

コンピュータ科学専攻各教員研究室紹介

教 授 萩谷 昌己

世界は、計算であふれている。「安心・安全」から「あやしい・あぶない」へ

本研究室では、従来からソフトウェアやプロトコルの検証に関する研究を行う（安心・安全）一方、自然界に内在する計算過程を解析し、それらを活用して人工的な情報処理システムを構築することを究極の目標として研究を行って来ています。

より具体的に、本研究室では、論理学をベースとして、新しい計算モデルの提案、各種計算モデルの解析・検証・合成、そのためのツールの開発などを行っています。特に、通常の電子計算機以外の物理・化学・生物現象を活用した計算（自然計算・非通常計算）の可能性を探求しています。具体的に、分子計算（分子コンピューティング）、量子計算、細胞計算などと、その応用（たとえば分子コンピューティングを分子ロボティクスに）の研究を展開しています。

<http://hagi.is.s.u-tokyo.ac.jp/rigakuru.pdf> ←リガクルの研究室紹介（日本語）

<http://hagi.is.s.u-tokyo.ac.jp/members/hagiya.html> ←萩谷のホームページ（英語）

最近の研究テーマの例：

- ・分散ソフトウェアのモデル検査
- ・IoT プロトコルのテスト
- ・衛生的マクロの理論
- ・フィールド計算のプログラミング言語
- ・量子暗号プロトコルの検証
- ・量子エンタングルメントの解析
- ・量子マネーのプロトコル
- ・量子回路の最適化
- ・生体分子反応のシミュレーション
- ・DNA ゲルによる計算
- ・DNA 鎖置換反応によるチューリング機械の実現

教 授 今井 浩

アルゴリズムと計算量理論を軸に、計算とは何かを追究し、新世代の計算の世界を切り拓き、現実社会の大規模な問題の解決から P vs. NP などの深遠な理論まで扱う研究室である。**量子コンピュータ**や**量子暗号**に代表される量子情報処理でのアルゴリズムと計算量の研究をはじめ、Web グラフや道路ネットワークといった**超大規模グラフ・ネットワーク**の問題を解く**高速アルゴリズム**や、囲碁など人間が遊ぶゲームから経済・計算論的ゲームまで幅広い**ゲーム探索・機械学習アルゴリズム**について、それらの問題の計算の本質的困難さを示す**計算量理論の研究**とともに研究し、計算の多面性にそった多様な展開とともに基礎理論の重要性を念頭に置いて取り組んでいる。以下のテーマに加え、新機軸の問題も扱っていく。

1. アルゴリズムの設計と解析：
 - ・ 幾何的なアルゴリズムを 2・3 次元から高次元で展開、結晶等の幾何構造を有するグラフの解析。
 - ・ グラフアルゴリズム、行列乗算といった基本的問題での超大規模化・現在の壁を打破する研究。
 - ・ SAT, TSP, BDD などの計算困難問題を解く指数時間アルゴリズムの究極的限界解明。
2. 離散数学、特に幾何構造を組合せ論的に抽象化したマトロイド・有向マトロイドの組合せ構造の研究と、その計算代数・最適化と融合させた統一的研究の推進。グラフマイナー理論の計算展開。
3. 計算量理論：
 - (ア) 決定性・乱択・非決定の各計算量の関係を明らかにし、計算量での未解決問題に取り組む。
 - (イ) 時間・領域計算量に加え通信計算量・対話証明・機械学習の回路理論を軸とした計算量解析。
4. 人間にとって面白い題材に対するコンピュータ科学からのアプローチ。囲碁プログラムから、インターネット通信での資源割当のゲーム、機械学習まで、種々の対象のアルゴリズムと計算困難さの解析。
5. 量子計算・量子暗号の研究。量子コンピュータ実現へ向けた設計支援のための基礎理論から、量子アルゴリズムの開発研究そして量子コンピュータの計算能力を解明する量子計算量理論まで。量子エンタングルメントなどの量子情報基礎の理論的解析と量子グラフ理論の展開。

<http://www-imai.is.s.u-tokyo.ac.jp/>

教 授 須田 礼仁

本研究室では、(1) 並列・高性能計算 (2) 数値アルゴリズムの 2 つを柱とし、大規模科学技術計算の高速化・高精度化・高信頼化を目標として研究を行っている。

並列計算では、グラフィックスプロセッサを汎用計算に用いて高性能計算を目指す「GPGPU」が一つのターゲットである。GPU は数百から千以上のコアを有しているので、アルゴリズムの並列性を最大限に生かした実装が必要である。また複数（数百にもなりうる）GPU を同時並列に用いる研究も行っている。

