データ科学コース

Data Science Course

https://www.ds.i.kyoto-u.ac.jp/

コース概要

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。

データ科学コースではデータ科学に関する研究拠点として、修士課程および博士後期課程の教育・研究指導を通してトップクラスのデータ科学に関する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。

本コースに配属された学生は、データモデリング・機械学習・信号処理・パターン認識などのデータ科学の中核となる手法に重点を置いて学びつつ、他コースの提供する科目を含め関連する実装・応用に関しても幅広く履修することができます。また、学内外の研究者によって最新のトピックが提供されるセミナーによって最先端の研究動向についても学ぶことができるほか、データ科学関連の研究所・企業等におけるインターンシップも一定の要件のうえ単位認定が行われるため、基礎領域から応用まで幅広い分野に対応した専門知識を学ぶことができます。

志望区分	研究室名
デ－１	統計数理
デ－２	統計推論
デ－３	情報論的学習
デ－４	信号情報処理
デ－５	応用機械学習（知能計算）
デ－６	応用機械学習（集合知システム）
デ－７	医療健康データ科学
デ－８	計算知能システム（連携ユニット）

Application code	Group
DS-1	Statistical Mathematics
DS-2	Statistical Inference
DS-3	Information-Theoretic Learning
DS-4	Signal and Information Processing
DS-5	Applied Machine Learning (Computational Intelligence)
DS-6	Applied Machine Learning (Collective Intelligence)
DS-7	Medical and Healthcare Data Science
DS-8	Computational Intelligence Systems (Adjunct unit)

統計数理分野　志望区分：デ－1

教授 下平英寿　准教授 本多淳也　助教 山際宏明

統計学，機械学習，データサイエンスの手法と理論を探求

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・機械学習は重要な役割を果たしています。とくにランダムネス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。研究室では、人工知能やデータサイエンスの基礎となる統計学や機械学習を融合する形でのデータ駆動型の帰納的推論の探求、すなわち「統計的方法論によるデータからの理解や思考の探求」を目指しています。

現実のデータにとりくんで，新たな理論を作る

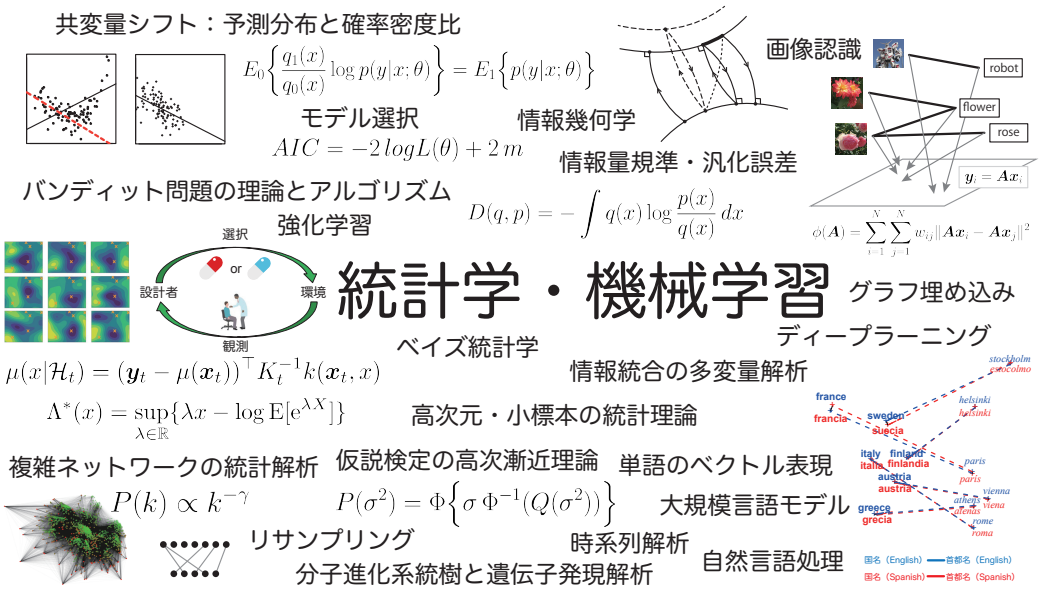
かつて遺伝学において R. A. Fisher が統計学を飛躍的に発展させたように、現実と向き合うことが方法論の発展をもたらします。近年ではデータ収集や処理が容易になり、大量のデータを扱える方法論の必要性が増しています。そして Transformer のようなニューラルネットによる言語モデルを学習させることによって理解や思考といった知能が実現しつつあり、ChatGPT などの普及によってもはや身近なものとなっています。言語に限らず画像やロボティックス等の様々な分野でこのような生成モデルによる革新が進行中です。しかし、ニューラルネットがなぜ優れた性能を発揮するのかについては未解明の部分が多く、さらなる発展のためにはその仕組みの理解が重要です。

