

Data Centric Science Research Commons 001-017

データに語る科学

データ中心科学 リサーチコモンズ研究紹介

丸山 宏

倉田のり

小原雄治

本山秀明

曾根原 登

中野純司

齋藤正也

才田聡子

岡 彩子

木村 暁

佐藤寛子

伊村 智

津田博史

片山俊明・山口敦子

中村卓司

椿 広計

北野 潤

データに語らせる科学

データ中心科学 リサーチコモンズ研究紹介

データに語らせる科学

データ中心科学 リサーチcommons研究紹介

Research View 2014 – 2015

著者:

情報・システム研究機構 URA(広報)編

取材・文:

池谷瑠絵

写真:

水谷 充

ブックデザイン:

畑 友理恵 + 大崎奏矢

発行:

大学共同利用機関法人 情報・システム研究機構

Inter-University Research Institute Corporation

Research Organization of Information and Systems

発行日:

2015年3月20日 第1刷

2016年2月22日 第2刷

印刷所:

シナノ印刷株式会社

ISBN978-4-909638-04-5 C0040

リサーチコモンズについて
北川源四郎 p.4

凡例 p.6

丸山 宏 p.8

倉田のり p.12

小原雄治 p.16

本山秀明 p.20

曾根原 登 p.24

中野純司 p.28

齋藤正也 p.32

才田聡子 p.36

岡 彩子 p.40

木村 暁 p.44

佐藤寛子 p.48

伊村 智 p.52

津田博史 p.56

片山俊明・山口敦子 p.60

中村卓司 p.64

椿 広計 p.68

北野 潤 p.72

プロジェクト紹介 p.76

編集後記 池谷瑠絵 p.82

研究者紹介 p.83

「データ中心科学リサーチcommons」について

情報通信技術(ICT)の飛躍的發展に伴い、ユビキタス化した情報空間に生きる私たちの生活からほとんど自動的に、時々刻々とデータが生みだされるようになり、これまでとは比較にならないほど飛躍的に大量の情報が蓄積されるようになりました。このようなデータをいかに組み合わせ、目的に合わせて有効に活用できるかが、ビッグデータ時代の産業、行政、科学技術など私たちの社会のさまざまな場面において、成否の鍵を握っているといっても過言ではありません。

科学技術の世界でも、地球環境や宇宙等の自然界、大学や研究所等の実験室で獲得されるデータがますます大規模化し、このようなアカデミック・ビッグデータの効果的利用がイノベーションの鍵となっています。このような背景から、これまでの理論科学・実験科学・計算科学に続く第4の科学として、「データ中心科学」が、学術・社会の課題を解決する研究戦略課題として、世界的に推進されています。

ところで、私たち大学共同利用機関法人情報・システム研究機構は約10年前、「情報とシステム」の観点から諸問題の解明と解決を図ろうという理念のもと、国立極地研究所、国立情報学研究所、統計数理研究所、国立遺伝学研究所という4つの研究所が集結して設立されました。設立以来、時代に先駆けて開拓・

発展させてきた、さまざまな新領域融合研究を背景に、この「データ中心」の考え方をまさに中心に据えたプロジェクト「データ中心科学リサーチコモンズ事業」を2013(平成25)年度から推進しています。これは、「データ基盤整備」「モデリング・解析基盤整備」「T型・π型人材育成」を3つの柱として、情報化時代を先導する先進的研究基盤と、オープンな国際共同利用・共同研究の中核的拠点を形成しようという、世界的にもユニークな研究事業とすることができるでしょう。

「データ基盤」では、まず地球環境、ライフサイエンス、人間・社会、学術情報、物理・化学などの科学的データ等をビッグデータ的に活用するための基盤の維持・整備を推進します。そして“読み解く知”としてのモデリング・解析の手法を開発し、共同利用を促進するような「モデリング・解析基盤」を整備・提供します。さらにこのような融合的な研究のプロセスの中に「人材育成」を組み込むことによって、持続的な研究発展のサイクルを築いていこうという計画です。2015(平成27)年3月現在、機構の4つの研究所および2つの直轄センター「新領域融合研究センター」、「ライフサイエンス統合データベースセンター(DBCLS)」が一丸となって、わが国の研究力強化に波及効果を生み出せるような共同利用・共同研究を進めています。

そして、このような共同研究・共同利用のモデルケースとして、全12の研究プロジェクト&研究データベースを推進してきました。また2014(平成26)年4月からは、「データ中心科学リサーチコモンズ」ウェブサイトを開示し、この最新研究成果を広く情報発信してきました。今回本書に収めたインタビュー記事は、このウェブサイト上で2014年度中に公開された連載記事「Research View」を集めたものです。

大学・研究機関との連携を深めながら、生命科学、地球科学、人間・社会等それぞれの領域でいま拓かれつつある新たな融合研究をご紹介します。情報・システム研究機構の大学共同利用機関法人としての役割と貢献について広く知っていただく機会となれば幸いです。



大学共同利用機関法人
情報・システム研究機構
機構長 北川源四郎

凡例 このブックレットのみかた

記事の通し番号•
記事タイトル•

Research View 001
**科学として「想定外」に
どう対応するのか？**

この研究者が紹介します•

統計数理研究所副所長・教授
プロジェクトディレクター 丸山宏

研究プロジェクト名または研究データベース名 ...•

Project
システムズ・レジリエンス

同じプロジェクトまたはデータベースに属する他の
記事で、関連項目がある場合は、ページ番号を•
示します。

.....Research View014 p.60➤

Research View 2014 – 2015



Research View 001

科学として「想定外」に どう対応するのか？

統計数理研究所副所長・教授

プロジェクトディレクター 丸山 宏

Project

システムズ・レジリエンス

Research View

INTRODUCTION



ちょうど4年前の2011年3月11日に起こった「東日本大震災」。以来、災害復興のためにICTをどう役立てるか、企業などでも対策が進んできました。

しかし過去の体験1つで、

将来の「想定外」の事象

Research View017 p.73

に対応することはできません。どうしたらこの問題に 대응することができるのか、もっとサイエンスとして解明すべきだ、というのがプロジェクトのスタートラインでした。お話は、プロジェクトディレクターの丸山宏教授(統計数理研究所)です。

