とBaxter 順列とが一対一に対応していることがすでに証明されている。実際にBaxter 順列からfloor plan を作るアルゴリズムも提案されている。

加えて、サイズn-1のBaxter 順列 からサイズnのBaxter順列 の生成が、left-to-right maxima または right-to-left maximaで定義される条件を満たすことで実現可能であることも説明された。このことを、確率モデルにしたい。つまり第n時刻に対し、次にn+1時刻のBaxter 順列を生成する確率過程を定義する。

これらを関係データにどのように使うかということになる。ノンパラメトリックベイズの文脈で現れるStick-breaking process（SBP）と呼ばれる確率モデルがある。長さ1の棒を β と1-βの比で折る。βはβ分布で生成する。同様にこの棒折過程を繰り返すとDirichlet過程になる。このことを利用して2次元化する。前述のfloor planでは、各部屋がメッシュ（離散サイズ）で作られる。この部屋の縦横のサイズを正の実数化すべくSBPを用いる。これによりBaxter PermutationProcess（BPP）と呼ぶ任意の長方形分割の新たなノンパラメトリックベイズモデルを実現した。理論的な証明もいくつか示している。

関係データに応用する際、ブロック（関係クラスター）は前述のBPPで作る。関係データの行列のi, j要素は、カテゴリカルデータになり、行列Xは非常に大きな行列になっている。 BPPを適用する際、行方向と列方向をサイズ1の正規化を行い、正規化した範囲で[0 1]一様乱数を考えて、一様乱数の行と列の座標値、それがfloor planのどのブロックに属すかが分かれば、予め定義したDirichlet 分布で カテゴリカルデータ（0 or 1）が生成される。これをデータの生成過程とする。

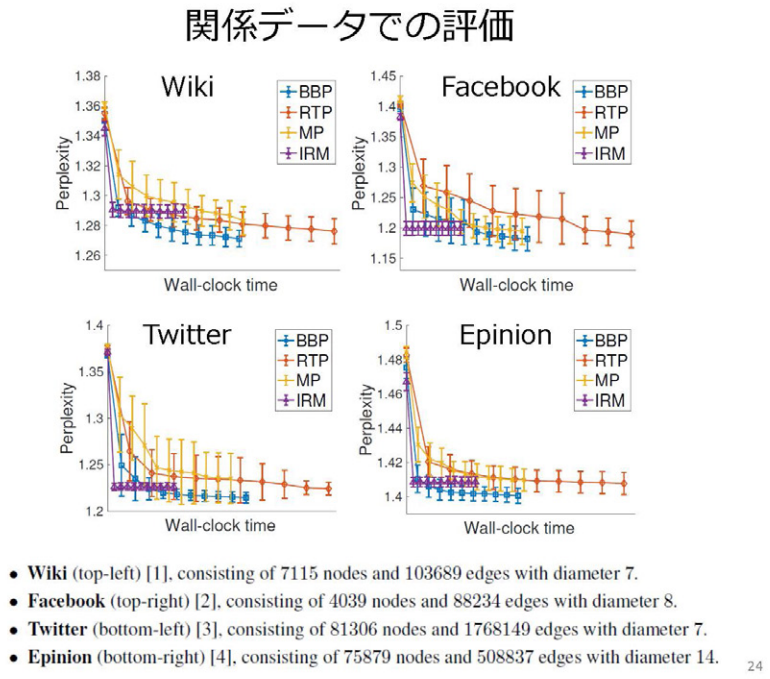
実際には同時分布として、

$$\mathrm{p}(\mathrm{X}, U^{row}, U^{column}, U, \beta)$$

を計算することになる。

学習するときには、同時分布を最大にするようにマルコフ連鎖モンテカルロ（MCMC）法でパラメーター推定を駆使しながら行う。

最後に実データの例で、FacebookやTwitterなどで関係データでの評価を説明された。これらのデータは、前述のリンク情報だと思ってもらえばよい。 どれとどれが繋がっているか、Facebookでも繋がり、Twitterは誰が誰に話しているなどのデータになる。ここでは、BBP、RTP、MPとIRMの比較を評価として説明された。



【質疑応答】

Q：COVID-19でSIRモデルってを使っている。講演の内容では、補足的にデータの変化点を見ることで、完全にデータ駆動な手法を与えていると思う。医療ということを考えたときに、徐々に病気になっていくとその揺らぎを見るということで分かることがあると思う。それはどっちかというと、力学的な視点だと思う。一方で心臓発作や脳溢血などは来るだろうと思いつつも、いつ来るかわからないということがある。それらにも今と同じように、微分的MDL変化を見て、兆候をつかむことができるのか。

A：予兆には、いろんな変化のタイプがあり、このタイプの変化については、この手法で予兆が見つけられるという類の整理が必要だと考えている。この講演の内容は、変化の開始点を求めるにはどうしたらよいかに限定した取り組みになっている。 病気でも徐々に症状が現れて、最終的に大きな変化につながることもある。気付かない時に症状が進んで、突発的にその疾病が起こるものは、情報自体が不足していると思う。より多くの情報を加えながら、別の見方をして考えればいいと思う。例えば、物理モデルでも相転移みたいなものもあり、相転移が起こる前の予兆というものを検知する話もある。現象をはらんでいるものは、様々な手法と結合させながら、変化予兆検知することは今後考えられると思ってい

る。ハイブリッドという感じになる。一概にこの手法だけで予兆が捉えられる訳では無く、時に応じた変化の予兆を現在整理している段階だと考える。（山西）

Q：例えば様々な予兆を見るときに、データにはたくさん独立な変数もたくさんあると思うが、その変数ごとに、講演の手法とダイナミカルな手法を分けて考えるというイメージになるか。

A：変数を独立的に扱うと言うよりは、変数が互いに絡んだネットワークを考えて、ネットワークの大域的な構造の変化を検知して行くことになる。いろいろと絡んでいる。例えば、ネットワークシステムであれば、それぞれのネットワークのごとに様々なシステムの端末がある。そういったところの挙動の相関を考えて、大きな相関行列を考えて、マクロな挙動としてスペクトラムなどに代表されるような性質を捉えていく形になると思う。（山西）

Q：後半の講演は、クラスタの変化や潜在的な構造変化を捉えるべきだという意味だと思うが、例えばTopological data analysisが行なっているようなパーシステント図のようなもので、大域的構造を見て行く事との関係はあるのか。 TDAでは、データを幾何学的対象だと思って、幾何学的なものを代数的に記述する。データというのは、いわゆるホモロジー群の変化を見て行くことによって状況を把握しようというものになる。手法としてテクニカルには違うのはよく分かるが、見ているところが似てる部分があると思って質問した。

A：代数的指標のようなものの変化も考えているのなら、大いに関係があると思う。この講演では、クラスター数というものが、代数的な指標に相当するので、例えば細かなパラメーターの変化は、トポロジカルには、代数的な指標にまで関わるような話でない。そのような大域的な構造の変化を捉えるという意味で共通性があるとすれば、多いに関係があると思う。（山西）

Q：完全にデータからのアプローチであるが、例えばデータからアラートが出たときに、アラートを信じるかどうかは、モデルの方に戻して両方付き合わせて考えるということになるのか。

A：理論的にはアラートの信頼性を評価する。第一種の過誤、すなわちfalse alarmであるかどうかということに対する確率を計算する。あるパラメーターで抑えられたときに、信頼性を担保できる閾値を求め、それを超えるとアラートを出すことにする。 そこまでの議論をすすめている。なので、ある意味その誤り確率というのを、統計的検定の枠組みの下で詰めた上で、アラートを出している。（山西）

Q：工場の事故の話があったが、どのようなデータを使っているのか？

A：工場のデータは詳しくは説明できないが、共同モニタリングしているセンサーデータになる。40カ所のセンサーデータを多変数データと考えて、多変数の中から異常を検知したことになる。（山西）

Q：多次元で見ているということか。

A：40次元になる。 COVID-19は、感染者数だけになる。（山西）

A：今日は省略したが、例えば死者数についても、20次元にして、各国の感染のクラスタの移り変わりも、実は解析することができる。動的モデル選択を使うと最初は中国で感染が多く、その2番手にイタリアがあったが、それがある時点で細分化され、さらにその細分化構造が変わり、最後にはそのクラスターや感染が蔓延している国とそうではない国に分かれていたという類の分析も出来る。その場合は、1次元だけではなくて多次元データになっている。このようなクラスターの変化およびその変化予兆も検知することができる。（山西）

Q：位相的データ解析の話が出た。クラスタリングを見たときに、パーシステントホモロジーの構成に似ていると思った。関連性はあるのか。

C：具体的な見方はわからないが、パーシステントホモロジーを使うときと雰囲気が良く似ていると思う。完全には分からないが、関係があるかと思う。

C：トポロジカルデータアナリシスは今まであまり注目したことがないので、大変有効な情報になった。（山西）

Q：交換可能過程ということだったが、完全に非可換の過程にすると出来ないとは思う。ただし、ある種

の非可換性というか、ある一定の条件を加味したモデルなら作れそうな気がするが、どうなのか。

A：どのような分割が起こるかというところをモデルとして、そういうことで非可換性を入れ込むこともできると思うが、構造としてはシンプルにしておかないと、不都合がある。（上田）

Q：マルコフ過程の話があったが、漸近的な表現論というのもある。例えば、対称群の全体のinductive limitを取って、無限次元の無限次対称群の表現論とかが確率過程としても研究されている。表現としても研究されているが、そのあたりの研究というのは、このBaxter permutationのカテゴリーでも研究は行なわれているのか。

A：それは我々がやっている。 Baxter permutationは、離散数学の世界では、確率過程的な議論を我々の知っている範囲ではなされていない。離散的にpermutationを効率的に数える方法というものだとか、Baxter Permutationの満たす性質や、符号化などでよく使われている。確率過程として着目しているのは我々の研究と言える。これを機に、掘り下げていきたい。確率過程的な理論的な性質も整理はしている。質問の点が非常に重要だと思っている。（上田）

C：ヤング図形などで図形の中で、いろんなパスを考えていくなど、そういう話も非常に深く関係しているような感じがして、大変面白いと思った。

Q：モンドリアンプロセスとか、四角形のところにこう埋めていく話があったが、例えば、他の与えられた形でも同じような議論が出来るのか

A：非常に的を射た質問だと思う。論文の査読者からもそのようなコメントがあった。 2次元化するだけでも、いろんなことを保証しないといけないので、大変な苦労があった。斜め方向とか、そういうようなことになると、さらにいろいろ検討しないといけなくなる。例えば2次元をtensorのように3次元化するとか、可能性は色々あると思う。簡単に一般化できる話ではないが、tensor的なフロアプランなど定義し、整理することは重要だと思っている。（上田）

C：この分野にはそんなにたくさんの研究者がいるわけではないので、論文を出すと査読者も興味半分で色々な事を言ってくるので、大変な状況にある。（上田）

C：山西氏の講演にも最近の興味深い様々な実例があったが、我々は関係データが道具としても重要な技術だと思っている。この分野の確立を発展させていきたい。（上田）

Q：関係データの評価は、グラフとしても確認する。

A：はい。（上田）

Q：関係データの評価の元のデータは全体か。Twitterとかでも。

A：元のデータは全体になる。（上田）

C：8万に納まるんですね。

Q：単純に、例えばフォロワー数が多いとか、そういうことを見ている。

A：これはあくまでこのデータでしかなくて、中身のプロフィールなどは一切見ることができない。あくまで、モデルのフィッティングが従来モデルと比べてどうかという比較をしている。（上田）

C：夢のある話な気がする。

C：個人情報が出た瞬間、研究が止められた。

C：様々な需要がありそうな感じがする。

Q：Dirichletプロセスというのをよく聞くが、どこを説明していたのか。

A：Dirichletプロセスというのは、無限次元の確率過程を作ることを言っている。その上位概念がノンパラメトリックベイズになる。パラメーターが無限に増えていく。我々が行っているのはBPPのような、新しい過程を提案したということになる。（上田）

3.10 データドリブン数理モデル構築（因果推論、情報量基準）

［概要］

CRDSの連続セミナーシリーズ、第10回は「データドリブン数理モデル構築（因果推論、情報量基準）」をテーマとして、二宮氏より「因果推論のための情報量規準」、伊藤氏より「経済学における因果推論 - データ分析を駆使して読み解く、人や企業の行動原理 - 」というタイトルでご講演いただいた。

二宮氏の講演は因果推論を行う際の基本的注意から始まり、Rubinの因果モデル、共変量を用いた推定が紹介された。次に、機械学習と違い医学統計などデータサイズが限られる分野では誤推定が生じ得るため、観測値に重みをつけ、欠測している情報の埋め合わせが可能になる「傾向スコア解析」が紹介された。因果を含めて、推定のためには用いる多項式回帰モデルの選択が必要であるが、過適合などの問題により妥当な選択は非自明な問題となる。そこで真の説明変数・目的変数の関係と、推定との誤差を調べる概念である「情報量規準」が次に紹介された。因果推論などに用いる情報量規準の理論は全くの未完成であり、日本人研究者も含めて注力される余地のある分野である事を述べられて講演は終了した。

伊藤氏の講演では、まず経済学の観点からの因果推論研究の歴史的背景を紹介され、実例による因果関係の誤解と、因果推論の基本的考え方が述べられた。その後、経済学における因果推論手法、特に「フィールド実験」「Regression Discontinuity Design」を実例を交えて詳しく紹介された。他に2つ、氏が研究に携わる手法が簡単に紹介され、データを「デザイン」して推論に応用する考え方の有用性を展開されて講演は終了した。

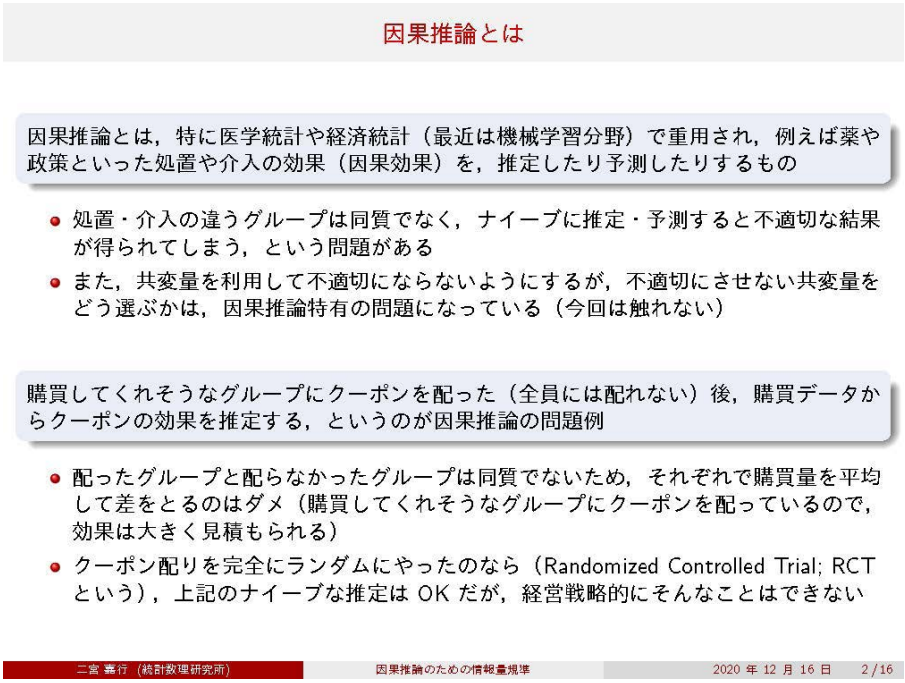
［Key point］

- 処置や介入の違うグループは同質でなく、これらを同列としたナイーブな推定・予測は不適切な結果を導き得る。（二宮）
- 発展が著しい因果推論における「情報量規準」は理論として全くの未完成で、現代統計版を開発する必要がある。日本のお家芸でもあるので、注力されるべき。（二宮）
- 社会科学データを用いた因果推論は、ビッグデータのような大量のデータ、統計学における高度なモデルの解析をもって解決できるものではない。いずれも「バイアス」の問題を解決できない。（伊藤）
- 経済学においてデータをうまくデザインするアプローチの研究が進み、これが因果関係を適切に抽出するのに大きく貢献している。（伊藤）

［二宮氏講演：詳細］

薬や政策という処置や介入の効果、すなわち因果効果を推定・予測する「因果推論」が、医学統計や経済統計、近年では機械学習分野においても多く用いられている。因果推論を行う際は複数のグループに対して処置・介入などによる変数値の差異を生じさせ、事象の変化とその間の関係を考察するが、これらの処置・介入の違うグループが同質である保証は一般になく、その異質性を深く考慮する事なく推定・予測をすると誤った結果を得る。また、グループやデータを分類するには「共変量」という基準を用いてグループ間の質を適切に修正するが、この選び方は非自明であり、因果推論特有の問題になっている。

因果推論の例題として、購買を促す（と期待される）グループにクーポンを配り、その効果を推定するという問題が挙げられた。クーポンの購買意欲向上効果を測るのに、「配ったグループ」と「配らなかったグループ」の間で購買量の平均を取るのは誤った推論を生む。その理由として、そもそも購買してくれそうなグループにクーポンは配るものであり、そのバイアスが単純な平均計算には考慮されていない事が挙げられる。一般に、クーポン配布の有無などによるグループ分けの操作を完全にランダムに実行したならば（これを「無作為化比較試験」、RCTと呼ぶ）、上記の単純な推定でも妥当な因果推定の結果を得られるが、一般に完全にランダム



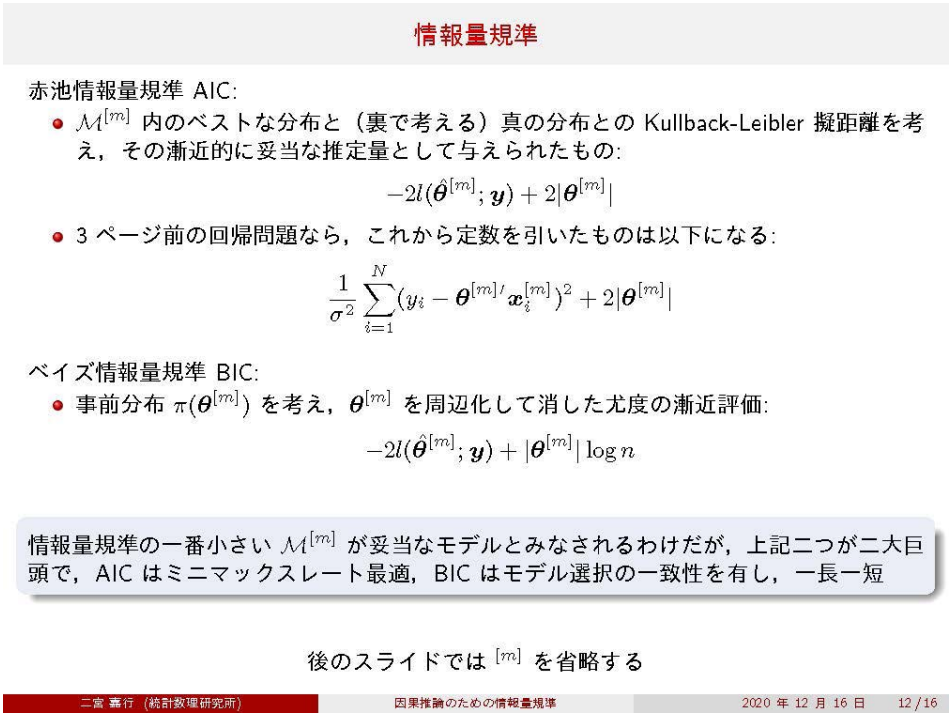
な試行は経営戦略などの観点から不可能である。

因果関係を測るモデルとして、「Rubinの因果モデル」が簡単なものとして知られている。実際に観測値に対して基本的にグループ分けのための介入の有無のいずれかしか観測されず、他方は欠測しているという反実仮想的な考え方をする。 RCTの場合は、このモデルの介入の有無の差の平均を取る事で、因果効果を測る事ができる。しかし一般には考えているグループが互いに異質であるというだけでは正しい推定はできない。そこで、異質度の説明をする「共変量」を導入する事で上記の考え方は一般化できる。

機械学習分野などでは回帰分析を用いる際、問題が高次元かつ複雑であっても（機械学習分野で代表的な方法の一つである）random forest を用いて直接モデル推定を行うが、医学統計などデータサイズが限られる分野ではモデルの誤特定、過適合、精度評価の困難といった問題が生じる。このような回帰モデリングを避ける方法として「傾向スコア解析」が導入され、因果推定に用いられる。観測値に傾向スコアと呼ばれる重みをつけることで、欠測している部分を埋め合わせる事が可能になる。

因果推論はここ20年ほど流行り続けている研究課題であるが、日本の数理統計専門家で因果推論を研究している者は非常に少ない。また何が何の要因になっているかを調べる「因果探索」も因果推論において重要な話題であるが、その考察には職人芸を要するため、研究者の数自体が多くない。

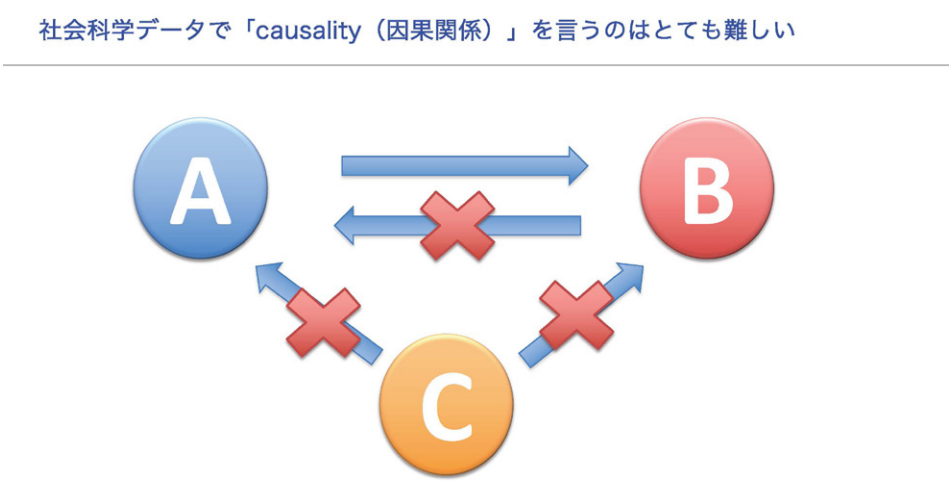
後半は、因果関係を考察するモデルを選ぶための回帰分析と、その妥当性を評価する情報量規準の話が展開された。説明変数と目的変数の関係を調べるのに、通常は多項式回帰モデルを用いるが、本来は推定の前にモデルを一つ定める必要がある。モデルに多項式を用いる場合、その妥当性は次数に依存し、一般には最大対数尤度を最も大きくするようなモデルを選ぶのが良いとされている。最大対数尤度は次数に対して単調増大するが、観測データに当てはまればいいというものではなく、当てはまりすぎて本来の関係と大幅にずれる過適合という現象が起こる。これが「最適なモデル」の選択を難しくしている。この困難を回避するための概念として、「赤池情報量規準（AIC）」、「ベイズ情報量規準（BIC）」が知られている。各々一長一短があるが、これらの情報量規準が小さい事をもって、モデルが妥当であるという判断がなされる。近年は細かい改良・一般化の話が多く、情報量規準そのものの理論はほぼ完成されたものと理解されているが、因果推論、高次元データ解析、特異モデル解析に用いるための情報量規準は、理論として全くの未完成である。現代統計版の情報量規準を開発する必要があり、もともと日本のお家芸とされるこの概念の研究は注力されるべきであるとのメッセージにより講演は締め括られた。



［伊藤氏講演：詳細］

事象Ａから事象Ｂが観測される場合、その間の因果関係がどのようになっているか、経済学においては2つの研究の流れがある。1つ目はＡからＢへの因果関係を生み出し得る（共変量に対応する）「第３の因子」を制御するという方向性である。そもそも「第３の因子」が何者か、存在するかも不明な状態で、1980年代は傾向スコアマッチングによってその制御が盛んに研究された。

近年では機械学習における制御も研究されている。一方、1986年ごろに経済学・社会科学における因果推論において「統計的手法は必ずしも有効ではない」事を示唆する研究が発表された。具体的には、ある経済・社会的対象で様々な統計スコア（因果関係の推定）をたくさん集めて、RCTに近づくかを検証したところ、



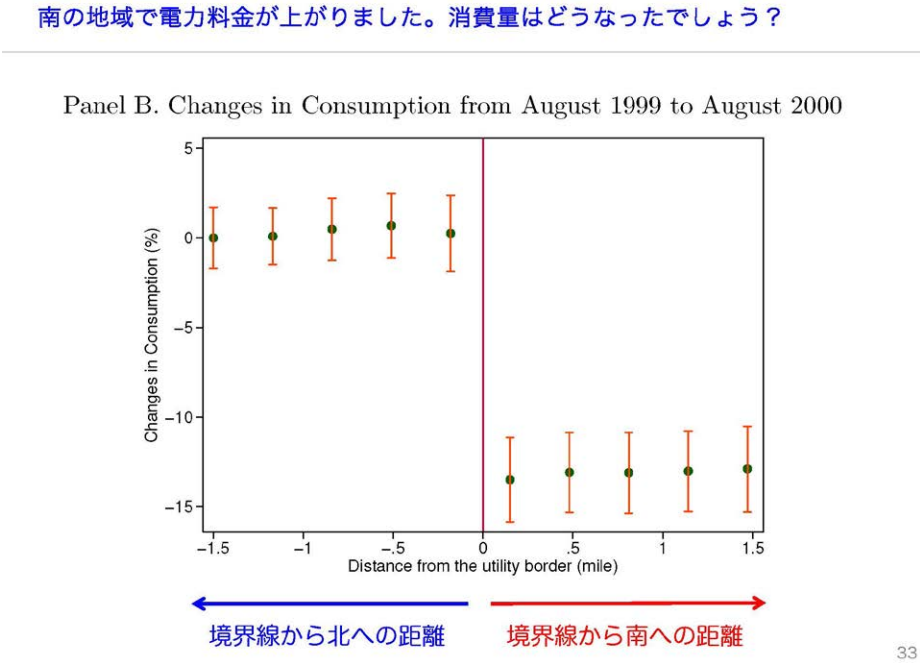
必ずしもRCTに近づかず、統計的手法が有効であると保証されない事が示されたというものである。これを機に、経済学では因果推論の研究において第2の方向性が生み出された。それは1. マッチングの改良、2. RCTをうまく使えるようにデータをうまく「切る」(デザイン・ベースド）アプローチである。

実際に「因果」を示すのは非常に難しい。「AとBのデータが一緒に動いている」という例をとると、(特に、Aを先に観測した場合)「Aが原因でBが起こった」と決め付けがちである。しかし、上の例では別の可能性もある。「Bが原因でAが起こった」という逆の因果関係もあり得れば、「別の因子CがAとBの両方に影響を与えただけで、AとBの間に直接の関係はない」という関係もあり得る。「Aが原因でBという結果が起こる」と結論づけるには、これら別の可能性を排除する必要がある。しかし、経済や社会的問題の中には「人の考え方」など、データとして存在しない、特に共変量として抽出できない因子も事象に介在し得る。これが、言及している「別の可能性」の排除を難しくしている。ビッグデータを用いるアプローチもあるが、データとして存在しない因子に対する「バイアス」の問題の根本的解決には至らない。

講演では、氏の研究テーマを含む経済学的因果推論アプローチが簡単に紹介された。

1.「フィールド実験」は社会で実際に実験をして、因果関係を考察するというものである。例として医学研究における「新薬の効果の検証」、「電力価格変化と節電行動の関係」が紹介された。 実際に多くの治験者、世帯に参加してもらい、ときには長年にわたる実験によってデータを入手し、因果関係を推定する。実際の検証からデータを取るので、ベストな方法ではあるが、コストや倫理的な面での制約が大きいのが難点である。

実際の実験が困難な状況において、あたかも実験が起こった状況を作り出して因果関係を推定する方法の一つが 2.「Regression Discontinuity Design（RDデザイン）」である。南カリフォルニアにおける電力消費量が例として挙げられた。南カリフォルニアの6都市では電力会社の境界線が存在しており、10年に渡りとても違う電力料金の変化を経験している背景がある。そのため、両地域において電力料金の変化が同様になっている場合のデータの蓄積がある。そこで、片方の電力会社だけ料金を変化させ、消費量を確認すると著しい違いが見られた。このとき、料金変化が消費量の変化を引き起こしたのだろうか？このケースでは、消費量の変化を引き起こし得る要因、例えば天候や経済的状況は、同じ都市での比較であるため消費量変化へ影響する可能性は低い。そのため、料金変化と消費量の変化の因果関係が妥当であると判断されるが、この判断を可能にしているのは管理電力会社の「境界線」である。このように事象の不連続な変化を引き起こし得る「境



界線」を見つけ、過去のデータも併せて因果関係を考察する手法がRDデザインで、フィールド実験と比較してコストも低く抑えられる。

他にもデータのヒストグラムと階段状の分布をうまく利用し、政策などの事象変化と起こった結果の因果関係を説明する3.「Bunching Analysis」、事象変化の「タイミングの差」を利用して事象間の因果関係を考察する4.「Staggered Event Study Analysis」も紹介された。

これらのアプローチは生物が起因となる事象の因果関係など、実験が容易でない対象にも適用が可能である。上記のようなデータをデザインするアプローチと統計的アプローチを組み合わせる事で、様々な因果関係を導く事が可能になると期待される。

余談　日本の新聞やテレビなどのメディアでは、現状データに対するリテラシーが非常に低く、様々な人間がデータを拾っては好き勝手な事を言っている印象を受ける。政府やシンクタンクにおける政策分析は、原因（の一つ）と結果の（時間に対する）ビフォーアフター「のみ」で決められているものが少なからずある（松江のコメント：因果関係に対する見識の浅さが露呈している？）。

【質疑応答】

Q：（二宮氏の講演における）「職人芸」とは？

A：いわゆる「潰しが効かない」テクニック（松江のコメント：他の理論などへの応用が効かないもの？）。

Q：介入の有無の（Rubinの因果モデルのような）「線形なつながり」が良く、二乗和のようなつながりを考えないのはなぜ？

A：期待値を取る事で因果効果を導けるため。

Q：「共変量」とは具体的にどういうものを当てはめる？

A：年齢や性別とか。例えばGoToキャンペーンが得策か否かも、年齢や性別で分けて考える。

Q：（因果推論において）共変量の選び方も重要になってくる？

A：その通り。一般に「無視できる割り当て条件」（二宮氏の講演スライド4ページを参照）を満たす共変量をデータから決めるのは不可能で、「いい選び方」を決めるのは重大な未解決問題。