研究室では、たとえば「自然言語処理のニューラルネットにおいて単語の意味がどのように表現されているか」を数理統計学の観点で解明して、言語モデルにおける「思考」を理解するための研究を行っています。また、ベイズや頻度論といった統計的方法論の基本的な研究に取り組んでいます。新薬の開発や新商品の推薦といった試行錯誤を要する動的意思決定問題にも取り組んでおり、バンディット問題とよばれる定式化を通じた理論限界の追求から強化学習を用いた実用的なアルゴリズムの構築まで理論・実践の両面から研究を行っています。

数学とプログラミング，どちらも重要

研究で最も重要なのはアイデアとデータです。そして数学とプログラミングは力です。定理の証明とコーディングは似た作業ですね。数学に自信のある人、Python，R，C++のスキルがある人は活躍するチャンスがあるし、やる気さえあれば研究を通して実力はつくものです。

最近の活動は研究室ウェブサイトを御覧ください　`http://stat.sys.i.kyoto-u.ac.jp`



統計推論分野　志望区分：デ－２

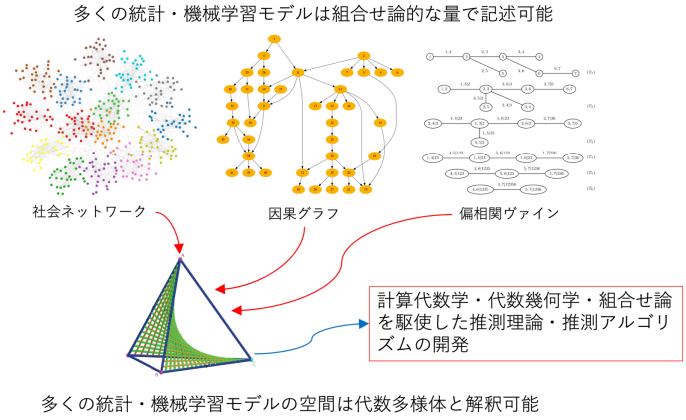
教授（国際高等教育院）原　尚幸

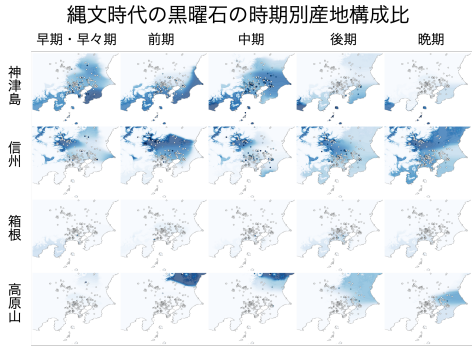
統計学・データ科学の数理

統計学・データ科学は，不確実性を伴う現象の帰結として得られるデータから，その現象のメカニズムを帰納的に解明するためのツールとして注目されてきました。近年では，高次元のデータから，複雑な現象を解明するための統計的推測手法の開発が望まれています。当研究室では，高次元の統計モデル・機械学習モデルの統計的推測理論や，推測アルゴリズムに関する研究を行っています。統計的因果モデルや社会ネットワークモデルをはじめとする多くの実用的な統計モデルは，変数間の相関や因果などの関係をグラフやネットワークのような組合せ論的な量によって記述したモデルと解釈することができます。このようなモデルの推測理論・推測アルゴリズムに関して，従来の数理統計学・計算機統計学の枠を超えて，代数幾何学，組合せ論，最適化法など，数理の異分野の知見も駆使して研究を行っています。最近は，ニューラルネットワークを用いた統計的推測手法の実装法や，関連するセミパラメトリック推測統計の理論研究 (Double Machine Learning や Targeted Learning など) にも興味をもっています。

異分野融合研究を通した新たな方法論の開拓

統計学・データ科学は，異分野との融合，新たな応用分野の開拓を通して発展してきた歴史があり，今後もそうあるべきと考えます。当研究室では，さまざまな異分野との融合研究を通して，統計学・データ科学の社会における役割を理解しながら，新たな方法論の提案を目指します。現在は，考古学・人類学分野における遺跡出土品の産地構成比の時空間的推定問題と，その古代の生活様式やモノの流通方式の時空間的変化や古人口の推定問題への応用，　認知科学分野 (心理学分野) における人の動きに関する高次元時系列データからの特徴抽出，源氏物語の写本の挿絵と粉本のマッチングによる類型化などの異分野融合研究を行っています。