レジリエンスの知恵を 集大成する

「レジリエンス」とは、もともと心理学やエコロジーの分野で使われていた用語で、最近は社会システムなどにもよく使われるようになってきました。社会、情報、金融、企業・組織、生態系など、さまざまなシステムに何らかの「擾乱」が起こった時、壊れにくく(Resistance)、また万が一壊れた後に素早く回復できる(Recovery)性質を「レジリエンス」といいます。そこでまず、さまざまな分野から知見をたくさん集めてきて、レジリエンスの知識の体系「Body of Knowledge, BOK」を構築しようというのが、私たちの目標です。実践的なマニュアルのようなものがあれば、たとえば海岸の堤防をあと1メートル高くするほうがいいのか、それとも堤防を越える津波が来た時になるべく早く復旧できるよう予算を投じたほうが有効なのか、といった問題に答えることができます。このように科学の知識と理論によって礎をつくり、行政などに活用してもらおうと考えています。

戦略としての適応性、 冗長性、多様性

プロジェクトも3年目を迎え、これまでの研究から、さまざまなシステムに共通のレジリエンスの戦略というものが見えてきています。1つめは冗長性で、システムにどれくらいの冗長性を入れて設計すれば、どのくらいの擾乱に耐えられるかを測ることができる。これは信頼性工学などの分野に研究の蓄積があります。2つめは適応性で、これは制御工学を使うと解明できます。ところが3つめの多様性(diversity)が、まだよくわからない。そして多様性は、ものすごくコストが高いんですね。たとえばコンピュータ・ネットワークに多様なマシンを入れると、すごく管理が大変です。また南極のようにシビアな環境下にある生態系は、最も適したものだけが生き残り、多様性はほとんどありません。一方で最近では、企業の競争にとって、多様性は実は有利なのだという議論が『ハーバード ビジネス・レビュー』などでもたいへん活発です。特にシステム開発の



大きなトレンドである「アジャイル」と組み合わせると、ビジネス環境の変化が激しい時代には、多様性によって「想定外」の事象に対応できるケースが多いのです。

大量ゲノムで 「収穫逡減の法則」との相関を探る

このような多様性をうまく組み込んでレジリエンスなシステムをつくるためには、「収穫逡減法則」との関係がひとつのキーになるとわれわれは考えてい

ます。「収穫逡減」はもともと経済学の用語で、あるシステムに投入量を増やしていくと、これに伴って増える収益量が、ある点を境に低減していく現象をいいます。そこでサブテーマ1共同研究者の南和宏特任准教授(統計数理研究所)は、コンピュータ上にロボットが住む仮想空間を作って、多様性や冗長性を組み合わせたさまざまなシミュレーションを行っています。またサブテーマ2共同研究者の明石裕教授(国立遺伝学研究所)は、アフリカに住むショウジョウバエの遺伝子という、当プロジェクトの中で最多量のデータを駆使して、DNAの中で同じアミノ酸を





生成する異なる塩基列に注目し、その遺伝子の状態を世代的に調べることによって、多様性と収獲遞減法則の関連を追っています。明石教授の研究によって、自然界でほんとうに収獲遞減の法則と多様性との間に相関があるのか、まさに解明されようとしているところです。

ンス・トライアングル(Michel Bruneau 他2003)」という指標が使われてきましたが、井上教授の成果は新たな指標・性質を加えており、名誉ある賞(Challenges and Visions Papers Award, Third Prize(AAMAS2013))を受けています。

理論から設計して 「レジリエンス」を定義する

さらに、われわれは「レジリエンスとはどういう性質なのか？」を理論的に定義するアプローチにも取り組んでいます。サブテーマ3共同研究者の井上克巳教授(国立情報学研究所)は、システムの現在の状態とその環境を数的に記述し、環境をシステムに対する制約として捉えた「制約充足問題」として解くことで「システムズレジリエンスモデル」をつくりました。システムに擾乱が起こった状態から、どのようなパスを通して復旧できるのかをモデル化し、これ以上破壊されたら修復不可能なポイント、復旧までの計算のステップ数、システムの機能低下の状態を測る指標などを定義しています。復旧までの過程については、これまでは主に「レジリエ



丸山教授が示しているのは、レジリエンスの「タクソノミー(分類学)」」。緊急時は必ず平時とは優先順位が変わる。レジリエンスを考えるには、時間的な局面を分類することも重要なポイントだ。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/b04/>

▶ プロジェクトウェブサイト

<http://systemsresilience.org/>



Research View

Research View 002

ゲノムが語る、 栽培イネはどこから来たのか。

国立遺伝学研究所・教授

プロジェクトディレクター 倉田 のり

Project

遺伝機能(生命)システム

INTRODUCTION



自分の目で確かめ、根気のいる実験を日々積み重ねていく——そんなサイエンスの伝統的なスタイルが主流だった生物学。そこへ何桁も違う大量データを生み出す次世代シーケンサーが登場し、さらにビッグデータ時代が到来します。時代の要請に応じて、ゲノム科学と統計学が協働する「新領域融合」研究としてスタートした「遺伝機能(生命)システム」は、すでにマウス、イネ、ショウジョウバエ、ゼブラフィッシュ等のモデル生物において、いくつかの重要な成果を挙げているプロジェクトです。今回は2013年の成果から、データ中心科学としてさらに発展しつつある「野性イネ」の研究をご紹介します。

002

生命の多様さとゲノムを いかにつなぐか？

われわれ研究者の前に、非常に多様な生物遺伝資源が、とにかくどんと大量にあります。この中にはまず、生物が持つ極めて多様な形態の情報があって、イネの場合なら花の形、実の大きさ、丈の高さなど、その生物を特徴づけている形態情報が大量に蓄積されているんですね。そして一方には大量のゲノム情報があります。ゲノムは単にアデニン(A)、グアニン(G)、チミン(T)、シトシン(C)の塩基配列(シーケンス)だけではなく、その中にコードされた遺伝子がいつどのように読まれて、発現するかという情報も含まれます。あるいは外からの刺激や環境等によって、生物がどう変化していくかという要素が入ってくることもあります。ところが、このように多様な形態の1つ1つが、いったい膨大なゲノムの中のどの情報(遺伝子)に対応しているのかが、わからない。これらを結び付け、本当に有用な関係性

を抽出する方法を探り出そうというのが、われわれのプロジェクトです。

融合研究じゃなければ 始まらない。

海外の大学では、生物に限らずほとんどすべての学部に統計を扱う部門があり、また生物学では特に「バイオインフォマティクス」と呼ばれる分野が、大きく進展しています。たとえば、われわれのプロジェクトでも遺伝研と情報研の共同研究では、マウスの持つ顎の形のような形質の情報を、画像処理によって理解しようという課題に挑戦しています。この場合に難しいのは、なんといっても、生物学者の目には明らかな形質の違いが、コンピュータの目にはなかなか区別できるようにならないという点です。人間はたぶん、全体の印象と細部の印象を複合的に認識できる点で、コンピュータより格段に優れているのだと思います。コンピュータに理解させるには、顎