もうひとつのターゲットは次世代の大規模スーパーコンピュータである。数百万プロセッサにおよぶ超大規模並列計算機ではプロセッサ間の通信の時間が膨大になると見込まれているので、通信の量や回数を削減した「通信削減アルゴリズム」が有効と考えられており、そのようなアルゴリズムを開発している。

高性能計算では、実際のマシンで試験実行をして最も性能のよい実装を選択する「自動チューニング」の研究に取り組んでいる。チューニングパラメタと性能との関係を数学的にモデル化し、ベイズ統計の考え方を用いて効率的に最適なチューニングパラメタを探索する手法の研究などを進めている。

数値アルゴリズムでは、計算量が少ない「高速アルゴリズム」を中心にこれまで取り組んできた。計算が大規模化・複雑化するに従って、倍精度では不足する事態も懸念されており、「高精度・高安定」なアルゴリズムも必要となっている。

<http://sudalab.is.s.u-tokyo.ac.jp/~reiji/sudalab.html>

教 授 小林 直樹

本研究室では、ソフトウェアに関する基礎理論とそのプログラム検証などへの応用についての研究を行っています。近年、交通システムや医療機器、電子商取引など世の中の多くのシステムがコンピュータソフトウェアによって制御されており、それらに欠陥があれば大惨事になりかねません。その一方でソフトウェアはますます大規模化・複雑化しており、テスト実行など従来のソフトウェア開発手法では品質の担保が難しくなっています。そこで、**数理科学的な手法を用いて機械的にプログラムの検証や変換を行うことによってソフトウェアの信頼性や性能を向上しよう**、というのが研究目的です。また、そのような目的を達成するためには、型理論、形式言語とオートマトン、定理証明など、理論計算機科学を深く学び、研究する必要があります。一見理論的興味の対象にしか見えない数学的概念が実は上記のような応用に結び付くことを実感できるのが本研究室で行っている研究の醍醐味でもあります。

以下、最近の研究テーマの例です。

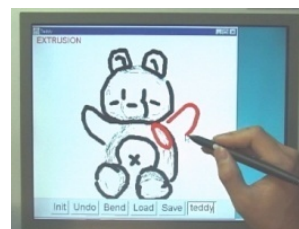
1. 高階モデル検査：システム検証技術であるモデル検査の拡張です。最近、世界初の高階モデル検査器の開発に成功して下の 2,3 のテーマに応用しています。
2. プログラムの自動検証：上記 1 の高階モデル検査を利用して、関数型言語 ML やオブジェクト指向言語 Java 用のプログラム自動検証器を作っています。
3. データ圧縮：文字列や木構造データをそれを生成するプログラムの形式で圧縮し、1 の高階モデル検査を利用して圧縮したままパターンマッチングなどのデータ操作を可能にする、というテーマです。
4. 型理論とプログラム意味論：1 の高階モデル検査などの基礎理論として、共通型の入った型システムとゲーム意味論などとの関係について研究を行っています。
5. セキュリティプロトコルの自動検証：インターネットショッピングなどで使われる、暗号を用いてパスワードなどの機密情報をやりとりする通信プロトコルの安全性を自動検証する、というテーマです。

詳細は <http://www-kb.is.s.u-tokyo.ac.jp> をご覧ください。

教 授 五十嵐 健夫

計算機を使いやすくするためのユーザインタフェースの研究を行っている。計算機が日常生活に密着して使われるようになってきており、この分野の専門家の重要性が増してきている。主に以下のようなテーマで研究を行っている。

(1) インタラクティブコンピュータグラフィクス: 個人が簡単に3次元モデルやアニメーションなどを作成できるようにする技術の研究開発を行っている。また、物理シミュレーションを援用して、家具や衣服など実世界で利用するモノを自分でデザインする方法などについても研究を行っている。



(2) ロボットなど実世界で動作する機器のためのユーザインタフェース: ペン入力やマルチタッチなどを利用した操作方法、ロボットに作業手順を教えたりするためのプログラミング手法などについて研究を行っている。ロボットの動作計画アルゴリズムの研究も行っている。



3) 効率的な情報アクセスのためのユーザインタフェース: アニメーション技術等を利用して、膨大な情報空間から必要な情報を素早く取り出したり、大勢の人との間で効率良く情報をやりとりしたりするための手法について研究を行っている。