Q：経済学において、（新旧含めて）数理科学はどの程度応用される？

A：RDデザインなどの発展は、数学的考察も同時に行われてきた。計量経済学（因果推論や機械学習）の研究者の半分ほどは数学に明るく、そのような研究者には日本人も（歴史的にも、現在も）多い。社会行動を経済モデルにする研究の理論は数学によって組み立てられ、研究者には数学科出身者が多く、経済学において数学者の活躍の場は多い。

Q：COVID-19と経済モデルの関連性について。医学としての見解と政策決定は同時に考えられるか？→経済学の観点からどのように考える？

A：一例を挙げると、（2020年の）大統領選において、トランプ氏が集会を行った地域でクラスタが発生しているという結果が論文として発表されている。この結果に基づく議論は、経済学の研究対象。日本での現状（2020年12月16日現在）は水掛け論に陥っている。エビデンスに基づいた議論がされておらず、全く生産的ではない。

なお、経済学では「人の動き」を考慮したモデリングも研究として行われている（因果推論とは少々異なるが）。

Q：2010年ごろの「日本の経済学」は数学をまともに扱っていなかった印象があるが、2000年頃は数学にめっぽう強い経済学部もあった。現在（2020年）の日本の経済学では数学の重要性を再認識する動きもあるが、「揺り戻し」が激しい。アメリカの経済学（部）において、（実感として）数学はずっと大事にされているのか？

A：（上記のような）ブレはない。確率・統計を始め、アメリカの経済学では学部から（理系と同じ扱いで）

数学の基礎をみっちり習得させる。日本とアメリカを比べると、高校までは日本のレベルが優っているが、大学にいる間にアメリカの大学に抜かれてしまう。

補足　経済学部卒の学生について、日本では「数学を習得している者」「していない者」の差が激しく、これが企業として採用を難しくしている一因となっている。
（松江のコメント：数学を習得している者が採用されやすいということか。）
同じ経済学部卒でも「アメリカ」「日本」で上記と同じ構図が起こっているとか・・・。
（松江のコメント：「アメリカの大学」卒が優遇されるという事と思われる。）

Q：モデリングやデータ解析において「ビッグデータをとって解析する動き」「人間の心理から演繹的にモデリングする動き」などがあり得ると思われるが、アメリカの経済学などでのモデリングやデータ解析に対するアプローチの仕方はどのようになっているのか？

A：「モデル化」と「データを見つめて検証」の2つを同時に行うのが経済学の良いところ。幅広くモデリングを考える事は、データを扱う際も「データが増えても解決しない問い」が生じた際に重要な動きである。
一例として、現在の経済学では「行動経済学的仮説」「別の仮説」によりモデリングを行い、どちらがより適切かをデータを使ってテストするという動きがある（データとモデリングの両方を大事にする動き）。

Q：日本のお家芸である「AIC」「BIC」の応用先はあるか？

A：因果推論でモデル選択をするという研究は非常に少ない（二宮氏の研究の一つ）。深層学習や機械学習からモデル選択ができるかという問いも、全員の悩みどころとなっている。

Q：複数の説明変数があって、「いくつかは因果推論に使える綺麗な変数で、残りは使えない変数」と区別する基準はあるか？

A：重要な問いになってはいるが、現状「無い」。

3.11 Machine Learningと数理モデル

［概要］

CRDSの連続セミナーシリーズ、第11回は「Machine Learningと数理モデル」をテーマとして、河原氏より「複雑ダイナミクスの理解への機械学習からのアプローチ」、渡辺氏より「数理モデルと代数幾何学」というタイトルでご講演いただいた。

河原氏の講演は「Koopman作用素」とその固有展開である「動的モード分解」を基礎とした力学系の推定の基礎的考え方とその応用、渡辺氏の講演は「代数幾何学」など多数の数学理論により構成された情報量規準により、多様複雑化している数理モデルに対する未知の事象の確からしさの評価が可能になる事が主な内容であった。ともに多種多様な応用がある一方、その発展には古典的かつ現代の数学理論が土台にあり、応用に資するものはそれらの援用なしには構築が難しい。理論自身の発展も促す、未来の基盤づくりとしての数学の重要性が改めて強調された。

［Key point］

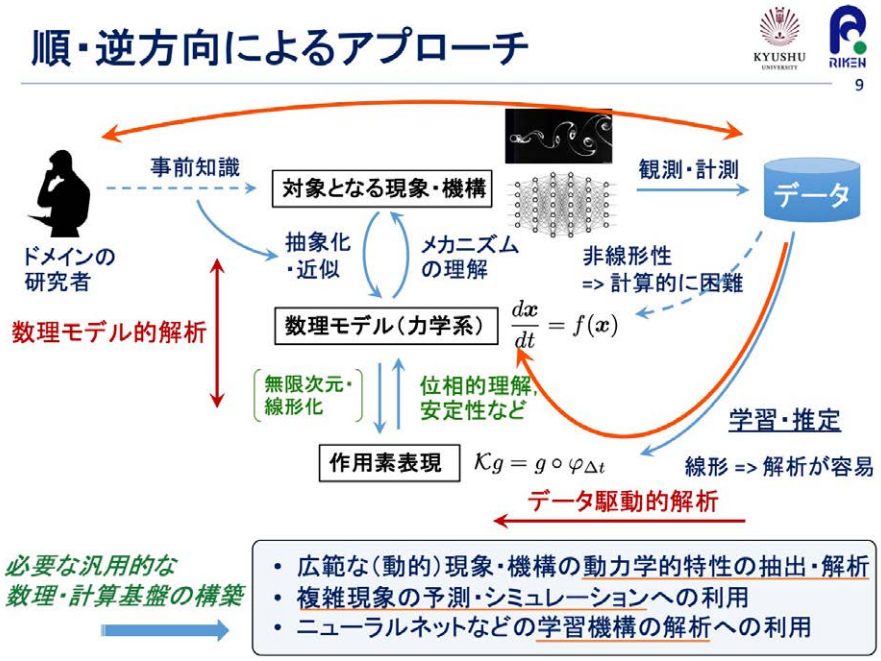
・データ取得のインフラ整備が確立され、AIの基盤としての機械学習も大きく発展している。現象を抽象化して数理モデルを導出、解析して現象を予測する順方向のアプローチに対して、計測データから現象やそれを記述する法則を見出す逆方向のアプローチが様々な応用とともに発展しつつある。（河原）
・数理モデルの妥当性を調べる方法が（古典的な）代数幾何学に基づく議論によって確立され、現実世界の問題解決に広く応用されているように、数学の基盤を作って発展させることは未来の問題解決や技術発展には不可欠である。（渡辺）

［河原氏講演：詳細］

機械学習はAIの基盤となる概念、技術である。近年はデータ取得のためのインフラ整備により、AIを用いた技術の発展が目覚ましい。河原氏の講演はAI技術を支える機械学習における研究課題、特に氏のグループが精力的に行なっている「複雑ダイナミクス」への機械学習の適用に関する話題が主となる。

系の「時間変化」が伴う力学系の理解は、従来は与えられた現象に対して（微分方程式や写像の反復で記述される）モデルの構成という抽象化がなされ、それを解析する事で現象を予測するというアプローチがとられていた。対してデータ駆動型のアプローチでは、現象から実際に起こった事象を表す具体的な計測データが先にあり、データから現象を記述する法則をモデルとして見出すというアプローチが採られる。前者を順方向とした時、後者は逆方向のアプローチと位置付けられる。現象にひそむ力学系の学習による発見は、カオス的振る舞いに隠されたローレンツ方程式などの決定論的法則の推定、ニューラルネットワークを通した系の近似的記述、特にエネルギーなどの保存量を（有する場合は）保つような推定などがあり、機械学習の国際的トップカンファレンスでも一大トピックとして展開される話題である。力学系の推定を試みるだけでなく、機械学習による予測に力学系の理論を援用し、得られる結果や議論を力学系理論を通して理解及び展開していくことは、方法論としての機械学習の発展にも資するものであると河原氏は強調する。

氏のアプローチの軸となるのは「Koopman作用素」を軸とした作用素論的データ解析であり、2021年1月現在、CREST領域における課題としてもこの研究が展開されている。Koopman作用素とは与えられた（非線形）力学系とそれを観測するために用いる汎関数の間に、ある種の可換関係を満たす「線形」作用素である。特に、解析が難しい非線形力学系の振る舞いに対して、線形力学系を通した解析と理解を実現させる。その特性から議論の土壌が有限次元から無限次元に拡張されてしまうが、線形力学系の議論に帰着させることの



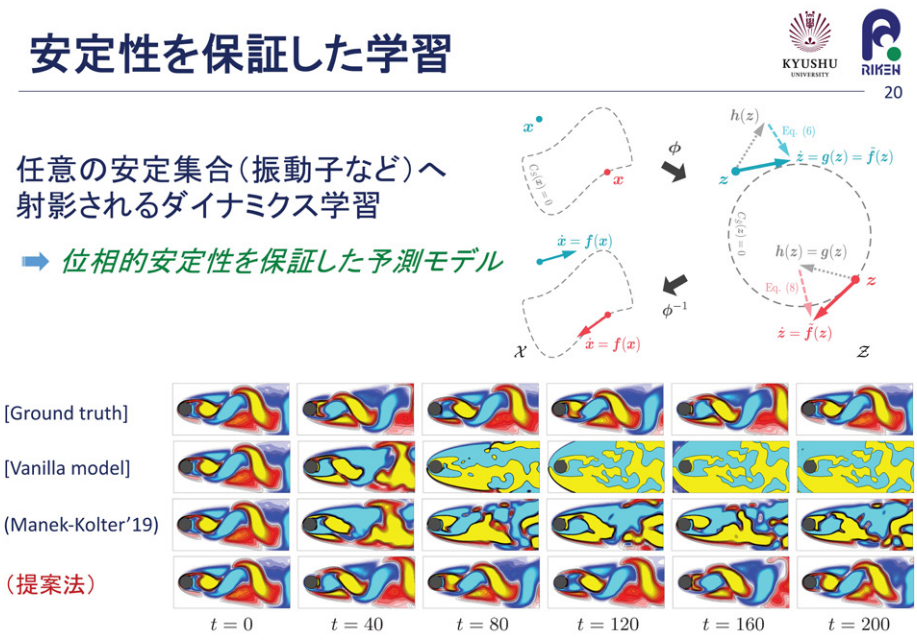
メリットの方が格段に大きい[1]。Koopman作用素を用いた力学系の推定は、「動的モード分解（DMD）」と呼ばれるKoopman作用素のスペクトル分解、すなわち固有モード[2]で展開し、作用素をその和として記述する方法が基礎となる。計算機で無限次元の情報を過不足なく抽出するのは典型的には不可能なので、有限のデータから有限個の（Koopman作用素の）固有モードを抽出して、力学系を推定するのが実際の応用で用いられるアプローチである。Koopman作用素の動的モード分解そのものにまつわる解析と合わせて、シリンダー周りの流体ダイナミクスの長時間推定など応用の有用性も紹介された。また発電システム、非線形系の制御、脳計測データにおけるコヒーレンスの解析、画像処理、感染症伝搬の解析など数多くの関連話題も紹介され、その汎用性の高さを伺わせる。

続いて、時系列データから系の特徴を抽出、分類するために、いかにして与えられたデータの「類似度」を定めるかという話題が展開された。これらはデータの違いを適切な「距離」で測ることに対応し、そのためにはデータが定義される空間の距離、特に空間構造（内積）を定めるカーネルをうまく定める事が要求される。例えば2つのデータの空間的な同調性を比較するカーネルの1つは、DMDが定める固有空間に対する部分空間の間の角度を使って定義される。これをGrassmannカーネルと呼び、ECoG信号によるブレイン・コンピューター・インターフェイスにより、（三角関数を空間基底とした）高速フーリエ変換では取り出せなかったシグナルを正確に判別することに成功している。一方、非線形力学系の予測においては、解の振る舞いだけでなく「解の族」の位相的性質も反映させる事が、長期予測では重要になっている[3]。そこで、系の位相的構

1　汎用性の高い解析方法が多数確立されているという意味で

2　固有値と付随する固有関数の組

3　保存量を持つ系の時間変化において、保存量を保つような推定をしなければ、系が本来有する振る舞いがうまく反映されないことに似ている



N. Takeishi, and Y. Kawahara, "Learning Dynamics Models with Stable Invariant Sets," in *Proc. of the 35th AAAI Conf. on Artificial Intelligence (**AAAI-21**)* (accepted).

造を記述するリャプノフ関数[4]を局所的に構成し、振動子などの不変集合の位相的安定性を保証した予測モデルを提唱している。これにより、渦度[5]を伴う流体ダイナミクスなどの長時間予測が可能となっている。

一連の研究は理論の発展とともに大きな応用の広がりを見せており、医療、航空機、プラズマ流、スポーツなど多種多様な分野への適用が試みられている。当該分野の大きな発展の可能性を強調されて講演は終了した。

［渡辺氏講演：詳細］

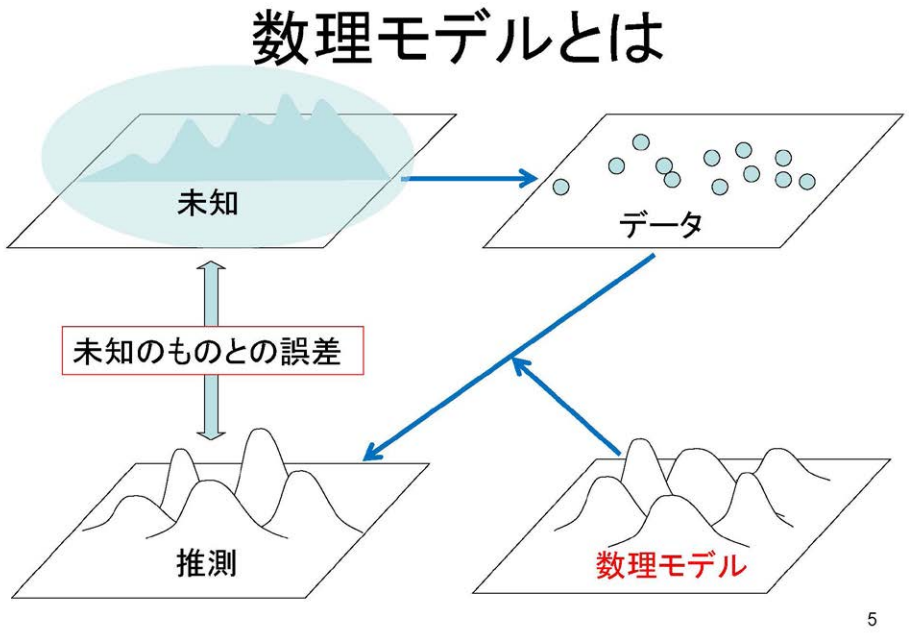
未知の事象を観測して得られるデータからその構造を抽出するための操作をモデリング、特に数学的概念に基づく場合には数理モデリングと呼ぶ。数理モデリングにおいて未知の現象を表現するための候補となるものを数理モデルと呼ぶ。数理モデルには確率モデル、統計モデル、機械学習によるモデルなど様々あるが[6]、本講演ではそれらの確率分布によって表現されるものに焦点を置く。

数理モデリングの適切さは、モデルに基づいて推定された結果と未知の現象とを比較し、その誤差が小さいほど良いと評価されるべきであるが、未知のものは不明であって、不明なものとの誤差を計算することはできない。このため、統計学や機械学習では、数理モデルをどのくらい信頼して良いのか、適切な推定が行われたと考えてよいのかを知ることができなかった。候補となるモデルは人間により作られたものであり、作ら

4　軌道に沿ってその値を単調減少させる汎関数。これを有する力学系は「勾配的」であるという。リャプノフ関数自身は相空間全体で考えられる事が多く、その特性より勾配的な系は「時間周期的、あるいは回帰的な軌道を持ち得ない」。各周期的な軌道で一定値を持つ弱い意味でのリャプノフ関数がここでは該当すると考えられる。あるいは局所的にのみ構成するという変位版の研究もある

5　流れの回転を表すベクトル。我々が渦として捉えている物は、流体力学において渦度という量を通して測られる。渦度ベクトルをつなげて作る「渦線」、流れ場に与えられた閉曲線を通る渦線を集めてできる「渦管」、その断面積を無限小とした「渦糸」など、いろんなバリエーションがある

6　講演中言及されていなかったが、微分方程式や写像の反復により表される力学系も１つの数理モデルであるが、ここでは確率分布で表されるモデルについて考えている



れたモデルによって推定された結果も異なるからである。すなわち1970年ころまでは、候補として作られたモデルの適切さを客観的に評価することは不可能であると考えられていた。この原理的な困難を数学的な法則を用いて解決するという考え方を世界で初めて提案し、実現したものが「赤池情報量規準」、通称AICである[7]。確率モデルがパラメーターにより表されているとき、未知の現象と推定された結果の誤差として定義される「汎化誤差」の平均値が「学習誤差」と「パラメーターの次元」の和との平均値と等しいという数学的な定理に基づいてモデルを評価する方法がAICであり、広範な数理モデリングに対して、その適切さを測る指標として世界的に広く活用されている。 AICは統計学の考えかたを根底から変革し数学の定理が統計学において必要であることを明らかにしたのである。

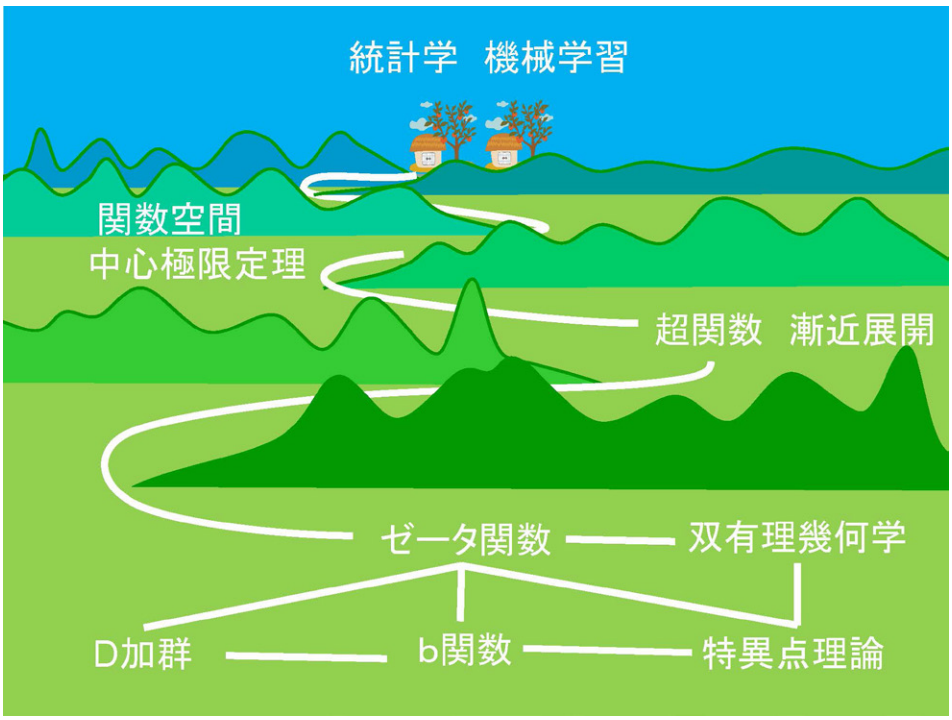
さてAICはモデルの「最適なパラメーターが1つ」であるという前提のもとで考えられているが、現代の数理モデルでは最適なものが唯一つではない場合も多い。階層構造や隠れた変数を持つモデルなども含め、現代の数理モデルは複雑かつ多様になっており、最適なパラメーターがひとつでない状況においても、「汎化誤差」を求めるための数学的法則を見出す事が、渡辺氏の研究の1つの動機となる。

氏が着目したのは、現代の数理モデルの最適パラメーターがなす集合が代数的集合[8]を形成している事である。統計学や機械学習で現れる代数的集合は極めて複雑であり、根性と人間力だけでは多様化・複雑化する対象に到底対処しきれず、数学によって作られていた基盤がなければ何もできなかったと、氏は述懐している。さて、代数的集合の構造を調べる数学の一分野に「代数幾何学」があるが、同分野における最も重要な結果として「特異点解消定理」がある。これは広中平祐氏により証明されたもので、任意の代数的集合[9]に対して、自己交差やカスプなどの「特異点」などがあっても、それらは有限回のブローアップと呼ばれる写像の合成で必ず特異性のない代数的集合の像となっている事を保証するものである。この特異点解消定理を基礎として、代数解析学、ゼータ関数、超関数論、中心極限定理など数学の中でも多岐にわたる分野の概念をつなぐこと

7　第10回、二宮氏の講演にて紹介されている

8　多項式の零点の集まり

9　例えば、標数0の体上の代数多様体が該当する



で「数理モデルのゼータ関数」が問題解決の中心にあることが解明された。これは上記の数学理論を基盤として、その最大極が数理モデルの推測精度を定めるような汎関数となっている。これを用いて、（最適パラメーターが代数的集合をなすような）現代の数理モデリングに対しても、「汎化誤差」が「学習誤差」と「汎関数揺らぎ」の和に等しいという定理が得られ、この定理に基づいて新しい情報量規準、通称WAICを導出することができた。

代数幾何学は階層構造や潜在変数を持つモデルの評価に対して重要な貢献をなしてきた。 WAICを中心とする氏の一連の研究成果は、医学、政治学、環境学、心理学、情報学など[10]、数理モデリングを必要とする分野で多数応用されている。こうした研究は、統計学の対象に潜む代数的構造に着目する数学の新しい研究分野「代数統計学」を形成する要素となっている。

数学研究そのものはもちろん何かに役立てるためのものではなく、暗い闇に隠されていた数学的自然に光をあててその美しい姿を描き出すものであるが、長い年月を越えて上記のように不可能に思われていた統計学の問題解決をもたらしたのであり、現代の科学や工学の発展の基盤となっている。「応用ができたから数学そのものの研究はもう不要」という考え方が一部で存在するようだが、未来における未知の問題の解決には、その基盤である数学はむしろよりその重要度を増している。数学の大切さが広く理解されることが必要であるという氏のメッセージとともに講演は終了した。

【質疑応答】

Q：「Grassmann距離」のGrassmannは、「あの」Grassmann？[11]

A：その通り。 Koopman作用素を考える線形（ヒルベルト）空間を定める「カーネル」により、Grassmann多様体上の距離を表現する。

10　最近はCOVID-19関連の研究も含まれる

11　幾何学で有名なGrassmann多様体の概念を提唱した人物。 Grassmann自身はGrassmann代数など、線形代数学の体系の確立に大きく貢献した人物としても知られている

Q：「ダイナミクスの表現」についてもう少し詳しく。

A：力学系を記述する支配方程式そのものを推定するのではなく、それに（表す現象が）近いと思われる「ニューラルネットワーク」でダイナミクスを表現する。実際の方程式（流体現象の場合はNavier-Stokes方程式）そのものの情報をどこまで拾い切れているかは不明。

Q：機械学習でも（振り子の）自励振動とか、時間的予測は無理だと思っていた。今回の話はKoopman作用素により、時間的変化も追え、未来の動きの予測もできるものであると理解している。この理解は正しい？

A：（ニューラルネットの）点で記述できるものでダイナミクスを近似しているので、その近似が精度良くできている限りはその通り。その記述には、Koopman作用素のスペクトルを用いている。カオスなどの時変的なもの[12]の長期予測は依然として厳しい。

Q：Koopman作用素自体は一意に決まるのか？実際に決める（＝推定する）時は基底の取り方や距離の取り方を工夫したりするのか？

A：Koopman作用素にまつわる数学理論により、もし存在すればそれは一意である事は証明される。実際に使う場合はKoopman作用素を「推定」する、すなわち良い近似として求める必要がある。この決め方は、Koopman作用素が定義される無限次元関数空間をうまく推定する必要があり[13]、Koopman作用素の良い推定はこの関数空間の推定に依存する。関数空間は、カーネル（＝積分核）やニューラルネットで表現する。もともと考えている系の知識も考慮して導入すると、より良い推定が可能となる。

Q：AICとWAICの違いは、未知のモデルとの誤差に「パラメーター次元」が効くか、「汎関数ゆらぎ」が効くかの違いと理解している。この2つの因子の関係は？

A：古典的な問題（最適解がただ1つに定まるもの）においては、両者、すなわちAICとWAICの概念は一致する。この意味で、WAICはAICの一般化と言える。

Q：この話題や必要な代数幾何学を理解するには、氏の本を勉強すればOK？

A：基礎となる代数幾何学の議論は、70年代までに盛んに研究された理論を使っているので、ここまでの知識を困難なくフォローできるならば、話題全体の理解も難しくないはず。

Q：特異点解消について。

A：代数的集合をより高次元のアファイン / 射影空間に埋め込み、特異点をほどく「ブローアップ」と言う操作を施す事が本質。特異点解消定理を提唱した広中平祐氏は、「ジェットコースターのレールがもともとあって（代数的集合のブローアップ）、我々が見ているのはその影（元の代数的集合）」という表現をされている。

Q：「現代」のモデルにWAICを適用できるようにするには、代数的集合やその特異点、ブローアップをどのように定めるかを上手に推定する必要があるのでは？

A：応用で実際に使うときは、代数幾何学の知識は必要にならない。あくまで定理の導出と妥当性の根拠に代数幾何学があり、ユーザーはそこに気を配る必要がない。ユーザーが代数幾何を習得している必要がないと言う意味でユーザーフレンドリーだか、その分、数学の大切さが実感されないことが残念なことでもある。実務でWAICが計算されるたびに、数学の輝きが誰からも知られることなく光っているのだと思いたい。

Q：古典的モデルから現代のモデルに合わせてWAICが提唱されたと言う流れがあるが、「より広いモデル」、特にこの議論でも扱えないもの（モデル）についてWAICの拡張など、何か見解はお持ちか？

12　定常状態や周期的振動など、系の局所的な構造だけでは記述し切れないものか？

13　つまり、線形空間の基底とか、内積と付随する位相などをうまく定める

A：WAICにまつわる議論には、いまのところ「データの独立性」が担保される必要がある。この仮定を緩める議論もある。数学的には、ゼータ関数の定義に用いたパラメーター依存する確率が、パラメーターに対する解析関数[14]である必要がある。この仮定を緩めるとどうなるかは、様々な研究者が多方面で研究している。

Q：河原氏の機械学習モデルにWAICの考え方は当てはめられるか？

A：WAICの考え方はモデリングに「確率モデル」（や統計モデル）を用いた場合に適用可能。ニューラルネットも確率モデルとして定義されている場合には適用できる。

Q：確率モデルじゃないものにWAICのような考え方は数学的にアプローチできるか？

A：ニューラルネットも、出力に雑音が加わるなどのモデルでは確率モデルにすることもできる。しかし、一般にパラメーターの次元が超高次元となる場合には、WAICを算出するためには大きな演算量が必要になる。

Q：量子アニーリングなどを用いて得られた「解」の検証にWAICなどの考え方は使えるか？

A：未知の現象を確率モデルを用いて推測する場合、対数損失関数を最小とするパラメーターを用いるとかえって推測精度は悪化する。未知の現象の推測には、対数損失関数を最小にするパラメーターひとつを用いる（絶対零度）のではなく、対数損失関数をエネルギー関数とするときの有限温度のボルツマン分布を用いたほうが推定精度がよく、WAICはその場合に適用できる。絶対零度と有限温度の違いなどはさらに研究を行う価値がある[15]。

Q：ダイナミクスの間の関係や比較は可能か？

A：ダイナミクスの間の距離を測ることで表現している。詳しくは2つのダイナミクスが表現する軌道の間の距離を見る[16]。

14 正の収束半径を持つベキ級数表示を持つ関数

15 統計学においては、絶対零度＝最尤推定法、有限温度＝ベイズ推定法、にそれぞれ対応する

16 力学系の間の軌道同値性や位相同値性を見ていると考えられる

3.12 数学と可視化技術

［概要］

CRDSの連続セミナーシリーズ、第12回は「数学と可視化技術」をテーマとして、高橋氏より「可視化におけるトポロジー」、安生氏より「映像メディアを支える数学」というタイトルでご講演いただいた。

高橋氏の講演では三角形のメッシュで表現された地形の形状から地理的な特徴を取り出したいという、可視化などでよく行われる問題から始まり、それに対して微分トポロジーにおけるMorse理論から、Morse-Smale complexやContour treeといった手法を用いて様々な可視化へ取り組まれているということが展開された。

安生氏の講演では、CGがこれからの映像表現は専門家やエンターテイメントの一部としてだけでなく、誰にでも使える基本メディアとして使われていくことを見据えて、現状を話された。 CGとは何かに始まり、制作プロセス、技術課題と展開され、単なるシミュレーションではなく、数理モデルに演出を加味するDirectabilityについて強調された。

［Key point］

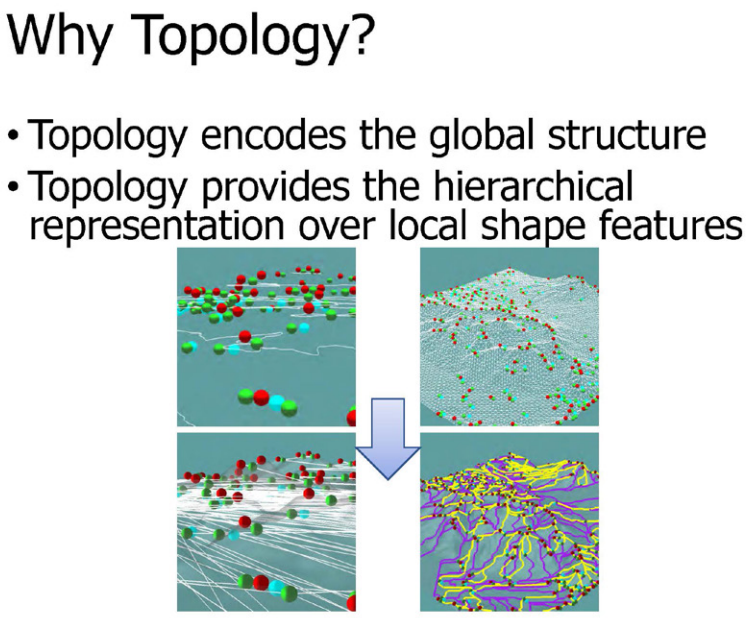
・トポロジーは、かたちの記述において局所的な特徴の上位階層の役割を果たし、結果として階層表現を提供してくれる。（高橋）
・可視化にはMorse理論などを含む微分トポロジーの考え方がよく用いられている。（高橋）
・CGはanalysisを基にしたsynthesisである。（安生）
・Directability「単なるシミュレーションではなく、演出の結果としての映像を生成する」ことが重要である。（安生）

