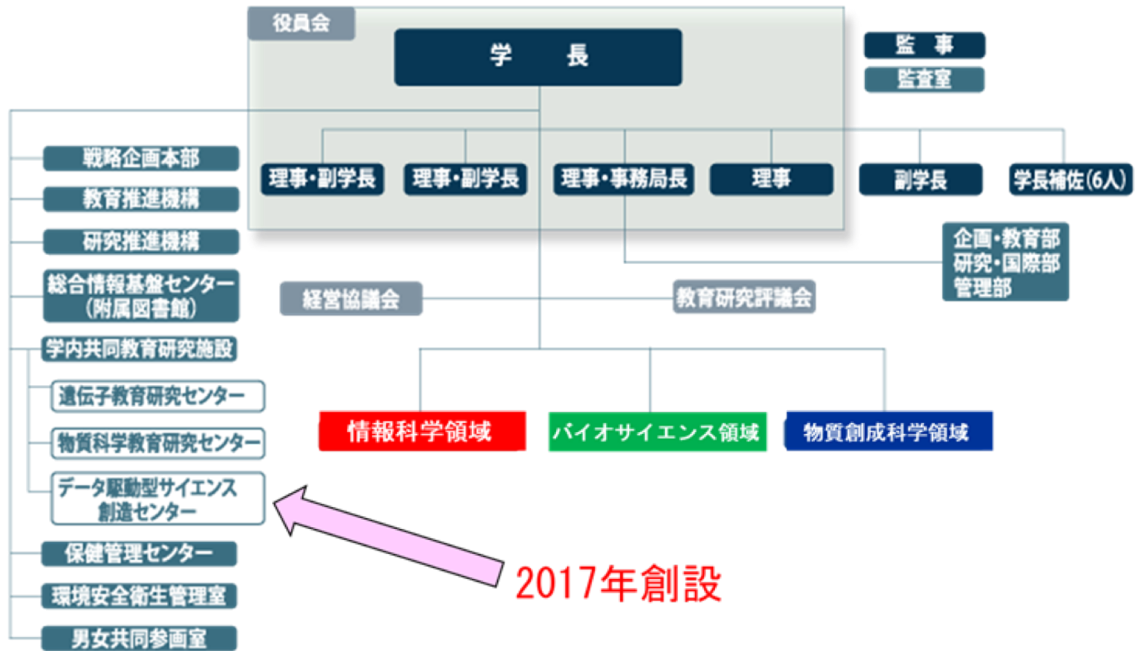


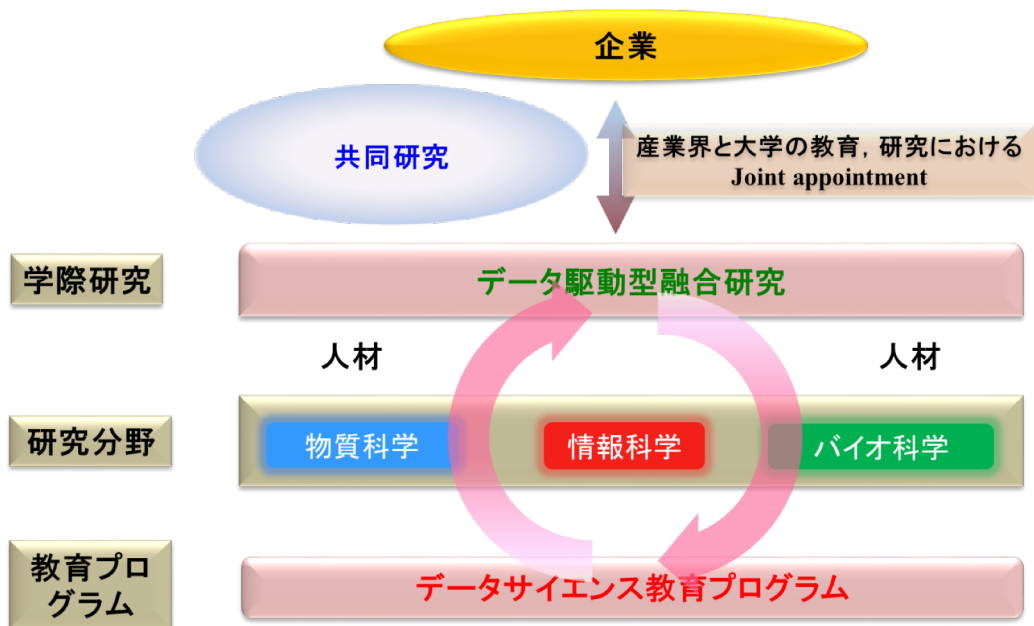
データ駆動型サイエンス創造センターの紹介



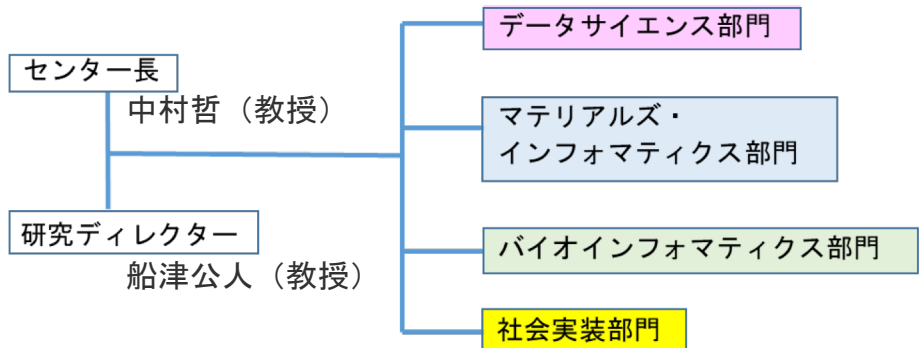
奈良先端大の組織図

データ駆動型サイエンス創造センターの取組

奈良先端科学技術大学院大学は、データ駆動型サイエンスを情報・バイオ・物質及びその融合領域に横断的に展開し、これらの領域の深化と新しい学際融合領域の開拓を行うため、2017年に、データ駆動型サイエンス創造センターを創設しました。データ駆動型サイエンス創造センターは、先端的研究の展開と、社会の要請に応える柔軟な教育を推進していきます。



データ駆動型サイエンス創造センターの組織図と構成員



◇データサイエンス部門

データ駆動型サイエンスの基盤的研究や人工知能（深層学習、推論モデル）を中心にデータ駆動型サイエンスの基礎となる理論を開拓します。

- 知能コミュニケーション研究室
中村哲（教授）、須藤克仁（准教授）、安田圭志（特任准教授）、田中宏季（助教）
- 計算システムズ生物学研究室
小野直亮（准教授）
- 自然言語処理学研究室
渡辺太郎（教授）、進藤裕之（助教）
- 数理情報学研究室
池田和司（教授）、福嶋誠（助教）
- 生体医用画像研究室
佐藤嘉伸（教授）
- ソーシャル・コンピューティング研究室
荒牧英治（教授）、若宮翔子（准教授）

◇マテリアルズ・インフォマティクス部門

データ駆動型の手法を物質創成科学に適用し、プロセスまで含めた新材料の探索や新規機能開発を行います。

- データ駆動型化学研究室
船津公人（教授）、宮尾知幸（准教授）、Jasial, Swarit（特任助教）
- 情報機能素子科学研究室
浦岡行治（教授）、上沼睦典（准教授）、藤井茉美（助教）、Bemundo Juan Paolo Soria（特任助教）
- ナノ高分子材料研究室
