NO. 153
AUGUST. 2021
THE INSTITUTE OF STATISTICAL MATHEMATICS

NEWS

統計数理研究所ニュース

大学共同利用機関法人　情報・システム研究機構

CONTENTS

02 響き合う人とデーター統数研プロジェクト紹介
第26回「数理知能表現による深層構造学習モデルの革新」

06 研究教育活動
新入教員紹介／特任教員紹介
2021年5─6月の公開講座実施状況
統計数理セミナー実施報告（2021年4月〜7月）

09 統数研トピックス
統計数理研究所オープンハウスを開催
「データサイエンスが描き出す『モノづくり』の未来シナリオ ～産学連携シンポジウム～」の開催
今田高俊客員教授が「瑞宝中綬章」を受賞

10 総合研究大学院大学複合科学研究科統計科学専攻関係
大学院説明会について

11 お知らせ
公開講座（一般講座）

11 共同利用
2021年度共同利用公募追加採択課題

11 外部資金・研究員等の受入れ
受託研究・受託事業等の受入れ
外来研究員の受入れ

12 刊行物
Research Memorandum（2021.4〜2021.8）
研究教育活動報告
Annals of the Institute of Statistical Mathematics

14 コラム

数理的アプローチでAIを真の社会インフラに

▲福水健次教授

AIが“新しい電気”となるのを阻むコストの障壁

「AI（人工知能）を電気のように社会の隅々まで行き渡らせるには、まずコストの問題を解決する必要があります」。統計数理研究所統計的機械学習研究センターのセンター長を務める福水健次教授はそう話す。

“電気のように”という比喩は、深層学習の研究者Andrew Ng氏による「AIは新しい電気だ」という言葉を受けたもの。電力供給が始まった当初は、扱える人が少なく給電方法も特殊だったが、やがて都市や産業の設計もその利用が前提となっていった。AIも似たような経過をたどり、電気と同じように一般家庭にも行き渡り、なくてはならない社会インフラになる、と予言したのだ。

しかし、現状はまだそうはなっていない。その障壁となっているものの一つが、研究や技術開発にかかるコストだと福水は言う。

“第三の波”と言われる今のAIブームの核は、ニューラルネットワークによる深層学習（ディープラーニング）だ。「ニューラルネット」は人間の脳の仕組みを模した情報処理システムで、外部からの信号を集めて処理し出力する「パーセプトロン」と呼ぶ計算単位からなる。深層ニューラルネットは、このパーセプトロンを何十層、何百層と重ねて計算を行い、答えを出力する仕組みだ。

深層ニューラルネットが注目されたきっかけは2012年、ある画像識別タスクのコンペティションで、8層の深層学習によるモデルが際立った成果を叩き出したことだった。それ以降、より多層な構造のネットワークが登場し、2015年には152層のモデルの誤識別率は3.57%にまで下がっている。また2016年に、AIによる囲碁ソフト「Alpha Go」が最強レベルのプロ棋士を破ったことで、深層学習の可能性が世界に知れ渡った。

インターネットによってビッグデータを容易に取得できるようになったこと、コンピュータの計算能力と性能が飛躍的に向上したことなども、深層学習の隆盛を後押しした。画像認識や画像の生成・変換、自動翻訳などの技術はさらなる進化を続けており、さまざまな分野での技術革新が始まっている。

ところが、現状でAIを存分に研究し、技術開発を進めているのは、自社の提供する検索エンジンやSNSのビッグデータを取得可能で、高性能コンピュータを保有し、優秀な研究者を抱える巨大IT企業に

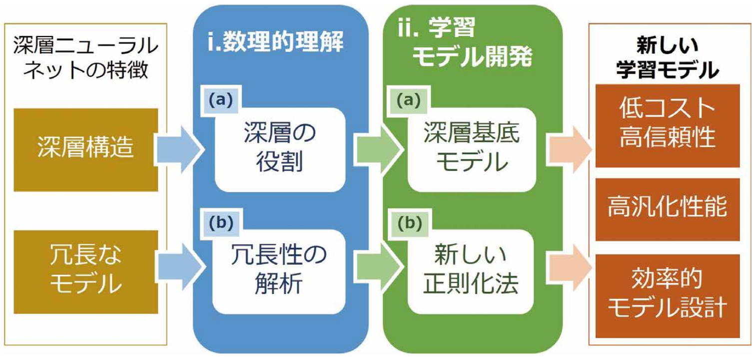
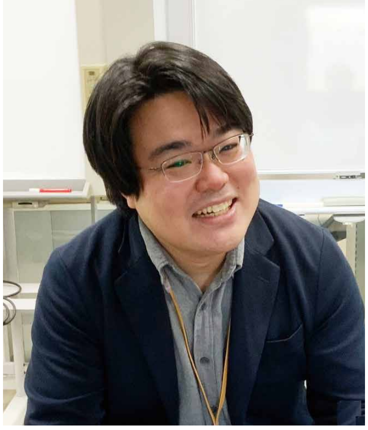


図1：新しい学習モデル開発のフロー。深層ニューラルネットを数理的に解明することで、高い信頼性かつ低コストの新たな学習モデルの開発につなげる。

へ

第26回「数理知能表現による深層構造学習モデルの革新」

深層ニューラルネットによるAI技術の開発は、欧米の巨大IT企業の寡占状態にあると言われる。その原因となっているのが、ビッグデータを取り込み膨大な計算をするためのコンピュータ設備と、それを扱う人材にかかる「コスト」だ。統計数理研究所を中心とするチームは、既存の深層ニューラルネットとは異なる深層構造モデルを開発することで計算コストを下げ、中小企業など、巨大IT企業でなくともAI技術を活用できる環境を整えるための研究を進めている。


▲本武陽一特任助教

限られている。アカデミアは企業の持っているデータを使うことができず、中小企業は計算資源も人的資源も足りず、深層学習を技術革新に活用できていないのが実情だ。「低コストで学習でき、信頼性も十分で、現行のモデルと同等以上の高性能なモデルを開発できれば、巨大IT企業だけでなく誰もがAI技術を扱えるようになるでしょう」と福水は説明する。