機械学習モデルの考古学への応用
同志社大学津村研究室との共同研究

キーワード：グラフィカルモデル，　統計的因果モデル，　ヴァインコピュラモデル，　時空間モデル，ベイジアンネットワーク，機械学習，　多変量推測統計学，　ベイズ推測，計算代数統計学，ニューラルネットワーク，人文社会科学研究との融合 etc

問い合わせ先：
京都大学吉田南 1 号館 307 号室 原 尚幸
hhara@i.kyoto-u.ac.jp

情報論的学習分野　　志望区分：デ－３

教授 田中利幸　准教授 小渕智之

―様々な情報処理システムを数理的な観点で横断的に捉える―

不確実性がある環境下での情報処理にまつわる多様な課題を，数理的な観点で横断的に捉えることで解明していくことを目指します．今日の情報処理の多くの課題は大規模な確率モデルにもとづいて定式化でき，確率モデルの大規模さから立ち現れる法則性を活用することが，高度な情報処理を実現する鍵となります．大規模な確率モデルの情報数理を統計力学との類比で議論する情報統計力学や，統計的機械学習や深層学習，データ科学に関わる理論的諸問題の検討などの主題群に取り組んでいます．

確率モデルにもとづく情報処理

不確実性を有する環境から意味のある情報をどう効率的に取り出すか，という問題は，機械学習をはじめとして多くの情報処理に現れる理論的課題です．この課題に対して近年注目を集めている，環境の不確実性を確率モデルによって記述し，それに基づいて推論，学習，適応を行う方法論について，統計科学，情報理論，統計力学，情報幾何学等の立場から領域横断的，多面的に研究を行っています．

圧縮センシングの数理と応用

「圧縮センシング」は，データの「スパースさ」に着目した新しい情報処理の枠組みとして，数理と応用の両面から近年大いに注目を集めています．圧縮の理論的限界の解明をはじめとして圧縮センシングの数理的性質を明らかにするための研究を行っています．さらに本学医学研究科や薬学研究科の研究室等と共同で，圧縮センシングの応用による MRI などの医用画像処理やメタプロテオミクスの高性能化に関する研究も行っています．

深層学習の数理

深層学習が様々な分野で注目を集めています．深層学習は，仮説に対する柔軟な表現系と，実用的な学習手段とを提供していますが，なぜ深層学習はうまくいくのか，深層学習に関連してなされている様々な提案はなぜ有効なのか，という問いには，満足な解答が得られていないものがたくさんあります．深層学習に関して数理的な観点からの研究に取り組んでいます．