網代広治（教授）
- 凝縮性物性物理学研究室
松下智裕（教授）、武田さくら（助教）

◇バイオインフォマティクス部門

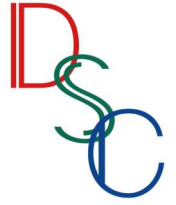
分子生物学の発達で急増してきた生物学データの解析にデータサイエンスの手法を適用し、生命現象の統計的な解明とその応用を目指します。

- システム微生物学研究室
森浩禎（教授）、武藤愛（助教）
- 分子医学細胞生物学研究室
末次志郎（教授）
- 計算生物学研究室
作村諭一（准教授）、国田勝行（助教）

◇社会実装部門

データ駆動型サイエンスの研究成果の社会的な実装と活用への積極的な取組を行います。

- 野島秀雄（特任教授）



NAIST データ駆動型サイエンス実習シリーズ 3:

「バイオデータサイエンスにおける奈良先端大における取組み」

- 日 時： 2020年8月28日（金）13:00-18:00
- 場 所： オンライン開催（Zoom）
- 参加費： 無料
- 主 催： 奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター
- 協 賛： データ人材育成関西地区コンソーシアム

■プログラム:

12:30 Online registration

13:00-13:10 挨拶・データ駆動型サイエンス創造センター概要
中村 哲（データ駆動型サイエンス創造センター長）

13:10-13:30 「コロナに負けず DB を作ろう！『テレワークでつくる KNApSAcK DB』」
金谷 重彦（計算システムズ生物学）

生物種とその生物種が生産する二次代謝物の関係を収録した KNApSAcK Core DB の開発を進めている (http://www.knapsackfamily.com/KNApSAcK_Family/)。意外にも、今回のコロナ禍の中、DB への収録を加速することに成功した。その成果を報告する。コロナ禍での大学での研究について多様なご意見をいただきたい。