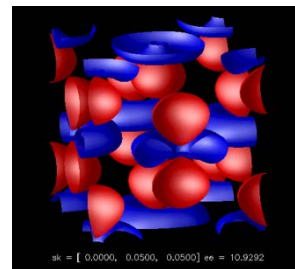
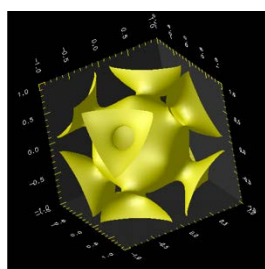
海外の大学や研究機関、医療機関や映像プロダクション、企業などと幅広く協力関係にあり、優れた研究成果であれば実用化していくことも十分可能な体制となっている。

<http://www-ui.is.s.u-tokyo.ac.jp/>

准教授 吉本 芳英

電子計算機の発明の動機の一つは、科学技術への応用でした。電子計算機が発明されて以降、半導体技術の飛躍的発展: ムーアの法則とともに計算機的能力は飛躍的に向上し、科学を計算によって推進する計算科学は大きな恩恵を受けてきたのですが、並列化技術など半導体技術の限界の顕在化に伴った計算機システムの複雑化のため、今後の発展には再び計算科学と計算機科学の間の密接な協力が必要になってきます。本研究室はこのような背景から2014年8月に発足した研究室で、計算機科学と計算科学の間をつなぐ教育研究を目指しています。(<http://www.cp.is.s.u-tokyo.ac.jp/>)

吉本は、半導体、金属、誘電体、磁性体といった物の性質を微視的な観点から明らかにする物性分野を専門としていますが、その中でも物性の大部分を支配している電子の量子力学をコンピュータで写實的にシミュレートする計算手法、第一原理電子状態計算を専門としており、このためのプログラム群 xTAPP を開発、公開しています。右図は電子状態計算の可視化例です。



(<http://xtapp.cp.is.s.u-tokyo.ac.jp/>)

本研究室は、計算機科学と計算科学の境界を立ち位置として電子状態計算をホームタウンとしつつも、

1. 幅広い計算科学分野について、各分野で個々に発展している方法論を計算機科学の観点から捉え直してそれらの相互交流につなげること。
2. 計算科学のニーズを計算機科学の観点から理解してより本質的な解決法を提案すること。

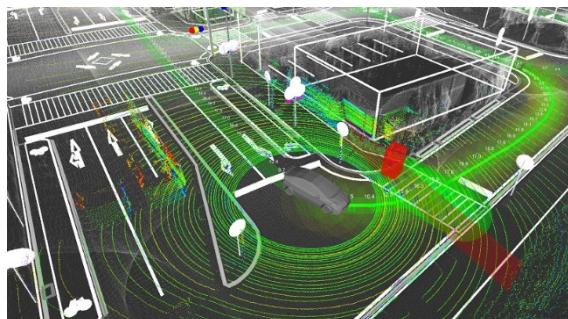
を目指したいと考えています。

准教授 加藤 真平

コンピューティングプラットフォーム研究室 (PFLab)

破壊的なイノベーションの原動力となるコンピューティングプラットフォームに関する研究を行っています。2016年4月に発足した研究室です。研究内容には例えば以下のようなテーマがあります。

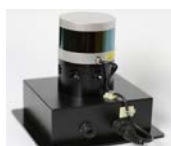
- 1つのチップに数万コアが集積されるメニーコアプロセッサ用のOS技術
 - ペタバイト級の実世界情報を実時間で扱う分散データ処理技術
 - 市街地を自律走行する完全自動運転システム用の組込みスパコン技術
- 性能100倍で電力100分の1、そんなコンピューティングプラットフォームの創出を目指しています。自動運転や3次元地図などのアプリケーション開発のために、AI、ロボット、センサー、クラウドに関する研究も進めています。海外の大学、メーカー企業、そして加藤が設立したベンチャー企業など、外部の研究チームとの連携を重視した研究室です。



3次元地図と完全自動運転システム技術



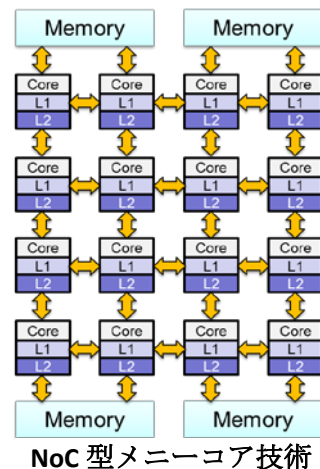
分散データ処理クラスター



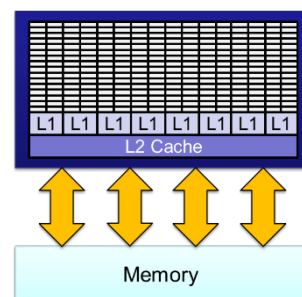
3次元センサー



組込みスパコン



NoC型メニーコア技術



GPU型メニーコア技術

教授 杉山 将

情報通信技術の飛躍的な性能向上に伴い、これまで人間にしかできなかった知的な情報処理が、コンピュータによって実現できるようになりつつあります。杉山研究室では、「コンピュータはどこまで賢くなれるのか」をテーマに、人工知能分野の機械学習とよばれる知的データ処理技術に関する様々な研究課題に取り組んでいます。