［高橋氏講演：詳細］

可視化とかコンピューターグラフィックスの世界でよくやる、地形の形状の表現で、三角形のメッシュで表現されたものから何かしら地理的な特徴を取り出したい、といった問題から講演は始まった。画像の特徴としても使われる、頂上、峠、谷底といった特徴点がある。これらは地形の起伏を表現するにはわかりやすいが、一つ一つ独立であり、ローカルな特徴である。可視化の枠組では相手がヒトなので、データの構造をみせたいという要望があり、尾根線や谷線を加えて特徴点のつながりを全体の構造として取り出すということをしたい。これはMorse-Smale complexと呼ばれるもので、地形の形状の起伏をネットワーク上で表現する一つのよく使われている手法である。一方、これらの特徴点は等高線が高さに関して変化をおこしている点であると理解でき、このような等高線の分岐併合を特徴としてとらえてやると、特徴点を頂点にもつグラフによる表現が可能となる。これをContour treeという。

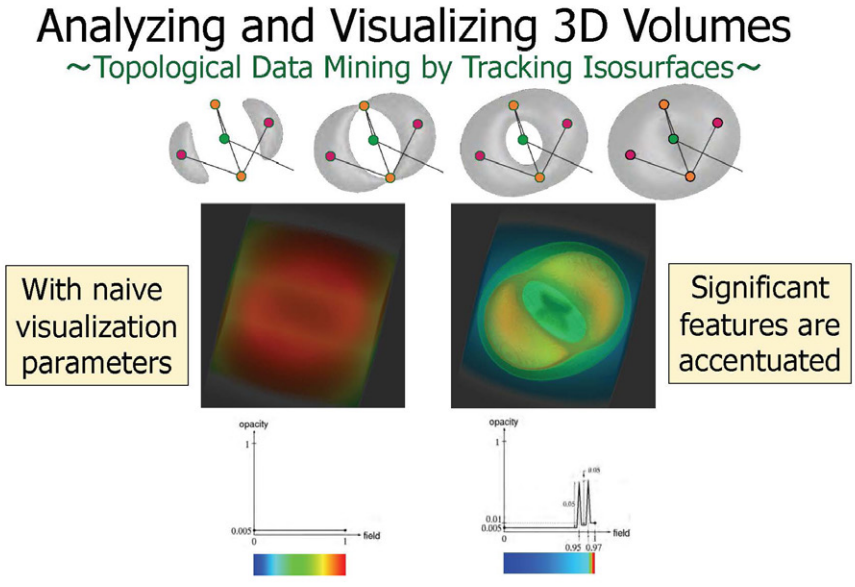
では、なぜトポロジーか？といえば、局所的な特徴だけでは全体の構造がつかめないからである。トポロジーはいろいろな道具を提供し、さらに局所的な特徴の上位階層として自然な表現を提供してくれる。可視化の分野では曲面のトーラスの穴を数えるトポロジーではなく、もうちょっと分解能の高い微分トポロジーを使う。 Morse理論により、さきほどの頂上、峠、谷底といった特徴点（特異点）に対して胞体が1対1に対応できる。さらに特異点のつながりも情報として加えるといったことが、可視化の世界ではやられている。

Morse理論を用いるためには離散サンプルという仮定の下で、Morse関数を定義する必要がある。地形の形状の解析では2次元から1次元空間への関数である。空間の点に温度（or 圧力 or 密度）を対応させる場合は3次元から1次元空間への関数を考えている。これはコンピューターシミュレーションにより空間分布のデータや医療3次元画像が対応するが、可視化の世界では重要な解析対象である。実際、何も考えないで色塗りするとぼやけた画像になるのが一般的である。一方でContour treeを使うと中の構造を浮き立たせて可



視化画像を作ることができる。例えば、水素原子に陽子がぶつかるシミュレーションデータにおいて、衝突直後のエネルギー分布の可視化の問題に Contour tree の考え方を適用すると、エネルギーがオーバーハングの形で戻ってきているということを見ることができる。これは可視化が新しい知見を提供した事例となっている。

Contour treeの作り方は英国Leeds大学のDr. Carrのアルゴリズムが基本となっている。例えば3次元定義域内にスカラー値をもつサンプルを考えて、四面体分割を施してスカラー値の線形補間を施すとする。このとき、スカラー値の高いサンプルから四面体分割の接続性をみてつなげていくとY字の分岐をもつグラフができ、また低い値からつむいでいくと逆Y字のグラフができる。これらを融合することでContour treeを作ることができる。一価関数で表現できないもの、例えばトーラスはグラフで表現すると、サイクルが含まれるよう



S. Takahashi, Y. Takeshima, I. Fujishiro, Topological Volume Skeletonization and Its Application to Transfer Function Design, *Graphical Models*, 66(1), 24-49, 2004. doi: 10.1016/j.gmod.2003.08.002

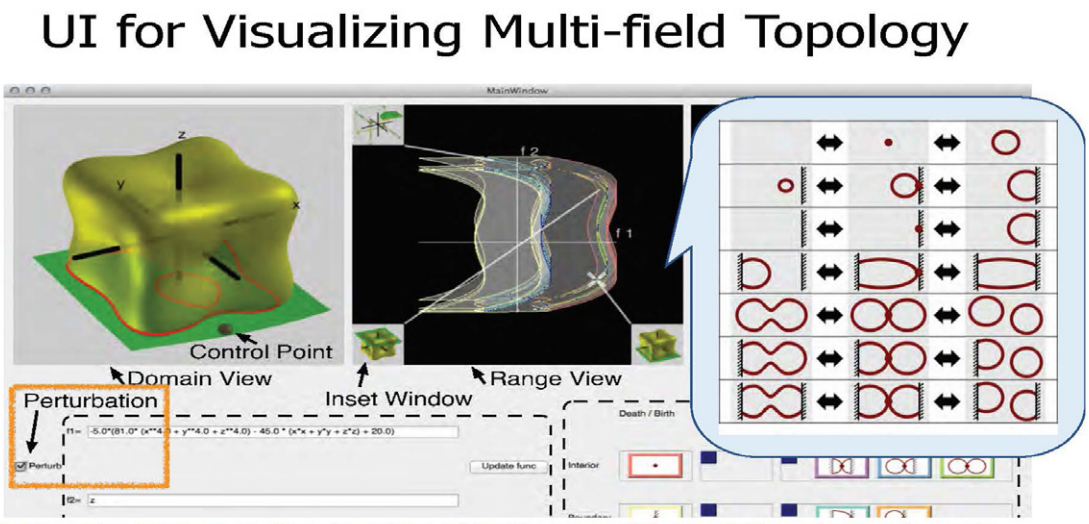
になる。こうしたものを総称してReeb graphといい、したがってContour treeはReeb graphの一部といえる。

Morse関数の例はたくさんあり、いろいろ手法が開発されている。例えばMorse-Smale complexは地形の形状や三次元の画像の構造を取り出す。特異点はノイズに弱いため、一般的には細かい構造を取り出した後に、特異点間のグラフ構造に対して簡単化をほどこすことで、より粗い構造を抽出している。この簡単化処理では、パーシステンスという概念を使う（質疑応答参照）。

他の手法として、曲面上にMorse関数をデザインしようという便利な考え方やContour tree系の手法で、形状の検索に骨格の情報を使うというものがある。また、燃焼効率をトポロジーの変化で評価する研究がある。それは特異点の数やタイプで燃焼効率をインタラクティブに評価していくものである。３次元空間のスカラー値の分布から抽出されるトポロジー情報に基づいて地形形状をつくり、これをメタファーだとおもってデータの構造を理解する手法なども提案されている。

需要として、Morse関数の定義域の次元を上げるということがある。高橋氏らの研究では、機械学習でよく知られている多様体学習の考え方を使って高次元空間に定義されるスカラー値のサンプルデータから、次元圧縮の結果としてContour treeやReeb graphなどを近似表現として取り出すということを行った。先ほどの水素原子に陽子がぶつかるシミュレーションデータの時間変化全体を、４次元の定義域のデータと考えてこの手法を適用すると、衝突の部分が分岐として取り出すことができる。

最近はMorse関数の値域の次元も上げる研究が進められている。空間の温度分布を考えて、等温面と等圧面をたどっていくとそれぞれ異なるContour treeができる。しかし、このままでは温度と圧力の関係は独立していて直接たどれないので、この二つの関係もわかるような可視化手法が欲しい。そこで、二つの関数値の等値面の交わりの部分（ファイバー）について着目し、その関数値に関する変化を追跡することで、特異点が存在するファイバー（特異ファイバー）のトポロジーの変化を解析する。具体的には、トポロジーの変化を値域で記述してみる。それが特異ファイバーの理論研究に直結することから、特異ファイバーの第一人者である九州大学佐伯所長と共同研究することになった。一例として、二つの関数のファイバーを追跡するインターフェイスを作る研究を行った。例えば、温度も圧力も変化するとContour treeをスイープしたようなシート状のReeb spaceを用いる必要がある。これは連続で計算するのは難しいので、離散の近似表現であるJoint Contour Net（JCN）を作る。これもDr. Carrによる比較的最近の提案である。この提案により佐伯



O. Saeki and T. Yamamoto, Singular Fibers of Stable Maps of 3–Manifolds with Boundary into Surfaces and Their Applications, *Algebraic & Geometric Topology*,16, 1379–1402, 2016. doi: 10.2140/agt.2016.16.1379

所長とDr. Carrとで共同研究することとなった。出来上がったインターフェイスはファイバーがどういう形になるか追跡できる。ファイバーの型の分類は佐伯所長が行った。コンピューターの扱いでは、データの定義域が有限のためファイバーに境界が生じるので、これが新たな数理的問題となり、それは佐伯所長側で論文化された。これはコンピューターサイエンスと数理の共同研究でお互いがうまみを得た事例となっている。

値域の部分のファイバーの領域を指定して対応する形状をみて解析をする方向についてはDr. Carrの提案がある。高橋氏も同じ方向で、福島原発周辺の線量の分布分析（JAEA眞田氏との共同研究）を行っている。

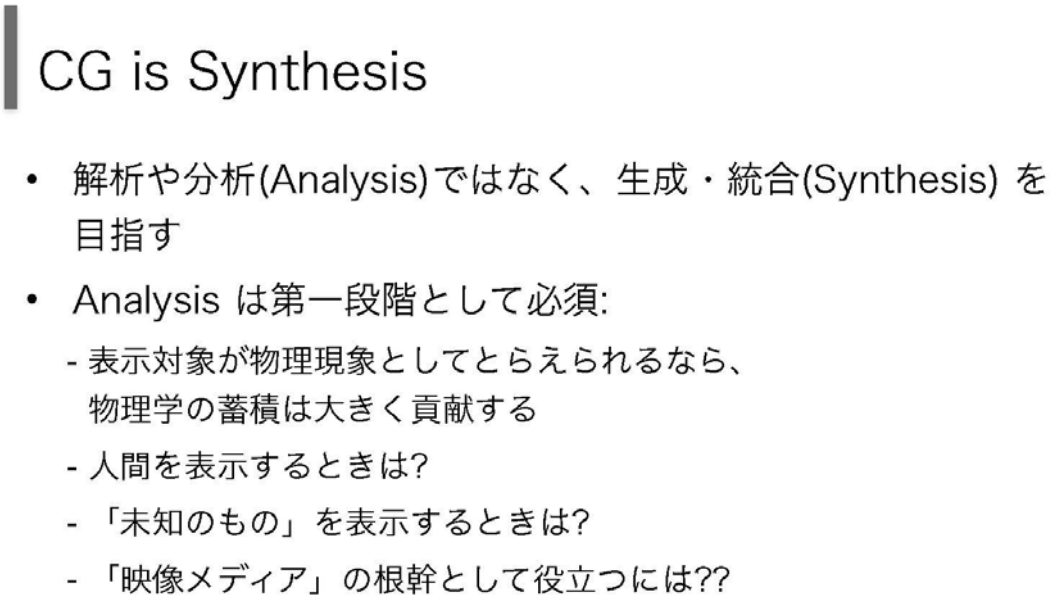
最後に、サーベイやトポロジー解析のできるツールの紹介があった。

［安生氏講演：詳細］

これから使われるCG映像表現は専門家やエンターテイメントの一部だけではなく、万人のための基本メディアとして使われていくことを見据えて、現状を話された。

技術的な意味ではコンピュータービジョンとコンピューターグラフィックスは対比されるが、現実を解釈するのがビジョン。 CGはどちらかといえば頭の中のイメージを可視化するもの。とはいえ相当リアリティの高いものも現在は作られるようになった。頭の中の映像化をするというsynthesisを行っているのがCGであるが、analysisは第一段階として当然必要である。対象を理解する部分で数学や物理を使う。それだけではなく今度は表示する対象を広げる、高橋氏のようにvisualizationということでデータを可視化するということも含めてのメディアである。いわゆる物理だけ、数学だけでは解決できない問題がいろいろある。「映像メディア」の根幹として役立つにはどういうことをしなければならないかを考えている。もちろん映像プロダクションの立場だが、半分ニュージーランドの大学で先生もされているということでこういう観点ももって考えているところであるとのことであった。

CGのプロセスで、モデリング（シーンを作る）とはここではナイーブな意味で捉えて、形を作るとか表面の材質をモデル化することも含む。レンダリングは、3次元シーンデータから、画像として表示するまでの処理全般を言う。カメラや本物または仮想のCG人間が動くことも想定して考えている（アニメーション）。講演で上映されたCG映像は、雲は一部実写からの画像だが、建物や樹木は大量のポリゴンで表され、木の葉っぱのテクスチャも画像データ化されている。大気の光と陰のシミュレーションやカメラワークも考慮され、1ヶ月程度を要したとのこと。すなわち、膨大な時間と手間が大きなボトルネック。海外でも同じだが、市場の規模などの要素があって、海外のほうが技術が進んでいることが多い。また表示対象が国や目的ごとに違うため、

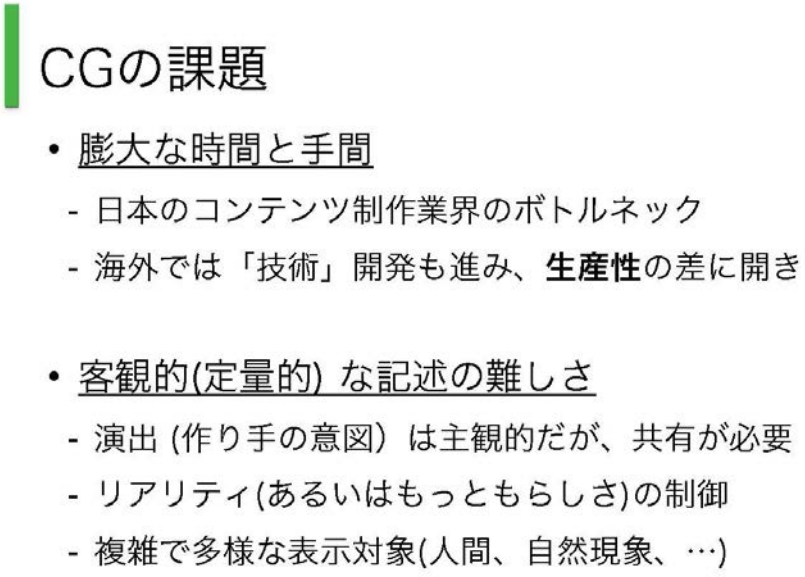


生産性の差に開きがある。もう一つの課題はリアルかどうか、演出（作り的な意図）がどう伝わるか。記述は人間がやるが、コンピューターが理解して映像化するステップは技術的に解決しなければならない。自然科学のアプローチだけではなくユーザーインターフェイスを考慮した数理モデルがCG分野の数学にたよっている大きな側面である。

モデリングとは何をやるのか。単純にメッシュを編集する。他にも例として、異なる対象物をつなぎ合わせてなめらかに表示するときに、つなぎ目のPossion-based gradient方程式といった 境界値問題を解く。今世紀入ってから、どんどん作られている。逆にそこまで長い歴史があるわけではない。

レンダリングは例えばレンダリング方程式という積分方程式を解いて計算する。現在もさまざまな状況下での高速化、ゲームなどのリアルタイム応用に技術開発が続いている。自然現象を表すなら、例えば流体ならナビエストークス方程式、弾性変形シミュレーションならそれを記述する微分方程式を考慮する。物理と数学を合わせたものを考える。レンダリング方程式でも映像にして最終的にはレンダリングプロセスに載る、ということまでやる。その視点からみて何が見える見えないというVisibilityの判断など、いろんな処理が入るので単純に方程式を解くより手間がかかる。そこにも数学的工夫が必要となる余地がある。そういうことを皆、産学一体となって研究している。

人間の表示、動きと流体は数年前にCREST（研究総括：西浦教授）にて行っていた。　人間と流体の表現は、一番難しく応用が多い。映像メディアでいろんな人が使えるようにしたいので、どういう工夫が必要かということを考えるCRESTをやっていた。例えばキングコングやアバターの表情は人間の表情データを加工した。これに関連する技術ではリターゲットというものがある。例えば、ある（スライドではピンクの顔の）キャラをインプットとして、他の異なる3つのキャラに表情のエッセンスを移植するということである。一つの理由は効率化。さらに、おおまかには類似な表情だが、細かいところは個性を出すためエディットという人間が介入するという作業を受け入れる技術を作る必要がある。このときにおこなったのは、target表情の時間変化はもとのsource顔のときとあまり変わらないというのを数学的にモデル化して、あるノルム空間の最適化問題を解く。解法はスタンダードだが、CGで作り上げる人間の特徴をどう表現するかという分野にも数理モデルが役立つというのが安生氏の言いたいこと。もう一つは流体。シミュレーションの流体の煙がだんだん曲がるが、曲げてみてもそれらしく見えるというところが人間が演出いれたときに対応できる技術である。粗いCGを高精細なCGにするのに2次元のデータベースを用い、また障害物との干渉も考慮する。これも効率性をあげるために数学的な考え方をいれている。また単にシミュレーションしているのではなく、作り手の演出を引き出すDirectabilityの実現を目指した。そのような演出表現が数学的なformulationによる数理モデルに組



み込めるかということがポイントである。実際に演出する作業自体は作っている人が行うのだが、それをなるべく直感的な操作で取り入れることができるようにモデル化している。単に物理現象を数理モデルとして解くというのとはもうひとつジャンプがある。

３Ｄモデルの顔にトゥーンシェーダー、アニメ的な陰影をつける話。フェイクな陰をつけたり、目の部分だけ明るかったり、ライトの変化が起こっても保てる。非物理的だが、アニメーションとしてはなめらかで自然な表現を作る。おおまかにはRBF補間を使う意味ではリターゲットと似ている。他分野で使われている技術がCG分野でも使えるというのが一つあるが、どちらかといえば発想として数学っぽいところが必要だという意味の重要性は感じるとのこと。技術的にどう解くかということに数学を入れるのはどの分野でもしているが、数学に落とし込めるところを考えるのが重要。物理的には正しくないシェーダーの例で、人の顔の陰影や鎧の肩の強調の映像があった。人の顔を下からフェイクな光を当てるという演出は実写映画でもよく行われているとのこと。

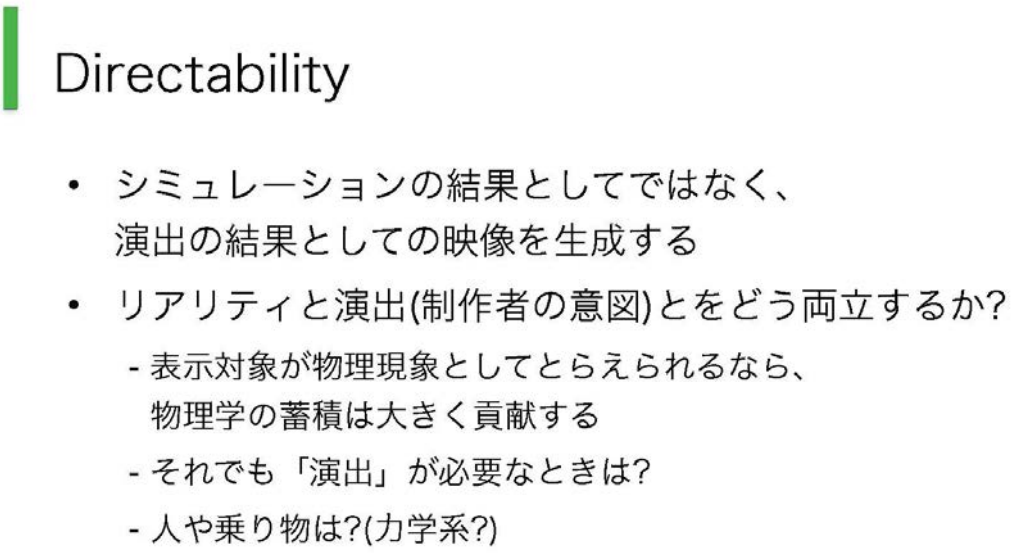
高橋氏の研究と近い話で、映像業界特有の表現だけではなく、何千万点のポイントクラウドのポリゴンで表現するということも研究されているとのことであった。もともとのデータは重いのでポリゴンリダクションをする。これは古典的な問題だが、特徴的な箇所は詳細度を保持したままポリゴンリダクションが必要となる。

ポリゴンリダクションに関連する４年間のLimperらのアイデアは各点での平均曲率のエントロピーをとる。しかもローカルなエリアを何種類か限定して、その中でのエントロピーを比較して、粗くするところ詳細に保つところを決める。大容量なポイントクラウドでは計算量が実用的でないので、その改良を考えているとのことであった。その工夫は法線情報をうまくつかって擬似的な曲率を計算するということで、高速計算やガウス曲率や平均曲率と同じような効果をこの目的（特徴的な領域を保持するポリゴンリダクション）のために有効に使える。それのエントロピーをとるということを考えている。冒頭と同様に、映像メディアとしてCGを使うということをエンタメ業界に限定せずにいろんなところで使えるようにしたいので、総合的に研究したいとおっしゃっていた。

まとめとして、可視化と人間がどう操作して結果を出すかを数学的アプローチでうまく表現するということが重要、物理的シミュレーションや最適化問題などの工学的アプローチだけでは不十分、人間の目から見た世界をCGで実現するには新しい数学の発展が必要、といったことが話された。

【質疑応答】

Q：今日の講演の研究にあたってデータ量はそれなりに必要だろうが、一つの例として病気（慢性病）と



か時系列データが取りにくいときに、今研究されていることが役立てることがあるのか。

A：時系列データは手に入るものと手に入らないものがある。最初はシミュレーションのデータからアクセスすることのほうが多い。計測されたデータはノイズがのりやすいのでなかなか難しいところがあるが、ただサンプリングの取り方は適応的に行うことで問題を解決する手法も開発されている。粗ければ粗いなりに密なら密なりに解析できるような下地がだんだんできあがってきているというのが私たちの理解である。

Q：特異点を見るということは、平常点をいくら見てもわからないことはよくわかるが、イベントの予測とかに使っている試みはあるか。

A：特異点の振る舞いの予測はデータの補間に対応する。一般的には高次の補間はしない。
いわゆるpiecewise-linearな、線形補間からスタートすることになっているので、データから得られる以上の情報の予測は難しいとされている。例えば、２つのスカラー値の空間分布があり、それぞれ特異点の分布が異なっているときに、その間で生じるべき必須のトポロジーの変化はすぐ計算ができ、それに基づいてスカラー値分布の補間をする枠組みは可視化でも形状処理でもやられている。具体的には、特異点を含む形状のモーフィングみたいな感じ。そのような形のデータの扱いはそれなりに研究されていて、さらにそのようなトポロジーの変化自体をデザインをするという枠組みもいくつか研究が提案されている。

Q：エネルギーランドスケープの分析は最適化みたいな話もかかわってくるので、アニーリングマシンではないが、イジングモデルに基づいた解析をするということは取り入れられたりするのか。

A：最適化の対象の関数の可視化という意味ではある。最適化の関数の特異値が最適値になったりすることが重要になるので、そのような解空間における最適化対称の関数の振る舞いを、可視化のツールを使って解析しに行こうとする枠組みは結構なされている。時間がなくてお話しなかったが、多目的最適化という枠組みがあり、そこではパレート解を計算する必要が生じる。ところが、パレート解は特異値の部分集合になるので、特異値の解析の枠組みでパレート解の解析をしようとチャレンジをしている方もいる。そのような枠組でとらえるのがトポロジー側からのアプローチになるのではないか。

Q：パーシステントホモロジーを用いているところがある、ということをちらっとおっしゃったが。

A：ホモロジーではないが、トポロジカルデータアナリシスのほうで使われている概念と結構似ている。たくさんある特異点から、ここで使われているのは　例えば２次元の場合だと頂上と峠（or谷底と峠）のペアで、頂上のスカラー値と峠のスカラー値の差が小さければパーシステンスが小さいという考え方をすることで、そのような特異点のペアを除去するというアルゴリズムが一般的に使われている。最近は高度化されているが、基本的な考え方は特異点のペアを刈りこむことでMorse-Smale Complexを簡単化していこうことである。１次元高いものでもだいたい似たようなアルゴリズムになっている。これはTDAなどで使われている考え方に方向性は似ている。
EdelsbrunnerがTDAの創始者とひとりとも考えられるが、彼は可視化のほうも結構研究していて、特に理論的な考え方は、彼が提案しているものが多くある。それをうまく計算機のアルゴリズムに落とし込むということを、コンピューターサイエンスの研究者が結構やってきた。そこの理論と実際の計算のギャップを埋めるのは結構大変で、そういうことを中心に研究が行われてきた経緯がある。この分野では彼の若いときの論文がたくさん出版されている。我々もよく参照して、これをアルゴリズム化するということをよく考えている。

Q：理論家の方々が可視化の分野に参入していないのか。

A：参入していると思うが、微分トポロジーの世界は関数が連続であることを仮定していることが多く、それをコンピューターの離散のアルゴリズムに落とすのには結構なギャップがある。ファイバーの話はさらに二回切るということになるので、なかなか離散のところに落としにくい。なので、研究が素直にうまくいっているというわけではない。結構皆いろいろ悩んで、これでいけるだろうという形でやっている。

最終的に不変量を保った形で情報が抽出できないとただのヒューリスティックになる。そうならないように連続の世界の性質を上手に離散の計算へもっていくというところがテクニックとして大変である。退化した臨界点の扱いもきちんとやっていくことは重要で、裏で結構苦労しているというのがある。

Q：Contour Treeについて、点群からのsimplexのようなもので面をはるとき、パーシステントホモロジーのように局所のノイズを平滑化する機構があるのか。

A：さきほどお話したことに関連するが、まずは与えられた離散サンプルからすごくノイズに起因する臨界点を含めて詳細なContour Treeをつくってしまうということを先にやる。その後、ここの臨界点が全体の構造として重要かどうか判断して、必要に応じて除去するのが我々の世界のアプローチの仕方である。仮にContour Treeを作らずに、特異点だけ相手にして取り除こうと思っても、臨界点それ自体は局所的な情報なので、それが全体としてどのくらい重要なのか判断できない。まずはContour Treeを作り、そこから得られる大局の構造を見ながら、個々の臨界点が重要かどうか判断して削除している。臨界点の削除の具体的な基準としてはパーシステンスみたいなものを使うが、Contour Treeの方からパーシステンスをみて取捨選択するというアプローチになっているところが、普通にローカルな情報だけをつかって臨界点を取捨選択しContour Treeをつくるのとはちょっと違う。

Q：数学の人たちと共同研究するときに、苦労した経験や印象は？

A：二つぐらい言うことができて、コンピューター科学者、数学者ともにopenmindedじゃないとなかなか難しい。我々の側からすると、応用に対して広い心がある数学者を見つけることのほうが大変であり、そこが最初のキーと思う。また、実際のところバックグラウンドがだいぶ違うので、コンピューターにおける離散的なデータの扱いは、やはり説明しないと数学者の方にはわかってもらえない。数学の側の考え方も、なかなか我々コンピューター科学の研究者も理解が及ばないところもあって、結局のところ何回か手を変え品を変えて出し方を工夫しながら情報を交換しつつ、お互い協力していくというスタンスをとっていた。その点ではやっぱり少し辛抱は必要だが、お互いの研究分野に対してある程度配慮や尊重があれば、それはそんなに大変ではないのかもしれないというのが、もう一つのことになる。
実はこういった研究をしている方は、さきほどのDr. Carrもそうだが、実は数学からコンピューター科学へ来ている人が結構多い。学部は数学で学位をとってコンピューター科学へ来たという人が可視化の研究をやっているケースも多いので、やはりそれなりの素地が必要と思う。可視化はもともと融合分野と考えられていて、いろいろなひとと研究してみると様々なバックグラウンドをもつ人がいるが、なかでも数学のバックグラウンドをもって活躍している場合が多い。

Q：数々強調されていたとおり、synthesisやDirectabilityが大事だとおっしゃっているが、物理ベースだけでは無理で、作り手の意図が反映されて、見る側（観客、受け手）にとってうんとうなずけるかが重要。そういうことを、数学を使いながら、古典物理でとらえられないような人間自身を研究しているのではないか。そこに数学が、というのは結局なにをやっていることになるのか。

A：だんだんそうなってくる。そんなにたくさん新しい数学が出てきたわけではないが、今日の話も、例えばRBF補間という技術であればいろんな分野ですでに使われているが、同じ概念がCGの分野ではこうやって使えるという経験がCG自身の技術発展、あるいは数学的な技術そのものの発展に寄与してきている。ポイントクラウドからうまく特徴的なところを残したポリゴンリダクションするというときの曲率の定義は今のところ調べた感じでは他ではやってない。そういう風に数学が出てくると面白い。もちろん役に立つという意味ではいろいろな既存の数学を使うのも面白いが、他の分野でも使われていない定式化になるものが出てくるのが楽しみで期待している。できるかどうかわからないが、その場合は数学者に助けてもらう。何をやっているかのスタイルは定式化した問題がいままでの数学に当てはまればよいが、そうではないものも色々と出て来ているので、そういうことを数学の人たちと一緒にやる、ということである。