機械と人間のあいだをつなぐ数学の言葉

福水は以前から共同研究などで親交のあった東京大学大学院情報理工学系研究科の鈴木大慈准教授と、実世界知能の実現技術に詳しい東京大学先端科学技術研究センターの原田達也教授にも声をかけ、「数理知能表現による深層構造学習モデルの革新」のテーマで2020年度の戦略的創造研究推進事業「CREST」に応募し、採択された。チームには、統数研の日野英逸教授と持橋大地准教授、本武陽一特任助教、今泉允聡客員准教授（東京大学）も参加している。

このプロジェクトのポイントは、数理的知能表現を活用することで、現行の深層学習に使われている「深層ニューラルネット」とは異なる深層構造を持つモデルを生み出そうとする点だ（図1）。「数理知能表現」とは知能や知識を表現するための数理的方法のことで、特にこのプロジェクトでは関数空間、代数構造、微分方程式などによるアプローチを研究している。「数理的アプローチによって、AIを誰もが使いやすい技術にしたいと考えています」と福水は力を込める。AIを本当の意味で社会インフラにする、大きな挑戦だ。

目的を達成するための研究課題は「深層構造の理論的理解による効率的で信頼性の高い数理知能表現モデルの研究」「深層構造モデルの最適化アルゴリズムの研究」「実世界問題への実践的応用による方法と理論の深化」の三つ。福水をリーダーとする統数研メンバーによる数理深層モデリングG、鈴木准教授をリーダーとする数理表現・最適化G、原田教授をリーダーとする実践的学習Gがそれぞれの課題を担当する。

▲東京大学先進科学研究機構
今泉允聡准教授（統数研客員准教授）

このプロジェクトにおける福水のグループの課題の一つは「数理構造による深層モデリング」の研究だ。例えば、既存の深層ニューラルネットによる画像生成技術では、「サングラスをかけた男性」の学習データから「男性」のデータを引き算し、さらに「女性」のデータを足し算して、元のデータにはなかった「サングラスをかけた女性」の画像が生成される、といった興味深い現象が知られている。だが、その理論はブラックボックスになっている。代数的なアプローチによって、その数理構造を解明することが研究課題の一つだ。

二つ目の課題は、「微分方程式と深層モデリングの融合」。自然現象や機械の動作の記述などのように時間的な構造を

持っているデータは、数学では微分方程式で表されることが多い。その時間的な「法則」を深層学習と組み合わせて、予測やCGシミュレーションなどの精度を向上させる技術を開発することが目標だ（図2）。「例えば気象データや地震データなども時間的構造を持つデータですから、災害予測や防災にも役立つことが期待できます」（福水）。

この課題については、福水研究室のメンバーでもある本武が中心となって取り組んでいる。本武は物理学の専門家で、物理現象の解明に深層学習を援用する研究をしていた。「機械学習は複雑な物理現象を内挿的に（学習データの数値の範囲内で）モデル化するのは得意ですが、外挿的、つまりリンゴが木から落ちるのを見て万有引力の法則を思いつくというニュートンの物語にあるような大胆なアイデアでは人間に及びません。両者のよい面を融合するためにはあいだをつなぐコミュニケーションが必要で、数学的モデル化がその言語の一つとなるのではないかと期待しています」と本武は話す。具体的には、時間的構造を持つデータから深層ニューラルネットで保存則などの法則をモデル化したり、逆に人間が持っている知見を深層ニューラルネットに導入したりといったことを考えている（図3）。

億単位のパラメータでも過学習を起こさないことを数学的に証明

鈴木准教授のグループが研究しているのは、より数学的な「数理表現・最適化」だ。深層学習の研究は今、世界中で活発に行われているものの、非常に複雑で数理的な扱いが難しく、分からないことがたくさんある。未知な部分を数理的に解明することで本質を捉え、より低コストな学習を実現するのがこのグループの狙いだ。

このグループのメンバーである今泉准教授は「深層学習では層が増えるほど、またパラメータが多いほど高い精度を発揮することが実験で確かめられていますが、これらは今までの数学の常識とは矛盾しており、

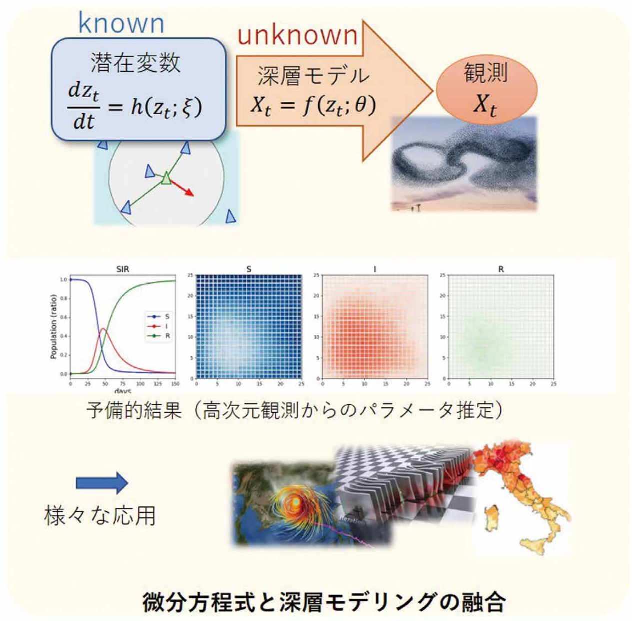


図2：微分方程式と深層モデリングの融合。気象データや地震データといった時間的構造を持つデータを微分方程式で表し、その法則を深層学習と組み合わせることで、予測やCGシミュレーションなどの精度向上に役立てる。