(1) 学習理論の構築

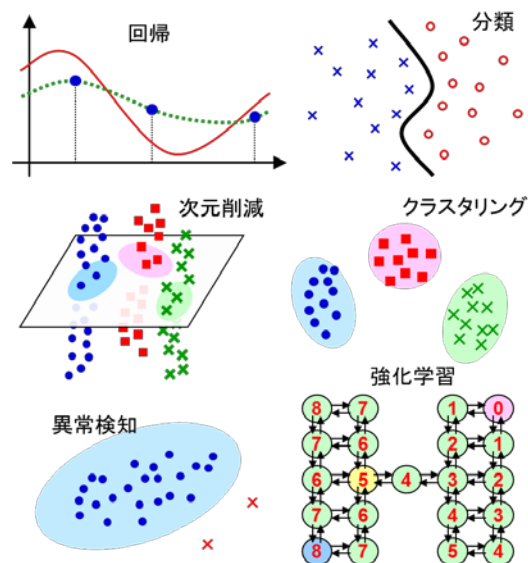
汎化とは、学習していない未知の状況に対応できる能力であり、コンピュータが知的に振る舞うために不可欠です。本研究室では、主に確率論と統計学に基づいて、汎化能力獲得のメカニズムを理論的に探求しています。

(2) 学習アルゴリズムの開発

機械学習分野には、入出力が対になったデータから学習を行う教師付き学習、入力のみから学習を行う教師なし学習、環境との相互作用を通して最適な行動規則の獲得を目指す強化学習など、様々な課題があります。本研究室では、理論的な裏付けを持ちつつ、実用性の高い機械学習アルゴリズムを開発しています。

(3) 機械学習技術の実世界応用

インターネットやセンサー技術の発達と普及に伴い、文書、音声、画像、動画、電子商取引、電力、医療、生命など、工学や基礎科学の様々な場面で膨大な量のデータが収集されるようになってきました。本研究室では国内外の企業や研究所と連携し、最先端の機械学習アルゴリズムを駆使して実世界の難問解決に挑戦しています。



<http://www.ms.k.u-tokyo.ac.jp/>

講 師 佐藤 一誠

『社会基盤としての統計的機械学習を目指す』

統計的機械学習は、大量のデータから機械が知的な処理を行うためのルールを自動的に抽出するための技術です。例えば、現在スマートフォンに搭載されている顔認識、オンラインショッピングサイトの推薦システム、近年注目を集めている車の自動運転など、様々な実社会の中で機械学習が重要な役割を果たしています。本研究室では、機械学習に関する以下のテーマを研究しています。

(1) **数理モデリングの研究** 統計的機械学習では、データの性質や解きたい問題に応じて数理モデルを構築する必要があります。本研究室では、データがもつ隠れた性質を表現する潜在変数という確率変数をもつ統計モデルの研究をしています。例えば、購買履歴を用いてユーザの嗜好にあった商品を推薦する問題では、ユーザの嗜好は明示的には現れないためユーザのもつ隠れた性質といえます。このような問題では、ユーザの嗜好を潜在変数としてモデルに組み込むことが有効です。

(2) **学習アルゴリズムの研究** 数理モデルが決まるとそのモデルのパラメータをデータから推定する必要があります。本研究室では、ベイズ推定や確率的最適化を基にした大規模データからの高速な学習アルゴリズムを研究しています。

(3) **科学実験を補助する統計的機械学習の研究** 研究者は、実験をデザインし、実験結果を分析し、実験設定の試行錯誤を繰り返し研究を進めています。このプロセスを機械によって支援もしくは自動化することができれば、研究分野全体の進歩に貢献できるはずです。本研究室では、実験のデザインは研究者が行い、実験およびその分析、実験設定の試行錯誤は機械が担当するという環境を構築するための機械学習技術について研究しています。

(4) **社会応用** 本研究の成果や近年の機械学習技術を基に社会へ積極的な還元を行います。例えば、東大病院との共同研究では、医用画像の読影支援システムの開発などを行っています。