Q：岩波新書などのレオナルドダビンチの手記をみると、数学について書いてあることの一つに「工学は数

学の楽園、果実が実る場所である」と書いてある。それ以外に自分のやっていることを理解するために数学的素養が必要とある。私たちはすぐに見る側と思っているが、作る側の操作性の問題がある。操作性自身は昔ならCGではなく漫画家が絵を作っていた。それに代わることを一部されている。そういう意味では人間がこれが良いということを操作性をよくして、かつデータの大きさも理にかなった範囲内におさめないといけないという最適化の問題が出てきますよね？

A：何でもできるというより、データを効率的にコンパクトにする。映像メディアとしての数学と考えることはリアルタイムということを暗に含む。結果は誰でも見ることができるが、一般のひとたちがメディアとして活用しようとするとリアルタイムは大前提。結果がでるまでに何日もまっていられないでしょう。データの効率的な圧縮はいつになっても、許容量は増えるだろうが、やはり重要な問題である。そこがコンピューターと人間からくる要請とどういうバランスをとれるのか、数学だけではないが、考えなければならない大きな問題である。

Q：映像の制作について、現場でアニメを作っておられる人たちの幾分の肩代わりをAIがするということもやっているのか？

A：それもやっている。今のところやっているのは、例えば色の塗分けの問題。アニメなら想像がつくと思うが、境界部分は黒い線でできているが、あれは結構途切れている。キャラクターが動いたときにどうやるか結構手間がかかるので、そこはAIを使って効率を上げるようとしている。なかなか実用まではいかないが、世界的にも学会で発表できるくらいの成果にはなっていると思います。

Q：リアリティーについて、最終的に人間がみて判断しているのか。

A：アニメの場合は監督の判断になる。演出に耐えうるものかどうかの判断は客観的ではないが、これでOKという話ができる。実際のシミュレーションした映像、実写映画の中の大洪水の映像があると思うが、音や迫力におされてすごいと思っているとは思うが、実際には客観的な指標は作りにくい。 CGの作り手側の視点からいうと、昔から言っているが、どれぐらいの人が使っているかが指標になります。リアリティーは１００人見たら１００人すごいということは実写の映像でもなかなかない。どれだけ信頼性をもった技術となっているかという評価だけはできるか、リアリティーがどれだけ高いかというところはなかなか難しい。作った場合はそっちの評価。その評価は結構時間かかる。今日の顔の話、人間の皮膚の変形の技術は２０００年ぐらいに発表されたが、２０年後アカデミー科学賞をとった。それぐらいの時間をかけると技術の信頼性があがる。真実を求めるなら、ある意味真実を語れればいいが、皆が使えるものになっているねという評価になるとそれなりに時間がかかる。そうなれば、その上につみあげていく新しい技術はすぐ現場に受け入れられていく。やがて普通の人が使えるようにしたいのが技術的な意味での目標である。

Q：例えばスターウォーズで死んだ俳優さんが出演されていてあれはすごいリアリティーだと思うが、皆さんが使う家のゲーム機で使ったりすることを目指すのか？

A：そうだと思うが、どこまでクオリティーが必要なのかはひとによって違うので、一般的にはわからない。

Q：最近だとディープフェイクのように、データから映像を作ることがある。結構よいものができているが、あのようなデータ駆動のものとモデルベースのものとを組み合わせるという可能性もあるのではないか？

A：いろんなところでそういった質問があった。冒頭でビジョンとグラフィックスを言ったが、ビジョン的な意味でのいろんな情報を、日本ではないが大手のプロダクションでカメラを数台用意して映像を作る。そのシーンにある３次元的なデータをいっぱいとって、でも映画自体はステレオグラフィックスで３Ｄにしたりもするが、基本的には２次元で、中にCGをいれたりいろんな加工をするときに実際は３次元的な情報をうまくいかして、２次元の映像やステレオスコピックを作るなりする。データドリブンでやるということは当然考えている。そうするとコンピュータービジョンの技術が、あるいはAIが山ほど必要になると思っている。

Q：さきほどポリゴンリダクションの話で、ある程度精密なポリゴンと粗いポリゴンはあったが、この二つの関係というか、定量的な評価して、この特徴はうまく保っているけど、見た目としてはだいたい同じように見えるという方針で考えているのか？

A：今日のスライドで紹介した部分は曲率のエントロピーを計算するという話なので、ある程度客観的である。ただそれでも局所エントロピー、ある範囲で限定した中で、どれぐらい顕著な曲率的な情報量をもつかという話なので、ではその領域をどう大きさを決めるのかは人間である。それとは別にここはとっておくという使い手側の基準による重要度があるので、両方考えることになる。今日のお話したのはどちらかといえば客観的に近い方。

Q：いまだされている論文では客観的な、ここの数値を保てるリダクションとここだけを置いておきたいということもできるようにしているのか？

A：現場的にはそうです。サイエンスっぽくないが。

Q：そういう（恣意的な）視点は新鮮。数学やっているにはどうしても客観的をもとめがち。恣意的に逆にしたいというのは面白い観点だが。

A：さきほどのトゥーンシェーダーの話は現実よりも大げさにする、スタイライズすることだ。 CG技術は、（演出意図を反映することや）恣意的な表現をも実現可能にするというユニークさがある。

3.13 シミュレーションとデータ科学

［概要］

CRDSの連続セミナーシリーズ、第13回は「シミュレーションとデータ科学」をテーマとして、吉田氏より「データ科学の視点からみた計算科学との価値共創の在り方」、三好氏より「ビッグデータ同化 ～ゲリラ豪雨予測から、予測科学へ～」というタイトルでご講演いただいた。

吉田氏の講演では現実世界における実験、計算機などにおける仮想空間における実験の良し悪しを踏まえた「理論・実験・データサイエンスの融合の在り方」の1つのアジェンダを、3節構成で具体例を交えて紹介された。実世界の諸問題（物質、生命、社会、情報など）との結びつきが非常に強い学際的分野であるデータ科学において、応用分野の見識もさることながら実問題の解決のアシストだけでは不充分であり、「データ科学自身がフロンティアを切り拓く」という気概を持った活動が求められる事を強調、それを実証する成果を多分に盛り込んだ内容となった。

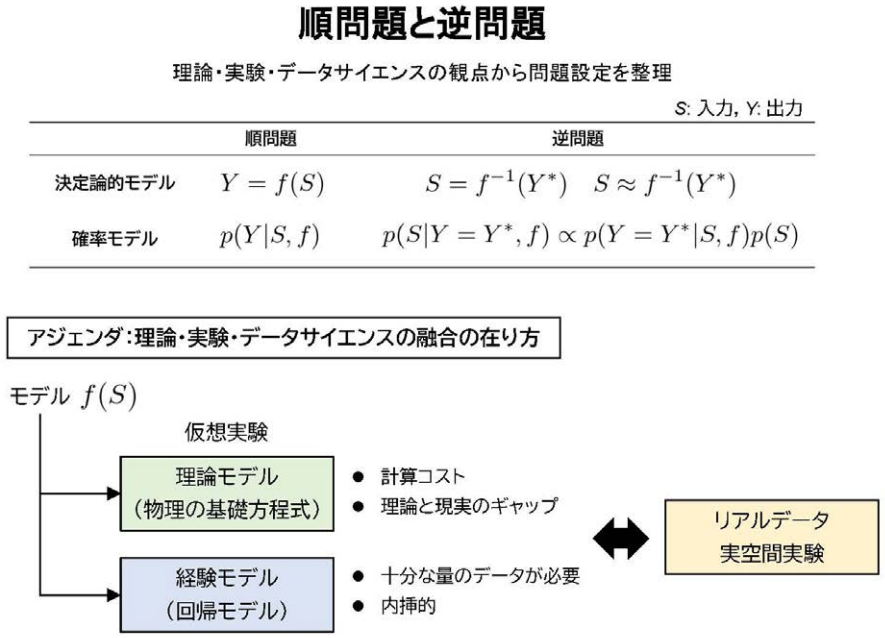
三好氏の講演では、比較の際の乖離が避けられない実観測データとシミュレーションによるデータを結びつけ、双方の情報を最大限に抽出し、精度の良い予測を実現する技術の1つである「データ同化」、特にビッグデータの利活用も考慮した「ビッグデータ同化」に関する話題を、氏の中心的研究課題である気象学をはじめ、多方面の分野への貢献も交えて紹介された。多くの実問題への応用が注目される一方、その議論の源流は数学・数理科学にあり、これらの理解も含めて応用・理論を広げていく事で、「経験」「理論」「計算」「データ」を中心とした科学の潮流がデータ同化を通して人間の予測能を向上させる「予測科学」という新たな潮流に遷移され得る事を強調された。

［Key point］

・データ科学では、「理論（数理）なき実践（実装）」「実践なき理論」のどちらが欠けても通用しない。（吉田）
・データ科学、数理科学の研究者が実問題の解決のアシストをするだけでは不充分で、データ科学研究者には「データ科学自身がフロンティアを切り拓く」という気概を持った活動が求められる。（吉田）
・データ同化には様々な応用があるが、その原理は数学・数理科学に基づいており、応用にはデータ同化自身の理論的発展も欠かせない。（三好）
・データ同化は「経験」「理論」「計算」「データ」を中心とした科学の潮流全てを包含し、人間の予測能を向上させる「予測科学」という新たな潮流を生み出しうる。（三好）

［吉田氏講演：詳細］

予測・推論を行う科学的手法には、大まかに2つのアプローチがある。1つは入力データがあり、これを何らかの規則に当てはめて出力データを求める「順問題」のアプローチ、もう1つは出力データが先に与えられていて、何かしらの規則に基づいてこれらのデータが出たと仮定して、この出力を実現する入力データを求める「逆問題」のアプローチである。上述の「規則」は「モデル」として記述される。大まかには決定論的モデルや確率モデルがあるが、このモデルの決め方も様々なアプローチがある。1つは物理などの理論をベースとした「基礎方程式」としてモデルを与える事。これはモデルの形を明確に与える事が可能であるが、実際の現象の再現性には議論の余地が残されており、計算コストとのトレードオフが常に付き纏う。モデルを定める別のアプローチとして、与えられたデータを学習し、回帰などの確率的、統計的手法により構成する「経験モデル」がある。基礎方程式などが陽に与えられなくても入出力の関係を与える事ができるが、関係を精度よく再現するには一般に豊富なデータが必要になる。また、経験モデルでは得られているデータの「外」の情報が何も与えられないため、未知の現象の予測が非常に難しい。一方、実験などで得られる現実のデータは、やはり未知の予測が非常に難しい。そもそも実験そのものの実施にも制約があるものが少なくない。



これらの「仮想・実空間実験」の良し悪しを理解した上で、「理論・実験・データサイエンスの融合の在り方」の１つのアジェンダを提唱するのが、吉田氏の講演の趣旨である。

まず氏が具体的に取り組んでいる課題として「新規ポリマーの発見」が挙げられた。現存する高分子の構造や物性を数値化し（「記述子」の導出）、その性質を学習させて、要求する物性を持つ構造を探索するというものである[17]。この課題に対する一連の研究成果に関連して、氏の講演では３節構成でアジェンダの詳細が紹介された。

1節「理論とデータの乖離、限られたデータの壁を乗り越える」

理論モデルによるシミュレーション結果と実験などで得られたデータの間には（モデルの精度などに依存して）一般に正確さにギャップが生じる。他方で、実験の困難さなどが原因で実データが非常に少なく、シミュレーションなどで豊富な計算データが得られる場合もある。前者は実・仮想の間の乖離をどのように埋めるか。後者は限りのあるデータをどのように統合するか。この２つの問題にデータ科学の手法を適用し、様々な問題解決に貢献する。氏の講演では例として、結晶構造と物性の相関を予測する問題が紹介された。材料と物性（講演の例では熱伝導率）のサンプルデータが45個という非常に少ない状況で、材料の構造と熱伝導率の相関を求め、望む熱伝導率などの物性を持つ構造の推定を試みる。まず、第一原理計算で得られるSPS（散乱位相空間, Scattering Phase Space）に関する320個のサンプルデータを加えてデータを増やし、SPSと熱伝導率の間の高い精度＝強い相関を得る事で熱伝導率の予測を試みたが、思ったような結果を得られなかった[18]。そこで氏の研究グループでは「転移学習」を適用した。豊富に用意されたSPSのデータに対して、ニューラルネットワークの事前学習を行い，元の入力記述子の縮約表現を合成する。この事前学習モデルの特徴量を熱伝導率の45個のデータを用いてファインチューニング（転移）することで、極めて少ないデータにもかかわらず、高精度な予測モデルを導くことに成功した。以上が本問題における転移学習を適用した解析の流

17　講演では、最初の例として高分子のガラス転移温度と融点を指標とした構造探索の例が紹介された

18　相関が弱いという意味で学習結果の「精度が低くなる」

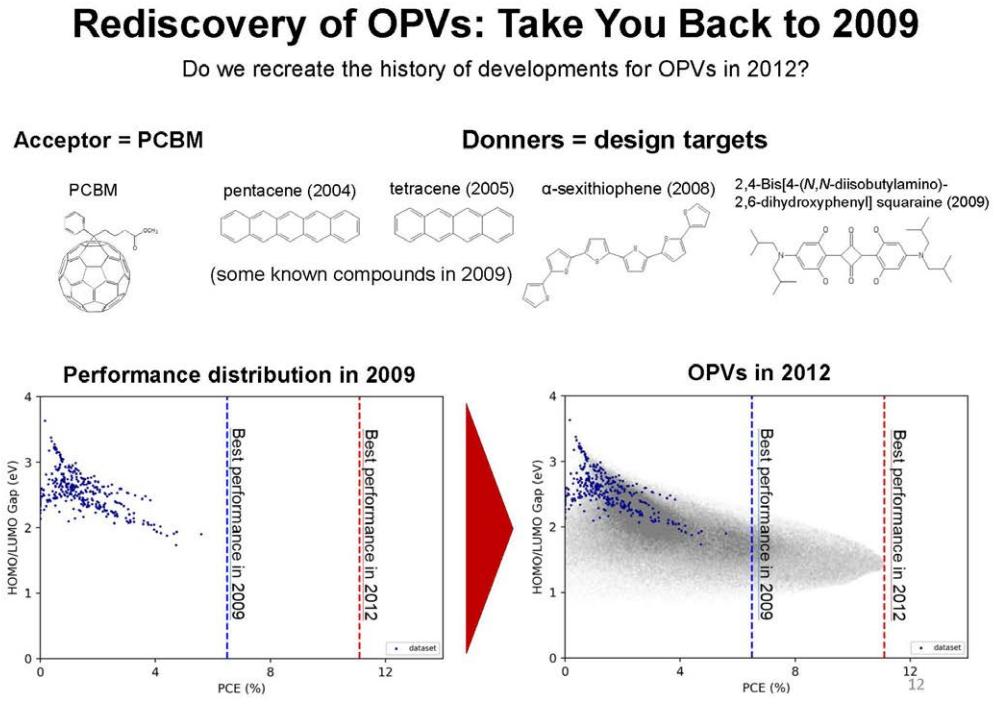
れである。実際、保有する45個のサンプルデータはおよそ300（W/mK）という熱伝導率の情報しか持っておらず、これらをいくら学習させても近い値の物性値を持つ構造しか得られない[19]。しかし転移学習を適用する事で、上記45個のサンプルデータから3000（W/mK）近い熱伝導率を達成する結晶構造を推定する事に成功している。この話題は、限られた数しかないデータ群に対して、シミュレーションにより豊富に得られたデータ群が持つ構造を転移させる事で、少ないデータに対する外挿的学習を実現させた1つの好例となっている。

2節「データ科学の内挿的予測の限界を乗り越える」

次の話題は実験や計算で得られたデータの「外」にある情報を得るための試みである。

例として、有機薄膜太陽電池の化学特性（HOMO-LUMOギャップ）とパワー変換効率の相関関係に関する考察が挙げられた。計算・実験手法の改良などによって2009年に得られたデータと比較して、2012年ではより詳細かつ広範囲の変換効率と化学的構造の相関関係を得られているが、ここでは『2009年に得られたデータだけから2012年に得られたデータを推定できるか』を問う。

上の話題そのものは一種のベンチマークであるが、2009年のデータにとって2012年のより広い範囲をカバーするデータは未知の領域（ブルーオーシャン）であり、この問題を解決することはデータの「外挿」の可能性に大きく貢献する。機械学習で予測していく手もあるが、精度の高い結果、あるいは望む物性値を達成する結果まで到達できないのが現状である。そこで、氏のグループは「適応的実験計画法」を組み込むことにより、第一原理計算の計算結果のサンプルを絞り込み、得られたデータのモデルによる最適値を発見、それを用いてモデルの修正を随時行う一連のワークフローをコンピューター上で実現させる枠組みを構築した。この方法により、こちらで与えた未取得データ領域の結果を得ることに成功している。

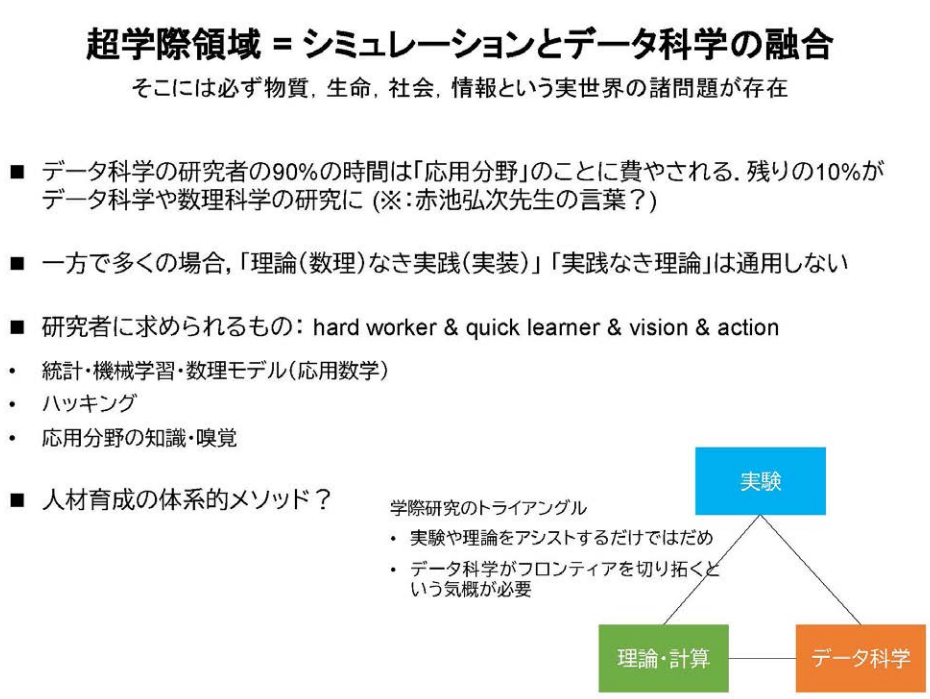


19　この意味で、経験モデルによる方法は「内挿的」であると言える

3節「データ駆動型研究に資するデータを作る」

最後は汎用性の高いデータベース作成に関する話題である。無機化合物では第一原理計算をベースに大量の化合物の物性値をまとめたデータベースの開発が進んでいるが、有機化合物、特に高分子材料は構造が多種多様であり、実験、計算ともにデータ取得のコストが極めて高いため、データベースとしてまとめられるだけのデータが揃っていないのが現状である[20]。MDの自動化やハイスループットスクリーニングなどにより効率的なデータ収集を行っているものの、高分子構造の多様性により1ラボで運用するには（計算資源、人材などのコスト面で）限度がある。そこで氏のチームでMDの自動計算システムと機械学習モジュールを開発、これを様々な産学連携機関とシェアし、各機関で計算し、全体でデータを取りまとめて巨大データベースを構築するというプロジェクトが進行している。

データ科学が絡む課題は超学際的であり、実世界の諸問題（物質、生命、社会、情報など）との結びつきが非常に強く、取り組みのためにはこれらの分野に対する見識も不可欠となる。「データ科学研究者の90%の時間は『応用分野』の事に費やされ、データ科学、数理科学そのものの研究は残りの10%しか時間を割けない」という提言もあるくらいである。しかし、吉田氏は「理論（数理）なき実践（実装）」、「実践なき理論」のどちらが欠けても通用しないと主張する。データ科学、数理科学の研究者が実問題の解決のアシストをするだけでは不充分で、データ科学研究者には「データ科学自身がフロンティアを切り拓く」という気概を持った活動が求められる事を強調されて、講演は締め括られた。



20 データの豊富さの評価は議論の余地があるが、データ量だけで評価するならばPoLyInfoというデータベースが、数万個の高分子化合物に対して、最大100個程度の物性情報を有しており、講演内容のみを閲覧する限りこれが高分子関連で世界最大のデータベースと見受けられる。ただし、複数の物性が同時観測されているデータはかなり少なく、データの偏りについても議論の余地が残されている

［三好氏講演：詳細］

実験などにより得られた観測データと、シミュレーションで得られたデータには乖離があるのが一般的である。これらの情報を結びつける事で双方の情報を最大限に抽出する技術に「データ同化」がある。近年ではビッグデータをスーパーコンピュータに取り込み、データ同化を適用する事で（ビッグデータ同化）、膨大なデータを有する現象、効率的エネルギー運用や生命現象、インフラ整備や、環境に関連した諸問題に大きく貢献できる事が期待される。三好氏とその研究チームはゲリラ豪雨などの気象に関連する諸問題を始めとして、ビッグデータ同化に関連した実用化も含めた先駆的な成果を多く挙げている[21]。

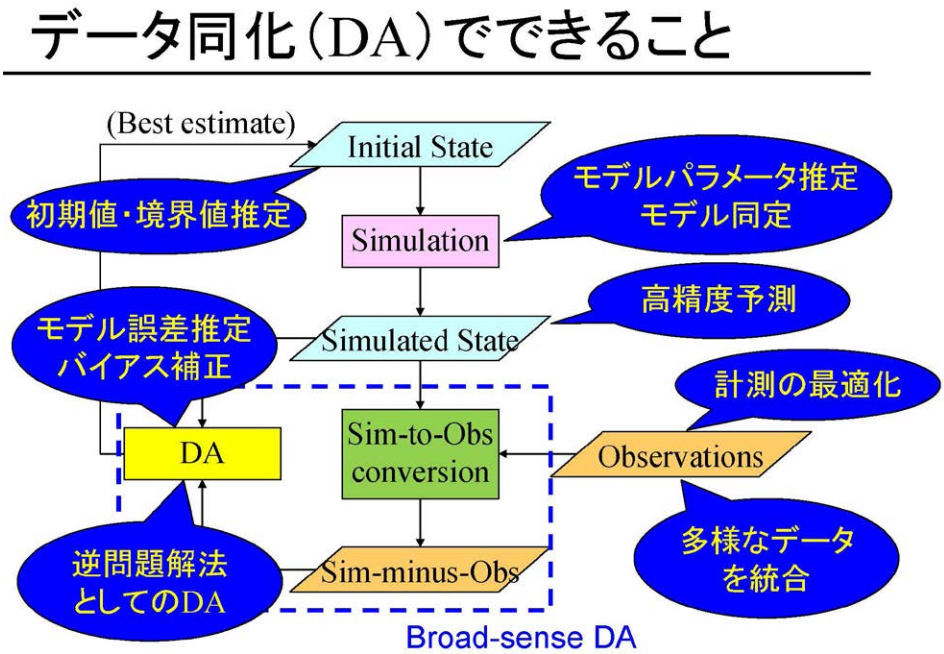
氏の講演は「天気予報」の技術的課題とデータ同化の歴史より始まる。天気を記載する天気図は従来、様々な観測地にて気象データを観測し、プロットする事で広範な地域の気象情報を得る。リアルタイムで広範な地域の情報を得るためには、非常に良い「通信環境」の整備が欠かせない。戦後間もない頃まではモールス信号などで離れた観測地点での間の情報のやり取りがなされたが、通信のタイムラグや観測地点の有限性による「『リアルタイム』『広範囲』での天気の精度」には限度があった。1959年頃より、日本でも数値シミュレーションによる天気予測がこれらの困難を克服する技術として期待され導入されたが、当初は精度が高くなく、実観測に替わるものとしては不充分なものであった。30年ほどして計算精度も高くなり実用性も高まってきたが、より良い精度を出すために実際の観測データとの関連、「データの取り込み」が課題として浮上し始める。これが気象学とデータ同化の関連の始まりであり、実観測データとシミュレーションデータ双方の「活かし方」を探究する契機となる。データ同化は数値シミュレーション結果に実測データを取り込む事でシミュレーションの軌道を随時修正、精度を高める技法である。近年では膨大なデータの取得が可能となった一方、従来のデータ同化ではその処理が追いつかなくなってくる。そこで、氏の研究チームではビッグデータとデータ同化の手法をスーパーコンピュータで取り扱えるよう高度化し、リアルタイムで急激に変化する天気、例えばゲリラ豪雨の精度の良い推定を可能とした。実際、氏のグループは30分後のゲリラ豪雨を予測する手法を開発し、

21　この成果の概要はYoutubeで閲覧可能である。

スマートフォンなどのアプリ開発[22]などに大きく貢献している。ゲリラ豪雨はわずか10分の行動の違いが生死に関わる劇的な差を生む[23]。氏のグループによるリアルタイムの天気予測、それを実現したビッグデータ同化の技術は人々の生活に非常に大きな恩恵を生んでいる事は疑いないだろう。

静止軌道に気象レーダを打ち上げたら天気予報がどの程度改善するかをJAXAと考えた。宇宙太陽光発電という新技術の考え方があり、巨大な太陽光パネルを静止軌道に置く必要がある。それに向け、３０m四方程度のパネルで小規模の実証実験を検討する。発電施設としては小さすぎるが、裏面を気象レーダとすると有益である。仮にそのレーダを用いて台風を観察するシミュレーションを行い、データ同化を行うと予報が改善する結果を得た。このように、データ同化を用いることで、今あるシステムで天気予報をよくするだけではなく、今はない仮想的な観測データの効果を評価できる。また、どういう観測装置を設計すれば天気予報に効果的か、仮に１億円あったとすれば同じ１億円で何をすれば効果的か評価できる。上記以外にも森林の分布のシミュレーションを衛星観測データとデータ同化する事による制御への応用や、金属材料のプレス加工にデータ同化を応用する例もある[24]。このように、ビッグデータ同化は天気予測以外の多くの場面でも貢献している。

データ同化は実問題に非常に応用の効く技術であるが、その源泉は「力学系」「統計数理」「UQ（uncertainty quantification）」などの数学・数理科学にある。例として「カオス的」な振る舞い[25]の同期が紹介された。系の初期値摂動に対する不安定性が非常に強い振る舞いに対して、観測データをシミュレーションモデルに取り込む（＝データ同化する）事で、真の系の振る舞いと同じ振る舞いを再現する事が可能と



22　株式会社エムティーアイ「3D雨雲ウォッチ」

23　2008年7月28日、神戸市都賀川における水難事故などが挙げられる

24　例として、金属板をプレスしたときに起こる「スプリングバック」の程度を評価する問題が挙げられた。非常に複雑な力学的特性の解析が必要であり、モデルの数値シミュレーションでは予測・制御が困難なこの現象を、（プレス加工中の測定データを援用した）データ同化を用いる事で予測精度を上げ、スプリングバックを防止する最適なプレス加工の実現に貢献している。同技術はクラックの防止にも使われている

25　第５回、國府寛司氏の講演も参照されたい

なる。講演ではカオス的振る舞いが発現する1次元ロジスティック写像モデルに対してこの特性が紹介された。ノイズのある「観測データを拾い」、シミュレーションモデルやデータにこの観測を「教える」事で、観測と数値シミュレーションの「誤差を制御する」事がデータ同化の1つの数理的本質である。微分方程式や写像の反復で表現される数理モデルと、確率モデルを組み合わせることで、最適な誤差の制御が可能となる。この際の確率モデルはベイズ推定に基づいた最適化理論である。観測を何らかの方法でシミュレートできる限りデータ同化のアイデアは適用可能であり、多種多様な分野でその威力を発揮する。

「経験」「理論」「計算」「データ」を中心とした科学の潮流があるが、データ同化はこれら全てを包含し、人間の予測能を向上させる「予測科学」という新たな潮流を生み出しうる事が三好氏の講演の締めくくりのメッセージとなった。

【質疑応答】

Q：（実在データに）不足している部分を補って乗り越える方法論を提示し、適用する事が今日の（吉田氏の）話題であると理解している。
不足分を補うために「これくらいのデータがあるといい」という評価は、トライアンドエラーという形ではなく体系的にできるものなのか？言い換えると、どの程度の転移があれば充分か、ある程度予測は可能なのだろうか？

A：1節の議論で用いたサンプル：320という数は意図せず選んだもので、これで結果が得られたのはある意味運が良かったと言える。
（体系的予測という意味で）賢いやり方があるとすれば、「実験計画法」を組み込む事だろうか。どのようなデータを作れば良いかのワークフローを、実験計画法の議論も併用して組む。デザインの方法はデータ科学分野でできつつある。他方でシミュレーション、実データ「のみ」を使って何かしらの結果を得るには適用範囲が限られてしまうので、両方をうまく使う事が大事。

Q：（吉田氏の講演）3節で紹介されたやり方では、計算で得られたデータ「のみ」を使ったのだろうか？

A：ここでのデータはMD（分子動力学）計算で得られた高分子構造の情報を用いている（特に、計算で得られたデータのみを用いていると言える）。
CREST[26]では、計算データと計測データの両方を用いて様々な物性を調べている。計算で得られた物性と、実際の物性の対応関係は、取り組んでいる最中の課題である。

Q：汎用プラスチックは様々な種類がある。ポリエチレンだけでも（結晶性とか、非定形なものとか）相当なものである。理想的な状況で構築できるものだけしか実現できないのだろうか？現状はどのように？

A：アモルファス[27]や混合系など、現在はデータを貯め始めたところ。データ取得、計測でもアプローチは様々であり、物性データベースの作成については無機化合物では多くの動きがあり、成功しつつあるようである。対して高分子材料は（上記のように多種多様な構造もあって）とにかく計算や計測を始め、データの収集が大変である。高分子の実験家の中には「計算だけでデータを集めても意味がない」と思いがちな人もいるらしいが、その空気に一石を投じたい。

Q：AIとデータ同化の融合は可能だろうか？

A：（三好氏の公開スライドの）37ページの図で挙げた項目全てに取り込む研究がある。しかし、AIのテクニックを単に取り込むことから始めるが、その先には理論や方法論、アルゴリズムレベルでの融合へと向かうことが大事だと思う。