13:30-13:50 「自然言語処理による知識処理」
渡辺 太郎（自然言語処理学）

科学技術に関する膨大な知識や情報の多くは学術論文として蓄積されていますが、論文の内容を理解し、そこから必要な情報を探すためには膨大な手間と時間を要します。生命科学や材料科学分野を対象として、論文に含まれる図表や実験条件・実験プロセスに関する記述を解析し、専門分野の知識の体系化や高度な論文検索を支援する自然言語処理技術を紹介します。また、グラフ構造として体系化された知識にもとづいて、論文など専門的な文書を要約あるいは翻訳する、といった応用分野における取り組みについても紹介します。

13:50 – 14:10 「臨床脳科学における数理情報学研究」
池田 和司（数理情報学）

データ科学の応用分野の中でも、脳情報科学はデータサイズが比較的小さく、何らかの工夫が必要となる。ここでは数理モデルを構築するてんかん研究と機械学習によるうつ病研究を紹介する。

14:10 – 14:30 「深層学習による生体医用画像解析」
佐藤 嘉伸（生体医用画像）

深層学習（人工知能：AI）による生体医用画像からの解剖構造の認識、AIによるフェイク画像生成技術の医療・バイオへの応用について、CTやX線画像を用いた応用事例を中心に説明する。特に、AIの問題である、ブラックボックス的性質や教師付き学習データ作成の手間への対応について、焦点を当てる。

AIのブラックボックス的な性質への一つの対応として、結果の信頼性（すなわち、AIの結果をどの程度信じてよいか）を出力するAI技法を紹介する。具体的に、Bayesian U-netと呼ばれる画像認識AIの枠組みを用いて、CT画像から解剖構造（筋骨格や腹部臓器等）の領域を認識する際に、結果の不確実性を数値化し、結果の精度を予測可能であることを示す。これにより、例えば、AIが「自信がない」と判断する（不確実性が高い）結果を、人間がチェックする等、することにより、人間とAIの効果的な共同作業を行える可能性が開ける。

もう1つの問題は、AIの構築に必要な「教師付き」学習データの作成の膨大な手間である。この問題に対して、ある撮影方法での画像（例えば、CT）による教師付き学習で構築したAIを、フェイク画像生成技術を利用して、「教師データ無し」で、他の撮影方法（MRI、X線）の画像認識に、効果的に拡張した事例について紹介する。

以上のAIによる生体医用画像解析への応用について、ベイズ推定を中心とした統計的予測の基礎・古典との関連を踏まえ、直感的理解を重視して説明する。

14:30 – 14:50 「基礎細胞生物学へのデータサイエンスの適用の試み」
末次 志郎（分子医学細胞生物学）

基礎生物学では、研究者の経験に頼る部分が大きく、特に細胞の写真の判定には研究者の経験が必要としてきた。我々バイオ研究者が、データサイエンスの専門家と共同で、基礎生物学にデータサイエンスを適用した例を紹介する。その結果、深層学習によるデータの分類は、これまでに蓄積のあった知見と整合性があることがわかった。今回は、培養細胞に薬剤処理を行い、薬剤の標的タンパク質の細胞内の分布を捉えた画像を得た。これらの画像を用いて、標的タンパク質の細胞内分布の変化が、深層学習により識別できることを確かめた。さらに、標的タンパク質と関連するタンパク質の細胞内分布が、薬剤処理により間接的に変化することが、深層学習により識別できるかどうか、調べた。その結果、「機能的」関連性に応じて、関連タンパク質の分布変化も、深層学習により識別できることがわかった。すなわち、深層学習による画像認識は、細胞生物学研究者がこれまで得てきた知見と概ね一致する結果を得ることができると考え

14:50 – 15:10 「オンチップ細胞検索エンジン」
細川 陽一郎 (生体プロセス工学研究室)

マイクロ流体チップで、1秒間に数1000個以上の細胞を個々に検出し、操作する技術(細胞検出エンジン)について紹介し、細胞より遥かに小さく難易度の高いウイルス等への適用の可能性とそのデータ解析について、議論したい。

15:10 – 15:30 「N-ビニルアミド系ポリマーによるガスハイドレード生成防止剤」
網代 広治 (ナノ高分子材料研究室)