講 師 本多 淳也

大規模なデータからコンピュータにより複雑な規則を見つけ出し重要な情報を取り出す機械学習の技術は最近では様々な場面で応用されています。一方で現実では、例えば新商品の推薦や新規材料の開発のように対象について十分なデータのないまま意思決定を行う必要がある場面も数多く現れます。本研究室では、このように環境から動的に情報を収集・学習しつつ意思決定を行う多様な問題におけるアルゴリズムの研究を行っています。

(1) **知識探索・活用のためのアルゴリズム構築**

インターネット上で広告配信を行う場合、各広告をある程度の回数表示（知識探索）してみない限りはそのクリック率を精度よく推定できませんが、実際に利益を得るためにはクリック率の高いと推測される広告を集中的に表示（知識活用）する必要があります。このように知識の探索と活用をバランスさせる問題は広告だけでなく価格の決定プロセスや投資バランスの設計など様々な場面で現れ、それらに対する効率的なアルゴリズムを構築するとともに、達成可能な効率についての理論限界を解析します。

(2) **あいまいな観測の活用のためのモデル化と探索**

映画や食べ物といった人間の嗜好に関する推定、あるいはスポーツやゲームにおけるプレイヤーの順位付けといった場面では、各候補の絶対評価が困難であり、相対的な優劣の比較のみが可能である場合が多くあります。このような直接的な観測が得られない状況下で最も「優れた」候補を適切にモデル化し、隠れた効用を最大化するための理論とアルゴリズムを構築します。

(3) **実問題のモデル化に基づく探索アルゴリズム**

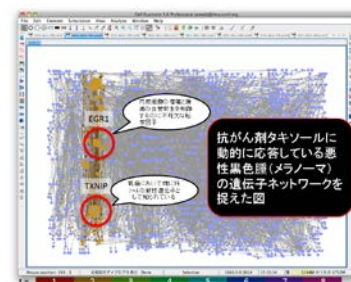
創薬や新規材料の開発といった場面では、一つの化合物の候補の物性値を計算するのに例えば数日といった長い時間がかかりますが、その計算を行うべき候補には膨大な数があります。そこで、それらの類似度を適切にモデル化しベイズ最適化といった手法を用いることで限られた計算資源で優れた候補を発見するアルゴリズムを構築します。

教 授 宮野 悟 ・ 准教授 山口 類

本研究室では、ゲノムデータをはじめとする大量個別の生体データから、がんを中心とする疾病のメカニズム解明および治療・予防に有用な情報を抽出するための数理的方法論を開発し、国内外の様々な分野の研究者、患者さん達と協力しながら、情報科学の力で未来の医療を切り拓くことを目指して研究を行っています。

がんは、DNAに蓄積された変異が原因となっているゲノムの病気であり、生命システムとしての統合的制御から逸脱した異常な細胞集団であることがわかってきました。しかしながら、それらの細胞集団は同一個体内であってさえも不均一であり、また異常な振る舞いを獲得するメカニズムおよび生長・生存戦略は多様かつ複雑であり、未解明なことが数多くあります。例えば、腫瘍細胞に蓄積したDNA変異の違いに加えて、個人ごとに異なっている環境要因などによりエピゲノムというDNAやヒストン修飾の状態に変化が起こり、これらが統合してがんの悪性度や治療応答性、治療による副作用の出やすさなどに影響することがあります。またがん細胞は、自身を攻撃しようとする免疫細胞からの監視を逃れる術を獲得することもあります。更にがんの細胞集団は環境に応じて時間的・空間的に進化していきます。

このように複雑な生体システムの理解、振る舞いの予測制御のためには、急速に進歩する最先端の計測機器から産生される多様なデータ（数値、文字列、画像、時系列）を解析し統合する方法論が不可欠です。当研究室では、ベイズモデリングに基づく変異検出、がん細胞集団シミュレーション、データ同化技術、人工知能による臨床シークエンス変異解釈法等の研究を進めています。

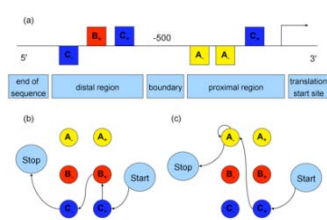


研究室：<http://dnagarden.hgc.jp>

システム癌新次元：<http://neosystemscancer.hgc.jp>

ポスト京重点課題 個別化・予防医療を支援する統合計算生命科学：<http://postk.hgc.jp>

教 授 中井 謙太 ・ 講 師 PATIL, Ashwini



本研究室は白金台キャンパスの他の3研究室と同様、いわゆるバイオインフォマティクスを専門としている。白金台キャンパスでは日本でも有数のスパコンシステムを利用できる環境が整っており、コンピュータ科学畑からの意欲のある学生の参加を歓迎する。教授の中井は、生命の情報がどのような形で一次元の文字列（配列）として記述されているのかを解き明かしたいというのが興味の原点である。生命の情報は基本的には各細胞がゲノムという形で保持しているDNAの化学的構造として記されているが、その情報は必要に