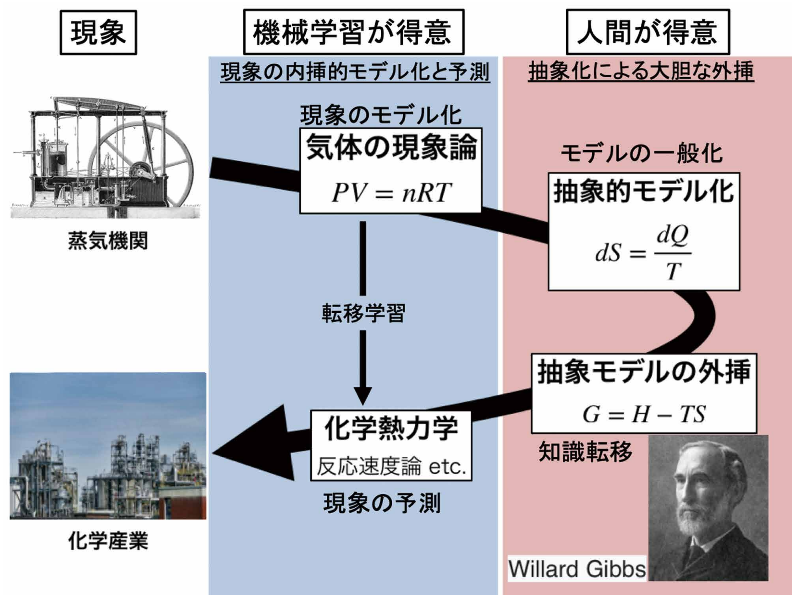


図3：抽象化と外挿による飛躍。機械学習が得意とする「内挿的モデル化」と人間が得意とする「抽象化による大胆な外挿」のあいだを、数学を言語としてつなぐ。

なぜそうなるのか分かっていません。私たちにとってはこれが解き甲斐のある魅力的な“謎”なのです」と微笑む。

過剰なパラメータの謎について理論研究者の間では最近、「二重降下」説が注目されているという。パラメータが増えると一旦は誤差が大きくなるものの、再び降下し始めるのではないかというのだ。今泉准教授は、深層学習においても特定の条件を与えれば、パラメータを億単位まで増やしても過学習に陥ることなく、誤差が小さくなることを初めて理論的に示した（図4）。

一方、原田教授のグループは「実践的応用」の研究を担っている。動画と音声、画像とテキストなどマルチモーダルなデータの認識や生成といった、社会とのインターフェースとなる重要な部分だ。「原田教授の既存研究には興味深いものが多く、今回のプロジェクトでも理論と応用を結ぶ成果を期待しています」（福水）。

例えば、人が自転車に乗っている動画を正しく判定することは現状のAIにもできるが、原田教授の開発した技術ではその動画に「スケートボードではなく自転車に乗っている。なぜならば座った姿勢をとっているからだ」と、理由までテキストで提示することができる。また現状のAI技術では、実写の動画に映っている馬をシマウマに変換することはできても、ゼロから馬のデータを生成して動かすことは難しい。プロジェクトでは、それを可能にするような研究も行うという。

「画像や動画と音声や言語などを融合的に情報処理していく技術は、今後非常に重要になってくるはず。効率のよい深層構造モデルを使って複雑なタスクを解いていけるようにしたい」と福水は話す。

さらに、医療画像への応用も期待される。CTやMRIなどの検査画像は、超高解像度で3次元、時間を加えた4次元のものもあり、データ容量は莫大になる。プロジェクトが目標としている「よりコストのかからない深層学習モデル」ができれば、医療画像の解析には特にメリットが大きいだろう。

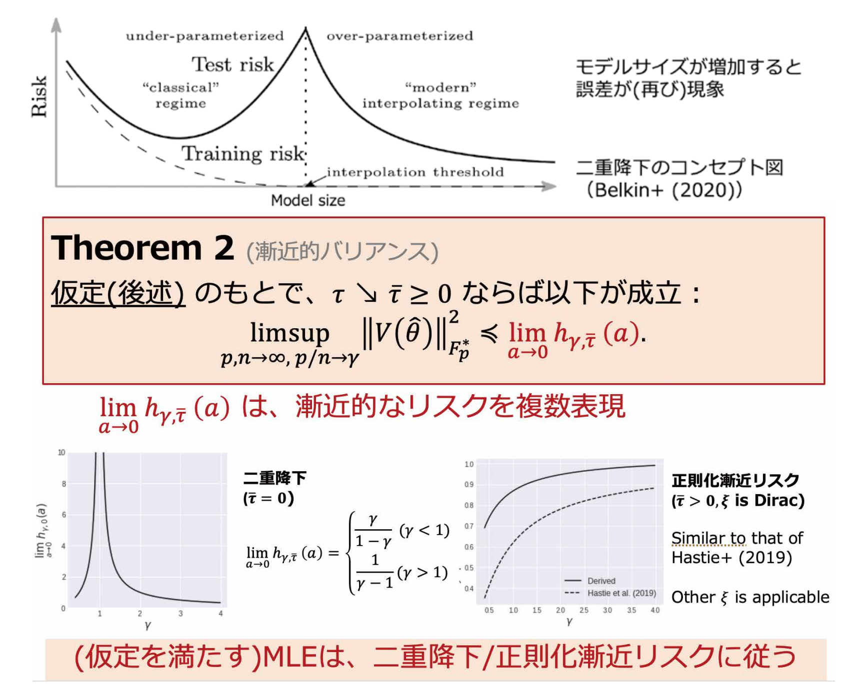


図4：「二重降下」の理論構築。パラメータの増加で誤差が再減少する「二重降下」の理論を数学的に示した。

日本の中小企業が巨大IT企業を追い抜く日

福水は、CRESTのプロジェクトの延長上に、さらなる展望を描いている。まず機械学習の面では、前述の「微分方程式と深層モデリングの融合」をさらに進化させ、自然現象など時間的に移り変わる事象の「法則」を自動的に抽出する技術に結びつけること。

また数理科学の面では、深層構造のための「関数空間の理論」を確立すること。関数空間の理論は数学界では古くから研究がなされてきた成熟分野だが、深層学習は工学や統計学の分野から生まれたため、その研究にぴったりの理論があるわけではない。深層学習を解析するために必要な数学理論を開発していくことが欠かせない。

