

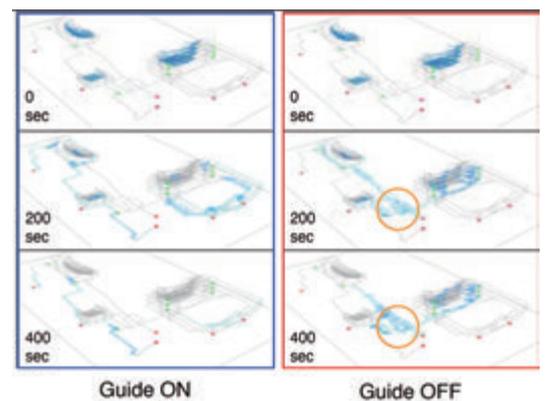
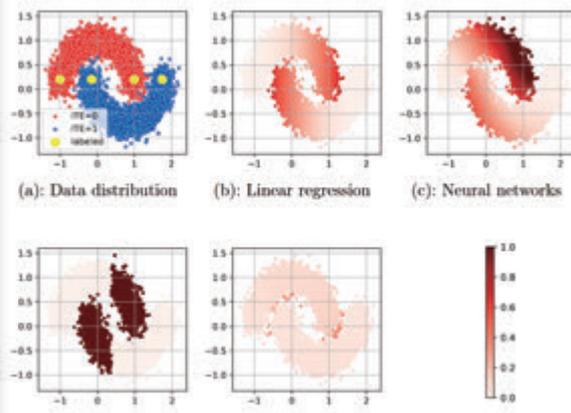
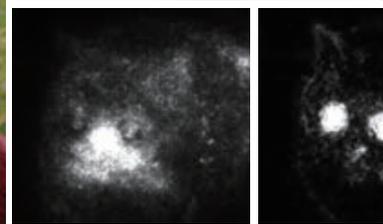
データ科学の、その先へ。

データ科学は、データから帰納的にモデル・仮説を構成する方法論として、近年その重要性が大きく増しています。

これは医療・教育・防災、農学・経済学・言語学などの他の学問分野に波及しており、またそれに伴って農業・製造業から流通・小売から金融・娯楽に至る様々な業界においてデータサイエンティストのニーズが高まっています。

2023年4月に新設されたデータ科学コースは、数理・データ科学・機械学習に関する教育研究拠点として、データ科学に関するトップクラスの専門家を育成します。

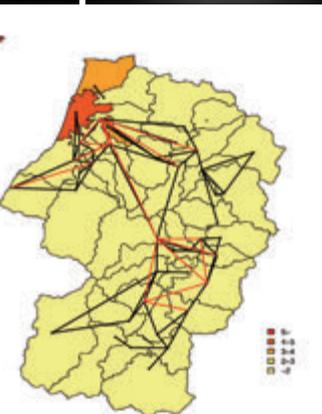
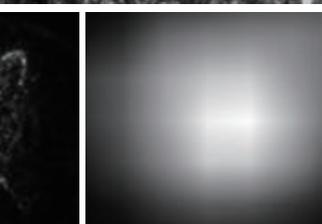
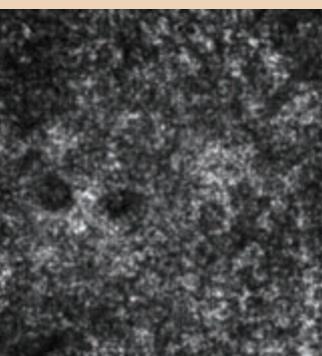
また、周辺分野の学生や研究者を積極的に受け入れることで新たな学際的な学術分野の創出に寄与します。