確率モデルにもとづく情報処理に関する研究成果の検証を主な目的として，空間データ解析，多次元データのクラスタリング，通信における情報推定，強化学習，ゲーム理論，などの具体的な問題群についても研究を行っています．我々の研究活動の特徴は「領域横断的であること」だと考えています．現在取り組んでいる研究テーマも，多岐に渡る学問分野（情報理論，通信理論，確率論，統計科学，統計力学，機械学習など）に関連しています．

研究室ホームページ　http://www-adsys.sys.i.kyoto-u.ac.jp/

信号情報処理分野　　志望区分：デ－4

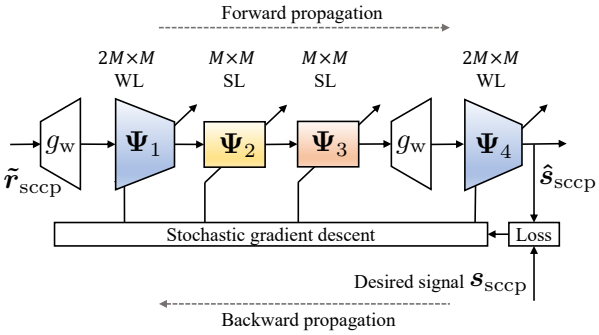
教授（国際高等教育院）　林 和則

概要

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を体系化した統計的信号処理を武器に，主に移動体通信や光ファイバ通信，IoT などの情報通信分野を中心としたシステムの問題に取り組んでいます．最近の研究テーマの例を以下に記します．

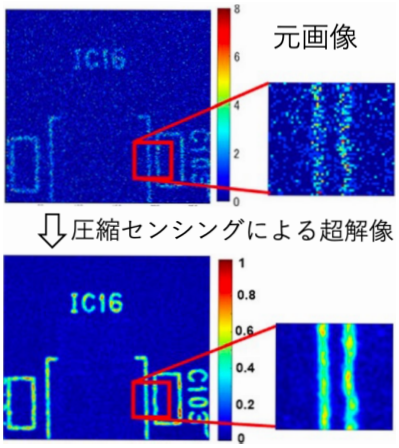
モデルベース機械学習に関する研究

様々な分野の問題において AI や機械学習の活用が検討されています．無線通信や光通信などの情報通信分野もその例外ではありません．情報通信のための信号処理では繰り返しアルゴリズムがよく利用されますが，繰り返しアルゴリズムはその処理を時間軸方向に展開することでニューラルネットワークによく似た構造をもつ計算グラフが得られます．この計算グラフ内の各ステップの処理が微分可能であれば，計算グラフに対して誤差逆伝播法や確率的勾配法を適用することで，そこに含まれるパラメータを最適化することができます．このアプローチは**深層展開**と呼ばれ，従来からのドメイン固有の知識を利用しつつ，いくつかのパラメータをデータに基づいて決定することで大幅な性能改善が期待できるため大変注目されています．本研究ではこれまで情報通信分野で蓄積されてきた知見を活かしつつ，データに基づく機械学習技術を利用することで新しいアルゴリズムや信号処理手法を開発しています．



スパースモデリングに関する研究

圧縮センシングはスパースな未知ベクトルをその見かけの次元よりも少ない線形観測から再構成するための理論的な枠組みで，スパースモデリングとも呼ばれます．本研究では未知ベクトルがスパースのときだけでなく，**各成分が離散値をとるベクトル**の場合にも劣決定の線形観測から実際に観測された未知ベクトルを再構成可能であることを明らかにし，通信路等化や MIMO 信号検出，IoT 信号検出など様々な無線通信の問題に適用しています．さらに，圧縮センシングの適用可能領域を拡大するために，高速かつ低消費電力の光アナログデバイスを利用した**アナログ圧縮センシング**の実現に向けた研究も行っています．



研究室の web ページ　**https://kazunorihayashi.github.io**
問い合わせ先　**hayashi.kazunori.4w@kyoto-u.ac.jp**

応用機械学習(知能計算分野)

教授 山本　章博　　特定准教授(国際高等教育院)　FLANAGAN, Brendan

志望区分：デ－5

本年度は募集せず.