応じて、RNAという分子にコピーされ、さらにさまざまなタンパク質を合成するために用いられる。従って、生命情報の研究は、DNA、RNA、タンパク質の3つのレベルからアプローチできる。

現在、我々はその中でも主に、ゲノムDNAにコードされた遺伝子の情報が必要に応じて読み出されるための制御情報（当然、これもゲノムDNAにコードされている）の解読に興味を持っている。上の図は、同じ種類の細胞で読み出される遺伝子の制御領域がもつ共通構造を、マルコフ連鎖を用いて確率的に表現したモデルである。このモデルを用いれば、同様の制御を受ける未知遺伝子の存在を予測することができ、それを実験で検証することもできる。

このように、研究室のカラーは基礎生物学志向が強い。また、留学生が多く、国際色が豊かなことも特徴の一つである。研究室内では様々なプロジェクトが同時に進行しているが、基本的には、各研究者がやりたい研究を推進しながら、お互いの研究にも興味をもつような方向を目指している。

研究室：<http://fais.hgc.jp>, スパコン：<http://supcom.hgc.jp>

教 授 井元 清哉

現在、1000 ドル程度のコストで個々人の全ゲノム配列は、数百 GB のデータとして得られるようになりました。このコストはやがて 100 ドルを切り、誰もが自身の全ゲノム情報を知ることが出来るようになるでしょう。このような背景のもと、数百万人の全ゲノム情報からなるゲノムビッグデータを個々人の疾患の予知・予防、並びに健康の維持に繋げるための方法が求められています。我々は、全ゲノム、トランスクリプトーム、エピゲノム、腸内細菌叢をシークエンスしたメタゲノムなど多様なゲノム関連のビッグデータとレセプト情報・特定健診等の情報など時間軸を有する健康・医療に関する大規模データを用い、統計科学的なデータ解析技術にスーパーコンピュータを用いた大規模計算を合わせ、国内外の企業、研究者と協力し、この問題に取り組んでいます。

研究テーマ 1：シークエンスデータ解析技術の開発

ヒト白血球抗原（HLA）、T 細胞受容体レパトア、腸内細菌叢メタゲノムなど免疫システムの多様性に関わるゲノム領域の解析、がんゲノムの解析を行う新たなベイズモデルの開発を行っています。特に、がんに対しては、国際がんゲノムコンソーシアムにおいて約 3000 人の全ゲノムシークエンスデータを用いた研究を行っています。

研究テーマ 2：生命システムのモデリング技術の開発

遺伝子やノンコーディング RNA を含む数万種類の分子が形成する生命システムのモデリングから、抗がん剤の効果予測を行う方法を開発しています。

研究テーマ 3：臨床シークエンスの取り組み

がんゲノムの解析から、数千から数万箇所のゲノム変異が同定されます。これらからがんの原因のゲノム変異を同定するには、論文によって公開されている情報が用いられます。しかし、がんを対象とした論文は、2016 年だけで 20 万報（年々増加）あり、その全てを網羅することは既に人知を越えています。この問題に対して、人工知能を用いた研究を行っています。

准教授 渋谷 哲朗

2003 年に 30 億塩基対からなるヒトゲノム（ゲノム＝全 DNA 塩基配列情報）が、国際プロジェクトにおいて世界中の多くの研究者の努力の末、莫大な費用と期間をかけ解読された。その一方、当時の計測技術では、個人や個々の細胞のゲノムに潜む未知の様々な種類の変異を網羅的に知るためのデータを得ることは、費用的にも時間的にも不可能であった。しかし、ここ数年に「次世代シーケンサー」あるいは「次々世代シーケンサー」と呼ばれる新たな技術が登場し、状況が一変した。これらの技術によって従来に比べてきわめて安価かつ短時間にゲノム情報を得ることが可能となり、個人個人のゲノム、あるいは、細胞レベルに異なる各細胞のゲノム、あるいは、これまで顧みることのなかった多種多様な様々な生物種のゲノムなどまでがシークエンスされるようになった。さらには、DNA 修飾情報などゲノム配列情報にとどまらないさらなる高次情報も網羅的に大量に得られるようになっている。そしてそれらの技術によって得られるデータを元に生命科学研究方法は今まさに革命的に変わろうとしている。しかしながら、シーケンサーによってもたらされるデータ量は桁違いに膨大である。それらのデータは時にムーアの法則を超える速度で増加しており、従来の情報科学的手法のまま解析を行うことはきわめて困難になりつつある。また、個人ゲノムの解析では、プライバシーなど、これまでは考慮してこなかった概念を考慮しながらの解析も必須になっている。そのような状況のもと、これからのビッグ・データ時代の医学生物学的研究においては、情報科学技術のまったく新しいパラダイムが必要となっているとまで言われる。