さらに、この研究を進めることには、大きな社会的意義もある。巨大IT企業でなくとも容易に適用できるAI技術を創出することによって、冒頭で述べたように、AIが電気と同じように社会に普及する世界を創出することも夢ではなくなるのだ。

「欧米に比べて周回遅れとも言われる日本企業のAI研究が、一気に挽回して前線に躍り出るチャンスもあると考えています。巨大なコンピュータでなければ出せない性能をパソコンや機械に内蔵するチップに組み込めるようになれば、状況は大きく変わるでしょう」と福水は期待を込める。

AIを電気のような社会インフラにする――。そのためのイノベーションの第一歩が今、踏み出された。　（広報室）

新型コロナウイルス感染拡大防止のため、本インタビューはオンラインで行われました。

新入 1
教員紹介

野場　啓

統計思考院　助教

4月に統計思考院に助教として配属されました、野場啓と申します。Lévy過程の挙動に関する基礎研究および応用研究を行なっております。特に最近はLévy過程を用いた確率制御の問題に興味をもっており、企業の支払う配当金を最大化する配当戦略を考える最適配当問題や、製品の在庫の保持等のコストを最小化する戦略を考える在庫制御問題の研究に力を入れています。統計数理研究所での研究や教員の方々との触れ合いを通じて視野を広げ、さまざまな道具を用いた研究ができるようになるよう精進していきたいと思います。よろしくお願いいたします。

特任
教員紹介
1

藤田　茂

モデリング研究系　特任教授

4月よりモデリング研究系特任教員を兼務することになった藤田茂と申します。本務はデータサイエンス共同利用基盤施設データ同化支援センターです。私は気象庁気象大学校で、地球大気の外側領域（外圏大気）のモデリング研究を行ってきました。外圏大気は人工衛星が飛んでいる領域であるため、この領域の変動（宇宙天気）は、衛星障害を引き起こし、社会活動に障害を与えることがあります。それを防ぐため、WMO（世界気象機関）が中心となり、主要国にて外圏大気環境の予測（宇宙天気予報）が始まっています。一方、数値モデルを宇宙天気予報に応用するにはデータ同化が欠かせませんが、この試みは世界的に開発途上です。また、外圏大気の研究にもこの数値モデルは重要です。統数研や気象庁のデータ同化専門家と協力して、宇宙天気数値モデルの開発を進めていきたいと思います。

特任
教員紹介
2

力丸　佑紀

リスク解析戦略研究センター　特任准教授

4月1日付けでリスク解析戦略研究センターの特任教員に着任いたしました力丸佑紀と申します。裏に連続的な広がりを想定できる空間データに対し、その構造の理解に必要となる空間データモデリングの研究を行っています。これまでは、格子点上の空間データに対するモデルのひとつである同時空間自己回帰モデルにおいて、最尤法によるパラメータ推定の確立を中心課題としてきました。現在は、空間経済学などでよく使われるSLMやSEMのようなモデルの一般的な使われ方と、誤差が弱定常性をもつ回帰モデルとの関係を整理した上で、それらのモデルにおけるパラメータ推定の問題に取り組んでいます。これまで理論的な研究を主としてきましたが、ここに着任できたことをきっかけに実際のデータの扱いも学べればと思っております。どうぞよろしくお願いいたします。

特任
教員紹介
3

TRAN Duc Vu

リスク解析戦略研究センター　特任助教

I joined ISM in April 2021 as a Project Assistant Professor. I have been working on COVID-19 related problems including COVID-19 infection spread modeling at ISM now, and COVID-19 related knowledge extraction from scientific papers as a researcher at Japan Advanced Institute of Science and Technology

before. Having experience of working with textual data, I am keen on investigating a methodology to utilize textual information(scientific papers, news, SNS, etc.) for enriching COVID-19 infection spread modeling (including performance and explainability). I am looking forward to thrilling research activities ahead.

有竹　俊光

統計的機械学習研究センター　特任助教

特任
教員紹介
4

今年の3月に早稲田大学の博士課程を終了し、4月より特任教員に着任いたしました。これから統計数理研究所という素晴らしい環境で研究ができることを大変うれしく思っております。これまではスパースモデリング手法の開発や実問題への応用を行ってきました。具体的には基底のシフト・スケーリング構造を考慮した辞書学習法の開発を行ったほか、応用研究としては超解像蛍光顕微鏡の画像処理アルゴリズムを圧縮センシングの問題として定式化し、深層学習を用いた高速かつ高精度な処理方法の開発を行いました。

現在は機械学習における転移学習の枠組みの1つであるドメイン適応に関する研究などを行っています。今後は知識を幅広く吸収し、理論、応用の両面から研究の発展に貢献できるよう努力して参る所存です。これから、どうぞよろしくお願いいたします。

坂本　航太郎

統計的機械学習研究センター　特任助教

特任
教員紹介
5

'Can machines think?' Barabási先生の言う「理解・予測・制御」する対象としての、そして less is more と more is different の間にある知的情報処理という不思議な自然現象に深く興味を持っています。AutoMLの基盤技術となる、能動学習について、情報幾何学、などの理論から、ベイズ最適化、強化学習、などの様々な方法論まで幅広いキーを駆使し、深層学習を始めとする機械学習の高性能化と効率化に取り組んでいます。ニューラルアーキテクチャ探索（Neural Architecture Search: NAS）や説明可能AI（Explainable AI）に対する能動学習の意義についても検討しています。新しいパラダイムとしての機械学習をあらゆる方面で実装していけるように実践家として理論と実験の橋渡しを目指して、産業においても科学においても、広く深く高く奔走していきたいです。よろしくお願いします。

中島　秀斗

統計的機械学習研究センター　特任助教

特任
教員紹介
6

4月より統計的機械学習研究センターの特任助教に着任しました、中島秀斗と申します。専門は数学の表現論という分野で、群の持つ対称性が、作用する空間にどのように影響するのかということに興味を持っています。特にこれまでは、正定値対称行列の空間を一般化した空間について、不変微分作用素やゼータ関数、固有値分布問題などを通して、様々な視点から研究をしてきました。現在は深層学習における注意機構に関する数学的理解に向けた研究や、群の不変性・共変性を利用した機械学習の研究を行っています。どうぞよろしくお願いします。