応用機械学習（集合知システム分野）

教授　鹿島 久嗣　　講師　竹内 孝　　助教　新 恭兵

志望区分:デー6

概要

当分野ではデータ解析を基礎とした人工知能技術の発展と、これらの実世界への還元を目指して研究を行っている。特に、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、様々な分野の課題解決に向けデータ解析の立場から取り組む。また、人工知能だけでは解決できない困難な課題に対して、人間と人工知能の両者の力を合わせることで解決するための方法論の研究も行う。

研究テーマ

・新しい機械学習・データマイニング手法の開発：
機械学習は近年著しい発展を遂げているとはいえ、その適用範囲は未だ限定的であり、現実世界で直面する様々な課題やデータ形式に対して適用できない場合も多々ある。このような新しい機械学習の問題設定を発見し定式化を行うとともに、アルゴリズムの開発を行う。たとえば、グラフなどの複雑な構造や時空間的な広がりをもつデータを対象とした予測モデリングや、因果関係に基づく意思決定の手法開発などが挙げられる。

・機械学習・データマイニングの先進的応用：
マーケティングやヘルスケアをはじめとする様々な分野で利用されている機械学習技術であるが、現実世界にはまだ多くの未開拓の応用先が潜んでいる。企業や国・自治体など様々な協力者とともに新たな機械学習の応用を見つけ取り組み、データ解析技術を実世界でインパクトのある課題の解決に結び付けることを目指す。

・ヒューマン・コンピュテーション：
最近、様々な知的タスクにおいて人工知能が人間を凌駕する力を示している一方で、抽象的で複雑な問題解決や臨機応変な対応など、人工知能だけでは成しえない困難なタスクが未だ多く存在する。人工知能には困難な、さらには一人の人間では解くことのできない問題を、両者を適切に組みあわせて解決する「ヒューマン・コンピュテーション」の世界を開拓することで、人工知能の新たな展開を目指す。

問合せ先

京都大学総合研究 7 号館 304 号室 鹿島 久嗣 (Tel: 075-753-5394)
電子メール：kashima@i.kyoto-u.ac.jp

医療健康データ科学

教授　田村寛

https://ds.k.kyoto-u.ac.jp/member/tamura/

志望区分:デー7

概要

国際高等教育院附属データ科学イノベーション教育研究センター（CIREDS）所属の田村寛教授の研究室です。21世紀の基礎教養としての情報学・統計学を基盤とするデータ科学の教育と研究を行う組織に属します。医師でもあるため、医学領域での経験が長く、健康関連データ科学、医療DX、視機能のデータ科学、データヘルス等の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

主に眼科領域での観察研究を主体とした臨床研究で、発展著しいマルチマダリティー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を行っています。

・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

平成25年閣議決定の日本再興戦略で掲げられている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を行っています。

・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

病院経営改善計画の中で、病院情報システムの改修・改善・二次活用は重要性が増しており、IOTや人工知能などの新しい技術を実務に取り入れるための課題の整理・実装検証に関する研究を行っています。

・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を分析に実装した「視線分析型視機能評価機器」では、視野を中心に視力や眼球運動といった複数の視機能の評価を目指している。また、再現性高く他覚的に評価できるため、視機能評価研究の促進も同時に見込んでいる。さらに、アイトラッキング結果等のデジタルフェノタイプと認知機能の相関を明らかにし、認知症及び軽度認知機能異常（MCI）のスクリーニングを行う医療機器としての開発も目指しています。

問い合わせ先

〒606-8507　京都市左京区聖護院川原町54　京都大学医学部附属病院（第2駐車場北隣）
先端医療機器開発センター 404号室
視線誘導型視野計開発プロジェクト
学内内線19-87666
E-mail:　htamura@kuhp.kyoto-u.ac.jp