の角度を決めるのはこの要素、先端の形状を決めるのはこの要素というように、とにかく1つ1つの要素に分解しなければなりません。ただいったん数値化できれば、要素を組み合わせたり再構築したりしてどんどん発展させることができるのが、大きな魅力ですね。

野性イネはなぜ解明できなかったのか？

イネの研究はこれまで、進化学をはじめとする遺伝学的な解析がほとんどで、また農業においては栽培化というテーマなどに沿って進められてきました。栽培イネは遺伝的な継承が明らかで、また遺伝子と形質の対応関係がある程度わかっているのですが、野生の集団はバラバラで、多様な自然変異を蓄積しています。また野性イネの系統は、現在のイネとはかけ離れて進化してしまっているため、簡単には交配させることができません。このような理由から、野性イネは、これまでの遺伝学的方法で扱うのは困難なのです。そこでわれわれは遺伝研が保有する

約450系統の野性イネの形質データとゲノム配列を、アソシエーション解析という方法を使って、統計的に結びつける研究を進めています。たとえば丈が高い系統から低い系統へ段階的に変化する形質データに対して、ゲノムの配列で同じような推移を示しているものはないか、と照らし合わせるような作業を行ったのです。

いよいよデータ中心科学へ向けて

このようにして調べた形質とゲノムの関係に加え、どの野性イネの系統が歴史的に先に出現したか、また地域的な分布に照らしてどのような関係があるか、といった縦軸・横軸の関係性を提示したわれわれの論文は、2012年にNature誌に掲載されました。遺伝的データをより詳しく探求していくために必要なステップとして、野性イネについて人間が持つ知識を広げたという意義が大きかったと思います。またいろんな意味で影響力があり、たとえば私が参加しているナショナルバイオリソースプロジェクト(NBRP)へ、国内外の研究者から野性イネの種子を配布して





ほしいという依頼が、これまでの2~3倍に増えるといった反応もありました。しかし、実はこれからがいよいよ本当の研究なんです。たとえば統数研や東大との共同研究による解析方法の開発が、日々進捗しつつあります。形質の特徴をどのような要素に分解するのか、どう統合するのかについても、まだまだ足りないものがある。もっともっと方法を洗練させたいと考えています。



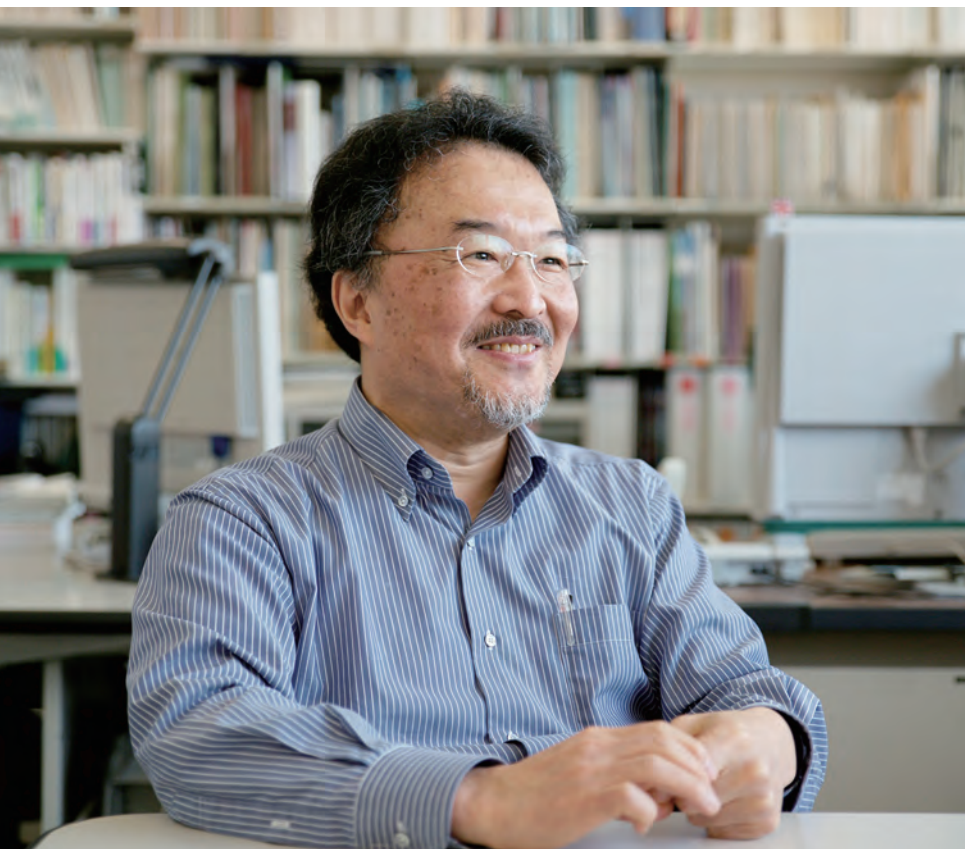
少しでも触れるとばらばらと実が散ってしまうという形質的特徴を持つ、野性イネの1系統。収穫が難しいため、栽培に向かないことは明らかだ。栽培化に寄与する遺伝子が、ゲノムのどの領域にあるかも、すでに解明されている。

▶プロジェクト概要

<http://rc.rois.ac.jp/research/b02/>

▶プロジェクトウェブサイト

<http://tric.rois.ac.jp/life/>



Research View

Research View 003

**国際標準化されたデータが、
生命を語り始める。**

国立遺伝学研究所・特任教授

DBCLSセンター長 小原 雄治

Database

ライフサイエンスデータ (DBCLS)

INTRODUCTION



ゲノムという膨大なデータで科学の進歩やイノベーションが見込まれるバイオ分野は、いち早くデータベース化に取り組んできました。アメリカではNCBI (The National Center for Biotechnology Information)、ヨーロッパではEBI (Part of the European Molecular Biology Laboratory, The European Bioinformatics Institute)そしてこれらの日本版であるDDBJ (DNA Data Bank of Japan)が、遺伝情報のデータベースの一元化に成功しています。これらとも連携して、日本の大学・研究機関・医療機関などが蔵する多種多様なライフサイエンスのデータベースの統合化を担うのが「ライフサイエンス統合データベースセンター (Database Center for Life Science: DBCLS)」です。プロジェクトディレクターでありDBCLSセンター長の小原雄治教授(国立遺伝学研究所)に、その取り組みをご紹介します。