「データ科学」が注目を集めています。

情報通信技術の進展を基盤として、かつてない規模で多様なデータが収集、蓄積されています。また、それらのデータから帰納的に知見を引き出す試みが、様々な分野で行われるようになってきています。データからの帰納的推論を支える枠組みとしての統計学や機械学習の理論も様々な側面で新たな展開を見せており、これまでは扱いづかったタイプのデータが系統的に扱えるようになったり、新しいタイプのデータ解析が可能になったりしています。そして深層学習です。大規模なデータの蓄積が進んでいる画像やテキストなどの分野を皮切りに、多層ニューラルネットワークを適切に学習させることにより、今や様々なタスクで「人工知能」と見紛うばかりの優れた能力をもつモデルが構築されています。こうした深層学習の顕著な成果は、翻ってデータからの帰納的推論の枠組みに対して新たな問いを投げかけています。なぜ深層学習はこれほどまでにうまくいくのか？最適化理論によれば、深層学習で典型的に現れる凸でない目的関数の最適化は難しいはずですし、統計的学習理論では、多層ニューラルネットワークなどの複雑なモデルには高い汎化能力は期待できないとされており、いずれも深層学習がうまくいくことを説明できていません。この問いへの答を模索する中から、深層学習、ひいてはデータ科学の新たな体系や方法論などが生み出されていくことでしょう。

こんな刺激的なデータ科学の世界に、みなさんも飛び込んでみませんか。



田中利幸
TANAKA Toshiyuki

大学院 情報学研究科
データ科学コース 教授

1988年東京大学工学部電子工学科卒業。1993年東京大学大学院工学系研究科電子工学専攻博士課程修了、博士(工学)。1993年4月東京都立大学工学部電子・情報工学科助手。その後、同大学大学院工学研究科電気工学専攻助手、同講師、助教授、および首都大学東京システムデザイン学部システムデザイン学科情報通信システム工学コース助教授を経て、2005年10月より京都大学大学院情報学研究科システム科学専攻教授。2023年4月より現職。

概要

■ 分野一覧

分野名	担当教員
統計数理	下平 英寿 教授 本多 淳也 准教授
統計推論	原 尚幸 併任教授 (国際高等教育院)
情報論の学習	田中 利幸 教授 小淵 智之 准教授
信号情報処理	林 和則 併任教授 (国際高等教育院)
応用機械学習 (知能計算)	山本 章博 教授 瀧川 一学 特定教授 (国際高等教育院) FLANAGAN Brendan John 特定准教授 (国際高等教育院)
応用機械学習 (集合知システム)	鹿島 久嗣 教授 竹内 孝 講師
医療健康データ科学	田村 寛 併任教授 (国際高等教育院)
計算知能システム (連携ユニット)	岩田 具治 連携教授 (NTTコミュニケーション科学基礎研究所)

■ データ科学コースカリキュラム

博士(情報学)		
博士論文		
3年	研究指導	
2年		
1年		
コース開設科目(計6単位)		
	データ学科特別セミナー E	
	統計モデリング特別セミナー A、B E 機械学習特別セミナー A、B E	
	応用データ科学特別セミナー A、B E (各2単位)	
修士(情報学)		
修士論文		
コース開設科目(他コース開設のコース推奨科目を含む)		
2年	研究指導科目 (必修10単位)	
1年		
コース専門科目		
	データ科学セミナー(修士)(必修2単位)	
	統計的信号処理論、統計的学習理論E、計算論的学習理論E、統計的システム論、情報論的システム論(各2単位)	
他コース開設の推奨科目		
	(シス)計算知能システム論 他28科目	
コース基礎科目		
	統計科学基礎論(2単位)	
	デジタル変容実践論、データの二次利用実践論(各1単位)	
研究科共通科目		
プラットフォーム学展望(2単位)		
研究科共通展望科目 (選択必修2単位以上4単位以下)		
	計算科学入門(2単位) 計算科学演習A(1単位)	
	情報と知財(2単位) イノベーションと情報(2単位)	
	情報学展望1 情報学展望2 情報分析・管理論(2単位) 情報分析・管理演習(1単位)	
	情報学展望3E 情報学展望4E 情報学による社会貢献E(1単位)	
	情報学展望5E (各2単位) 情報学におけるインターンシップE(1単位)	
研究科が提供する その他科目		
	データ科学特殊研究2E (修士2年、5単位)	
	システム科学特殊研究1E (修士1年、5単位)	
入学前	学部で学習する程度の 各自のコース学術基礎 等	
微積分	線形代数	確率統計