26 熱制御領域「高分子の熱物性マテリアルズインフォマティクス」（代表：森川 淳子，東工大）

27 結晶や準結晶などのような「並進・回転対称性」を一般に持たない分子構造を指す。典型的には「ガラス」に見られる構造で、第16回のセミナーにて詳しく紹介されると思われる

Q：（三好氏の講演について）結果の「近さ」の評価はどのように？
A：これは簡単ではない。天気予報でも自明ではなく、その難しさは「『天気予報が当たる』ことの定義が自明でない」事が根本にある。現実的なのは様々な指標を導入、見比べて偏向を見て判断する事だろう。
Q：（先の質問に関連して）AICなどは使える？
A：ようは「近さ」の定義をするだけの話（これが自明でない）。参照データの選び方次第であり、ユーザー次第でどうとでもなると言える。「ある指標」を固定してその土俵で議論はできるだろうが、ユーザー次第となると「近さ」の話題は社会科学分野の問題として考えなければならないかもしれない。
Q：「1つの場面でAIを使うより、いくつかの手法を統合した方がいい」という議論はある？
A：手法の統合については、いくつかのやり方が既に議論されている。計算速度を上げたり、精度を上げたり。その使い方はワークフロー中に色々な考え方があり、関連研究も様々である。データ同化のテクニックは最適化理論に基づいており、これはAIでの手法ともリンクしている。表に見える部分ではなく、データ同化の原理としての最適化理論を見るなど、「原理に近い部分」でのアイデアの融合により、改善はありうるのではないか。
Q：吉田氏のような分野で活躍する「人材」はまさにスーパーマン。「人材育成」となるとスーパーマンを育てるような感じになる気がするが、学際分野での研究開発をするのは、チームでやるのは限界があるのだろうか？一人で学際分野全てを賄う力が必要なのだろうか？
A：関連する学際領域全てをカバーできる人材を大量に作るのは無理。実際は場所・分野ごとにリーダーシップを発揮し、プロジェクトでは相補的に結果や学識を共有する。このような融合研究を続ける中で次第に他の分野の見識も増え、マルチリンガルになっていき、指摘のような人材が出来上がるのではないだろうか。
（CRESTではどのようにされているか？という質問に対して）データ科学という分野そのものがそもそも学際的な分野である。すなわち、統計数理手法＋応用先の分野両方を押さえて研究する分野であるので、自然とマルチリンガルになっていく。（吉田）
教育に尽きる。データ同化という分野は気象学において一大分野になっていて、気象学におけるデータ同化のエキスパートには「自分はデータサイエンティストというより気象学者だ」という気概を持つ者も少なくない。もともと気象学には2つの流れがあり、数式で気象を記述し予測する「モデラー」と、実測などで気象を予測する「フィールドワーカー」がいる。この2つは分離しており、連携は簡単でない。データ同化はこの2つをつなぐものだが、どちらからも爪弾きに合うこともある。とはいえ、現在データ同化の理論と技術は両者の橋渡しとして欠かせないものとなっている。（三好）
Q：（上の人材育成に関する質疑応答を受けて）JSTのようなFunding Agencyにできることはないだろうか？
A：京大でデータ同化の講義を実施し、また理研でデータ同化スクールを実施していて、データ同化の理論と技術を身に着ける場を定期的に設けているが、情報をリーチできる範囲に限りがある。興味を持つ人にうまくリーチする方法があればいいが、ここにJSTなどの協力があると変わってくるかもしれない。データ同化に関しては、「データサイエンス」という分野がそもそも幅広すぎるため、「興味を持つ人」がどういう人かを判別するのも大変かもしれないが。（三好）
Q：三好氏の手法に吉田氏の手法を取り込むことは可能だろうか？
A：AI活用の話と近い。計算速度や精度向上のために、データサイエンスの手法を取り組むという研究も多い。
C：人材育成の質疑応答で出た、三好氏が学位取得しテニュアトラック教員として勤めた「メリーランド大学」の話を少し。
メリーランド大学には物理学、数学、応用数学などの諸科学分野の研究室・研究者が1つになった「IPST（ Institute for Physical Science and Technology ）」という組織がある。ここでは上述の

分野の研究者の他に（カオス理論を含めた）力学系の研究者、気象学者、DNA分野の研究者もいて、異分野協働が自然とできる環境にある。データ同化に限っても、気象学、数学の両面で教育できる環境が整っている[28]*。

対して、日本ではデータ同化の教育が非常に遅れている。気象学を扱う日本の大学ではほぼデータ同化の専門教育が行われていない。他の（数学科などの）組織からの教育外注も無理がある。この現状には、日本の大学の縦割り構造が背景にあると考えられる。（データ同化のような）学際的トピックの教育を広げるには、「横串を通す」枠組みが必要。

28 （松江's コメント.）むしろ、互いに異分野であるという考え方すら存在しないようにも感じる

3.14 統計とスパースデータとAI

［概要］

CRDSの連続セミナーシリーズ、第14回は「統計とスパースデータとAI」をテーマとして、椿氏より「統計数理科学産業応用の類型と近年の動向」、木虎氏より「スパースデータに対するＡＩ構築アプローチ」というタイトルでご講演いただいた。

椿氏の講演では日本でも統計家が活躍していただきたいとの考えのもと、統計やデータサイエンスと産業界との双方向的な関わり方が展開された。学術と産業界に関する統計の歴史的経緯から、CAPDoとPPDACのデュアルサイクル、統計数理の４つの原理。そして、デュアルサイクルのプロセスの中に産業界は必要な先端的な数理、データサイエンス、AIの技術を投入すれば良いということでまとめられた。

木虎氏の講演ではHACARUSのご紹介から始まり、機械学習のおさらいがあった。そしてスパースデータに機械学習を適用したときに起こりうる過適合の問題を解決する有効な方法の一つであるスパースモデリングが紹介され、応用例を４つ示された。最後には企業として数式に関するリテラシーをもつ人材を必要としており、今回のこのセミナーの活動からそのような人材が多く世の中に出てくることを期待しているとのメッセージがあった。

［Key point］

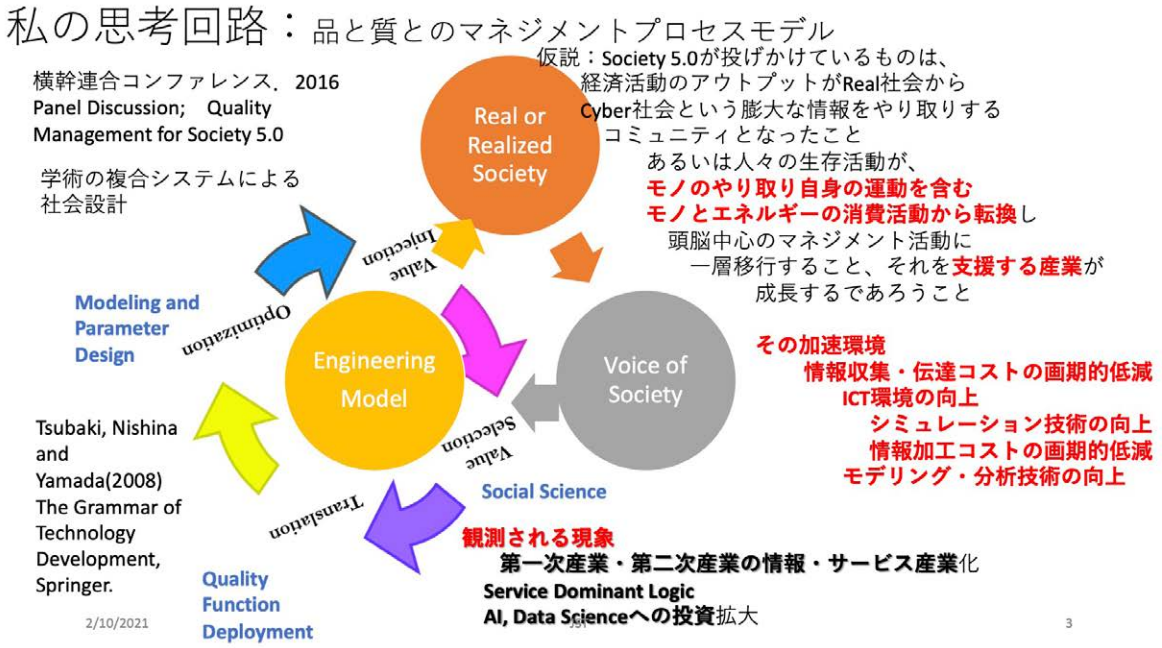
・日常管理のCAPDo（PDCA（Plan, Do, Check, Action））のサイクルと問題解決改善のサイクルのPPDAC（Problem, Plan, Data, Analysis, Conclusion）サイクルのデュアルサイクルとして産業界はこれまで統計学を古典的に使ってきた。（椿）
・統計技術を貫く数理の原理は４つある。それは予測と推論、発見、決定、学習である。（椿）
・機械学習はデータからパターンを導き出す帰納的プログラミングである。そのためデータ量がある程度ないとモデリングの精度が落ち、すなわち過適合、過学習が起こる場合が多い。（木虎）
・過適合、過学習やスパースデータに対して有効な方法の一つがスパースモデリングであり、それは「推論結果のラベルに大きく影響する特徴量は少ない」という仮定に基づく。（木虎）

［椿氏講演：詳細］

椿氏は、現在統計数理研究所の所長、品質工学会会長や自殺総合対策学会の理事長のstatisticianであるが、統計学者よりは統計家であるとの自己紹介であった。アメリカではmathematicianは3000人、大学ではなく社会で活躍して、大変高給をとっていることがアメリカ労働統計局から発表されている。日本でも統計学者より統計家というものがいろんな意味で活躍していただきたいと考えておられ、これまでどういう風に統計やデータサイエンスに関わるか、どういう形で産業界へ関わってきたかこれからかかわるべきかといったことを以下で話された。

最初に、統計数理研究所が産学連携にいたった経緯を話された。統数研に戻ったのは約２年前。それまでは総務省の外郭団体である独立法人におり、公的統計の仕事を４年ほどされていた。戻ってきて驚いたことは、これまで統計数理研究所は日本の統計科学に関する学術分野と共同研究が主体だったが、急速に産学連携研究が進捗していたこと。2018年にAI・ビックデータ・IoTの社会実装が強く言われるようになり、データに基づく産業構造の変革が世界的に進んでいる状況の中で、学術だけではなく産業に対して統計数理科学分野が寄与することは必須の状況になった。この事情は、ビックデータというよりはデータの価格、あるいはデータの解析に関する価格が破壊して、いわゆる限界費用が極めて小さくなってきたのが遠因ではないかと考えられているとのことであった。

次に、横幹連合コンファレンス（2016）では、Sociaty5.0においてどのようなことを考えなければならな

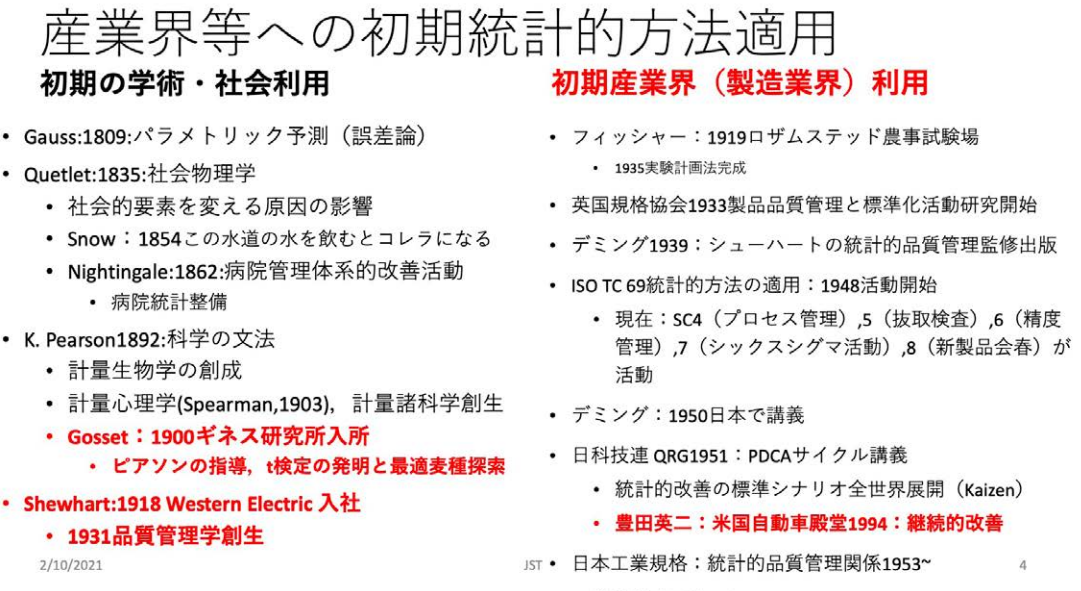


いかを当時品質管理学会長の立場で説明したことがあったとのことであった。基本的に情報収集や伝達コストが画期的に低減され、それからリアルデータではなく数理的シミュレーションに基づくデータ分析が圧倒的に大きくになった。それから情報加工コストに関しての画期的低減、モデリング・分析技術向上。これも以前に比べたら圧倒的に安くなった。フリーのソフトのPythonやRでかなり重要な分析ができるようになってしまった。これに基づいて、データや情報に基づいてもの作りに付加価値を与えるということ、いわゆる第一次産業や第二次産業が情報という意味で、サービス産業化した。世界がそれを目指しているので、AIやデータサイエンス、統計数理研究所に対しても投資が拡大してきたと信じているとのことであった。

こういう状況の中で、統計は過去にどういうものをやってきたか、産業界とどう関わってきたかというと圧倒的に学術に対する貢献であったということの歴史を順に話された。統計自体は数学者のガウスによるパラメトリックなモデル（いわゆる最小二乗法誤差論）が近代の創成に大きく寄与した。その後、社会とかかわるという意味では、天文学者であったQuetletが社会物理学を創成して社会的要素を変える原因の確率をきちっと議論しようとした。 Quetletの提言に基づいて国際統計学会、国際統計協会、今日の国勢調査というものが立ち上がってくる。その後、Quetletの方法は演繹的理論的な立場ではないのだが、1854年にロンドンでコレラが大流行したときに、Snow博士が「この水道の水を飲むとコレラになる」と完全にデータだけで説明した。当時、コレラ菌、細菌の概念もなく、細菌学がまだできてなかった時代でも、データに基づいて対処できた。次にNightingaleが1862年に病院管理の体系的な改善活動を行った。すなわち、病院に入った方の原因と結果のデータを取るという病院統計整備を英国で行った。これに基づいて病院の改善活動を行った。今日のナースコールも、このデータに基づく活動の中で出た。これら個別の活動に対して、1892年にK. Pearsonが科学の文法という概念を提唱して、ある意味で統計を応用数学の一つとしてではなく、数理科学をどのように応用するプロセスを形成するかが熱心に語られた。 K. Pearsonはナレッジマネージメントのパイオニアという科学史の中での評価もある。その後20世紀は学術が計量化、統計化する時代があって、K. Pearsonは計量生物学、さらに指導を行ったSpearmanが計量心理学を作った。そして計量経済学というものが出てくる。ここも学術に対する貢献、データに基づく学術を作るという運動であった。その当時、Gossetがギネス研究所（1900年設立）に入所して、K. Pearson の指導を受け、t検定を発見（or発明）し、それに基づいて6年間にわたってビールにおける最適麦種、あるいは香味、賞味期限の延長を統計的に開発した。これは品質改善活動のパイオニア的なものである。一方、Shewhartが1918年にWestern Electricへ入社後、

1931年までに品質管理学を創生した。彼は生産を計量科学にするというパイオニアである。これは産業界に対して非常に大きな影響を与えた活動である。さらに、農業分野に関しては実験計画法のパイオニアあるいは尤度原理を確立したR. A. Fisherが1919年にロザムステッド農事試験場に入って、非常に大きな貢献、数理統計学自体を作るということを行う。これは現場としての農事試験場があった。なお、ギネスビールのGossetはロザムステッド農事試験場のような優秀な農事試験でやった品種推薦はスコットランドでは全滅したので、もっと一般化可能性のある推論ができなければならないと提唱しており、その解答をFisherは実験計画法という形で与えた。1933年にはShewhartの影響を受けて、英国の規格協会が製品の品質管理と標準化活動（統計的な活動）の研究を開始する。日本のデミング賞にも影響を与えたDeming がShewhartの統計的品質管理の監修を行って、1939年に出版した。1943年にはアメリカで統計的品質管理の講義を全国的に行った。 WWII後の1948年にISO の中に統計的方法の標準化が行われ、今日にも続いていて、プロセス管理、抜取検査、精度管理、改善活動に関するシックスシグマ、新製品開発というものを行っている。 Demingは1950年に日本で講義を行い、これが日本における品質管理活動の大きな転機になった。1951 年には日本の中で、世界で初めて計画実施チェックアクションである（Japanese）PDCAサイクル、こういうサイクルで問題を解決しなければならないということ、が生まれる。これが日本の中で、非常に初等的な指導要領の中で、極めて単純なものといわれているものの中で改善活動が展開された。すなわち、今の日本の小中等統計教育レベルでの改善活動が行われた。統計的改善の標準シナリオが世界的に展開された。豊田英二が1994年に米国自動車殿堂入りするが、これは統計的な方法を使った継続的な改善が評価されたことになる。日本の中では、統計的な方法の産業界利用のための標準化活動が1953年から出てきている。

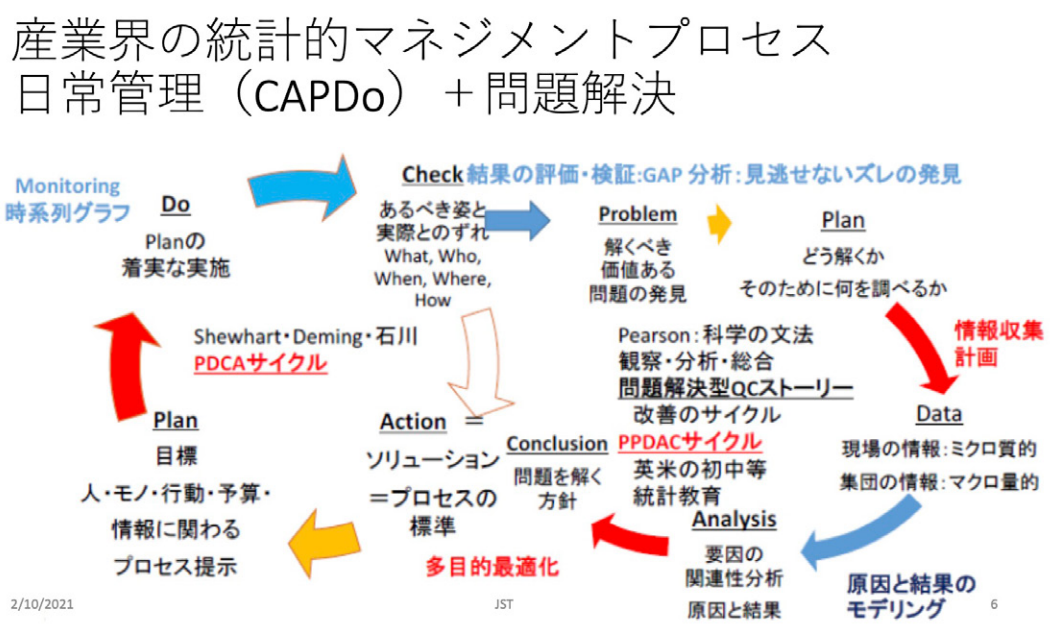
統計をどのように使うかの類型、プロセスモデルが非常に重要であると産業界では歴史的に考えられている。Pearsonの科学の文法は現象の分析→法則形成→批判的検討→現象の分類ということだったが、それをShewhartが目的の規定行為→目的の達成行為→目的が達成したかを検証する行為というPlanDoSeeというものを提唱した。これが日本に伝わってアクションというものが出てきた。ISI（国際統計協会）会長のミシガン大学Nair先生がリオデジャネイロ総会講演2015で統計学の最大の産業貢献はDeming-Ishikawaによる改善の標準シナリオであるPDCA（Plan, Do, Check, Action／CAPDoとも言う）を示したことだと評価した。この後、国際標準化に関してもISO、イギリスや中国が幹事国であるシックスシグマの中で、世界的流布しているDMAIC（Define, Measure, Analysis, Improve, Control）といわれているものがある。新技術開発において、IDDOV（Indentify the Opportunity, Define the Requirements, Develop the Concept,



Optimize the Design, Verify Conformance）、デザインフォシックスシグマというものの標準化がすでに確立しているが、国際標準にしていこうとしている。これについてはアメリカの委員会であるASI（American Supplier Institute）の田口伸氏がタスクフォースリーダーである。

産業界の統計的マネージメント、Deming-Ishikawaの貢献とはCAPDo（PDCAとはあまり言わない）である。 CAPDoとPPDAC（Problem, Plan, Data, Analysis, Conclusion）のデュアルサイクルは、あるべき姿と実際のずれからスタートする異常検知で、そこで異常があったら、解くべき価値のある問題の発見と考え、情報収集し、モデリングを行って、問題を解く方針を決めて、標準化してソリューションとして、新しいプランにつなげて、日常的に管理して、また日常の管理の中で、アウトライアーが起きたら、解くべき問題ではないかと考えて、問題解決のサイクルへつなげる。そのデュアルサイクルとして産業界はこれまで統計学を古典的に使ってきた。今日の大学の統計教育というより、欧米圏の小中等数理科学統計教育では、PPDAC教育が徹底している。ある意味そのパイオニアは日本の改善の標準シナリオであったとのことであった。

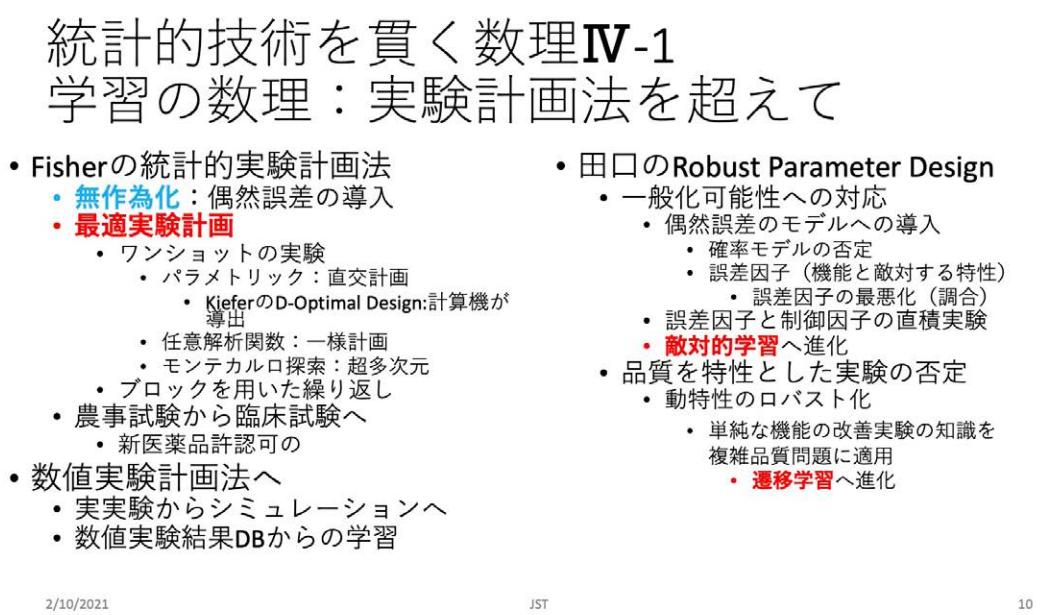
ここから、統計技術を貫く４つの数理について語られた。統計学はサイクルの上で考えることは非常に重要だが、問題解決するときのモデリングを貫いているものは、予測と推論の数理である。誤差論以来、AIや深層学習などで結局予測をしている場合では、条件付き期待値関数を最良近似するかに過ぎない。それ以上は錬金術である。パラメトリックモデルみたいな単純なモデルは有効な推定量を与えてその中で、必要なパラメーターを与えてスパース化する。統計数理研究所の赤池元所長が行ったAICはまさにそういうものである。セミパラメトリックモデル（ノンパラメトリックモデル）というちょっと見えにくいモデルになったとき、その種の予測最適化をCross validationみたいな経験主義的なものにしていくことで、コンピューターでいろんなことができるようにする。さらに射影追跡回帰はノンパラメトリック項をいれていくことで、今日の深層学習のようなニューラルネットを含む統計モデルがすでに20世紀にFriedmanらによって提示された。最適予測の原理ともう一つ、母集団を推測する、観測されていないデータの外のものを推測するという統計学の大きな流れがある。これには単純に標本調査論や、観測していない情報に基づく推論、ノーベル経済学賞に一番近い統計学者Rubinの因果推論がある。後者が今の人工知能、機械学習系でどう単純な予測に吸収するか議論されている。統計数理研究所の中でも、ものづくりデータ科学研究センターの中では、藤澤副センター長を中心に、欠損のデータから何か要因を特定するという東芝との共同研究がある。まさに母集団の推測と最適予測の結合の技術、スパースモデリングみたいなものが行われなければならいと考えておられるようであった。



第二の数理は、発見の数理である。さきほどのデュアルサイクルの問題解決のトリガーになる、何が異常であるか、外れているか。これは人工知能ですでに実現していると考えている。これは基本的に予測原理がきちんとできれば、予測と実際が乖離があることから発見される。金融関係ではcorrelation break downという異常現象を発見して、いつ店を閉めるかは今でも昔でも重要である。株価指数の中でどこに異常があるかを後からではなくリアルタイムで発見する技術が極めて重要となっている。そのためにはモデルを分解して、モデルと実際の予測値の残差を時系列的なモデリングをして、その中で外れているデータを発見するということが重要である。

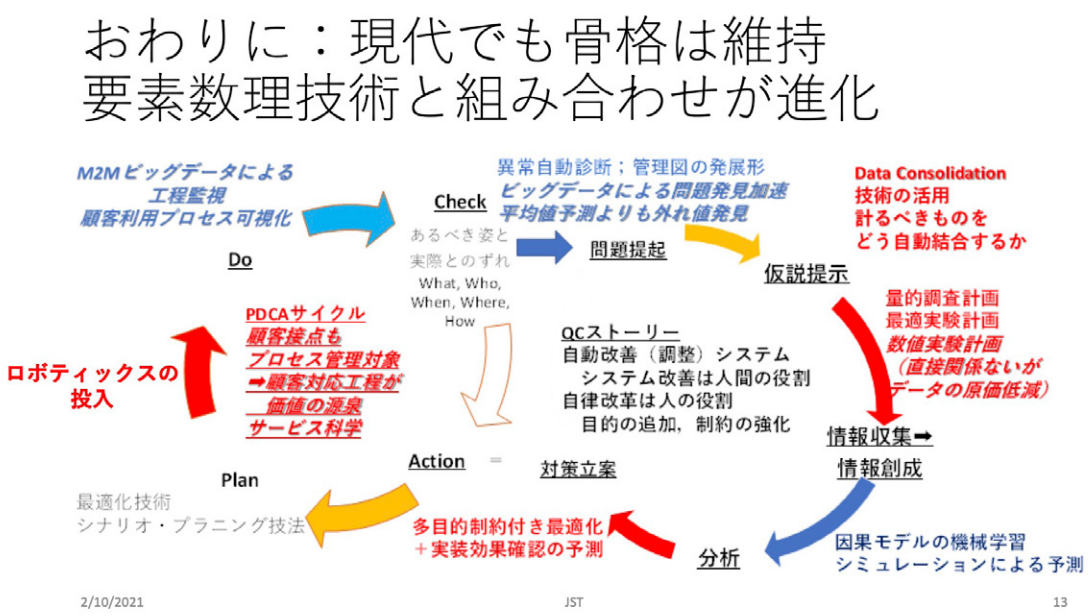
もう一つ、統計技術を貫く数理の中には決定の数理がある。これはどちらかといえばマーケティングや経営戦略論で用いられた。アメリカのビジネススクールで当初から教育されていた方法であった。損失関数を最適化するということで、予測の数理を包含するものになっている。2つの戦略があり、一つは期待リスク最小化戦略、いわゆるベイズ的意思決定。これは機械学習の中でも正則化技術に極めて密接に関連している。もう一つは、事前分布は未知なので、ミニマックス的な戦略、ゲーム論的方法（最も不利な状況、相手が自分にとって一番悪い手を打ってくる）を使うということ。統計的機械学習の中で、敵対的学習という形で、すでに統合されている。古典的な例だが、アメリカの化学業界などではシナリオプランニングで、最適化戦略をすでに使っている。それはビジネススクールの必修科目になっている。

重要なのは学習の数理という、古典的な実験計画法を超えるもの。 Fisherの統計的実験計画法の中で、誤差との戦いということで、無作為化や最適実験計画という方法があった。我々の知りたいことを最適に抽出するか、最小のデータで最大の情報量を得るということ。これには日本の産業界のお家芸だったパラメトリックモデルに対する直交計画がある。それから、任意解析関数の近似である一様計画をKai-Tai Fang教授が提唱した。フォードが600万ドルで買い取った技術になる。超多次元ではモンテカルロ探索。それらが農事試験から臨床試験（新医薬品許認可）に使われていた。今日、リアルデータのデータ解析ではなく第一原理計算に基づく数値実験計画が極めて重要になってきている。すなわち、model basedの開発ということで各製造業において使われているのは、実験計画法の枠組みである。日本が作ったものとしては田口のRobust Parameter Designは重要な業績である。統計を超えていて、偶然誤差をモデルに投入してしまう。確率モデルではなく、意図的に誤差因子（機能と敵対する特性）を作る。これは1950年代から行われていた。誤差因子の調合と呼ばれる最悪化、これは敵対的学習へまさに進化したものだが、実験計画自体は1960年代



から1990年代までには確立して、企業の中で使われていた。品質を特性とした実験を否定して、ある意味単純な機能に関する実験知識を多く出し、それを実問題に類推として適用する。これは遷移学習に進化した。統計数理研究所の吉田教授が1月21日（当セミナー第13回）にて、すでに話されたが、逐次実験（Wald流の逐次推論の最適化、ベイズ流逐次最適実験、能動的学習）、その上で逆問題の予測、遷移学習により実際に新しい化学物質を予測に可能とした（XenonPy）。こういうことが敵対的学習、実験計画法や田口method（Robust Parameter Design）の進化形として、今日機械学習の中に埋め込まれつつある。そのあたりの総合的な理解があるかどうかは別であると考えておられるとのこと。