統合データベースとは？

ゲノム情報、遺伝子発現の情報など、ライフサイエンスのさまざまなデータが日々蓄積されています。また過去の研究プロジェクトから生みだされたデータなどが研究所や大学のサーバーに大量に保管されていたり、場合によってはメンテナンスできずに眠ったままになったりしていることもある。つまり、どこに何があるかまったく分からないんです。そこでばらばらに存在しているそれらのデータベースを、一か所に集めなくても、みんなが使えるようにしましょうというのが、統合データベースという考え方です。使えるようにするには、データの形を標準化する必要があります。たとえば現在公開中の検索サービスでは、国内すべてのライフサイエンスのデータベースを全部探してきて、串刺し検索ができるようにしています。ただそれだけではGoogleで検索するのとはほぼ同じですね？ 実は、われわれの開発は、

いよいよこれからということなのです。

RDFという国際的な 共通フォーマット

今、私たちがGoogleなどの検索エンジンを使って得られる情報は、ウェブページという「文書」のかたちをしています。その後は人間がその文章を読んで意味を理解して取捨選択し、レポートを書いたりする。ウェブは今、このようなWeb of documentに代わって、Web of dataへと進化しようとしています。たとえばウィキペディアの次世代版である「DBpedia」では、機械がいろんな情報を組み合わせ、別の情報、つまりこれまでのレポートに近いものまでつくることができるんです。このようなセマンティックウェブを実現するために最適な情報のフォーマットとして、今「RDF (Resource Description Framework)」が世界標準となりつつあります。RDFはデータを、主

語(英: subject)述語(predicate)目的語(object、述語の値)が一義的に定義できるような形で書き、これを「トリプル(triple)」と呼んでいます。データをトリプルで揃えることで、意味を含めた高度な検索やグラフ化など、データの使い途が飛躍的に広がります。また同じ言葉でも分野が変われば意味が異なるため、「辞書セット」とも言える各分野の意味の地図にあたる「オントロジー」を整備し、検索結果のさらなる精度向上が世界的に目指されています。

データベースの国際標準化を担う「バイオハッカソン」

そこで片山俊明 特任助教(DBCLS)たちが中心となって、このようなデータベース開発の将来的なビジョンや仕様のディテールを巡り、協働作業と議論を通じて国際的な標準をつくっていくというワークショップ「バイオハッカソン(BioHackathon)」を毎年開催しています(DBCLSとNBDCの共同開催)。短期の合宿形式で集まって、今抱えている問題を解決してしまおう、そんなワークショップのスタイル

を「ハッカソン」といい、非常に生産性が高いため、今IT業界などでも盛んに採り入れられている手法です。 「バイオハッカソン」

では、今第一線で活躍しているデータベース、オントロジー、アプリ開発など現役の技術開発者たちが集まって、通常ならば1か月以上もかかるような課題を、わずか1週間で解決してしまう。さらにこのイベントを通じて知り得た最新の知識やノウハウを普及するために開催する、国内版のイベントも、4回目を数えます。まさに継続は力なり、と言えますね。

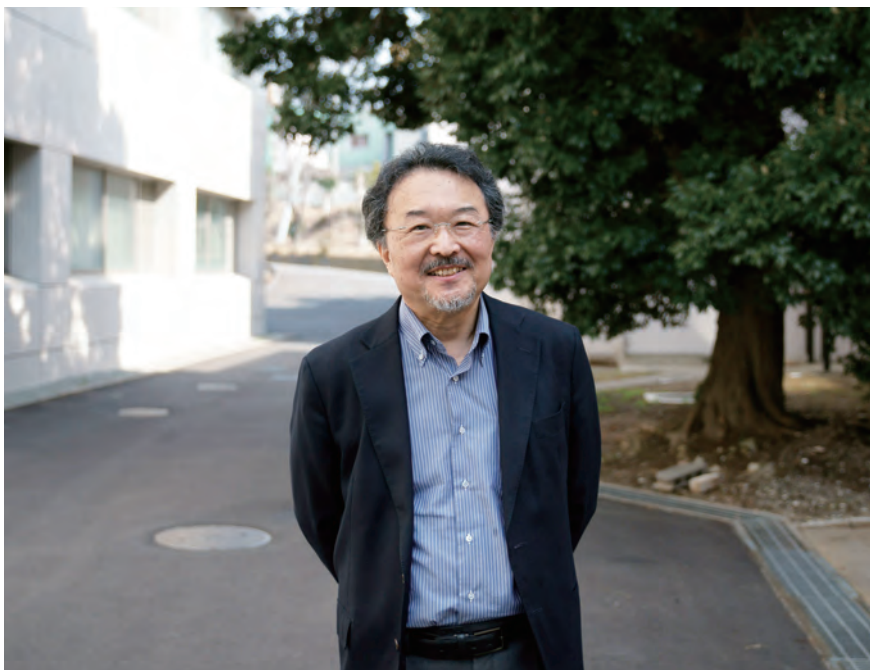
自然な文章でたずねれば、高度な検索結果が返ってくる

このような技術的な課題のなかでも、データベースにどのように問い合わせれば、うまく情報を引き出せるかという「Q and A」は、最も先端的なもののひとつです。金津東 特任准教授(DBCLS)の最近の研究では、英文で「〇〇病にかかわる遺伝子は何か？」といった自然文を入力することで、RDFデータへの

Research View014 p.60

Research View





標準の問い合わせ言語である SPARQL (SPARQL Protocol and RDF Query Language) を自動生成し、「〇〇病」や該当する遺伝子に関わる文献情報、ゲノム情報、分子構造など、意味的に正しく関連づけられた情報を一元的に表示する技術を開発しています。難しいと言われる SPARQL をライフサイエンスの科学者などが広く使いこなせることができ、一方、医療の現場などでも役に立つための性能評価の基準づくりも進めています。この他、動画の整備も始まっており、細胞の動きの解析結果を見て生物学者が意味を発見するといった「データ中心科学」の実現も期待されます。まさにそんな将来へ向けて、継続的な努力を続けていきます。