天然ガスパイプラインを閉塞するガスハイドレートは、爆発の原因ともなるためにパイプライン中において結晶成長を抑制することは重要である。ガスハイドレートの結晶成長を阻害する化合物は、水にも油にもなじむ両親媒性高分子材料が効果を示す。我々はこれまでにN-ビニルアミド系ポリマーの様々な誘導体を分子設計してきた。ここではこれらのポリマーをガスハイドレート生成防止剤へ用いることを検討し、分子構造とガスハイドレート阻害能の関係を調べた。これまでにM位に様々な置換基を導入した新規モノマーを約10種類合成し、これらを用いて共重合体などの新規ポリマーを合成した。

ガスハイドレード生成防止能については、天然ガスおよびテトラヒドロフランを用いた結晶化実験により評価した。アルキル基を導入した場合には、N位に導入したイソプロピル基よりもノルマルプロピル基の方が高い効果を示すことが分かった。一方、ブチル基の場合では、イソブチル基の方がノルマルブチル基よりも効果が高い傾向が認められ、わずかな構造の違いが性能に影響することが分かった。オリゴエチレングリコール鎖を導入した場合には、水素結合により環状のコンフォメーションをとることが可能な置換基が高い効果を示すことが分かった。これらのことから、単純な親水性疎水性のバランスのみならず、主鎖から枝分かれまでの距離・コンフォメーションなど、かなり細部にわたる構造が性能にかかわることが示唆された。

15:30 – 15:50 Q & A

15:50 – 16:05 休憩

16:05 – 16:10 パネル討論会の主旨説明

- 1) 今後の奈良先端科学技術大学院大学・データ駆動型サイエンス創造センターとしての社会への還元に向けた接点の模索
- 2) Corona virus 関連での経験をもとに、現代社会の抱える問題とデータサイエンスからの解決への道筋

パネラーの自己紹介と研究内容の紹介

16:10 – 16:25 政府専門家会議：
「ソーシャルメディアを用いた COVID19 対策」
荒牧 英治（ソーシャル・コンピューティング）

今、医療が変わりつつあります。カルテなどに集積される医療ビッグデータ、それを用いた人工知能による診断支援、さらには、スマートフォンやスマートスピーカーといった新たなデバイスからの情報の収集など、様々な材料、技術が登場しています。中でもまったく新しい情報源として注目されている技術がソーシャルメディアデータです。Facebook による近況報告、LINE による交信、ショッピングサイトによる商品購入、Instagram による写真の共有。今や、国民の大多数が何ら化のソーシャルメディアを用いています。特に、日本では、東日本震災以降、Twitter などのソーシャルメディアが電話やテレビと並ぶインフラとして認知されつつあります。ソーシャルメディアは、個人のデータに紐付いたデータを扱え、なおかつリアルタイムな情報という従来ない特性を持っており、現在、これを利用して、医療情報の収取や公開を行う試みが始まっています。本パネルでは、スマートフォンを用いたクラスターの把握、つばやきを用いた流行予測の試みなど、ソーシャルメディアを利用した感染症対策の研究事例を紹介し、議論したいと考えています。

16:25 – 16:30 質疑応答

16:30 – 16:45 奈良県庁の依頼：
「コロナ感染者数の時系列解析」
作村 諭一（計算生物学）

感染症の流行は第1波、第2波と呼ばれるように感染者数が急激に増加したあと減少に転ずる。SIR モデルは感染者という実体を考慮してその広がりを微分方程式で記述し、感染者数の増加と減少を再現する。SIR モデルは再生産数などのマイクロなパラメータによりその挙動が決まる。一方で、感染者数の増加と減少というガウシアン的な変化がほとんどの場合で当てはまるのであれば、その増減の形そのものを定式化し、マクロパラメータで時系列フィッティングすることも可能である。本解析では、ガウシアン様の増減を積分した波形の近似としてシグモイド関数を用い、シグモイド関数の重ね合わせによる感染者数のフィッティングを行った。過去の感染者数系列によって得られるパラメータで将来予測が可能な場合、パラメータに制約を与える必要な場合、得られた関数から計算される必要病床数等について結果を紹介する。また、本解析において感染者数のデータのみから得られる情報、解析法の限界、詳細な解析手法への展開の可能性について議論する。

16:45 – 17:00 奈良県庁の依頼：
「CoViD-19 の感染者数予測モデルの構築と評価について」
小野 直亮（計算システムズ生物学）

奈良県庁からの依頼を受け、必要な病床数を見積もるための回帰モデルによる 感染者数予

測と SIR モデルによるピーク予測を行なった。データ解析に必要な 客観的なデータを集める
ことの重要性と問題点について議論したい。

17:00 – 17:10 質疑応答

17:10 – 17:55 全体としての討論会

17:55 – 18:00 Concluding remarks

■連絡先： 奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター
Email : dsc-info@dsc.naist.jp Tel : 0743-72-6056
HP : http://www-dsc.naist.jp/dsc_jp/