本研究室では、上記のような様々な生物・医学分野のビッグ・データ時代の超大規模データに対する超高速検索や高機能検索、あるいは高精度・高機能解析などのアルゴリズム設計・解析の研究を通して、実際の生命科学研究にインパクトを与えることを目指して行っている。その一方で、現在、生物・医学分野に限らないあらゆる分野においてビッグ・データに対応することが求められており、そのようなビッグ・データ情報科学への貢献も目指している。この分野は、情報科学、医学、生物学、物理化学等様々な分野の学際分野でもある。ひとつの分野だけでなく多くの分野の最先端の話題を扱いながら、様々な分野の基礎研究への貢献を行うと同時に、実際に世の中に役に立ちインパクトを与えるような研究も同時に行うことも目指せる刺激的な分野でもある。また、このような学際分野はまだ未開拓の領域が多く、未踏の領域を新たに切り込んで開拓することも常に目指したいと考えている。

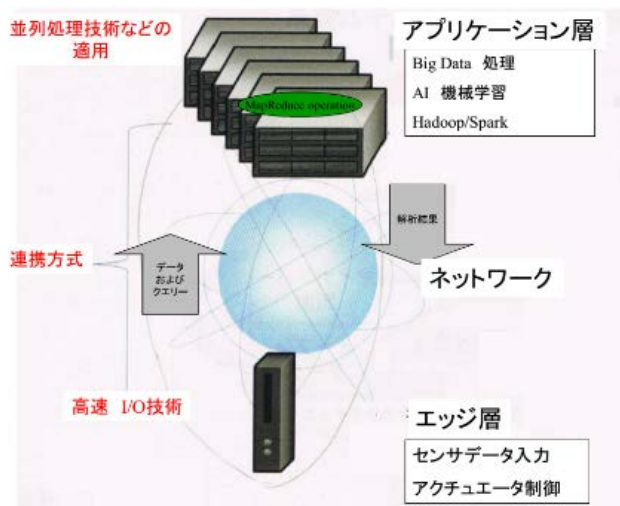
ホームページ → <http://shibuyalab.hgc.jp/index-j.html>

教 授 中田 登志之

実世界とサイバーシステムを結合するサイバーフィジカルシステムの社会実装により、現在国民が瀕している少子高齢化、食料自給率の低下、エネルギー危機、労働人口の減少などの諸問題に対応できることが期待されている。一方従来のサイバーフィジカルシステムは比較的小規模の問題には対応するものの、国民的課題を処理するような大規模システムの実装はいまだ着手にかかったところである。

本研究室では、大規模なデータ処理に対応する高性能なサイバーフィジカルシステムの構成方式について、種々の観点から研究を遂行する。以下に研究テーマの例を挙げる。

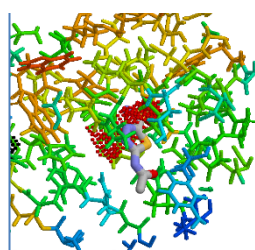
- Big Data 処理の効率的な処理を行う方式の研究。
Hadoop や Spark のような大規模データ解析用プラットフォームを用いた Big Data 処理の方式に関して、並列処理や、GPGPU の導入による処理高速化の研究
 - 上記と連携した機械学習・AI 処理の高速化の検討。
 - クラウドシステムと、センサ・アクチュエータを制御するエッジノードとの間の効率的な連携方法などのシステムアーキテクチャの確立とその評価。
 - Big Data 処理と Open Data 活用連携に関する方式の検討とその有効性の検証
 - 他の研究室と連携して実世界のモデリングとそのシミュレーションの効率的手法の評価
- 高速 I/O 技術の検討と評価