千野　雅人

統計思考院　特任教授

特任
教員紹介
7

この4月に着任し、文部科学省の公募事業である「統計エキスパート人材育成プロジェクト」に取り組んでいます。この事業は、大学統計教員を育成し、持続的に統計人材が育成されるシステムを全国の大学に構築することを目指すものです。私は、これまで30年以上にわたり総務省に勤務し、統計局長などの経歴を持つ実務経験教員です。同じ「統計」ですが、「統計数理」とは異なり、「統計法」など統計制度の企画立案や統計調査による統計データの作成など、「統計行政」と言われる仕事を行ってきました。偏りのない信頼できるデータを作成することは、とても手間のかかる泥臭い作業です。例えば、「国勢調査」は、年間700億円の予算と全ての都道府県・市町村の職員、70万人の統計調査員を動員して実施します。国家が全ての国民の情報を漏れなく完璧にデジタル管理しない限り、このような調査が必要となります。広く統計の発展に貢献したいと思います。どうぞよろしくお願いいたします。

中西　寛子

統計思考院　特任教授

特任
教員紹介
8

思考院に特任教授として着任しました、中西寛子と申します。統計数理研究所の新プロジェクト「統計エキスパート人材育成コンソーシアム」のメンバーとして採用されました。若いころの専門は多変量解析（特に判別分析）でしたが、その後、様々な分野に関心があり研究を進めておりました。最近では、統計検定をはじめとする統計教育にも関心を持っています。元職は成蹊大学経済学部で25年間教鞭をとり、統計と情報を融合した教育に専念しておりました。今までの経験を活かし、研究を進めるとともに、プロジェクトを始めとし、様々な取り組みでお役に立てることができればと思います。どうぞよろしくお願いいたします。

室田　一雄

統計思考院　特任教授

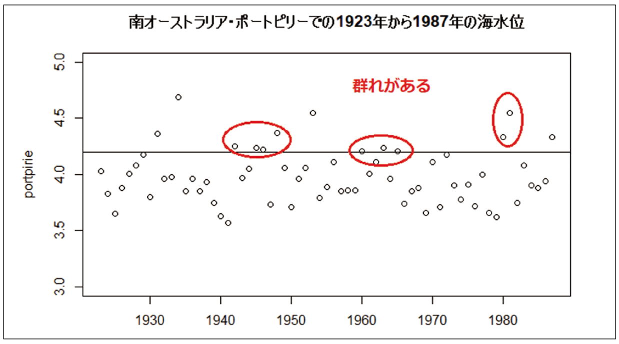
特任
教員紹介
9

本年４月に統計思考院の特任教授に着任しました。最適化、数値計算、群論的分岐理論など、数理工学のいろいろな分野に興味をもって研究活動をしております。中心は、離散凸解析という分野です。離散凸解析は、整数格子点やグラフ構造の上で定義された関数を、凸解析と組合せ論の両方の視点から考察する理論で、私自身が提唱してきたものです。最近では、ゲーム理論、数理経済学、離散最適化、オペレーションズ・リサーチ、離散幾何などへ応用が広がってきています。昨年まで、40年間にわたり、複数の大学の工学部、理学部、経済学部などに在籍して、いろいろな分野の学問に触れ、研究交流をしてきました。今後も、学問を続けていく所存です。どうぞよろしくお願い致します。

Report

2021年5-6月の公開講座実施状況

5月25日（火）に、山梨大学の西郷達彦准教授による「Rによる極値統計学」がオンラインにて配信されました。講義では、極値統計学の基盤となる数理的な背景から始め、複数のデータ取得デザインに対応した推定法と、Rのパッケージを用いた実際の解析までを解説しました。質問も多数寄せられ、口頭での回答だけでは時間が足りなくなったため、テキストでの回答も併用しました。（情報資源室）



公開講座「Rによる極値統計学」資料より

統計数理セミナー実施報告（2021年4月～7月）

毎週水曜16時から所内研究教育職員および外部の方が1人40分ずつ、1日に2人の講演を行っています。2021年4月～7月のセミナーは下記の通り行われました。新型コロナ感染拡大防止のため、統計数理セミナーは現在オンラインで開催しています。

日　程	氏　名	タイトル
2021年　4月14日	田中 未来	2錐間の大域的最小角を求めるための分枝限定法
4月21日	齋藤 翔	位相的二面体ホモロジー
4月28日	福水 健次	背景ノイズにロバストな位相的データ解析
4月28日	釼持 尚輝	敵対的生成ネットワークの核融合プラズマ研究への応用
5月12日	坂田 綾香	表現型空間の進化的次元圧縮
5月19日	吉田 亮	データ駆動型材料研究の諸問題
5月26日	星野 伸明	情報保護のための多項サンプリング
6月　2日	小池 祐太	高頻度金融市場における多重先行遅行関係のウェーブレット解析
6月　2日	藤澤 洋徳	L1 正則化を利用した転移学習
6月　9日	前田 忠彦	温故知新：戦後間もない時期の社会調査再訪
6月23日	松井 知子	DICEモデルを用いたCOVID-19の気候と経済への影響に関する一考察
6月30日	山下 智志	データ結合技術とプロファイリングの現状、社会的影響
7月　7日	ウ ステファン	地盤工学のビッグデータ：粘土特性予測のケーススタディ
7月　7日	小森 理	Kolmogorov-Nagumo平均によるk-means, fuzzy c-means, Gaussian mixtureモデルの統一的定式化
7月14日	二宮 嘉行	事前分布強調型情報量規準
7月21日	日野 英逸	能動学習の停止基準とその応用
7月28日	村上 大輔	疎なカウントデータのための地理的加重回帰の安定化
7月28日	加藤 直子	新興科学技術の日常生活への応用に関する意識調査：情報提供と知識との関連