※Eと記された科目は英語だけでも修得可

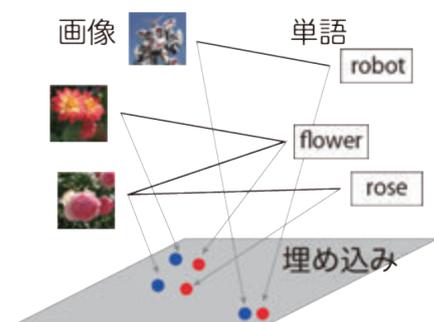
統計数理分野

統計学と機械学習の理論と応用

ビッグデータ、データマイニング、人工知能の流行を支える理論的基盤として統計学は重要な役割を果たしています。ランダム性を考慮してデータから帰納的推論を行う方法論を提供することが統計学の大きな特徴です。これを基礎として、大量のデータからの

情報抽出から少量のデータ下での意思決定まで機械学習は近年急速に発展しています。この転換期において、数学とプログラミングを原動力に現実のデータに向き合うことで新たな方法論の発展を目指します。

[下平 英寿・本多 淳也]



ニューラルネットを用いたグラフ埋め込みでデータの次元削減を行い、画像や単語の相互検索を行う。

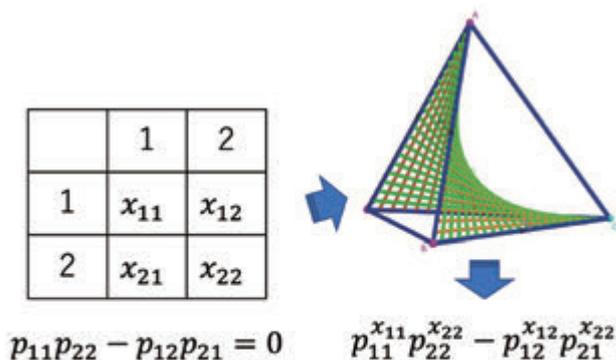
統計推論分野

ビッグデータ時代の統計的推測の数理と応用

ビッグデータ、機械学習、人工知能の隆盛により、高次元複雑システムにおける推測理論や推測アルゴリズムの重要性が高まっています。大きいサイズのデータの入手は可能になりましたが、解明したい現象の複雑さはそれを超えているため、信頼のある推測を行うためには小標本においても安定的な推測手

法の開発が必要になります。古典的な推測手法だけでなく、組合せ論、代数幾何学、最適化法などのさまざまな数理を用いて、高次元推測手法の開発と、考古学・人類学・認知科学などの分野への応用に関する研究・教育を行います。

[原 尚幸]