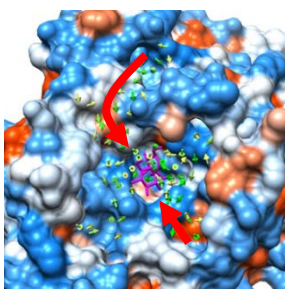
教 授 清水 謙多郎

私たちの研究室は、情報科学の手法を使って生命現象の解明をめざすバイオインフォマティクスの研究を行っています。研究内容は、タンパク質の構造・機能に関する研究から、ゲノム、トランスクリプトーム解析まで幅広く、その手法は、機械学習、統計解析、分子シミュレーションなど様々です。高度なバイオインフォマティクスの理論および手法の開発と、バイオインフォマティクスを用いた生物の新しい知見、社会に貢献できる成果の獲得の両面から研究を行っています。具体的には、タンパク質の高精度の構造予測、タンパク質と他の分子（タンパク質、核酸、糖鎖、脂質、金属、低分子化合物など）との結合部位の予測や、ドッキング予測（結合した状態の複合体構造の予測）、さまざまな機能予測、また、タンパク質のダイナミクスを解析し、他の分子が結合する過程や自由エネルギーなどの物理化学量を求めたり、生体内での詳細な原子間相互作用を解析したり、さらにタンパク質がフォールドする過程をシミュレートしたりする研究を行っています。最近では、タンパク質の機能に重要に関わる天然変性や膜タンパク質のさまざまな構造特徴の解析・予測に関する研究を手がけています。実験研究者との共同研究では、酵素の改変、医薬品の開発、機能性食品、環境浄化など、医薬や食・環境などに関わるアグリバイオの研究に関わっています。本教員のほか、助教1名、さらに特任准教授2名、ポスドク数名が協力して指導にあたります。ホームページ: <http://www.bi.a.u-tokyo.ac.jp/>、場所: 農学部キャンパス（弥生地区）

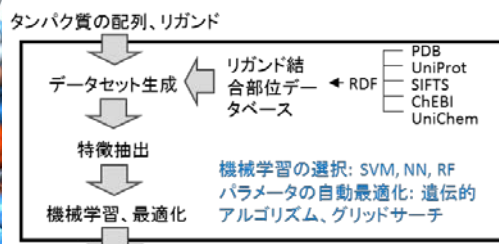
タンパク質・リガンド結合部位予測（赤色が予測部位、灰色が実際のリガンド）



タンパク質・リガンド結合の粗視化分子動力学シミュレーション（リガンドがタンパク質表面に沿ってポケットに入り込む動きを解析）



リガンド結合部位予測ツールの自動生成



教 授 高野 明彦

- (1) 「プログラミングの代数」：科学的なソフトウェア作成法の確立には、正しさが数学的厳密さで保証された高信頼なプログラム部品の蓄積と、それらを組み合わせて所望の機能と性能をもつソフトウェアを構築する方法の開発が必須である。この観点から、データ構造や制御構造の代数的性質を利用してプログラム部品を融合するプログラム変換、部分計算、データ変換などについて研究している。
- (2) 「連想の情報学」：人間が膨大な電子情報と創造的にインタラクトするには、人間の連想能力の活性化がカギとなる。それらを基礎づける計算機構は、膨大な情報を対象とする連想計算（類似性計算）と考える。数千万件規模の文書集合に対する連想計算を高速実行する汎用連想計算エンジン GETA を研究開発して公開している。この連想計算を用いて創造的相互作用の場を実現する、情報空間との対話技術について研究している。

これらの研究成果を応用して、実用的な情報サービスを立ち上げることに注力している。これまで構築した代表的なものに次の4システムがある。

- ・ Webcat Plus (<http://webcatplus.nii.ac.jp/>)
 - ・ 文化遺産オンライン (<http://bunka.nii.ac.jp/>)
 - ・ 新書マップ (<http://shinshomap.info/>)
- 想・IMAGINE (<http://imagine.bookmap.info/>)

教 授 相澤 彰子

本研究室では、言語テキストを中心とするコンテンツとメディアに関して研究を行っています。中心となる研究テーマは以下の通りです。

- (1)言葉の意味の獲得と利用
情報の同定による記述と実体の対応付けや言語資源の自動構築
- (2)コンピュータによる言語理解
言語解析による意味構造の抽出や知識の獲得
- (3)人の言語活動のモデル化
テキストを介した人間やコミュニティ活動の計測とモデル化

具体的な研究課題として、言語理解タスクの分析、文圧縮と要約、文書論理構造の解析と活用、数式および数学概念の検索と理解支援、視線検出装置を用いた読み方モデルの構築と応用、情報同定基盤技術、研究者のための情報推薦および執筆支援などに取り組んでいます。また、言語処理や情報検索に関連した新しい研究テーマへの挑戦も歓迎します。

本研究室では、国立情報学研究所が有する大規模な情報資源や分散計算機環境などを利用した研究が行えます。研究所内でのセミナーや輪読会への参加、学際色・国際色豊かな研究者との交流、所内外の共同研究プロジェクトへの参加を通して、研究室という単位にとどまらず、一人の研究者として活動の幅を広げることを目標としています。

研究室ホームページ：<http://www-al.nii.ac.jp/>