セミナーの開催予定はホームページにてご案内しています。https://www.ism.ac.jp/　　（メディア開発室）

統数研トピックス

統計数理研究所オープンハウスを開催

2021年6月18日（金）、統計数理研究所オープンハウス“計るを統べる統計科学”を開催しました。昨年に引き続き新型コロナウィルス感染拡大防止のため、オンライン形式での実施となりました。

公開講演会は「味と香りのデータサイエンス」をテーマに、宮里義彦副所長の司会のもと椿広計所長による開催の挨拶から始まりました。

旭酒造株式会社 桜井博志会長による講演「ピンチはチャンス！～山口の山奥の小さな酒蔵だからこそできたもの～」では、日本酒「獺祭」の酒造りの中で逆境をチャンスに変えて成功した体験談や、データを読み取ることのできる人間とその奥にある酒造りを追求する気持ちの大事さが語られました。

公開講演会の質疑応答の様子

続いて農業・食品産業技術総合研究機構 農業情報研究センター　AI研究推進室 確率モデルユニット 藤岡宏樹上級研究員による講演「香りの分析とデータの活用について」では、センサ値から香りを情報化する質量分析装置、香りセンサーを使った評価法の事例と課題について紹介されました。

最後に本研究所 数理・推論研究系 間野修平教授による「源氏香はクラスタリング～ベル数とその周辺～」では、香道の源氏香を例として紹介し、数値で与えられないデータ解析について講演されました。

本講演会はZoom Webinarで配信を行い、772名の方々にご視聴いただいた盛況な講演会となりました。

その他、統数研の研究紹介として研究内容ポスターの

PDFが109名分、内容を紹介する動画が14件、統数研と3研究系（モデリング研究系、データ科学研究系、数理・推論研究系）の紹介動画が公開されました。

Zoomで行われた統計よろず相談室は9件の相談を受け、総研大統計科学専攻の大学院説明会には43名が参加し、修了生による講演会は41名が視聴しました。

（広報室）

「データサイエンスが描き出す『モノづくり』の未来シナリオ ～産学連携シンポジウム～」の開催

2021年6月17日に、統計数理研究所オープンハウス2021の連携イベントとして、産学連携シンポジウム「データサイエンスが描き出す『モノづくり』の未来シナリオ」を、統数研と連携を深めてきた大学、企業のデータサイエンスの専門家を迎えて開催しました。当日は、本研究所の北村浩三特命URAが司会を務め、九州大学マス・フォア・インダストリ研究所梶原健司副所長による挨拶の後、本研究所椿広計所長がシンポジウムのオリエンテーションを行いました。その後、花王（株）エグゼクティブフェロー、Preferred Networks Fellowの丸山宏先生による「ものづくりとソフトウェア―DevOpsとSoftware2.0」、（株）デンソー生産技術部F-IoT室担当次長・技師の吉野睦先生による「データサイエンスと品質管理」、（株）三菱ケミカルホールディングス執行役員の浦本直彦先生による「製造業におけるデジタルトランスフォーメーションの実践」、休憩を挟んで本研究所教授、ものづくりデータ科学研究センター長の吉田亮教授による「データ駆動型材料研究の諸問題：現状と展望」の各講演が続きました。

シンポジウム当日の様子

その後、椿所長がモデレーターを務めるパネル討論を行いました。話題は、マテリアルズインフォマティクスからデータの共有、DX、製品の品質、企業内及び大学でのデータサイエンス教育の現状、人間系とデータサイエンスの共存等の広範囲におよび、活発な意見交換が行われ、最後に各講演者がデータサイエンスへのこれからの期待を述べて閉会となりました。

本シンポジウムは新型コロナウィルス感染拡大防止のためオンラインで開催し、1,202名の事前参加登録と、当日は911名の視聴があり大変盛況なイベントになりました。

（URAステーション）

今田高俊客員教授が「瑞宝中綬章」を受賞

令和3年春の叙勲において、調査科学グループの今田高俊客員教授が「瑞宝中綬章」を受賞しました。

瑞宝章は、日本の叙勲制度内の勲章の一つで、「国家又は公共に対し積年の功労ある者」に授与すると定められています。具体的には「国及び地方公共団体の公務」または「公共的な業務」に長年にわたり従事して功労を積み重ね、成績を挙げた者を表彰する場合に授与されます。

社会学分野において「自己組織性」を備えた社会システム論を世界に先駆けて提案したことや、地位の非一貫性という視点から新たな社会階層論を展開したことが評価され受賞となりました。（広報室）

総合研究大学院大学複合科学研究科統計科学専攻関係

大学院説明会について

令和3年6月18日（金）に、令和3年度第1回大学院説明会をオンラインにて開催しました。前半は統計科学専攻修了生による講演会を行い、「大学院時代と現在」の内容で修了生4名より講演いただきました。参加者は41名でした。後半は入試ガイダンス・カリキュラムの説明・質疑応答・教員マッチング等について説明を行いました。参加者は43名でした。

（総務課･大学院係）

●公開講座（一般講座）

一般社会人・学生を対象に、下記の公開講座（一般講座）をオンラインにて開催いたします。

Ⓒ Rによる時系列解析入門 ㊔

日時：10月22日(金) 10時〜16時45分(5時間45分)
講師：川崎 能典(統計数理研究所)
申込受付：9月6日(月)10時〜9月13日(月)10時
受講料：5,000円
定員：100名(応募者多数の場合は抽選)
講義レベル：中級

講座内容：

定常性とデータの変換／自己共分散関数とスペクトル／最小二乗法とモデル選択／自己回帰移動平均(ARMA)モデルの性質、推定／季節ARIMAモデルの識別と推定／見せかけの回帰と単位根検定／多変量自己回帰モデル：これらをフリーソフトウェアRで実行し、出力を理解できることを目標に解説を加える。講義内で行うデモに関してはソースコードを配布する。(※時間の制約上、実習と銘打つほどの時間は確保できないので予めご了承の上お申し込みください。)