Medical and Healthcare Data Science Group

Professor: Hiroshi TAMURA

https://ds.k.kyoto-u.ac.jp/member/tamura/ Application Code: DS-7

Description

Professor Hiroshi TAMURA, belonging to the Center for Innovative Research and Educaion in Data Science (CIREDS), Institutes for Liberal Arts and Sciences, leads the laboratory. CIREDS is an organization that conducts education and research in data science based on informatics and statistics as a basic education for the 21st century. As he is also a physician, he has a long experience in the medical and healthcare field and is mainly engaged in research on halthcare-related data science, healthcare digital transformation, data science in visual function, Data Health. The research projects are mainly conducted within the Kyoto University Hospital.

CIREDS
Center for Innovative Research and Education in Data Science

Research Topics

Epidemiological studies using clinical data, including imaging and genetic information:
Clinical research, mainly observational research of ophthalmology, that combines rapidly developing multimodality imaging and genetics to elucidate pathological conditions and predict prognosis.

Epidemiological studies using claims data including the NDB:
Epidemiological research based on the secondary use of claims data, etc., which plays a central role in the Data Health Plan set forth in the Japan Rebuilding Strategy approved by the Cabinet in 2013.

Research on the improvement of hospital information systems and the implementation of new systems aimed at improving hospital management:
The renovation, improvement, and secondary use of hospital information systems are becoming increasingly important in hospital management. Research on organizing and verifying the implementation of new technologies such as IOT and artificial intelligence into practice.

Research on the development of visual function evaluation devices based on gaze analysis:
"Gaze Analyzing Perimeter," which implements eye tracking and AI for analysis, aims to evaluate multiple visual functions such as visual acuity and eye movement, focusing on the visual field. It is also expected to promote research on the evaluation of visual functions with objective and highly reproducible method. Furthermore, by elucidating the correlation between digital phenotypes such as eye tracking and cognitive functions, we aim to develop a screening system for dementia and mild cognitive impairment (MCI).

Contact Office for the Gaze Analyzing Perimeter Project
#404 Clinical Research Center for Medical Equipment Development (CRCMeD),
Kyoto University Hospital.
Shogo-in Kawahara-cho 54, Kyoto, 606-8507 Japan.
Telephone only within Kyoto University: 19-87666
E-mail: htamura@kuhp.kyoto-u.ac.jp

計算知能システム(NTTコミュニケーション科学基礎研究所連携)

志望区分：デー８
連携教授：岩田具治

概要

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．

研究テーマ

・メタ学習

通常の機械学習手法では，あるタスクで学習したモデルは，そのタスクでのみ利用できます．新しいタスクが与えられたときは，そのタスクに関する大量のデータを集め，再度モデルを学習する必要があります．この問題を解決するアプローチとしてメタ学習が近年注目されています．メタ学習では，新しいタスクが与えられたとき少量のデータでも性能が高くなるように，様々なタスクから学習の仕方を学習します．メタ学習を活用することにより，人間のように，多様な経験から学習し，未知のタスクに対しても少ない試行錯誤で高い性能を達成できる機械学習手法の実現を目指しています．具体的には，メタ学習に基づく少量データでの空間解析や異常検知の研究に取り組んできました．また，人間のように言葉も活用して学習を効率的に行うメタ学習の研究も行っています．

・帰納バイアスを活用した学習

ニューラルネットワークは任意の関数を近似することができます．その柔軟性ゆえに大量の学習データが必要となります．そこで，タスクに関する事前知識を帰納バイアスとしてニューラルネットワークに導入することにより，少ないデータでも適切な関数を学習することを目指しています．例えば，データの生成過程をニューラルネットワークに組み込むことや，入出力が特定の制約を満たすようなニューラルネットワークを設計することで，帰納バイアスを導入できます．また，生成過程を組み込むことにより，データに潜む隠れた構造を抽出する研究も行っています．