統計数理にはいくつかの基本原理がある。重要なことは、統計学は科学の文法から出発して発展してきたものであるから、一種のプロセスモデル、統計的な方法、数理的な方法をどう使うかは非常に関心がある。今日の進化、1930年代、1970年代の中でもCAPDoとPPDACのデュアルサイクルのモデルを維持してきたわけだが、CAPDoのCheckの中で当然AIが使われる。数値実験、因果モデルなどでAIや機械学習が新しい要素技術として埋め込まれる可能性がある。多目的制約付き最適化がActionを決めるところで要素技術として数理の先端的技術が当然使われることがある。日常管理においてもロボティクスが投入され、プロセス管理の対象としてはロボティクスだけではなく顧客がどうつかうか、小松製作所がやっているような顧客利用プロセスを可視化する。椿氏の考え（もしくは信じていること）としては、こういう形でデータサイエンスが発展していくと思うが、このプロセスの中に産業界は必要な先端的な数理、データサイエンス、AIの技術を投入すれば良いということであった。それを支えるもともとの数理的原理をそこに繋がっているということを意識することも有用とのことであった。



［木虎氏講演：詳細］

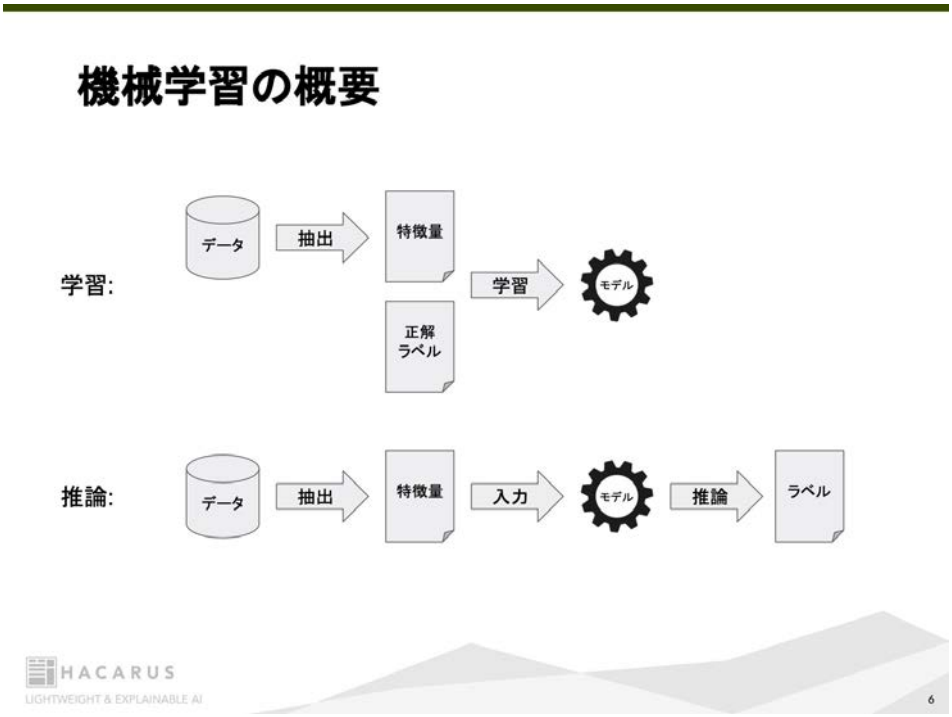
木虎氏はソフトウェア開発、機械学習やデータ分析をバックグラウンドに持っているとのことであった。本題の前に、HACARUSの説明があり、2014年にヘルスケアのベンチャーとして創業、途中何度かのピボットを経て、8年目の現在、いわゆるAIベンチャーといわれるスタートアップ企業とのことであった。あべのハルカスの“ハルカス”ははればれとさせる意味の古語から名付けられているのに対して、ヘルスケアベンチャーとして起業したHACARUSは創業当時スマートキッチンメータを作っていたのだが、栄養素の摂取量を“計（はか）る”ことから名付けられたとのことであった。ロゴも計測の計をモチーフにしたものである。HACARUSは、

現在は主に医療と製造業を中心にした産業に、AIのサービスを提供している。

AIと機械学習は厳密に言うと違うものであるが、昨今AIというとディープラーニングなどを始めとする機械学習を指すこととなっている。機械学習をおさらいして、スパースデータに相性のよいスパースモデリングというコンセプト、最後にスパースモデリングの応用例について以下で話された。

機械学習にはこれと決まった一つの定義があるわけではないが、木虎氏のしっくりくる定義とは「明示的にプログラミングすることなく、問題に合わせて選んだ手法とデータを与えることで、利用者が求める結果を得られるようにする技術」ということである。従来の手法との違いを比較した例は以下の通り。従来のシステムは演繹的プログラミング。入力に対して求める結果が得られるように仕様に基づいてプログラマが処理を記述する。一方、機械学習のシステムは帰納的プログラミング。データサイエンティストが選択したアルゴリズム、与えられたデータに基づいて、コンピュータが入力のルール（モデル）を見出す。すなわち、問題に合わせて選択したアルゴリズムにデータを与えてモデルを構築する。

機械学習は大きく分けて、学習と推論というフェーズに分かれる。学習フェーズにおいてはオリジナルのデータから抽出された特徴量と人によってつけられた正解ラベル（正常や異常、人や自動車など）を学習してモデルを得る。推論フェーズでは学習フェーズと同じ方法でデータから特徴量を抽出して、学習フェーズで得られたモデルに与えることで推論結果を得る。（さきほど上で述べたように）機械学習はデータからパターンを導き出す帰納的プログラミングである。そのため学習に使ったデータがいろいろなパターンを網羅していた方がより推論精度が高くなる。すなわち、ある程度データ量が必要だということ。逆に言うとデータ量がある程度ないとモデリングの精度が落ちる。具体的な例として、平面に点がばらまかれているデータを考える。データ数が十分あれば、近似の直線をひいたときのばらつきが少なくなることは直感的に理解ができるが、一方、同じ例でもデータ数が十分になければ、十分にある場合の直線と同じ物を引くことは少ないと思われる。またこのとき、学習に使ったデータへの当てはまりの良い曲線を引くことができるが、その曲線は学習に使った以外のデータに対しては当てはまりが良くない。すなわち、過学習や過適合である場合が多い。こういった場合に指針となるのが、オッカムの剃刀である。つまり「ある事柄を説明するのに必要以上に多くのものを仮定すべきではない」という指針である。この指針は14世紀の哲学者神学者のオッカムが多用したことで有名になったと言われている。原文は「必要が無いなら多くのものを定立してはならない。少数の論理でよい場合



は多数の論理を定立してはならない。」というもの。オッカムの剃刀の指針に則った方法論がスパースモデリングである。スパースモデリングでは観測値に影響する要因はシンプルに説明できるという仮定を置いている。機械学習と同じ図（言い方）に沿うと、推論結果のラベルに大きく影響する特徴量は少ない。データから本質を抽出するのがスパースモデリングということもできる。スパースモデリングも理論は数学に基づくもの。簡単には、100変数の連立方程式の場合、方程式が10000個あったとしても、そのうちの100個（全体の1％）の方程式さえあれば、連立方程式の解は求まる。仮に連立方程式が10000変数の場合でも、非ゼロの数が100個であれば、100個の方程式があれば解は求まる。非ゼロの数が100個ということは重要な要素が全体の1％、すなわち、スパース（疎）であるということを仮定できれば解を求められると言える。スパース性の仮定をさまざまな形で表現したり利用したりすることで、いろいろな問題に対応できる。スパース性の仮定も数学に基づいており、数式で表すことができる。

次にスパースモデリングの応用例を4つ紹介された。1つ目はブラックホール撮像への応用。約2年前、2019年4月に話題になった。これは地球上の6カ所にある8つの電波望遠鏡を用いて、ブラックホールシャドウを観るという試みである。電波望遠鏡のデータだけでは観測量が不足していて方程式が解けずにブラックホールシャドウを復元することは困難だった。無理矢理画像復元してもデータが足りていないので、不自然な結果にしかならなかった。ここにスパースモデリングを適用することで、少数の本質を特定して撮像に成功した。

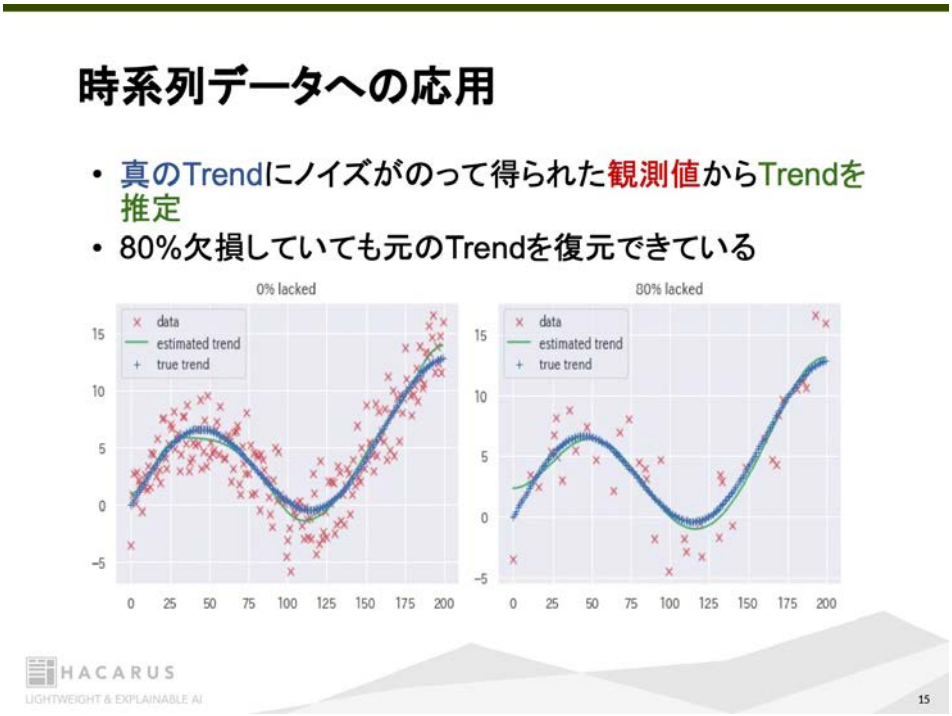
2つ目はマテリアルズインフォマティクスへの応用。これはスパースデータへの応用というわけではなく、スパースモデリングの応用事例である。マテリアルズインフォマティクスというのはさまざまな種類の材料からそれらを混ぜ合わせて目的の特性をもった素材を開発したいということ。人間の試行錯誤では膨大な実験回数であったり時間がかかったりするので、データから材料の候補をしぼって実地作業を効率化するモチベーションとしたもの。リチウムイオン二次電池の負極のとなる有機材料の例では、膨大な材料の候補からまず研究者のドメイン知識を使って、16個の重要な材料に絞り込んでから、電池の負極となりうる有機材料の観測において、負極の容量を決定づけるのに重要な要因は少数であると仮定してスパースモデリングを適用することで、容量を予測するために重要な因子を抽出した。

3つ目は時系列データへの応用。バイタルデータなどを始めとして、ウェアラブルデバイスなどのセンサー

から時系列データを得やすくなっている。一方で、ウェアラブルデバイスなどでは肌との間に隙間ができるなどの理由でデータに欠損が生じるときも珍しくない。バイタルデータなどの時系列データからTrendをみるケースも多いが、欠損の生じたスパースデータからは正確なTrendを計算することが難しくなることがある。そういった場合もスパースモデリングを適用することで、うまくTrendを推定することができる。実際、80%欠損していてもうまくTrendを計算可能であることが具体例をもって示された。

最後に4つ目、画像への応用例。オリジナルの画像に対して、ランダムに50%欠損をさせた画像のみを用いて元画像の復元させることができる。ここでは、画像をいくつかの小片（パッチ）に切り分けて、小片を構成する画像要素を学習によって得るという手法を使っている。より詳細にはある小片を構成するのは少数の重要な画像要素であるという仮定をおいて、スパースモデリングを適用している。これにより欠損していない部分のみから画像の構成する重要な構成要素を欠損していないものと同程度に求めることができるようになる。

まとめとして、今日の流れが話され、最後は以下の様に締めくくられた。「本日はあまり数式を使わずにスライドを構成したので、伝わりにくかったところもあるかもしれないが、スパースモデリングも数学的な理論に基づいたものであって、論文などでは数式を伴って説明されている。当然のことながら、数式が読めないとスパースモデリングを実際につかったりつかいこなすことができなかったり、プログラムの実装もできないので、数学的な理論を構築する側でなく使う側ではあるものの数式に関するリテラシーをもつ人材を企業としては必要としている。今回の活動からそのような人材が多く世の中に出てくることを期待している。」



【質疑応答】

Q：子供の頃から統計といえば数理統計だが、統数研はあえてmathematical statisticsではなく、前にstatisticsを入れて、後ろにmathematicsをいれるその心は現在でも大きく生きているのか？

A：統計数理研究所は昭和19年にその名前で出発した。発足当時、終戦直後になっても当時の社会的な問題、（戦時中は辛辣な問題であったが、）現象や行動、意思決定にかかわる実際の、現場の問題に取り組む現場主義というものが統計数理研究所の中で脈々と文化として培われていった。現象や行動を実際に解明するために、統計学、数学をつかう精神が統計数理研究所のパイオニアが築かれたもの。

それが由来に関する第一点。一方。 Statistical mathematicsと英訳すると英国では数理統計に近い概念になると聞いており、日本の統計数理研究所はそういう研究所，「統計数理」が外国に対して誤解を与える可能性がある。 ISIのNair元会長は「日本人は名字と名前を逆に言うからStatistical mathematicsはmathematical statisticsのことだ」と本当に言ったことがある。「それは困ります」と申し上げた。現場主義、現場から統計や数理統計の問題が出てくる。仮説検定、実験計画法はすべて数理統計学の基幹となった問題だが、ほぼ間違いなく現場の問題解決の中から生まれてきた。

Q：産業界とかいろんな統計学の活躍している場所を話されたが、今、いろんな社会的データがあり、AIと同じで、下手に使うと危険なことがたくさんあるとの指摘がある。恣意的に危険な状況もあればそうではないこともある。そのときに統計学が果たす一番大きな役割はどんなものがあるか？

A：指摘は重要。二つある。データが間違っていたら、AIは大変なことになる。データの推論や予測を自動化する技術なので、原料となるデータの質は大変大きな問題になる。統計学の中に、データの品質、精度を扱うものが画然とある。さきほどのISOのTC69SC6の精度管理はデータの品質を表示し、そしてどのような実験や調査をすることでそれを表示することができるかを国際標準化するもの。であるから、データのクオリティーコントロールは今日の最も大きな問題といえる。もう一つの数理統計学側の貢献は母集団の推測、偏ったデータからいかに偏らない推論をするかという技術である。今日、ものすごく重要視されている。機械学習を生業にしている企業などでもこういう部分とこれまでの予測技術と、偏りがあるデータからきちっと推論する技術もAIに中に作れないかという話が出てきていると承知している。

C：因果推論や母集団の推測は大変気になるところで、数学側への問題を寄越していると感じている。学問としても豊かなものと思っている。

C：数学にもいろいろ助けていただく、教えていただいている。

C：逆のことも多いのではないか。

Q：今日の椿氏の話は、明らかに文系の学生にとって、いかに数理が大切であるかを解説していただけたと数学の立場から思う。これを高校レベルぐらいで文系の学生に数学がないと将来困るということを伝えたいが、高校レベルの学識の中で、こういうことを伝えるにはどうしたらいいか？

A：私は中教審の中ではむしろ初中等数学の担当であった。今回の新しい指導要領では数学の中でも「生きる力」ということを非常に強く話していて、問題解決のサイクルは欧米から逆輸入されたPPDACといったものがすでに日本の高校の数学教育の中に入ってくる。これは統計教育の枠組みではなく、数学全般を使う実際の問題を解決するためのサイクル、ものの考え方が重要だということが日本の新しい数学教育の中に入っている。

Q：PDCAやPPDACといったサイクルがもちろんわかるが、そこと数学が一般のひとには結びついていない。このサイクルを回すのに今勉強している数学が必要になんだといういくつかの例を見せたい。

A：私は統計上ばらつきの多い人生で、統計数理研究所来る前まではビジネススクールにいて、インドやバングラデシュ（ダッカ工科大）の方とかいろんな方が入ってくるが皆、社会人であるのに数学が優秀。日本人のほうが高校までの数学の成績は良かったが、私は社会人大学院にいて、日本人の方はほとんど数学がなぜ必要かわかってもらえず、かえってインドやバングラデシュの人たちは数学はこんなに役に立つにし、面白いのになんで日本の社会人はなにをやっているんだという言われ方をしていた。これも数学や統計の機能がなんであるか、数学においては解析学、代数学、統計で用いる線形代数学でもほとんどきちんとした機能をもっていて、世の中に対して仮に役に立つとしたら、機能に役立つ部分があった。文系の方へは数学がどういう機能をもっているか明確に意識させなければならない。おっしゃるように、中等教育、高等教育では数学問題を解く楽しさ美しさも非常によいが、高校の教程に入っている数学自体がどういう機能をもって、社会的や日常的な問題、別にビジネスの問題なんて世知辛いことは言わなくていいが、そういうものにかかわるか明確に教員が教えられなければならない。逆に

そういう高校教員が多数派なのかどうか。与えられた受験問題的なものを解く技術は非常に長けている。それに対するプロフェッショナルは重要なのだが、それ以外の部分の教育をやっていただける、それを数理的な探求科目でやっていただかなければならない。

C：大学で離散数理なんかで、離散数学からベイズ、機械学習に結びついていくことと大学一年生に教えると、いろんなアンケートをとると数学が使えるということを今まで教わってなかった。つまり、高校生の数学がそんな難しい、大学一年生ですぐわかることがたくさんあるのに、それを全く教えていない。組み合わせ論を教えるならベイズ、機械学習まですぐいくのに、何にも教えていない。その辺りを意見していきたい。

C：そうです。今度の数学Aがチャンスで、今までの順列組み合わせと確率の中に、数学Aのところで、条件付き確率、条件付き期待値、場合によってはベイズの定理まで、と学習指導要領の解説に書かれていて、その部分だけとらえるとチャンスがある。

C：そういうチャンスととらえたい。

Q：少ないデータとスパースモデリングと言ったときに、100変数の連立方程式があったときに方程式がたくさんあったとしても1％の方程式があれば解が求まるといったことや10000変数あってもほとんどがゼロだったら、解は求まる。これはまさしく、今日はスパースデータという言葉でご説明いただいたが、圧縮センサーなどの考え方もそういうことですね？

A：そうです。

Q：（100変数の連立方程式10000個あっても100個で事足りるという方程式の話のような）過剰決定かもしれないものから、100個を選び出して得た解が良い答えであるということをどうやってある程度保証するのか？

A：難しいところではあるが、現場においてはやはり数学だけで、計算だけで出したものだけではなくある程度のドメイン知識を使って判断することが多い。

Q：時系列のデータで、８０％消えていても復元できるのはすごいなと思ったが、欠損の仕方が満遍なく欠損した場合は問題なくできるだろうが、偏った場合はどうか？

A：それでも他の手法、例えば移動平均に比べたらうまくいくというのはある。

Q：欠損の仕方がどうなっているか見るということはできないわけですよね？また、（欠損は）椿氏の講演にあるデータの評価にからんでくると思えば良いか？

A：おっしゃる通り、データそのものだけを見て、どの程度欠損しているか評価するのは無理である。これがバイタル、ウェアラブルデバイスから取られたデータであるとわかっていれば、例えば、5秒間に1度のペースでとられているはずだということがわかるので、そういったところから欠損の程度はわかる。

Q：つまり、（欠損の評価は）ドメイン知識を使うということですね？

A：そうですね。

A：（今の一連の欠損の質問に関することで）Trend復元できる条件は、欠損がRubinの定義しているmissing at randomという状況のときである。これは多くの問題で出てくる。 missing at randomでないとは観測値自体の大きさが欠損の確率に影響する。もともと偏ったデータがおきる状況でなければということがmissing at randomということ。例えば、機械学習の入力変数に欠損は依存するが、出力変数には依存しないという状況であれば（スパースモデリングによる）Trend復元はうまくいく。問題はmissing at randomでない偏ったデータに関してはいろんなsensitivity analysisの方法が提案されている。ただ、非常にスパースモデリング自体は有効な方法だと思う。あとは人間に対して単純な解釈を与えるということが圧倒的に優れていて、そうでなければリッチ回帰、ベイズ縮小という原理が有効だが、木虎氏がおっしゃられていたとおりで、人間が非常に単純なモデルで思考を進めていくということに対してすごく有効である。実はベイズ的な縮小との使い分けは極めて難しい問題だとは思っている。

Q：ウェアラブルデバイスの話があったが、（時系列データのTrend復元は）バイオインフォマティクスや薬物、医療データに関してどれぐらいやっておられるか？

A：医療データへの適用例に関しては、有名なものはどちらかというと時系列データへの応用というわけではなくて、画像の欠損補間に近いものや超解像といわれるものといったものである。 MRIの撮像時間を短くするためにスパースモデリングを使って、超解像して、高精細な画像を得る試みはされている。

Q：データ同化の研究をしているが、missing データ, imperfectなデータ、ノーズがあるデータから推定されているということで、その推定誤差（もしくは推定の分布）、イメージでいうとピクセル毎の値に誤差幅があるが、その誤差幅の推定はできるのか。

A：誤差幅の推定はわからないが、オリジナルとの評価という点では、一般的にPSNRやSSIMといった指標をつかって評価している。

C：さきほどまであった（一連の欠損に関する）質問にも関係するが、データが欠損するほどおそらく推定誤差が大きくなると思うので、例えばデータ同化の手法だと推定誤差も推定する、最近だとuncertainty quantificationという分野が広がってきていて、ある推定値に対して、それ（推定誤差）がどういう分布を持っているかということを推定する。それができるとデータが欠損したときにどれぐらい推定できているのかがわかるのではないかと思ったので質問させていただいた。

Q：スパースモデリングに関して興味があるのは例えば行列とか線形代数をどういう風に教えるかを考えるときに、行列の基本変形を教えるのにスパースモデリングは良い例かと思うが、現実的にそれを学生にうまく教えるのは難しいか？今の行列の基本変形を教えるというと線形方程式を解くアプリケーションがメインだが、いろんなアプリケーションがあると思うが、こういうAI系はインパクトが大きい。スパースモデリングは実は行列を普通の変形とは違うある種の変形をすることだということをうまく教えたいのだが、それについて何かsuggestionがあればいただきたい。

A：例えば辞書学習といわれるものも（画像の欠損補間の）バックエンドで使っているが、おっしゃる通り行列の計算になる。時系列データのTrendも行列での計算といえるかとおもう。一度、実際にこういう風に応用できるのだということを体験してもらうというのが興味を引くポイントになると思う。

Q：ソフトウェアを使って体験してもらうのがいいか、手で計算してもらうのがいいか？

A：Trendの復元や画像の欠損補間は手でやるのは厳しいので、コンピュータで実際自分の目に見える形で体験していただくがよいのではないか。プログラミングというハードルはあるが。

Q：小さい行列でこういうことをやっていると見せて、あとは先生がプログラミングで見せるという感じですかね？

A：そうですね。

Q：最後に数式へのリテラシーを持つ人材を欲しているとおっしゃったが、毎年１人ぐらいとっていく計画があるか？定期的にそういう人材の採用があると仕事がはかどるという意味で。

A：年に１人というレベルではなく、優秀な方がいれば１０人、２０人のレベルで採用したいと思っている。

Q：現状としては（数式へのリテラシーをもった）人材をもっと欲しい状況にあるということでよいか。

A：そうですね。

Q：高度人材育成事業をやっておられると思うが、統計高度人材となると諸分野となると思うが、どのような分野を想定されているか？

A：これまで統計数理研究所樋口前所長がすすめていたのは産業界のリーダーシップをとれるような方ということに対して、古典的な統計学だけではなく機械学習へ繋がっていく統計学をやっていく。これがleading data analytics talentと言われている事業であった。2015年以降、日本全国にデータサイエンス学部やデータサイエンス学研究科の設置が出てきた段階で、統計数理研究所次期中期の中で、どういう風にそういう方々と連携していくかを考えなくてはならないところ。つまり、産業界に対する、アメリカの統計学科の修士課程、いわゆる専門職と認められるぐらいの人材を投入することに関しては、

10年後はデータサイエンス学研究科のミッションになってくるのではないかというのが私の感覚である。それに対して、統計数理研究所はそういうところと連携して、世界的な標準的なレベルにおけるデータサイエンスないしは統計数理科学とはこういうものであるという協定を皆でネゴシエーションしていく中核組織になっていけばと考えている。一方で、統計数理研究所が現在抱えているものは日本の中では「卵が先か鶏が先か」という論争があって、エキスパートレベルの方々を創出するための大学における教員が圧倒的に不足している状況がある。統計数理研究所自体、全国の大学と連携して大学において統計数理科学の基幹理論を教育できる人材を創出するというものを次期中期にむけた方針として考えているところである。これは地味に思われるかもしれないが、例えばディープラーニングの、2014年ぐらいに出ている敵対的学習を作ったGoodfellowさんのMITのテキストをみると、第一部入門編と第三部今後の研究はほとんど数理科学や数理統計のモデリングに近いところが書かれている。そういうことを考えると、日本の中で今後データサイエンスやAI分野の基礎研究開発能力を高めるためには意外と数理統計や統計、基幹理論に近い人材をそれなりの数をアメリカや中国のようにそろえておかなければならないというのが私どもの問題意識。大学でエキスパートレベルを教えられる、しかも共同研究、学内のコンサルテーションをできる人材を統計数理研究所のメインターゲットとすることを我々は議論している。

Q：今の話の続きとして、スケールとしてはどれぐらいの人数のエキスパートレベルを大学で教えられる人が日本にとって適切か？

A：アメリカは現在、統計家や数学家がどれくらい大学に毎年入ってくるかの推計がアメリカ労働統計局で行われていて、博士を取った統計家が大学に、最低今後10年間毎年100人ずつ入る、1000人増えるだろうというのがアメリカの政府の推計である。一方で、アメリカはご承知のように各大学に結構統計学専攻があって、もともと統計学者が各大学に配置されている中でそういうことがおきていることがあって、日本も人口比からすれば、本来年間５０名数理的統計ができる方々が学内に配置される状況を考えなければならない。ほぼゼロからの出発になっている大学が多いなかで、非常に難しいことになっている。今般の政府予算に文科省が統計エキスパート人材、数理的な統計で修士ぐらいまでを担当できる方を今後５年間で３０名程度育成することを挙げている。この方々が日本の中で核になって、孫弟子を大きな規模で作っていくということが今国会で審議中である。５年間３０名はまだまだ我々としたら難しい規模だが、全国の大学と組めばできるのではないか、と考えている。来年度から５年間３０名以上の大学教員、しかもそれは現在の経済学、医学、工学分野で学位を取られている研究者の方々に少なくともそういう共同研究ができる人材にしようということで、走りながら考えているというのが実際のところである。

3.15 科学・工学・医学における数理

［概要］

CRDSの連続セミナーシリーズ、第15回は「科学・工学・医学における数理」をテーマとして、水藤氏より「臨床医学と数理科学の協働」、中川氏より「東京大学大学院数理科学研究科 / 日本製鉄 社会連携講座『データサイエンスにおける数学イノベーション』が目指すもの」というタイトルでご講演いただいた。

水藤氏の講演は自身の研究プロジェクト活動を踏まえた、臨床医学と数理科学の協働の社会的重要性と困難さの紹介より始まる。氏のチームにおける「血流現象」、「画像診断」、「数学的基盤」の研究成果が紹介され、医療特有の問題である「患者固有の問題」に対処するための研究課題のシフトの流れの紹介と続く。上記の研究は、数学を通して「熟練医の技」を「客観的に」特徴付ける試みと位置付けられる。これを踏まえ、続く氏の研究プロジェクトではAI援用で「熟練者の技」を学習、モデル化する枠組みを構築し、現場の負担の軽減を目指す。異分野間の意思疎通には「通訳」の存在が欠かせず、協働の原点を常に意識しながら研究開発を続けなければ異分野協働は実現し得ない事が氏の講演の締め括りのメッセージとなった。

中川氏の講演は、「普遍性」と「一を知り、全体を知る」事が可能となる数学の強みを活かした数学イノベーションを実現する取り組みの紹介で構成される。氏が携わる『東京大学 / 日本製鉄 社会連携講座』では数学イノベーション、及び産業の問題解決を図る「学際連携の場」の創造を目標とした活動が行われている。実例として「温度分布の外挿決定問題」「逆問題AIの技術」「鉄道の状態監視技術の高度化」、さらに学生など若手数学研究者向けのStudy Group Workshopにて提示された課題として「結晶格子図形の表現方法」の教育研究の取り組みが挙げられた。これらの取り組みを通して、数学イノベーションの精神を体現するための本質的な姿勢が説かれる。最後に数学イノベーション実現のために必要な事、実現に向けた日本の強みを認識し、活かすことの重要性を解かれて氏の講演は締め括られた。