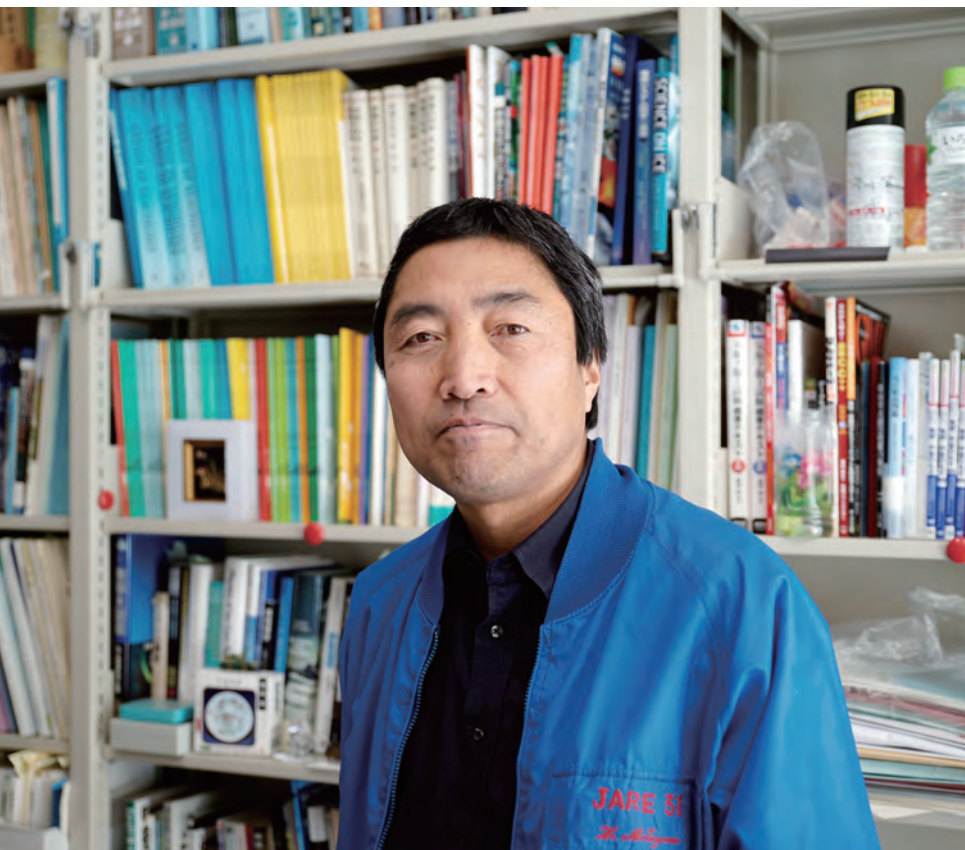
晴れた日には富士山が一望できる、三島市の国立遺伝学研究所にて。DBCLS は2014年春、東京大学(本郷)から、センター本隊は千葉県柏市へ、一部は国立遺伝学研究所へ移転した。

▶ プロジェクト概要

<http://rc.rois.ac.jp/database/d02/>

▶ DBCLS ウェブサイト

<http://dbcls.rois.ac.jp/>



Research View

Research View 004

**南極に積もる雪から、
太古の生命と環境を削り出す。**

国立極地研究所・教授

プロジェクトディレクター 本山 秀明

Project

地球・環境システム

INTRODUCTION



「氷床コア」と呼ばれる稀少な氷柱を削り出す場所は、標高3,810m、年間平均気温マイナス54℃という南極ドームふじ基地。すべての水が凍り、液体の水が存在しないため、その氷上には生命活動がないといえます。人だけでなく多くの生命にとってもサバイバルが厳しい極地の環境下では、生物の多様性が極端に少ないこと、ウィルスがないので風邪をひかないこと……等々、ふつうの生活からは想像できないさまざまな特色があります。そして降り積もる雪の下に封じ込められた氷は、なんと70万年から現在に至る私たちの地球環境や生命を、今、いきいきと語り始めているのだそうです。お話は、プロジェクトディレクターの本山秀明教授(国立極地研究所)です。

004

氷床コアとコケ坊主、
ふたつのターゲット

われわれのプロジェクトには、大きく氷床コアとコケ坊主という2つの対象があります。極地研が標本を獲得し、極地研や遺伝研の研究者が遺伝子情報を導出し、その解析に統数研・情報研の研究者にも参加してもらうというバランスのいい融合研究を、10年近く推進してきました。そのうち、私自身は雪や氷を専門に研究しています。雪はどんどん降り積もりますから、上の方が新しい雪で下の方が古い雪ですね。南極の大陸を覆っているたいへん大きな氷の塊を「南極氷床」といいますが、この氷床をわれわれはずっと3,000m下まで掘って、70万年前にあたる氷を掘削・採取しています。この氷から過去の地球の気温や、二酸化炭素・メタンといった大気成分などが分かります。たとえば過去70～80万年前から地球は寒い時代が7～8万年、暖かい時代が1～3万年という10万年周期で大きな気候変動をしてお

り、今よりも地球平均で5～6℃低く、極地ではさらに寒かったことなどが分かっています。大きな気候変動は、海底の堆積物からも知ることができますが、時間あたりの堆積量が少ないため、雪がたくさん降る南極のほうのはるかに時間分解能が高く、年代を詳細に特定できる優位性があります。また過去の大気がそのまま残っている点も、氷河や氷床ならではの貴重なデータなんですね。

氷床コアのなかの
太古の微生物を解析する

氷床コアの分析は地球の気候や環境の復元だけでなく、その中に非常に微量の微生物が含まれており、氷床コアの中の大きな気候変動とともに、生物がどう対応し、どう進化し、またその遺伝子がどうなっているか、といったことが突き止められる可能性があります。しかしながら氷上に生命活動がないため、微生物が予想以上に少なく、解析が非常に困難

であることがわかってきました。ところが、実は氷の塊である南極氷床も、地球内部から伝わる地熱によって氷が融けているところが多くあります。われわれの探掘では岩盤直下まで到達していて、深部は雪や氷の重みによる圧力で融点下がりが、マイナス2℃の3,000m下には水脈のようなものがあるんです。そこで今、探掘してきたその水に集中して、解析作業を進めています。

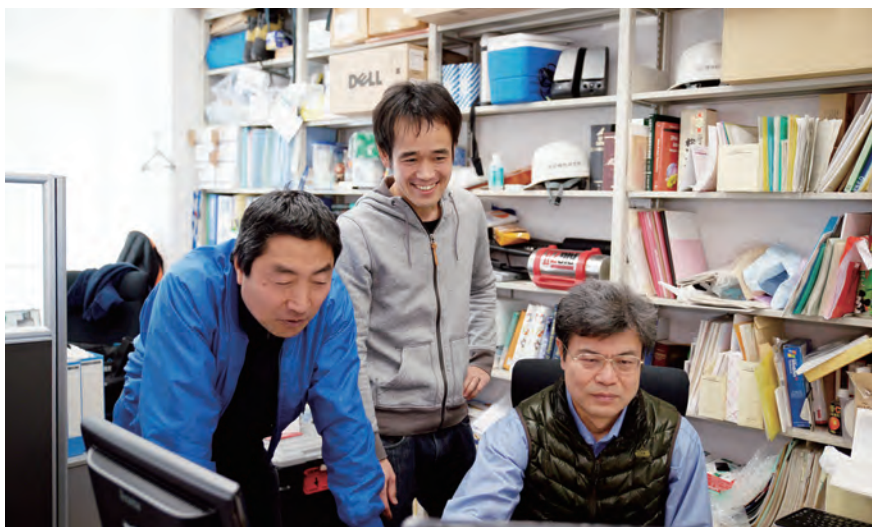
世界で一番難しい ゲノムのサンプル!?