受講者に期待する予備知識やレベル：

相関、回帰、推定・検定等の基本的知識に基づき、重回帰分析の実行や結果の解釈が一通りできるレベルにあること。

参考書：

北川源四郎 (2020)「Rによる時系列モデリング入門」岩波書店
田中勝人 (2006)「現代時系列分析」岩波書店

その他：本講座では状態空間モデリングは扱わない。

●一般講座の他、系統的な講座編成により現代的な統計科学の姿を示す「リーディングDAT講座」や、医学・健康科学分野のデータサイエンスを基礎から最先端まで詳説する医療健康データ科学研究センターの「系統的教育コース・公開講座」も開講予定です。

詳細は、以下のwebサイトをご覧ください。
https://www.ism.ac.jp/lectures/kouza.html

（情報資源室）

共同利用

2021年度共同利用公募追加採択課題

【一般研究1】1件

分 野	研究課題名	研究代表者（所属）
h5	深層学習による画像処理と自然言語処理の融合	持橋　大地(統計数理研究所・准教授)

（研究推進課・共同利用係）

外部資金・研究員等の受入れ

Report

受託研究・受託事業等の受入れ

委託者の名称	研究題目	研究期間	研究経費（円）	受入担当研究教育職員
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	求解軌道のマクロ表現によるアルゴリズム制御理論の創出	R3.4.1〜R4.3.31	19,279,000	数理・推論系 坂田　綾香 准教授
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	数理知能表現による深層構造学習の革新	R3.4.1〜R4.3.31	21,710,000	数理・推論研究系 福水　健次 教授
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	統計学的手法による宇宙画像解析	R3.4.1〜R4.3.31	5,200,000	数理・推論系 池田　思朗 教授
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	位相的統計理論の構築	R3.4.1〜R4.3.31	10,660,000	数理・推論研究系 福水　健次 教授
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	局所性に基づく計測対象のモデル化と高速化	R3.4.1〜R4.3.31	14,300,000	モデリング研究系 日野　英逸 教授

委託者の名称	研究題目	研究期間	研究経費(円)	受入担当研究教育職員
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	機械学習と計算科学に基づく熱物性マテリアルズインフォマティクス	R3.4.1～ R4.3.31	14,300,000	データ科学研究系 吉田　亮 教授
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	行動実態･変容データ解析、環境データ解析	R3.4.1～ R4.3.31	1,300,000	データ科学研究系 村上　大輔 助教
国立研究開発法人科学技術振興機構　分任研究契約担当者 契約部長　近藤　章博	メッシュ統計データの統計的観点からの特性評価と利活用	R3.4.1～ R4.3.31	650,000	統計思考院 岩崎　学 特任教授
独立行政法人　日本学術振興会 理事長　里見　進	統計科学関連分野に関する学術研究動向－COVID-19関連データの解析技術の調査－	R3.4.1～ R4.3.31	1,560,000	モデリング研究系 松井　知子 教授
国立研究開発法人産業技術総合研究所　理事長　石村　和彦	深層学習･転移学習等の適用による材料設計技術の高度化	R3.4.1～ R4.2.28	3,450,000	データ科学研究系 吉田　亮 教授
国立研究開発法人新エネルギー･産業技術総合開発機構 理事長　石塚　博昭	深層確率コンピューティング技術の研究開発	R3.4.1～ R4.3.31	8,591,000	数理･推論研究系 福水　健次 教授
国立研究開発法人新エネルギー･産業技術総合開発機構 理事長　石塚　博昭	自動機械学習による人工知能技術の導入加速に関する研究開発	R3.4.1～ R4.2.28	18,200,000	モデリング研究系 日野　英逸 教授
大学共同利用機関法人　人間文化研究機構長　平川　南	現代語の意味の変化に対する計算的･統計力学的アプローチ	R3.4.1～ R4.2.28	1,000,000	数理･推論研究系 持橋　大地 准教授
国立長寿医療研究センター 理事長　荒井　秀典	高齢者２型糖尿病における認知症予防のための多因子介入研究（J-MIND-Diabetes）	R3.4.1～ R4.3.31	50,000	データ科学研究系 野間　久史 准教授
一般社団法人いのち支える自殺対策推進センター 代表理事　清水　康之	行政における統計データの利活用の推進に関する研究	R3.5.14～ R4.3.31	5,000,000	椿　広計 名誉教授

（研究推進課･研究推進係）

外来研究員の受入れ

氏　名	職　名	研究題目	研究期間	受入担当研究教育職員	称号付与
山本　誉士	明治大学研究・知財戦略機構・特任准教授	動物の時空間分布動態モデリング	2021.4.8～ 2022.3.31	中野　慎也 准教授	
石辺　岳男	地震予知総合研究振興会・主任研究員	地震発生確率評価に用いるBPT分布のαについて	2021.5.1～ 2022.3.31	庄　建倉 准教授	
河村　優美	理化学研究所環境資源科学研究センター・テクニカルスタッフ	ケモインフォマティクス	2021.4.8～ 2022.3.31	吉田　亮 センター長	
Massimiliano Zamengo	東京工業大学物質理工学院・助教	Xenonpyを用いた高熱伝導ポリマーの開発	2021.4.8～ 2022.3.31	吉田　亮 センター長	
瓜生　寛堂	東京理科大学・学生	機械学習を用いた新規ハイパーマテリアルの探索	2021.4.8～ 2022.3.31	吉田　亮 センター長	
稲垣　佑典	総務省統計局統計調査部・統計専門職	公的統計データを併用した社会調査資源の統合的利活用法の研究	2021.4.8～ 2021.5.31	前田　忠彦 准教授	
張　俊超	新潟大学経済科学部・助教	擬似データの作成とその生成技術に関する研究	2021.7.1～ 2022.3.31	南　和宏 教授	