多くの統計モデルは代数多様体と解釈でき代数的な量が統計的推測に重要な意味をもたらす。

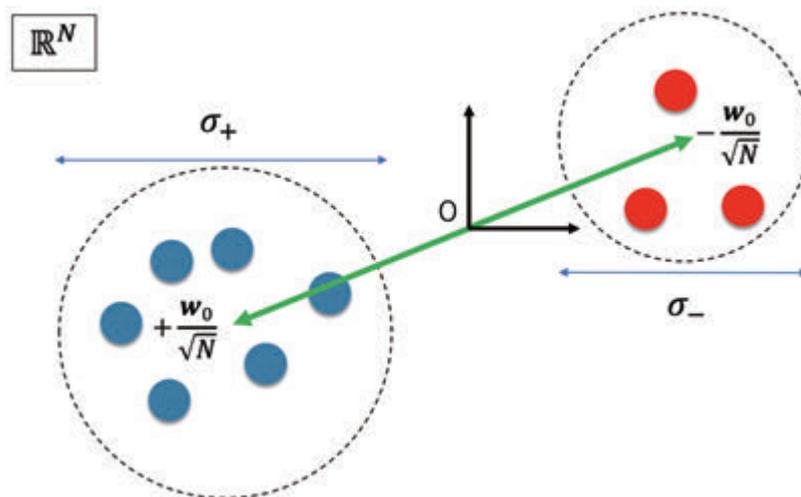
概要

情報論的学習分野

機械学習の情報数理

近年の深層学習の進展は目覚ましく、様々なアイデアが生まれ、多様なモデルが提案されています。また、深層学習だけでなく、データ科学の多様な課題を扱うことを目的に、機械学習の様々な手法が開発されています。これらの手法がなぜ有効にはたらくのか、どのような性質をもつのか、わかっていないことも少なくありません。情報理論や不規則系の統計力学などの知見も援用しつつ、そのしくみを明らかにすることを目指しています。

[田中 利幸・小淵 智之]

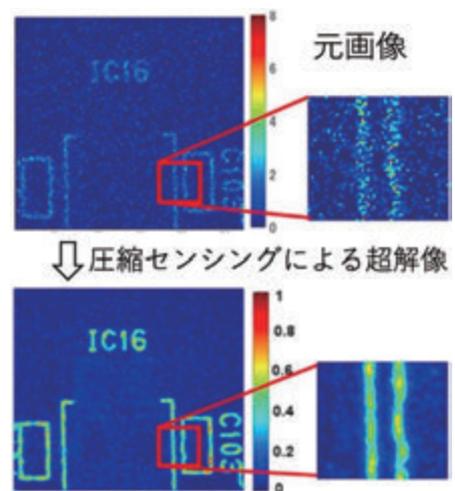


信号情報処理分野

統計的信号処理とモデルベース機械学習

信号情報処理分野では、観測された生の信号やデータから有益な情報を抽出するための方法論として、モデルベースのアプローチである統計的信号処理とデータ駆動のアプローチである機械学習を融合した、モデルベース機械学習に関する研究を行っています。無線通信や光ファイバ通信、IoTなどの情報通信分野を中心に、計測、画像処理、生体信号処理など幅広い工学の問題に取り組んでいます。

[林 和則]



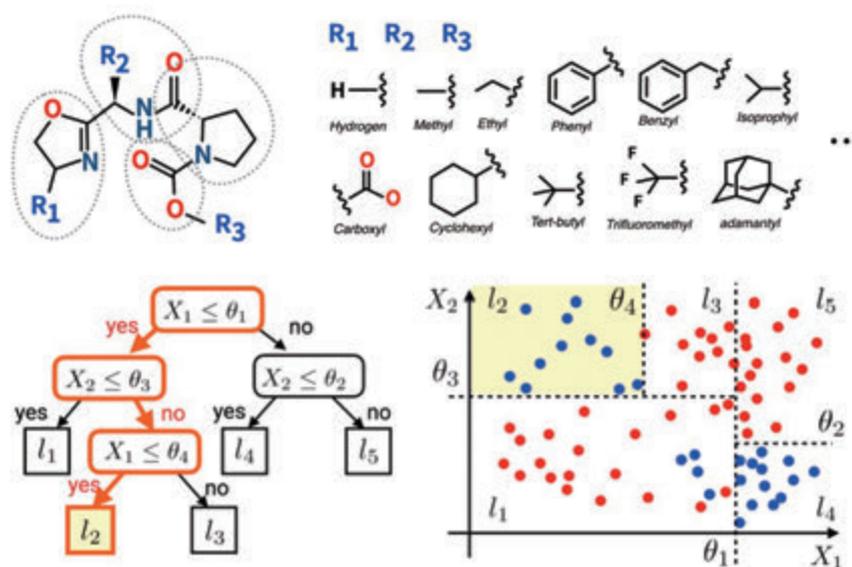
蛍光X線計測のための圧縮センシングによる超解像の例

応用機械学習分野（知能計算）

実問題を解くための機械学習・データ科学の実践

応用機械学習分野(知能計算)では、自然科学および教育学における実問題・実課題の解決のための機械学習・データ科学の実践研究を行っています。生命科学や化学では、ゲノム配列・分子構造・分子間相互作用など離散的な組合せ構造を機械学習に反映する必要があり、そのための技術研究にも取り組んでいます。教育学では私たち自身の学習の効果・効率を向上させる教育情報基盤や方法の開発に取り組んでいます。科学的発見・理解を通じた具体的実践、記号論理・知識表現・アルゴリズム設計など一般の離散構造処理との融合、人間の学習と機械の学習の対比を通じて、新たな知能計算の基礎を拓くことを目指しています。