この分析は、瀬川高弘 特任助教(国立極地研究所、新領域融合研究センター、写真中央)がサンプルからゲノム情報を取り出し、近藤伸二 特任准教授(国立極地研究所・新領域融合研究センター、写真右)が情報解析するというチームワークで進めています。たとえば光合成をするシアノバクテリアという生物の断片が見つかったら、現在地上にある似通った種の遺伝子配列と比較します。すると一致する度合いが低い、つまり未知のものが多く含まれているんですね。一方、難しい点は、まずサンプル数が非常に少ない

上に、年代が古いのでDNAがちぎれていて、なかなか長いひと連なりの配列を拾えないことです。もうひとつには「コンタミネーション」と呼ばれる技術的な精度にかかわる問題で、本当にその時代の氷から取れたものなのかどうかを証明しなければならない、ということなんです。

コケ坊主が 与えてくれる示唆を活かして

研究の進展につれ、最も興味深いのは、このような氷床コアの地球変動や微生物の研究と、伊村智教授(国立極地研究所)が進めているもう一つの対象であるコケ坊主が、地球生命システムという1つのテーマの中でどのように関わり、何を明らかにしてくれるのかという点です。「コケ坊主」というのは、南極昭和基地の周辺地域に点在する湖沼底に棲息する、大きいもので直径30cm、高さ60cmほどの微生物の共生体です。極地では陸上よりも水の中のほうが比較的穏やかな環境と言えますが、地球上のあらゆるデータベースと照合すると、コケ坊主の中にいるのと同じ菌類の仲間が、北極や中緯度にもいること





004

がわかってきたんです。すると、同じ種類の生物が、寒い場所と暖かい場所に生育しており、そこにどのような遺伝子の違いがあるのか、解明できる可能性があります。また氷床コアのほうでは、生物相の変化や遺伝子を年代順に追っていくことで、寒い時代には寒い耐性を持つ遺伝子が入っているとか、逆に間氷期の微生物を見ることで暖かい気候に適した遺伝子はどれかなどを導出できる可能性があります。地球環境の危機が叫ばれる現代、いかにサステナブルな社会を形成し、人間が絶滅を逃れることができるのかに示唆を与えられればと考えています。



直径10cmの円柱状で長さ約4m毎に削り出される氷床コア。3,000mの深さだと3時間以上を要する作業だそう。引き上げられたコアは、深さ決めや初期解析を終えた後、日本へ送られ、「上向き」を示す矢印表示とともに袋詰めされてマイナス50度で厳重に冷凍保管されている。

▶プロジェクト概要

<http://rc.rois.ac.jp/research/b01/>

▶プロジェクトウェブサイト

<http://tric.rois.ac.jp/earth/>



Research View 005

**イノベーションに役立つ
データ中心“政策”科学を。**

国立情報学研究所・教授

プロジェクトディレクター 曾根原 登

Project

社会コミュニケーション

INTRODUCTION



データ中心科学 リサーチコモンズで、
データベース「人間・社会データ」

と、プロジェクト「社会コミュニケーション(コミュニケーション情報学)」という2つのプロジェクトディレクターを務め、一体化させた研究を推進する曽根原登教授。人間と社会のリアル&サイバー空間に関わる研究を礎に、ビッグデータ時代の情報利活用技術を、地域再生やイノベーションを牽引する教育、医療・健康、防災、観光・農業分野へ活かす研究を進めています。

Research View016 p.68

500

経験や勘から データ中心の意思決定へ

政策決定や意思決定の立場から考えると、今までは人間・社会に関する合理的な判定を行うのは、なかなか難しかったと言えるでしょう。つまり部分的なデータや不完全な知識で、経験や勘に基づいた主観的な判断をしなければならなかったと思うんです。でも現代の日本社会は、ネットワーク環境やセンシング環境が急速に整い、スマートフォンやウェブから集めることのできるビッグデータがありますね。ならば、科学的データに基づいて価格を決めたり、さまざまな政策決定を行うことで、大いに経済再生やイノベーションにつなげていけると思うのです。

傾斜地の自殺はなぜ多い？

たとえば、日本では毎年約3万もの人々が自殺しています。高齢者の自殺も多いし、線路への飛び込みで通勤電車もよく止まります。科学が何かできないだろうか。そこでこの問題を扱っているサブテーマでは、いろんな公的統計データを駆使したり、調査したり、またSNSやツイッターで自殺をほのめかすつぶやきにアラートを出すというように多角的にアプローチしています。たとえばこれらのデータから、15度以上の傾斜地に住む人の自殺率がとても高いことがわかっています。なぜでしょう？—— 出不精になるからなのだそうです。家にこもって悩みを抱えてしまう。社会問題の原因をこのように物理的な条件で特定し、「こころの問題」にしてしまわないことが重要です。これを応用すれば、津波対策として検討されている高台への高齢者の移住にも、重要な示唆を与え得ることがわかるでしょう。

地域再生の肝「観光」に ITのパワーを

観光政策も、IT化による経済効果が見込まれる領域です。日本全国に約1,720の地方自治体があり、それぞれが別個に観光案内したり、人手による統計や調査を行っています。また100メートルしか離れていない観光地でも、もし市町村の境を越えたら調査や施策の対象外であるといった現状があるんですね。そこで一藤裕 特任研究員(新領域融合研究センター)が開発した「Web データ駆動の観光予報システム」では、さまざまなウェブサイトで公開されている膨大で多様な宿泊施設関連データを横断的に収集・蓄積・分析することで、行政区の枠組みを越えて観光地域の宿泊状況や料金を予測することができます。また2013年3月11日はどうだったかというように過去のデータも確認でき、京都市観光MICE推進室との調査研究協力によりデータの信頼性も検証しました。他の地域でも活用できるシステムになっているので、富士山の世界遺産登録で海外観光客の増加が予測される山梨でも、ペンションなどの小規模宿泊施設を対象としてこのシステムの応用化が始まり

つつあります。さらにこれを不動産情報に応用した分譲、賃貸、リフォームにかかわる意思決定を支援するシステムの開発も進めています。どんな物件が売れ残るリスクを持っているのか、リフォームではどこに資金を投じたらいいのか……