（研究推進課･共同利用係）

刊行物

Research Memorandum（2021.4～2021.8）

No.1215: Hoshino, N., Urn Models Closed under Resizing

（メディア開発室）

研究教育活動報告

No.50：総合研究大学院大学 複合科学研究科 統計科学専攻（編）, 2020年度 総合研究大学院大学 統計科学専攻 学生研究発表会 報告集（2021.2）

（メディア開発室）

Annals of the Institute of Statistical Mathematics

Volume 73, Number 3（June 2021）

Kangning Wang and Wen Shan
Copula and composite quantile regression-based estimating equations for longitudinal data ……441
Peng Lai, Fangjian Wang, Tingyu Zhu and Qingzhao Zhang
Model identification and selection for single-index varying-coefficient models ……457
Li-Pang Chen and Grace Y. Yi
Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error ……481
Ji Chen, Jun Shao and Fang Fang
Instrument search in pseudo-likelihood approach for nonignorable nonresponse ……519
Yuting Wei, Qihua Wang and Wei Liu
Model averaging for linear models with responses missing at random ……535
Haruhiko Inatsugu and Nakahiro Yoshida
Global jump filters and quasi-likelihood analysis for volatility ……555
Aki Ishii, Kazuyoshi Yata and Makoto Aoshima
Hypothesis tests for high-dimensional covariance structures ……599
Lei Wang and Wei Ma
Improved empirical likelihood inference and variable selection for generalized linear models with longitudinal nonignorable dropouts ……623

Volume 73, Number 4（August 2021）

Simon Clinet and Yoann Potiron
Estimation for high-frequency data under parametric market microstructure noise ……649
Yury A. Kutoyants
On localization of source by hidden Gaussian processes with small noise ……671
Mingyang Ren, Sanguo Zhang and Qingzhao Zhang
Robust high-dimensional regression for data with anomalous responses ……703
Jiasen Zheng and Lixing Zhu
Determining the number of canonical correlation pairs for high-dimensional vectors ……737
Pratik Misra and Seth Sullivant
Gaussian graphical models with toric vanishing ideals ……757
Shanshan Qin, Hao Ding, Yuehua Wu and Feng Liu
High-dimensional sign-constrained feature selection and grouping ……787
Byungsoo Kim, Junmo Song and Changryong Baek
Robust test for structural instability in dynamic factor models ……821

（メディア開発室）

コラム
COLUMN NO. 153

「お得なポイント」と個人情報プロファイリング

山下 智志　データ科学研究系

コロナの少し前に、北京に行く機会があった。中国は国家による個人情報の管理が進んでいることで有名である。例えば、地下鉄などの日常の移動の切符を買うにも個人情報の提示が必要となる。極端な例では、飲料の自動販売機においても個人認証が必要となっており、個人の購買行動もほぼデータベース化されている。

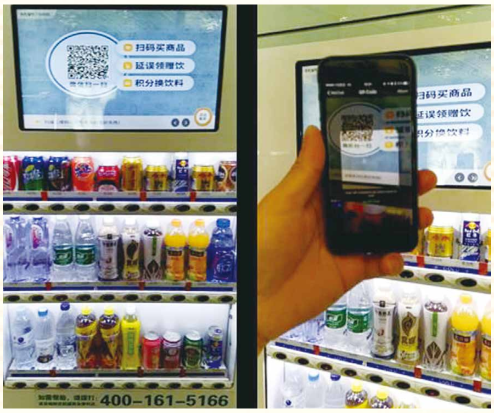
一方、日本では個人データのプロファイリングは、主に民間企業で行われている。マーケティングにおいてデータを積極活用した最も顕著な例は、セブンイレブンのPOSデータであり、店員が観察した顧客の状態と販売記録を合体させることによって、マクロの購買行動を把握するのに役立っていた。ただし、この段階では「個人」データではない。

このような日本の古典的マーケティングデータ分析に一石を投じたのは、TSUTAYAのTポイントカードである。これは当初、レンタルビデオの貸出についての管理カードであったが、住所、生年月日、性別などの個人情報と購買行動が密接に絡みついており、個人データベースを構築するのに結果的に有力なツールであった。特に重要なのはTポイントシステムで、消費者が個人行動の記録を積極的に提供するインセンティブを与えるものになっている。その後、TSUTAYAの運営会社であるCCCは、Tポイントシステムを他社にも解放した。特にファミリーマートの参加が契機になり、Tポイントカードの社会的普及が一気に進み、あらゆる購買行動を把握する強力なツールとなった。

ポイント制による個人データの把握は、マーケティングに大変有利であることに企業が気づき、その後一気に普及することになる。さらに、携帯電話の位置情報やおサイフケータイなどの決済機能を通じて、企業がもつ個人データベースがますます高度化されていった。最も象徴的な事件はソフトバンクによるCCCの買収であり、合法的にTポイントカードの個人情報をソフトバンクが入手することになった。これによりソフトバンクはTカード、PayPayなどのQRコード決済、携帯電話の位置情報など、個人の行動をかなり高い精度で把握することになる。

少し内容は異なるが、携帯電話会社の位置情報を有料で提供するサービスが行われるようになった。もちろんこれらのデータでは個人情報は削除されている。しかし、データを買う側がその地域の購買データを持っている場合、時間と場所の情報をのりしろにして、携帯電話の1データの個人をある程度の確率で特定することが可能である。一つ一つのデータは個人情報が保護されているとしても、複数のデータベースを参照することによって、そのデータが誰のものであるかということについて確率的に予測することが可能となる。このような処理を確率的データプロファイリングという。個人データがたとえ匿名性を保った状態で流通していたとしても、データをプロファイリングすることによって、個人が隠したいと思っている自分の情報が、ある一定の確率で他人に伝わるリスクが現代社会には存在する。

「お得」という言葉が広く流通する世の中になった。ポイント還元をしてお得感を消費者が楽しむことは決して悪いことではないが、それは同時に自分の行動の秘匿性が奪われていることを、意識して欲しいと願う。



北京のQR決済専用自動販売機

買収による個人情報の取得

統計数理研究所ニュース　No.153（2021年8月25日発行）
発行：大学共同利用機関法人　情報・システム研究機構
統計数理研究所　広報委員会
〒190-8562　東京都立川市緑町10-3
TEL 050-5533-8500（代）　https://www.ism.ac.jp/
制作：統計科学技術センター
●送付先変更は、Fax（042-526-4334）にてお知らせ下さい。
リサイクル適性Ⓐ
この印刷物は、印刷用の紙へリサイクルできます。