［Key point］

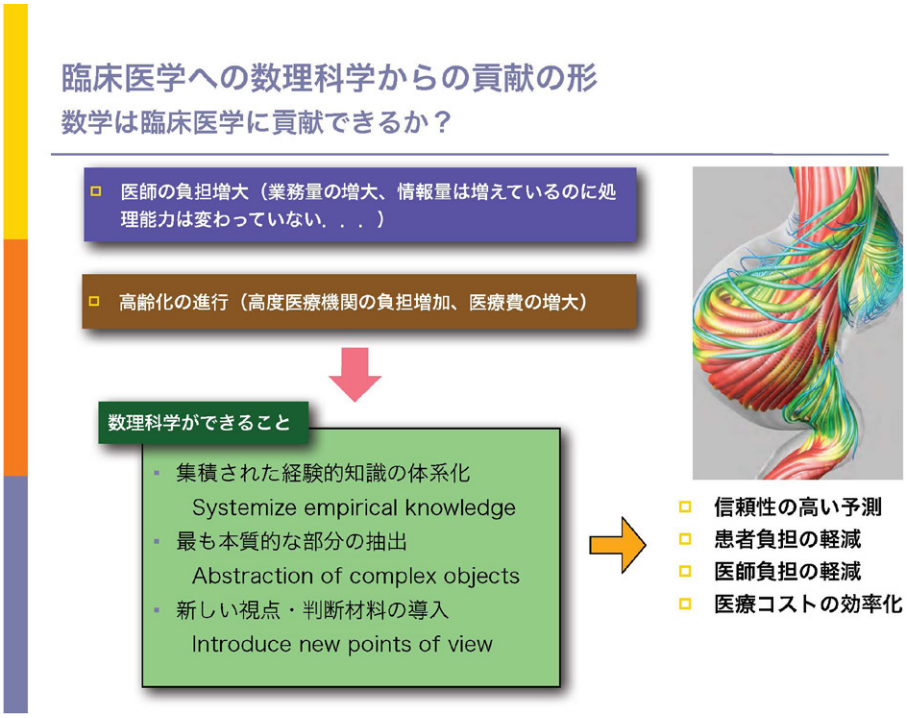
・異分野間の意思疎通には「通訳」の存在が欠かせない。医学と数学は大きな隔たりがあるように思われがちだが、従事者の思考回路が近い時はしばしばある事も注目すべき点である。「協働の原点」を常に意識しながら研究開発しなければ、異分野間の協働は実現し得ない。（水藤）
・個々の事象に依らない「普遍性」を持ち、事実の背後にある本質、理論体系を辿ることで「一を知り、全体を知る」事が可能となる点が数学の強み。「モデル駆動型」スタイルに内在するイノベーションにおける強みは、数学の強みとマッチする。数学サイドの研究者がニーズに向き合い、「最速で」納得を得る解決の提案をし続ける事、また「方法論をゼロベースから構築する」事が数学イノベーションには必要。日本においては「事の本質」を深く考える現場の技術文化を育んできた企業、数学者のレベルの高さという2つの優位性があり、これらを認識し、活かすべきである。（中川）

［水藤氏講演：詳細］

JSTがサポートする様々なプログラムを通して、水藤氏は数理科学と臨床医学との協働を様々に展開してきた。きっかけはさきがけにおける上記2分野の協働の提案である。その後CRESTにてその展開として放射線医学との協働による臨床診断の実現に向けた活動を行い、継続して関連する数理モデリング手法のさらなる発展も盛り込んだプロジェクトが現在稼働中である（2021年度まで）。

さらに未来社会創造事業で「熟練者の技」を解明するプロジェクトとして、一連の活動にて得られた知見を医学の外の領域に応用する動きや、今年度よりムーンショットプログラムにおいても数理モデルを基礎とした構造性の理解のプロジェクトも発足させている。

水藤氏の上記における一連の活動は、「数学・数理科学の臨床医学への貢献」という課題に基づく。業務



量の増大や、近年のビッグデータ社会化などによる情報量の増大に対する人間の処理能力の限界は、医療現場における医師やスタッフの負担の増大につながっている。また高齢化の進行により、高度医療機関の負担増大、医療費の増大が深刻化している。この社会的問題に対して、数学が臨床医学への貢献という立場から何ができるか？水藤氏は数理科学ができる事として「集積された『経験的知識』の体系化」「本質の抽出」「新しい視点及び『判断材料』の導入」を挙げ、病気などに対する信頼性の高い予測、医療コストの効率化とそれに伴う患者、医師双方の負担軽減につながると説く。数理科学と臨床医学の協働の実現にはフィードバックの積み重ねが必須であり、氏の研究グループはそれを実行するための「研究者ネットワーク」を構築している。数学・数理科学、データ解析、機械工学、内科、外科などその範囲は多岐にわたっている。数学と医学は非常に遠い立ち位置にあり、知識共有のためのコミュニケーションを取る事自体容易ではないが、氏が携わるネットワークでは両者の間に放射線科医が入り、翻訳家の役割を担うことで意思疎通を図っている。

この研究者ネットワークの紹介の後、主にCRESTにおいて展開された研究テーマの紹介があった。主なものとして「血流現象」「画像診断」、臨床現場にて適用される数理に対する「数学的基盤」がある。血流現象の解析は、例として高齢者に多い大動脈瘤、大動脈解離といった疾患の一般的特徴を同定し、多数の患者に共通する疾患の特徴を同定する課題が挙げられた。

血流の異常に伴う臓器への負荷は、血流が持つエネルギーを評価する事で測る事ができる。氏のCREST研究チーム（主たる共同研究者：植田琢也・東北大学、齊藤宣一・東京大学、滝沢研二・早稲田大学、増谷佳孝・広島市立大学）は数理科学の知見により血流の数値シミュレーションを実現させ、可視化により病態を議論する土壌を作った。これにより知識を蓄積し、予後予測に繋がる事が期待される。一方で血管や血液の物性など「測れない、あるいは測定が難しいもの」も非常に多く、数値シミュレーションの結果の妥当性にも議論の余地が残る。そこで氏のチームでは同時に血管壁の構造力学的解析や、上記を含めた課題に適用するための数理モデルに対するIGA（Iso-Geometric Analysis）法などの数学的基盤確立に向けた研究も同時に展開している。

上記は一般的な病態に対する研究であるが、一方で患者固有の問題も重要であり、病態の一般的特徴と同時に、個別の疾患に対する血流特性の関係を見出し、術後形状を人工的に（数値的に）作る事で、数値シミュレーションなどを通した手術計画検討につなげるという研究課題のシフトがあった事も紹介された。

他の課題として、画像から得られる気管支の構造を特徴づける研究が紹介された。画像から病態を判断するには熟練医の経験が必須となっているが、これを数理科学の言葉に翻訳する事で、診断に簡潔かつ「客観的」な指標を与える事が期待される。

上記の取り組みにおいて、数理科学の貢献が期待される成果の多くの部分は「熟練医の経験による診断の客観的判定」と関連している。この点に焦点を当て、未来社会創造事業において「熟練者の技」を解き明かす研究が開始された。人工透析や手術後の状態管理などが主な対象となっている。専門家の判断を学習したAIによる誤答の傾向を調べ、「学習阻害因子」を見つける。その原因を詳細に調べることで、「熟練者の視点」を見出す事ができるという考え方である。[29]

講演のまとめとして、臨床医学と数理科学の協働における障壁と共通点が挙げられた。異分野間の意思疎通には「通訳」の存在が欠かせない（翻訳するだけではダメである）が、双方の学識のすり合わせも醍醐味の一つとなっている。また医学と数学は大きな隔たりがあるように思われがちだが、従事者の思考回路が近い時はしばしばある事も注目すべき点である。しかし協働の原点はゴールに基づいた要求や疑問[30]を常に意識するべきという姿勢であり、これ無くして協働は実現し得ない事を強調されて、水藤氏の講演は締め括られた。

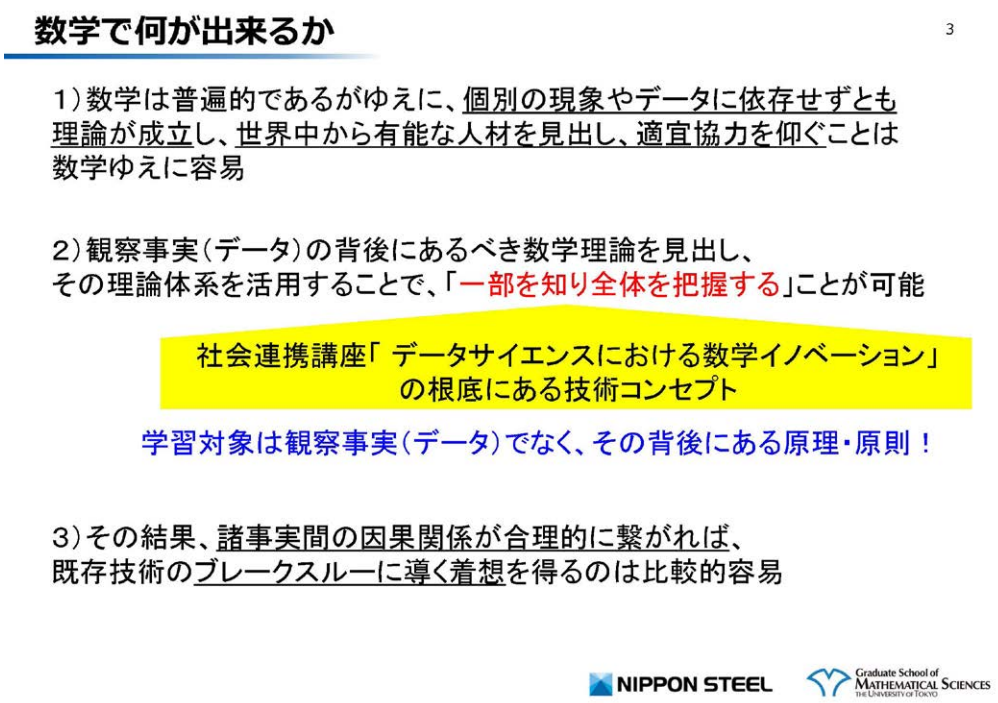
補足：講演においては、発足されたばかり、あるいは構想段階のプロジェクトも追加で紹介されている。

［中川氏講演：詳細］

中川氏の講演は（社会的課題の解決に向けて）「数学で何ができるか」という基礎的な問いに対する氏の考えを提唱されるところから始まる。氏は個々の事象に依らない「普遍性」を持ち、事実の背後にある本質、理論体系を辿ることで「一を知り、全体を知る」事が可能となる点が数学の強みであり、この強みが根幹に

29　松江's コメント. 医療従事者の技だけでなく、人口が減少している職人の技の継承にも貢献できると考えられる

30　医学との協働の場合は、ゴールは実際の医療に活かすこと。それを明確に示すのがclinical questionとなる



あるおかげで産業応用をはじめとした数学イノベーションを実現できると強調する。

ここで言及される「数学イノベーション」[31]とは、イノベーションの方法として数学を用いる動きを指し、「モデル駆動型」の研究開発スタイルであり、事の本質を深く考え、モデルを作り、データを用いてその適合度を検証するという特徴を持つ。また（数学においては常套手段である）最初に定義と計画を行うモデル駆動型は不連続、予知不可能な現象に関わる活動である「イノベーション」において強みが発揮されるスタイルである。データ駆動型スタイルとは補完関係にあり、両輪として活用されるべきであると氏は説く。特に氏が携わる『東京大学 / 日本製鉄 社会連携講座』は、前記「数学イノベーション」であり、自然科学の原理を「公理系」に、工学の先験情報を「モデル化」し、産業の問題解決を図る「学際連携の場」の創造を目標として展開される。主な具体的活動として、1. 企業との共同研究、2. 若手数学者育成のための共同研究がある。学際研究が主であり、対象は材料科学と数学・数理科学、経済学と数学・数理科学、数学の異なる専門分野間の研究など多岐にわたる。

講演では数学イノベーションの精神を体現した（社会連携講座における活動を含む）いくつかの事例が紹介された。1つ目は電気炉における温度分布の外挿決定問題。温度計で観測できるデータから、温度測定が技術上困難な稼働面（溶鋼）の温度を決定するというものである。氏の研究グループは観測データと背後の「原理」、数学理論を併用し、逆問題の枠組みでデータの外挿を実現し[32]、検証データとの一致を通して妥当性を提示、実際に製鉄所にて実用化された。実用化例として算出された熱流束による電気炉への投入電力量の制御による耐火物の溶損防止が挙げられた。この成果はコスト節約に役立つだけでなく、数学理論の適用により、要求される以上の情報を得る事ができ、製造における稼働系の原理という付加価値を与えた例となって

31　グレーヴァ - 三橋 共著「数学イノベーション：モデル駆動型イノベーションの特徴と事例研究から見る日本企業の可能性」、一橋ビジネスレビュー 2020, 68-1, p. 81-95 を参照している

32　背後の原理とは「非定常熱伝導の物理」。数学理論は熱伝導を記述する枠組みである「放物型偏微分方程式論」。外挿の方法として、与えられた地点での温度変化（放物型偏微分方程式の解）から、初期値と未知の境界条件を同時に求める偏微分方程式のコーシー問題に対する逆問題の理論を適用している

いる。

2つ目は逆問題AIの技術について。材料のミクロ組織構造など「原因」を記述するミクロな系のデータ集合と、材料性能など「結果」を記述するマクロな系のデータ集合を与え、これらの間の因果関係をモデル化する[33]という構造材料の最適設計が事例として挙げられた。機械学習などによる相関関係の定量化と異なり、因果関係を数式として表す事で事象の背後にある本質を普遍的公理系として記述、知識体系を蓄積していく。ミクロ組織構造と材料特性の背後にある物理の原理・原則の数学的定式化により、非平衡状態遷移及びマルチスケール構造を繋ぐ数学モデルの構築に多方面に亘る数学理論を総動員して取り組んでいる。

3つ目は鉄道の状態監視技術の高度化。鉄道車両における「モノリンク力」[34]などの計測データを状態変数として組み込んだ鉄道車両モデルを構築し、計測困難な他の状態変数をフィルタ技術を通して導出し新しい特徴量を創出、系における既存の特徴量、ここでは軌道不正量の高精度な予測を実現する。

4つ目は軸受疵の予兆検出技術が紹介された。固有値解析に基づいた本質的なシグナルの選択により修正モデルを提唱し、疵の予兆、微小な疵を生み出す固有振動数の正確な検出、異常の予兆検出に成功している。

最後に、Study Group Workshop（SGW）[35]における人材教育も視野に入れた物質・材料の未解決問題とその解決の紹介があった。現実世界において、物質は結晶などの構造を一般に有し、その中に含まれる格子欠陥や粒界などの構造より、強度や塑性的性質などの物性が決まる。原子配置とそれに対応する物質構造や物性の一般的記述は、「未知の材料」の構造と対応する物性を特定する一般的枠組みを与える第一歩となる。特に氏が提示する枠組みは結晶格子と結晶群、格子欠陥とアファイン多様体など、結晶構造と物性の代数学的・幾何学的側面にフォーカスする。具体例として、結晶格子図形の表現方法に関する取り組みが紹介された。これはSGWで提示され、参加学生やポスドクが氏のグループと共同で議論されたものである。各原子に対して、（最近接、第二近接など）隣接する原子数に着目し、それをgrowthとして定義した。これは材料科学で使われる配位数の数学的普遍化として提唱されたものである。様々な種類の格子構造に対しgrowthを定め、結合エネルギーの他の記述子による回帰分析により、growthの材料特性決定因子としての統計的な重要性を調べるなど、実用的な側面を蓄積する。さらに第n-growth[36]の表示を総当たりで調べる事で、「nをパラメーターとするgrowthの列は準多項式で記述される」という予想が提示された。この記述は純粋に代数的な枠組みであり、材料科学など産業の観点から提起された問題が数学の問題として認識、昇華されたエポックメイキングな結果となっている。後年、この予想は代数学を中心とした数学を用いて厳密に証明され、長年にわたる結晶学の予想とされていた問題が数学的に厳密な意味で肯定的に解決された。この結果は権威ある材料科学系の論文誌において非常に高い評価を受け、採択されている[37]。本課題は（学生、若手含む）数学者の妥協無き探究が材料科学の問題として提示された話題を数学の問題として昇華させ、普遍的な結果を導いた数学イノベーションの成功例の1つであると氏は強調する。またデータサイエンス分野に対する数学者の課題への向き合い方として、「数学者の専門性が活かせるようにデータの定義を変える」事の重要性が説かれ

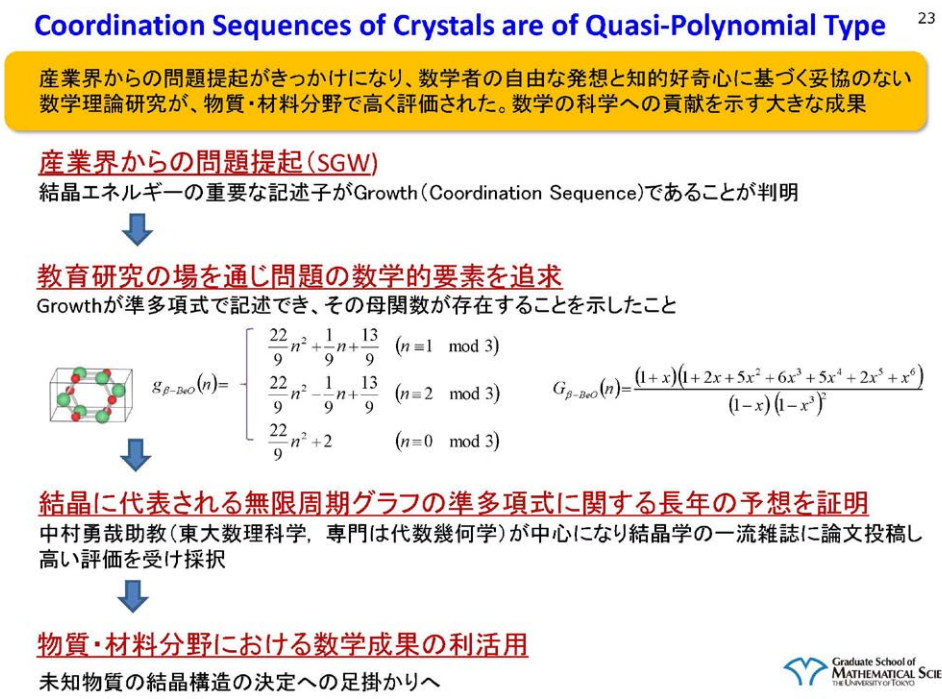
33 ミクロ組織の集合の間の熱履歴を伴う時間変化の対応（状態遷移則）、マクロ領域の構造から決まる材料の物性と内部のミクロ構造の関係を記述する因果関係（報酬則）をモデルにする問題が具体例として挙げられた。状態遷移則のモデルはランダムウォークで、報酬即における最終的なマルチスケール構造モデルの定式化は、エネルギー汎関数の変分問題として特徴付けられた

34 車両の台車、輪軸のヨーイング量（上下を軸にする回転運動量）の差から生じる走行方向の応力

35 九州大学マス・フォア・インダストリ研究所、九州大学大学院数理学府、理学研究院及び東京大学大学院数理科学研究科が主催している短期集中・課題解決型研究集会。1968年、英国のオックスフォード大学で開催された「Study Group」を源流とし、ヨーロッパだけでも500以上の課題が提示され、具体的解決に結びついたものも数多くある。九大・東大主催のSGWは2010年度から継続的に開催されている。近年では大阪大学、名古屋大学でも開催されている

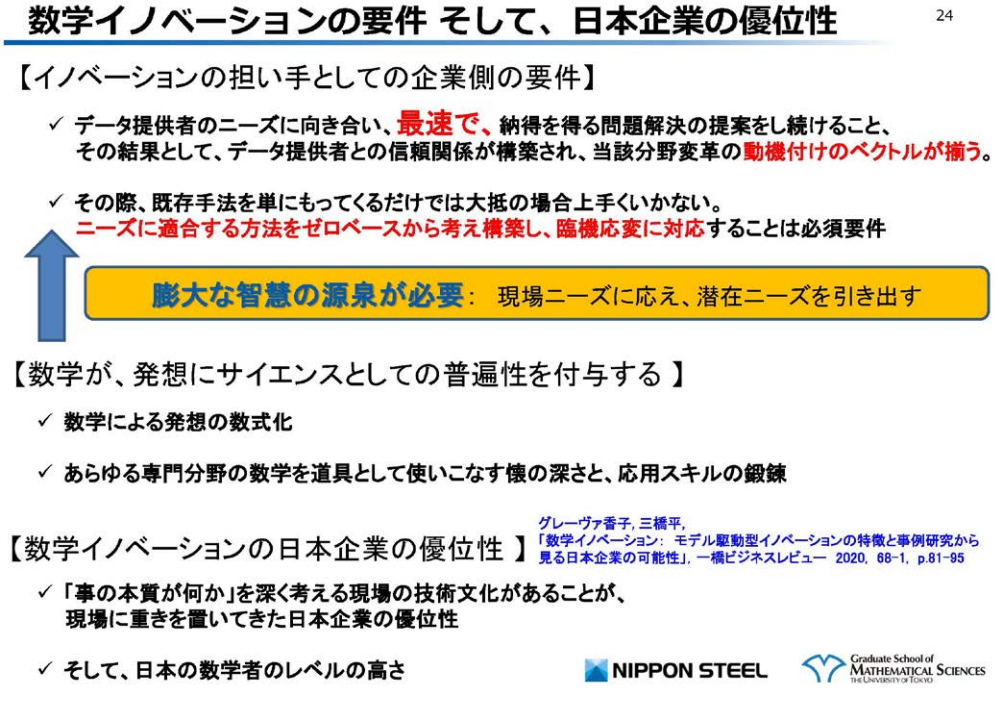
36 第n近接原子数により定義される

37 Y. Nakamura, R. Sakamoto, T. Mase, J. Nakagawa, Coordination sequences of crystals are of quasi-polynomial type, Acta Cryst.（2021）. A77, 138-148



た。実験研究者による結果の一連の考察もまたデータであり、これらも生かした全く新しい切り口からの斬新な発想がデータサイエンスの理論を一気に進化させることを可能とし、それがデータサイエンスにおける数学イノベーションであるとされる。本課題はそれを体現するものとなっている。

講演のまとめとして、「数学イノベーション」を起こすための必要条件が提示された。数学サイドの研究者が現場ニーズに向き合い、「最速で」納得を得る解決の提案をし続ける事、また単なる既知の（数学）理論の利活用に限らず、「方法論をゼロベースから構築する」事が肝要で、これらを満たして初めてイノベーションの担い手となり得る。数学が現場ニーズに応え、潜在ニーズを引き出すには「膨大な智慧」の源泉が必要で



あり、数学者及び高度な数学教育が受けた人材がその担い手となり得る。日本においては「事の本質」を深く考える現場の技術文化を育んできた企業、数学者のレベルの高さという２つの優位性があり、これらを認識し活かす事の重要性を強調される事で中川氏の講演は締めくくられた。

【質疑応答】

Q：（中川氏の経歴につき）数学に目をつけたきっかけは？

A：技術開発を行う過程で、「本質は何か？」を常に問いかけてきた。本質に関する仮説を立てて開発を行うプロセスが、数学の問題解決プロセスに非常に似通っていた（氏のアプローチは自己流との事）。転勤を機に、本格的に数学を取り入れ始めた。専攻が化学工学で、親近感があったのも大きいと考える。

Q：数学イノベーションに対する日本企業の優位性の１つに、「数学者のレベルの高さ」を挙げられていたが、数学を修めた学生（あるいはポスドク）が企業に就職すると、「企業の人材観」に個人の考えとズレがあると感じたという声を聞く。「企業の人材観」に関する中川氏の考えを伺いたい。

A：講演にて紹介した数学者のレベルの高さは、あくまでアカデミアでのもの。企業人として活動するには、「自分が数学者である」という考えを一旦捨てる必要があると思う。

（松江’s コメント：「自分だけで問題を練り、自分だけで解決していく」というスタイルを指すと思われる。）

企業での活動はアカデミアと比べると事が進むスピードが10倍以上速く、自分で何もかも実行していくのは不可能。課題となる「テーマ」を練り、解決のための「構想」を創造するプロセスが必要。「構想」具現化のために、アカデミアの数学者の智慧が必要になる。「数学の問題」を設定し、アカデミアの数学者達と連携し産学でＷｉｎ－Ｗｉｎの関係を構築できるのが理想である。数学と産業を繋ぐためにコーディネーター（調整役）が必要と謂われているが、必要なのは「構想」を創造し「数学の問題」設定ができる人と考える。

Q：水藤氏の研究者ネットワークでは、数学者と医師の間に「放射線科医」など、間に入る人がいた。中川氏の場合は本人が間に入る「通訳」的な役割も担ってきたと想像する。間に入るのはどのように実践された？

A：現場の人との対話、これに尽きる。企業の研究部門にいる人間は自分で動かなければいけない。式やアイデアなどは自分で組み立て、PDCAサイクルに載せて進展と修正を繰り返していく。現場の人間は、事の本質を定性的にとらえている（それを数式化し、課題解決に加え、科学的な付加価値を創出するのが数理サイドとしての仕事）。重要なのは課題解決の「動機付けのベクトルが揃う事」。（対話による実績を通して揃えていく事になるが）これが揃いさえすれば、数学的な技術面は業務を通して理解できる。

Q：「現場から数学に新しい洞察」をもたらす動きは起こり得るか？

A：Growthの話がまさにそれである。数学者は、課題が「数学の問題である」と「認識」したら勝手に突き進んでいく。

Q：数学は現場にどれくらい恩恵をもたらすだろうか？

A：現場が数学の結果を取り入れるか、そのポテンシャルを、具体的な成果を通じて採用してもらえるか次第。

（松江’s コメント：工学でも医学でもこの流れは共通している印象を受ける。）

数学的考え方が入ると、問題そのものが非常にクリアになる。その結果、問題を数式でシンプルに表現できるので、問題解決のスピードが劇的に向上し、現場と数学的成果を共有できる。AI適用の潮流があるが、単にAIに問題や課題解決を投げると、単なるAIのスキルの勝負になり、人間が「物を考えなくなる」。これは将来的に、日本企業の強みである現場自らが事の本質を考える技術文化の破壊につながるのではないかと危惧する。

Q：AI関連の質問。医療画像の解析でも病気を発見するなど成果が多く報道されているが、成功の裏には多くのうまくいかなかった例もあると考える。この例などに数学は貢献できるだろうか？

A：AIの定義によるだろう。先の「うまくいかなかった例」の多くはデータ駆動型の課題である。（紹介があった）軸受疵の予兆検出の話題は、データを扱ってはいるが基礎は数式で表現されたものであり、数学的エッセンスでデータを表現している（モデル駆動型と言える）。

Q：（関連して）予兆検出成果の発想はどのように？

A：現場との対話が大きい。異常箇所とセンサ間に多数の媒体が関与しているので物理モデルだけではどうしようもなかった。信号波形を見て、それを数学的に取り扱うことで特徴を抽出した結果が、最も現場の納得感があった（実空間の信号でなく、修正ＡＲ係数の空間で現象を観察するという発想）。（中川）

適切な「記述子」が入る、もしくはこちら側から入れる事で、コトの理解が進む。この「記述子」の導出こそ、数学が貢献できる大きな点だと考える。（水藤）

Q：SGWなどで得られた成果の知財への繋がりは？

A：（中川氏の個人的考えであるが）SGWなどの公に出す課題は、知財に繋がることのない学術的な問題であることを意識して出している。

Q：（上の質問に関連して）知財になりうる思いもよらない成果が出た時は？

A：最初の問題の出し方次第、これに尽きる。ここを間違えると公と知財の間でややこしい問題が生じる。スタート地点をうまく設定する事が大事。

3.16 トポロジカルデータ解析：数学的発展と諸科学への応用

［概要］

CRDSの連続セミナーシリーズ、第16回は「トポロジカルデータ解析：数学的発展と諸科学への応用」をテーマとして、平岡氏より「トポロジカルデータ解析の共通基盤化への試み」、平田氏より「アモルファス構造の隠れた秩序 - ホモロジー解析による構造抽出 -」というタイトルでご講演いただいた（以下、簡単のためにトポロジカルデータ解析をTDAと略記する。Topological Data Analysisの略）。

平岡氏の講演ではTDA、特にその一つであるパーシステントホモロジーの概略、その有用性と共通基盤化に向けた動きを氏の研究グループにおける研究成果、主に材料科学分野への応用に基づいたものを通して最初に紹介された。のちに社会還元の一つの形としてのソフトウェア開発、パーシステントホモロジー以外のTDAによる応用研究を通して、世界拠点、ネットワーク構築によるTDAの共通基盤化に向けた活発な動きが紹介された。

平田氏の講演ではガラスに代表されるアモルファス構造の解析におけるTDAの有用性を、自身の研究グループにおける研究成果と共に紹介された。パーシステントホモロジーを中心としたTDAは、一見ランダムとされているアモルファス構造に隠された秩序の同定に大きく貢献している。さらに近年開設された氏の現在の所属先では、大学院教育において材料科学と応用数学のどちらも組み込んだカリキュラムが組まれており、数学・数理科学と諸科学分野の協働を実践する若手人材育成に大きく貢献することが期待される。

［Key point］

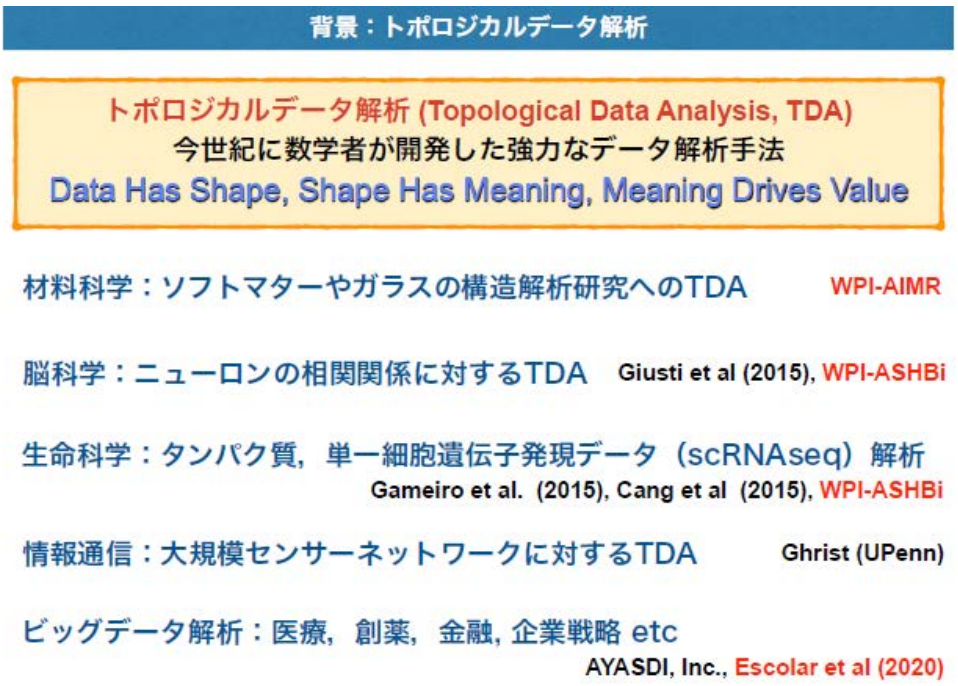
・パーシステントホモロジーは、データの形、特に穴の数、サイズ、形を記述し、さらにデータのマルチスケール構造を捉える。 TDAには数学・数理科学だけでなく、諸科学分野、産業界からも強い期待があり、予想外の応用分野への展開の可能性を秘めている。（平岡）
・アモルファス構造の解析において、ひとつの測定実験、モデリング、解析手法ではある側面しか見ることができない。様々な解析手法を併用して相補的な情報を集めることで、構造の解明は進む。（平田）