すべてデータに基づいて判断

を下せるようになるでしょう。

参加型アプリで、 防災・減災対策支援ツール開発

また非常用システムは、緊急時に特化したシステムよりも平時に稼動しているものがそのまま活かせるシステムのほうが役立つことが知られています。そこで「Web データ駆動の観光予報システム」は、実は平常時と同じ使い方で、災害発生時の帰宅難民対策等にも利用できるんです。また2013年に開催した産学連携による「スマートフォンアプリ開発 全国学生コンテスト」からは、帰宅途中などの生徒児童が今いる地点からどこへ逃げたらいいかをナビゲートしてくれるアプリケーション「てんでんこラリー(東北工業大学)」が生まれました。地域の子どもたちが

Research View013 p.57

Research View





005

大学生と一緒に町を回りながら避難地などを記録し、これを統合してハザードマップを作成し、防災に備えるツールです。こちらも今後さまざまな地域で、活用が広がっていくでしょう。

▶プロジェクト概要

<http://rc.rois.ac.jp/research/b03/>

▶プロジェクトウェブサイト

<http://tric.rois.ac.jp/human/>



Webデータ駆動の観光予報システム」と一藤裕特任研究員。

京都の約400の宿泊施設のウェブ上にある予約システムから日々の稼働率、料金、駅からの距離などの特徴、サービスのオプションなどのデータを取得・整備し、実際のデータと付き合わせて精度を上げていく。マップやグラフで視覚化された画面では、空室状況や月ごとの稼働状況などがひと目でわかる。



Research View

Research View 006

**データを使って
モデルをリアルに近づける。**

統計数理研究所・教授

プロジェクトディレクター 中野 純司

Project

データ同化・シミュレーション支援技術

INTRODUCTION



今回ご紹介するのは、データ中心科学「リサーチコミュニティ」の3つの基盤整備事業のうち「モデリング・解析」を担う、「データ同化・シミュレーション支援技術」のプロジェクトです。

「データ同化」

とは、ひとことで言えば、シミュレーション・モデルとデータ解析を結び付ける統計手法。この手法は物理系から生物系、マーケティングなどの経済系まで、およそあらゆる分野で役に立つ応用範囲の広さを、大きな特徴とします。また本プロジェクトを通じて統計学を学び、大量データの取り扱いやプログラミングに親しむことで、まさにビッグデータ時代に要請される人材が巣立っています。プロジェクトディレクターの中野純司教授(統計数理研究所)をご紹介します。

Research View010 p.46

現象の奥にひそむ本質的な動き

本プロジェクトは、もともと統数研樋口所長が始めたもので、われわれはこれを引き継ぎ、研究所にある「データ同化研究開発センター」を中心に、国立極地研究所や国立遺伝学研究所のグループと連携して研究を進めています。そもそも統計とは、理論だけではあり得ません。データを使うことによって、その確率的な現象の奥にひそむ本質的な動きを捉えたいという大きな目的を持っています。遺伝学者でもあったロナルド・フィッシャー(Sir Ronald Aylmer Fisher, 1890–1962)は、10、20といった少ない数の実験データから、いかにその背後にひそむしくみ(モデル)を見つけ出すかを考えるなかで、現代に連なる統計学を確立しました。今あるデータで、まだ手にしていないデータについて考えたり、予測したりすることができるのは、やはりモデリングの力だと言えるでしょう。

2種類のシミュレーション

現象が数式でほぼ記述できる分野では、昔からシミュレーション・モデルづくりが行われてきました。よく行われているのは1個のモデルを決め、これに必要な初期値を与えて未来の状態を生成するような「物理シミュレーション」です。これに対して「統計シミュレーション」では、モデルづくりのプロセスにおける「不確定性」が大きな役割を果たします。大量のデータがあったとき、それを説明する式をひとつには決めるに、という場合にどれが一番いいかを推定する。このためには、まずなんからの量を与えて、実際の現象との「ずれ」を計算します。データを全体として見たときに最も「ずれ」が小さくなるようなモデルを選び出し(パラメータを推定し)、そのパラメータの近辺でもう一度探すという作業を繰り返していきます。ちなみにランダムに発生する事象のデータを作成するには乱数が有効であり、またデータにフィッ

Research View007 p.33

トするモデルを探すには「粒子フィルター」と呼ばれるベイズ推定のアルゴリズムを活用しています。

「データ同化」とは何か？

このように乱数を使って作り出したデータを元に計算を積み重ねていったシミュレーション・モデルを、実際のデータと比べて見ると、やはり違うところがあります。そこで実際のデータを使って、その結果をなぞるようにモデルをフィットさせていけばどんどん精度が上がり、本物に近づいていくはずですね？ 実はこの手法こそが、「データ同化」です。人間社会の現象には、どうしても制御しきれないところがある。物質にしても、そもそも不確定なのだということを量子力学は示しています。誤差だけでなく、データの欠けなども避けることができません。そういった事象をうまく扱えるのが統計学であり、データ同化である——つまり、より現実的なのですね。

コラボレーションから “筋のいい手法”を探り出す

われわれのプロジェクトは「数理・計算」「モデリング」「データデザイン」の3チームから成り、それぞれ統数研、極地研、遺伝研が中心になって推進しています。とはいえ、そもそも統計学はデータを見ながら、あるいはデータをつくりながら、これが知りたいという目的に応じて手法をつくってきた学問ですから、研究所間の交流も盛んですね。たとえばこの実験ではある「違い」を絶対に区別することはできないとか、なるべく少ない実験で言いたいことをはっきり示すにはどんな実験が必要かといったことを、まさに統計学が提案できるのです。例えばデータデザインチームでは、このような共同研究を遺伝研・統数研のメンバーで進めています。そしてその際、われわれ統計学者は、なるべく汎用性の高いモデルをつくろう、と挑戦しています。なぜなら筋のいい手法は、幅広い分野で使い回しが利き、長く使われていくからなんですね。さらにいろいろな場面でデータ同化を使いやすくするために、適切な環境を提供することも重要です。そのため





006

Web上でのシステムの構築、統計ソフトウェア「R」への移植作業なども進めています。2014年度には、統数研が大学や研究コミュニティに提供するスパコン利用の新サービスもスタートし、「データ同化」の活用とサポートがいっそう拡充されます。

▶プロジェクト概要

<http://rc.rois.ac.jp/research/a03/>

▶プロジェクトウェブサイト

<http://tric.rois.ac.jp/math/>



統計数理研究所のデータ同化スーパーコンピュータシステムの前で。本システムは世界最大規模の共有メモリー型スーパーコンピュータであり、大規模なデータ同化並列計算を行うことができる。