[山本 章博・瀧川 一学・FLANAGAN Brendan John]

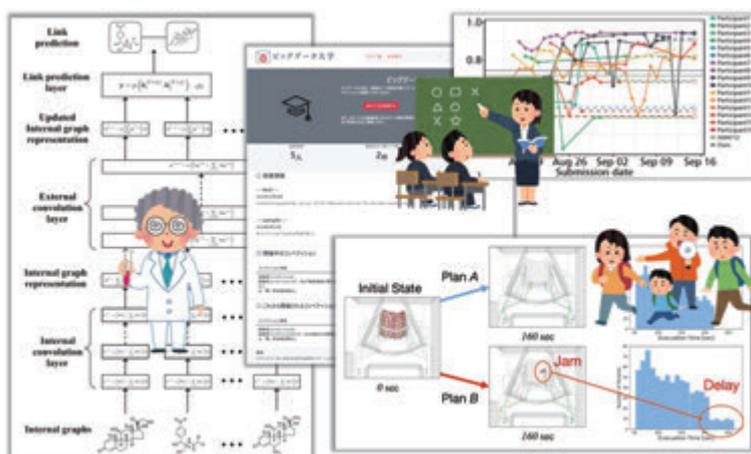


応用機械学習分野（集合知システム）

データ解析技術の新たな応用を開拓

応用機械学習分野(集合知システム)では、統計的機械学習やデータマイニング手法の研究開発を行うとともに、ヘルスケア、バイオ・創薬、教育、交通、マーケティングをはじめとする様々な分野の重要課題を、データ解析の立場から解決します。企業や国・自治体などの様々な協力者とともに新たなデータ解析技術の応用を開拓し、実世界でインパクトのある課題を解決することを目指します。

[鹿島 久嗣・竹内 孝]



様々な分野の課題をデータ解析技術を駆使して解決します。

概要

医療健康データ科学分野

医学領域での下に示すような研究を主に医学部附属病院内で行っています。

研究テーマ

■画像情報や遺伝情報も含む臨床データを活用した疫学研究

主に観察研究を主体とした臨床研究で、マルチモーダルな画像情報や遺伝情報を組み合わせ、病態解明・予後予測につなげる研究

■ナショナル・データ・ベース(NDB)を含むレセプトデータを活用した疫学研究

日本再興戦略で示されたデータヘルス計画の中核を担う、レセプトデータの二次活用などに基づいた疫学研究

■病院経営改善を指向した病院情報システムの改善、新システムの実装に関する研究

病院経営改善を目指し、病院情報システムの改修・改善・二次活用に、IoTや人工知能等の新しい技術も絡めて、課題の整理・実装検証を行う研究

■視線分析型視機能評価機器開発に関する研究

アイトラッキングとAIを実装した「視線分析型視機能評価機器」を用いた、認知機能も含めた多様な視機能評価システムの実装を目指す研究

[田村 寛]



計算知能システム連携ユニット

多様なデータからの知識創生をめざして

深層学習の発展により、機械学習手法は画像・言語処理などの分野で高い性能を達成しています。しかし、大量の良質な学習データが手に入らない場合は性能が低いという問題があります。機械学習をより広い分野で利用可能にするため、データが少量しかない場合や良質なデータが得られない場合など、望ましくない状況においても高い性能を達成できる機械学習手法の研究に取り組んでいます。具体的には、学習の仕方を学習するメタ学習、帰納バイアスを活用した機械学習などの研究を進めています。

[NTTコミュニケーション科学基礎研究所連携：岩田 具治]