［平岡氏講演：詳細］

インターネット・計算機・計測技術の発達による膨大なデータの蓄積、及び数理・データ科学・AIの発展に伴うデータの潜在的価値抽出のニーズが高まっている。価値の一つの表現として「データが持つ形」に着目し、それをクリアに抽出する数学的記述子及びその総称としての「トポロジカルデータ解析（Topological Data Analysis, TDA）」が近年爆発的に研究、応用されている。平岡氏の講演はその一つとして特に注目を浴びている「パーシステントホモロジー」を中心に、その有用性と成果例、分野としての基盤形成に関する一連の活動の紹介で構成される。

TDAは、21世紀に数学者が開発した強力なデータ解析手法及びそれを研究する分野である。データ解析の一分野として位置付けられているものの、氏が本講演で述べるTDAという分野は、トポロジーの視点からデータの解析を実現する「数学言語の開発」に主軸を置く。

21世紀に生まれた概念であるにもかかわらずその数学理論を始め、材料科学、脳科学、生命科学、情報通信や産業応用など、TDAの発展はあまりにも凄まじく、一大分野を築くまでに至る。その手法の一つに「パーシステントホモロジー」がある。これは図形に存在する穴の情報を代数的に表現する「ホモロジー」の発展版で、データに含まれる「穴」をその数だけでなく、サイズ、形状のマルチスケール構造の抽出を実現する。典型的には点データに対して上記の情報を抽出するが、数学的には「点の集まり」であるデータ点群



に穴は存在しない[38]。しかし、仮想的にデータ点を中心とした球を設置し、その球の半径を変動させる事で、球の和集合の穴の生成と消滅を観察できる。データ点群が作る球の和集合における全ての穴[39]の生成・消滅の情報を、対応する球の半径から抽出、集積する事で、データに内在する穴のマルチスケール構造を厳密に表現かつ可視化できる。この構造の可視化ツールは「パーシステント図」と呼ばれ、データの形の一つの記述子となる[40]。現在、与えられたパーシステント図からその構成要素となる情報が持つデータ構造を抽出する逆解析も研究が進んでいる。

平岡氏はパーシステントホモロジーの（逆解析も含めた）数学的側面の精力的な研究もさることながら、応用面で多くの成果を挙げており、TDAの理論、応用研究の先駆け的な存在となっている。特に氏のグループの主な研究成果として、材料科学におけるTDAが紹介された[41]。1つはガラス、1つは高分子、1つは粉体の構造解析。いずれも多数の原子、分子及び粒子から構成される物質で、隠れた構造を抽出するのは平均量をとるなどの1つのスケールで物を見たときに限られるのが一般的である。しかしパーシステント図は先も述べたようにマルチスケールスケール構造を同時に抽出、可視化する。ガラスに対する適用例は、平田氏の講演を参照されたい。一連の研究成果はSIP, MI²I, NEDO, CRESTなどの研究プロジェクトにつながっていくが、同時にTDAの共通基盤化の転機となる動きが生じる。それが「パーシステント図の機械学習法」と「ソフトウェア」開発である。

パーシステント図は点データ群が持つ形を抽出するが、それ自身が一つのデータとしての側面を持つ。（多数の）パーシステント図から統計的性質を抽出し、材料物性などデータの背後にある物理などの関連を学習

38 データ点が密集して穴っぽく見えても、各点が厳密には離れており、数学的にデータ点群は離散点の和集合でしかない。当然、点に穴は空いていないので、穴は無い。この事より、従来のホモロジーでは（点の集まりという以上の）データの「穴」や「形」を特徴付ける事はできない

39 各次元のホモロジー類

40 パーシステント図とデータの形の構造の数学的対応は現在膨大な議論と研究結果がある。「パーシステント図からデータの形を読み取る」応用の妥当性は、それらの蓄積により保証される

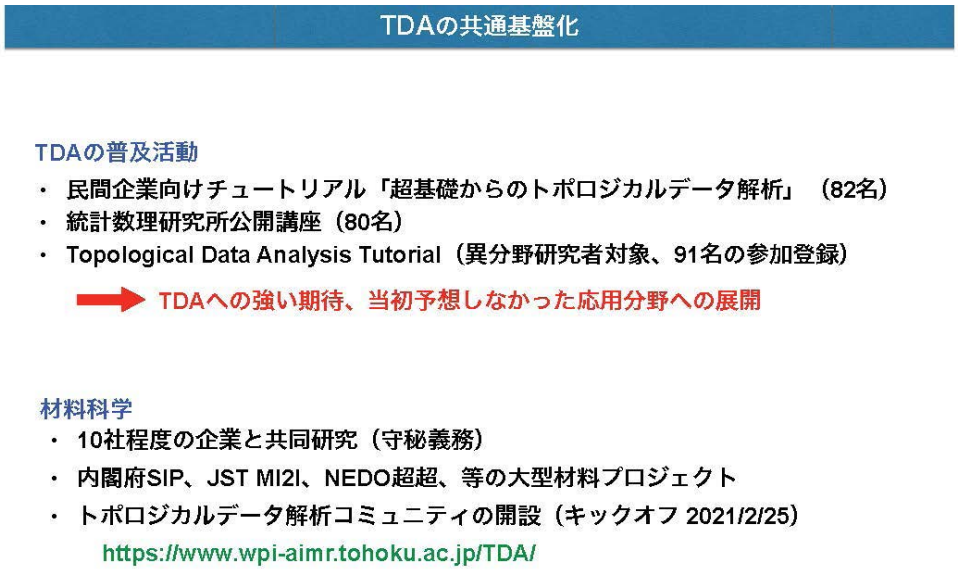
41 東北大学AIMRにおける研究に基づく。ここでの「チーム形成の実現」が成果に大きく結び付いたと平岡氏は述懐する

させて、材料開発などに貢献する機運が高まってくる。事実、パーシステント図から再生核ヒルベルト空間を形成するカーネルの構成や、パーシステント図に対する機械学習とスパース解析と逆解析を組み合わせ、データが元々有していた幾何構造や背後の物性の学習をさせる試みが成果として上がっている。これらはパーシステントホモロジーの（確率論、表現論、代数幾何学的側面と同様に）数学的側面を追求した研究である。

企業の共同研究など応用のニーズが増える一方、パーシステントホモロジーやパーシステント図の背景や計算には高度な数学的知識を要求される。そこで幅広い層の利用者に応用してもらえるよう、数学的予備知識を必要としない汎用ソフトウェア開発の機運も高まる。氏の研究グループでは、様々なOSに対応し、高機能のGUI、サポートを充実させたソフトウェア「HomCloud」を開発した（リーダー：大林一平氏）。数学理論と応用の一つの社会還元が結実した物である。

TDAの応用は材料構造解析に止まらない。最新の高度実験技術により実現された「単一細胞遺伝子発現データ（scRNA-seq）」の解析へのTDAの応用がその一つである。scRNA-seqは各細胞の遺伝子発現を表す超高次元データであり、最先端の計測技術は1細胞を識別するまで高度化している。しかし、これらの解析手法としての数学は1細胞を識別するための理論が確立されていないのが現状であった。氏の研究グループではMapperと呼ばれる（パーシステントホモロジーとは別の）TDAの概念と従来の数学理論を総動員して新たな解析手法を開発、始原生殖細胞や免疫系T細胞における新しい分化経路を発見した。超高次元のデータを扱いつつ、その中にあるマルチスケール幾何構造の抽出を可能にした画期的な結果である。

このように、TDAは数学理論、応用研究、社会還元全ての面で大きく飛躍を見せており、その潮流は世界で波及している。その中で、平岡氏は現在国内研究者の新規参入の促進、材料科学分野におけるTDAネットワークの確立、関連分野の国際論文誌の編集委員など、TDAの世界拠点とネットワークの形成に尽力している。代表的なものとしてアジア太平洋地域を中心としたオンラインセミナーシリーズ[42]、民間企業向けのチュートリアルや公開講座、TDAコミュニティの開設がある。これらの普及活動は多方面からのTDAへの強い期待が垣間見え、研究当初予期しなかった応用分野への展開がなされている。企業との共同研究や大型プロジェクト、ソフトウェア開発、学会におけるTDAの研究部会発足なども含めて、TDAの共通基盤化に向けた様々な活動の展開を提示、TDAのさらなる発展を期待させて氏の講演は締めくくられた。



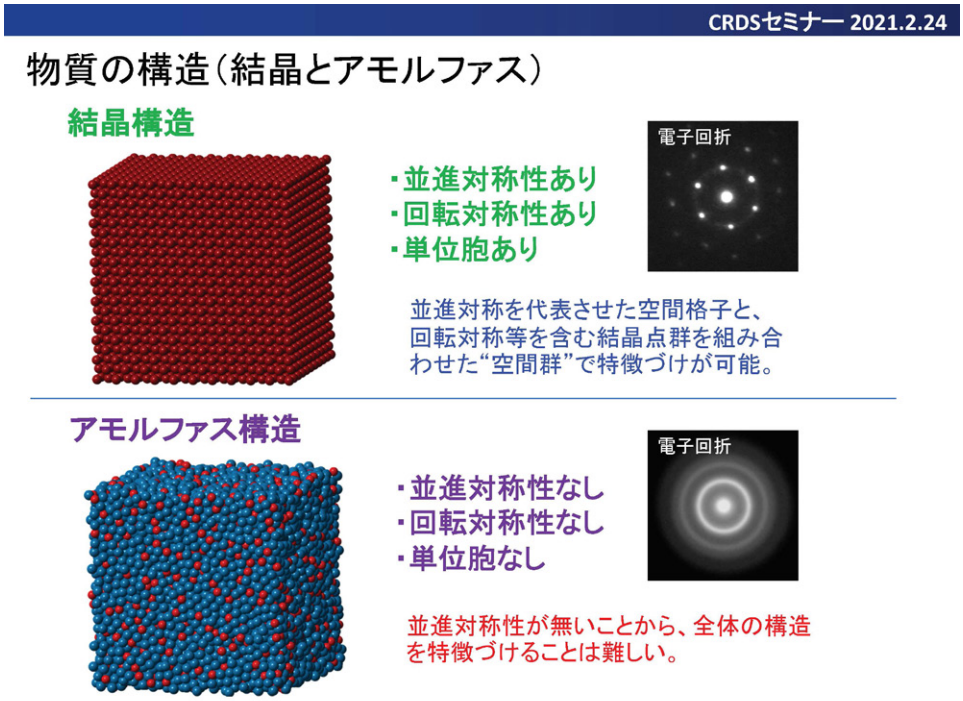
42 Asia Pacific Seminar on Applied Topology and Geometry（APATG）

［平田氏講演：詳細］

物質の構造を知ることは材料理解の基本である。平田氏が中心的研究課題に据えている「アモルファス構造」[43]は、結晶構造と違い並進対称性、回転対称性、周期性を特徴付ける単位胞（構造ユニット）を持たず、一見「ランダム」としか言えない原子配置をしており、局所構造と大域構造のつながりを記述する事が非常に難しい。アモルファス構造を理解するには、原子配置の「観察」から始める[44]。観測データから構造モデルを作成し、詳細を適切な数理解析手法で考察するのが典型的であるが、様々にある観測データから得られる構造モデルは「原子座標の羅列」に過ぎず、それは情報の得方で様々に変わる。また解析手法も、構造のどの特性に着目するかで座標の羅列から得られる情報が様々に変わる。情報取得、構造解析の方法論を氏は総称して「虫眼鏡」と例える。一見ランダムなアモルファス構造を理解するには、ある側面しか見出す事のできない1つの虫眼鏡を用いるだけでは不十分で、様々な虫眼鏡で見て相補的に情報を取得する事が重要になると氏は強調する。

観測データだけでなくアモルファス構造の「モデリング」の試みも多くある。歴史的に有名なものは1960年頃に作製されたハンドメイドモデルがあり、その後MD（分子動力学）を用いた多数の原子からなるシミュレーションモデル、その方法論が多く用いられるようになっている。しかし、得られるデータは実測と同様「原子座標の羅列」でしかないため、アモルファス構造を理解するための虫眼鏡の用意は依然として重要な課題として残る。

構造解析手法に関して、有名なものがいくつか紹介された。1つは「2体分布関数」。これは各原子に対して、平均密度（原子数）からのズレを、与えられた原子を中心とした半径の関数として表現するものである。この関数は構造モデルからも、回折などの実験で得られた実測データからも推定が可能であり、MDなどの構造モデルの妥当性を判断できる点が大きな利点となっている。しかし、あくまで2体の原子対のずれがデータとして得られるのみで、3体以上の多体系がなす構造を詳細に抽出するのは難しい。次に紹介されたのは「ボ



43　ガラスを形成する構造が典型的である

44　観測手法としては、X線回折、中性子回折、電子回折、EXAFS、NMRなどがある

ロノイ多面体解析」。局所構造をボロノイ図形による多面体構造で表現し、その構成要素[45]をボロノイ指数としてラベリングし、アモルファス構造全体（図形による全構造のパッキング）を記述する。これによりアモルファス構造を局所構造の族として記述し、その違いで分類が可能となる。特に、原子クラスターとして代表的なものを抽出する事で、アモルファス構造の「短距離秩序」を記述する事ができる。一方、ボロノイ多面体解析では「多様な局所構造がある」点が「多角形の数のみ」で大きく強調され、その構造が一般に多岐にわたる事より「秩序」の記述の適切さには議論の余地がある。他にも、ボロノイ指数による記述では各多角形の「対称性、歪み具合」は全く無視される。アモルファス構造は局所的な原子配置の歪み具合が本質であるという議論もあり、歪みを無視した「秩序」の表現の妥当性という観点からも、「1つの虫眼鏡で見る事の欠点」が強調される。ここで、氏は歪みに代表される「構造の不均一性、不完全性」を不変量として記述する概念として、トポロジー、特にホモロジーの適用に着目した。

最初のステップとして、観測、モデリングいずれかの方法で原子配置データを得て、これらを中心とした原子の「半径」を仮想的に変化させて、連結成分の数の変化を見る考察を行った。原子の球たちからなる集合に対し、対称的な多角形配置の場合はある半径での連結成分の数が急激に減少する。対して、対称性の低い原子配置に対しては、原子球の半径が増大するにつれて球のつながりが徐々に生成され、連結成分の数はなだらかに減少する。この連結成分の減少の仕方により、原子配置の「歪み具合」を表現する[46]。この手法の利点は原子数、また最近接の原子の対応に依らず、原子配置のみから半径の増大というシンプルなやり方で実現できる点にある。局所的な構造から遠い原子との（中間にある原子を介した）繋がり方も合わせた大域的構造も、歪みという観点から同時に得られる。実際に金属ガラスに対する原子配置のボロノイ多面体や観測で得られたクラスターを例として上記手法を適用すると、連結成分の原子半径[47]に対するなだらかな変化が観察された。特に、多面体構造の「歪み」が連結成分の変化という記述子で明確に記述された。さらに、200原子からなる構造モデルを局所ボロノイ多面体ユニットの族として生成して手法を適用すると、連結成分の変化が一様にフィッティングされる事が見出された。これは同様の「歪み具合」を持つクラスターが繰り返し並ぶ事で無秩序と見えるアモルファス構造を形成している事を示唆しており、歪み具合としてアモルファス構造の短距離「秩序」を明確に抽出できた例となっている。

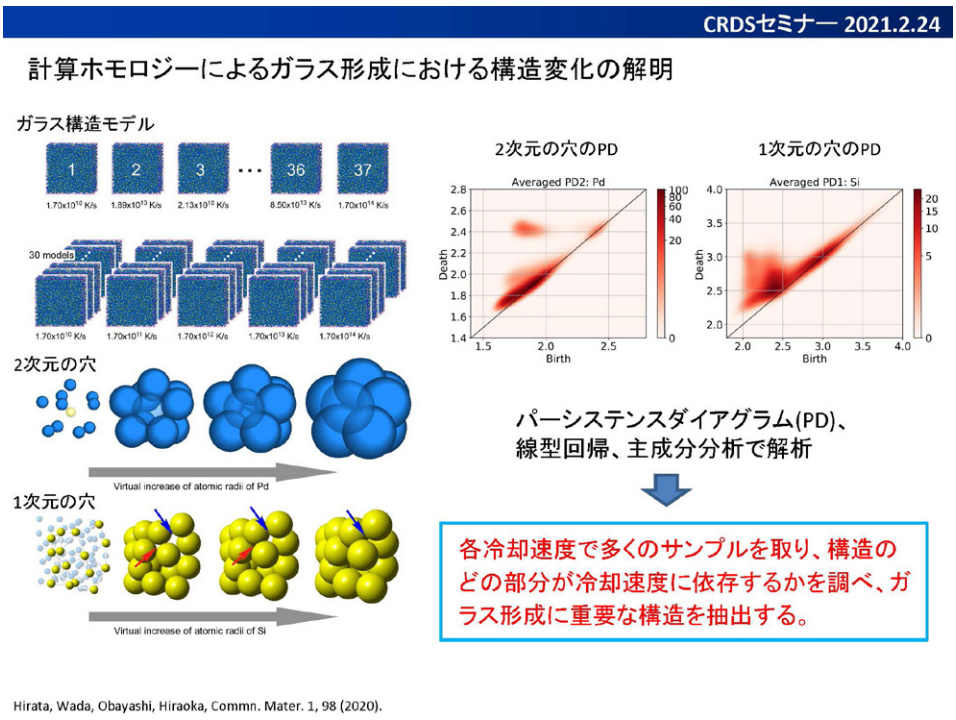
氏はこの知見をもとに、アモルファス構造の秩序の「階層」を抽出する研究に取り組む（先の平岡氏も参画する）。例としてアモルファスシリカの結晶、液体、ガラス構造が明確に分類される事が紹介された。例ではMDによるシリカの構造モデルをサンプルとして、ケイ素、酸素原子を原子球と見立ててパーシステントホモロジーを計算、パーシステント図を介して構造、特に「リング形成」機構に着目した。すると、酸素とケイ素原子の共有結合からなる大小のリング構造に加え、一部の酸素原子による非自明な拘束をもった分布が観察された。パーシステント図により、アモルファス構造内の最近接原子による短距離秩序だけでなく、酸素原子からなる中距離秩序の一端となる構造が同時に抽出される事が示され、より一般の中距離秩序、構造のマルチスケール性、特徴的な欠陥構造などの解明が期待される結果を得た。

上記のパーシステントホモロジーを用いた結果をさらに発展させ、氏の研究グループは「ガラス形成」に関連する構造変化の抽出に着手する。原子の族はなるべくエネルギーの低い安定構造、特に対称性の高い構造を取ろうとするが、ガラスのようなアモルファス構造の場合は対称性の高い構造を取ろうとする間もなく急冷させて、乱れた配置のまま原子が激しく動けないような状態を形成させる。「冷却速度」は異なるガラス状態

45 例えばアモルファス金属の場合は、1つのボロノイ多面体に対して3角形、4角形、5角形、6角形の数を数え上げ、1つの「記述子」とする

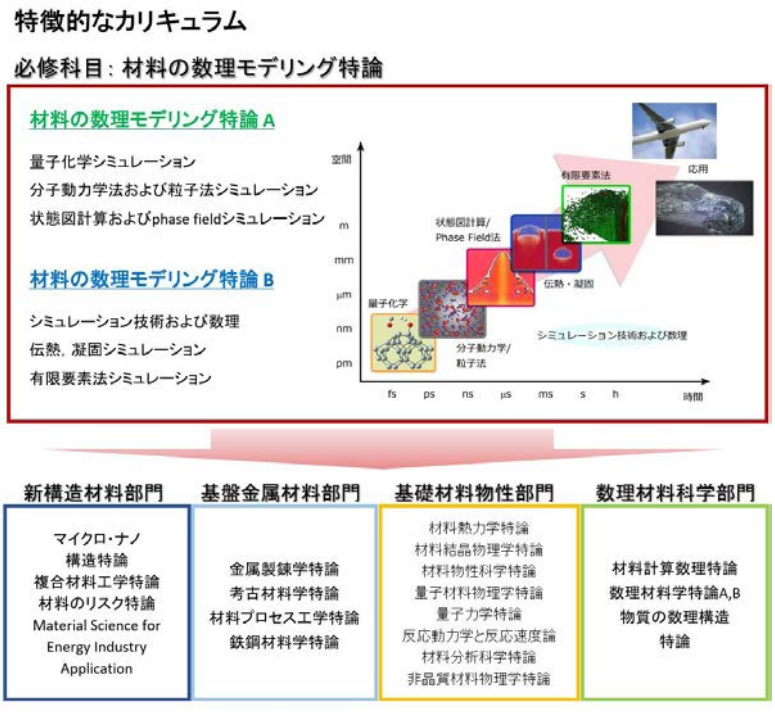
46 研究では0次ホモロジー群の半径依存性を調べる事で記述している。これがより詳細な構造解析における、パーシステントホモロジーの適用という発想につながる

47 異なる金属原子間の金属結合の特性を保つため、結合半径比を一定にする制約を課した正規化半径を導入し、実際の変動パラメーターとしている



Hirata, Wada, Obayashi, Hiraoka, Commn. Mater. 1, 98 (2020).

を形成する鍵となる因子であるが、この冷却速度に依存する局所構造が考察の対象である。先の例と同様、主なアイデアは原子配置データに対するパーシステント図の作成であるが、今回は各冷却速度で多くのサンプルを取り、パーシステント図の族に線形回帰、主成分分析を適用する事で冷却速度に最も依存する成分を統計的に見出す。これにより、パーシステント図に冷却速度に依存する明確な差異が確認された。この主成分の逆問題解析により、対応するパーシステントホモロジーを生成する構造を抽出し、局所構造の同定を行う事で、ガラス形成に関係した隠れた秩序を見出す事に成功した。このように統計的手法とトポロジカルなマルチスケール構造の抽出を可能にするパーシステントホモロジー、さらに近年研究が進んでいるパーシステント



ホモロジーの逆問題解析による原子配置の抽出技術の組み合わせが、一見ランダムとされるアモルファス構造の因子を抽出し、その解明に大きく貢献すると期待される。

平田氏の講演の締めくくりとして、氏の現在の所属先の紹介がなされた。早稲田大学にて2019年4月に開設された「理工学術院　基礎理工学研究科　材料科学専攻」は、材料科学をメインとしつつ応用数学の教員も構成員として参画しており、大学院教育より異分野、特に材料科学、機械工学と数学の協働の流れを生み出す事を目指す専攻である。材料の数理モデリングなど、学部時代に数学あるいは材料科学を学修した学生のいずれも学際的に多方面の分野を学べるカリキュラムが組まれている。数学の持つポテンシャルを研究と教育の両面から展開する氏の活動は、数学以外の分野にバックグラウンドを置く研究者による数学との協働の、一つのモデルケースになると考えられる。

【質疑応答】

Q：パーシステントホモロジーを使った一連の研究の流れは、当初はTDAと呼んでいなかった。
1.TDAと呼称する流れはいつから？
2.データを図形として見る事で、代数などの数学的理論と非常によくマッチングすると感じる。このような考え方や研究は他にもあるか？

A：まず、2つ目の質問につき。指摘の見方はパーシステントホモロジーの創始者の一人であるGunnar Carlsson氏のスローガン[48]が本質。他の方法としては「多様体学習」や、自然言語処理における「Word2vec」という技術が該当するだろうか。いずれも「データを大域的に捉える」事を主軸に置いている。現在「数理モデリング」と言うと解析学の枠組みで論じられる事がほとんどだが、幾何学の枠組みで数理モデリングを提唱したい[49]。
1つ目の質問につき。これは最初、「ホモロジーを計算する」事を主軸に据えていた。 Rutgers大（Konstantin Mischaikow氏ら）や京大（國府寛司氏。セミナー第5回の講師）を中心にその動きがあり、2000年ごろより、パラメーターに対するデータのロバストネスを測る道具として、パーシステントホモロジーが台頭し始める。その後もMorse理論をベースとしたMapperなど、いろいろな理論やツールが出ており、この流れ、研究分野を総称して「TDA」と呼称する事になったと考える。

Q：TDAのアメリカや中国における動きは？

A：アメリカは非常に活発。特に研究チームによるマンパワーがすごく、これは日本の数学コミュニティだと非常に難しい動きになる。（平岡氏の場合は）幸運にも東北大学AIMRにてチームづくりの機会を与えられた。そこでの成果を起爆剤として、TDAの理論、応用が一気に広がり、現在に至る。
中国では、なぜか（TDAの）主要な研究者や研究グループがほとんど存在しない。留学などで学ぶ者は多いが。現在はオンラインセミナーなどを開いて、中国を含めた諸外国での活動の活発化を促している。オンライン形式は非常に効果的である。

Q：研究開発の発展の流れとして、知財化、ツール化など民間に向けた動きがあると思うが、その点に関する（平岡氏の）見解は？

A：個人で民間どうこうという事は考えていない。 HomCloudやチュートリアル、共同研究などで展開する事を考えている。

Q：（早稲田大理工学術院の材料科学専攻に）学部で数学を学んだ人は入ってきているか？

48 Data Has Shape, Shape Has Meaning, Meaning Drives Value”
Carlsson氏は初期からこの考えを軸に、トポロジーを応用数学に展開していく方向で研究活動を進めていたとか。なお、Carlsson氏は元々代数的K-理論を研究しており、現在はベンチャー企業であるAYASDI, Inc. の代表である

49 このコメントは平岡氏の考え

A：応用数理から材料に進む道も用意されている。メインは「数学＋機械工学」という組み合わせだが、他の大学で学生が触れる事のできない講義や実習に触れられるのは大きな魅力。まだ創設2年目ということもあり、これから広がるものと思われる。

Q：（早稲田大理工学術院の材料科学専攻に）代数や幾何など、他の広い数学の履修は可能？

A：リストに挙げた教員の分野（精度保証付き数値計算や可積分系）はあくまでメインのもので、他の講義の聴講も可能。

Q：アモルファス構造の解析にTDAが非常に有用である事が示されているが、結晶構造や粒界（grain boundary）に用いて強度を調べるなどの研究はあるだろうか？また、可能だろうか？

A：（平田氏が）知る限りはないはず。一連の研究で、アモルファス構造解析にて得られた知見は多いので、それらを活かして（粒界も含めた）材料強度と構造の関係性の解明に取り組みたい。

Q：人材育成に関する展望を伺いたい。

A：チュートリアルなどでの活動が今はメイン。数学系での学生のモチベーションは非常に高く、諸科学分野への応用、データ解析も含めてTDAを勉強したいと直接訪ねてくる者も多い。

　しかし、この方向性の研究だけだと「数学系で学位認定できない」事が数学系学生の参入のハードルを上げている。これは制度上の事情であるが、数学の定理を作る事を求められる数学系において、理論構築とデータ解析、応用を全て期限内に習得、実践するのはハードルが高すぎる。

　他方で、オンラインセミナーはかなり機能しているので、間接的トレーニングを実践しながら育てる事が今は現実的。（平岡）

学部1年や3年の講義など、様々な場所で宣伝を行なっているが、応用数学からの進学がまだ少ないのが現状。とはいえまだ発足から2年ほどなので、これから。（平田）

Q：パーシステントダイアグラムなど可視化の概念や技術について、3次元ダイアグラムなどの他の可視化についての可能性は？

A：TDAにおける2つの大問題と関連する。

1つ目は「動く点群」のパーシステンスを、どのように代数的情報として還元するか。時間（やパーシステント図を作るときに用いられる半径以外のパラメーターに）依存する点群で形成される数学的対象からあるモジュライ空間（一種の分類問題の解）の不変量を抽出する問題に還元する事になるが、これは代数幾何学の問題として非常に難しい。

2つ目は「ランダムネス」の取り込み方。実際のデータにはどれだけ気を配っても、必ずノイズは入る。対してパーシステント図の作成、代数的には加群の分類に相当するこの問題は、100%正確な答えを求めようとする。ノイズが入る事を考えると、「ノイズに対して安定な構造だけを取り出せれば十分なのでは？」という考え方に行き着く。しかし、ランダムネスの代数としての取り扱いとその解釈が非常に難しい。いずれにしてもこの問題は材料科学など応用の側面から生じた数学的問題であり、日本初の問題意識として世界に提言している。

Q：（先の質問と関連して）冷却速度に依存するガラスの構造変化は、ある時刻でのスナップショットで判断しているのか？

A：基本的にはその通り。スナップショットを複数取り、その時間平均を使ってパーシステントな構造抽出をしている。ガラス形成は低温での議論なので、摂動などは物理的にノイズとみなして差し支えなく、現行のパーシステント図でうまく構造抽出ができている。対して、「液体」などの高温環境における構造は、分子の動きが本質的になってくるので、先の疑問にある時間依存するTDAの技術開発が待たれる。時間依存性を考慮したTDAは、液体構造以外にも緩和時間など平衡と非平衡をつなぐ現象の評価にも応用の可能性はある。

ワークショップ企画・報告書編纂メンバー

若山 正人	上席フェロー	（システム・情報科学技術ユニット）
高島 洋典	フェロー	（システム・情報科学技術ユニット）
吉脇 理雄	フェロー	（システム・情報科学技術ユニット）
松江 要	特任フェロー	（システム・情報科学技術ユニット）
宮西 吉久	特任フェロー	（システム・情報科学技術ユニット）

俯瞰ワークショップ報告書 **CRDS-FY2020-WR-09**

システム・情報科学技術分野
数学と科学、工学の協働に関する連続セミナー

WORKSHOP REPORT

System and Information Science and Technology Area
“A series of seminars on collaboration among mathematics, science, and engineering”

令和 3 年 3 月　March 2021
ISBN 978-4-88890-713-2

国立研究開発法人科学技術振興機構　研究開発戦略センター
Center for Research and Development Strategy, Japan Science and Technology Agency

〒102-0076 東京都千代田区五番町 7 K’s 五番町
電話　03-5214-7481
E-mail crds@jst.go.jp
https://www.jst.go.jp/crds/

本書は著作権法等によって著作権が保護された著作物です。
著作権法で認められた場合を除き、本書の全部又は一部を許可無く複写・複製することを禁じます。
引用を行う際は、必ず出典を記述願います。
This publication is protected by copyright law and international treaties.
No part of this publication may be copied or reproduced in any form or by any means without permission of JST, except to the extent permitted by applicable law.
Any quotations must be appropriately acknowledged.
If you wish to copy, reproduce, display or otherwise use this publication, please contact crds@jst.go.jp.

ISBN 978-4-88890-713-2

FOR THE FUTURE OF
SCIENCE AND
SOCIETY

https://www.jst.go.jp/crds/