Research View

Research View 007

**感染症は都市で
どう拡がるだろうか。**

統計数理研究所・特任助教

齋藤 正也

Project

データ同化・シミュレーション支援技術

INTRODUCTION



「データ同化・シミュレーション支援技術」のプロジェクトで活躍する齋藤正也特任助教(統計数理研究所)。インフルエンザなどの感染症が都市部でどう拡がるのか、「データ同化」という手法を使って、その予想に取り組んでいます。

最新の統計手法のひとつと

言える「データ同化」

を、感染症の拡大という具体的なテーマに応用すると、私たちの生活にどのように役立つのか、またどんな難しさがあるのか？——日々進化を遂げつつある研究開発の現況をご紹介します。

Research View006 p.30

データ同化で パンデミックに備える

2009年に起こった新型インフルエンザの大流行を、ご記憶の方もきっと多いことでしょう。同年6月には世界保健機関(WHO)から「世界的流行病」宣言が発表され、日本でも「パンデミック」という言葉が一気に広まりました。われわれはこの問題に何か対策できないだろうかと考え、当時から交流のあった東京大学医科学研究所のグループと議論を始めました。この新型インフルエンザウイルスは、日本ではまず関西地方で流行し、その後東京で感染者が見つかる、という経過を辿っていました。このようなウイルス拡大への介入策として、当時は、たとえば大阪—東京間の電車の全面運休もやむを得ないのか？ といった議論もありました。そこでわれわれは、電車を止めずに、ワクチンをうまく活用して拡大を緩和する可能性を探ることにしました。また、もし都市から都市へ感染することが予測できれば、

自治体個別の対策ではなく、共同で取り組む必要性を示唆できます。そこでこのための感染予測シミュレーションをつくったのが、本研究のきっかけでした。

2つの不確定性により 推定に時間がかかる

感染症データの難しさは、まず実験によって現象を再現することができず、情報が極めて限られている点です。そして人が接触したからといって、必ずしも感染や発病が起こるわけではないという不確実性があります。このような2種類の不確実性があることから、いろいろな場合を想定した複数のシナリオをシミュレーション計算に加える必要があり、するとスパコンを1日以上駆動してやっと予測できるような計算量になります。データ同化という手法は、このように取得できる情報量が限られていたり、複数の可能性を並行して追っていかねばならないような場合に、威力を発揮してくれます。

「適当にどこか」を 再現するには乱数が利く

一方、設定をごく簡単にして、鉄道でつながっている3つの町に、それぞれ会社・学校・家庭・公園などがある場合を考え、通勤・通学などで人がゆききしながら、感染がどう拡がっていくかを可視化した動画も作成しました。このようなシミュレーションでは、一人一人の動きを逐一記述せず、たとえば電車が混んでいる状態とか、週に約2回公園へ行くといったおおよその動きを、乱数を使って再現します。感染についても、人が接触しても感染したりしなかったりするけれども、全体としてみれば一定の確率で感染しているので、同様に乱数を使って感染拡大の様子を構成しています。このシミュレーションを使って、たとえば会社員全員にあるタイミングでワクチン接種したら、いかに効果的に感染緩和できるかといった予測も行っています。

実データを使ってより具体的な 予測を実現する

このような予測ができるようになってきたものの、このシミュレーションにはまだ、実際の感染者に関する調査結果が反映されていません。そこで実データにぴったりと合わせるデータ同化の手法を、現在模索しています。またこの際、会社や家庭といった項目よりもやや抽象度の高いモデルを使うことにより、予報に使えるようなシステムが組めるのではないかと考えています。そして最終的なゴールとして、たとえば東京一大阪間を行き来する感染者数などのデータに基づき、たとえば来週どれだけ感染者が増えるかといった予想につなげていきたい。2009年にはいわゆる水際防衛という政策が採られましたが、感染には潜伏期間の問題があり、のちに感染症の専門家が、最悪の場合で感染者の9割が防衛をすり抜けてしまうという研究結果を出しました。このような場合、統計が活躍したはずであり、今後は、





政策決定やひいては人々の生活に使える道具をつくるために、がんばっていきたいと考えています。

▶プロジェクト概要

<http://rc.rois.ac.jp/research/a03/>

▶プロジェクトウェブサイト

<http://tric.rois.ac.jp/math/>



統計に関する書籍を広範に集め、広く一般に供している統計数理研究所図書館にて。大学・研究者の閲覧はもちろん、一般の方々にも公開している。



Research View

Research View 008

オーロラはどこに降る？

統計数理研究所・特任研究員

才田 聡子

Project

データ同化・シミュレーション支援技術

INTRODUCTION

オーロラという現象を
計算で再現する

上の写真は、地球の磁気圏を磁気流体力学方程式で解いたシミュレーションを表しています。地球と太陽の重力が釣り合う「ラグランジュ点」を周回する衛星が、太陽風のプラズマの密度等を観測しており、これを入力してコンピュータに計算させるとこのような図を描くことができます。太陽から地球へ、「太陽風」と呼ばれる高エネルギーで高速のプラズマが押し寄せています。緑〜赤色で表されている帯状のものは、このプラズマがどのくらいの強さで地球の磁気圏に当たっているかの圧力を示しており、たくさんの線で描かれているのは、地球がつくる磁場の様子です。地球は、その内部にある核が自転の影響を受けながら熱対流することにより、コイルに電流を流したときのように磁場がつくられます。プラズ

天体の動きが高精度に予測できる現代、オーロラという現象は、たとえば天気予報のように予測できるのでしょうか？ また南極や北極などの高緯度な地域で、しかも夜だけに見られるオーロラは、私たちの生活にいったいどんな関係があるのでしょう？——今回採り上げる

のは「データ同化」

の手法を採り入れ、オーロラ現象解明と同時に、生活に役立つ知見を目指す研究です。「データ同化・シミュレーション支援技術」プロジェクトの第3回目は、国立極地研究所との人材交流から生まれた、才田聡子特任研究員(統計数理研究所)の研究をご紹介します。

マは磁力線を横切ることができない性質を持っているため、地球を直撃することなく、はね返されているのです。

太陽からのプラズマが降り込む夜

ところで地球の磁場は、南から出て北へ向かっています。太陽風の磁場も同じ方向を向いているときはよいのですが、もし逆向きになって太陽風の磁力線と地球の磁力線が接近すると磁場の拡散が起こり、2つの磁力線がつながる「磁気リコネクション」という現象が起こります。すると地球の磁力線は太陽風と一緒に流されて、夜側にひっぱられてしまいます。夜側に引っ張られた南半球から出る磁力線と北半球に入る磁力線はお互いに逆の方向を向いているので、ここでもう一度「磁気リコネクション」が起こります。すると太陽風の中にあったプラズマが地

球の磁力線に乗り、磁力線に沿ってそのまま南極や北極に降り込んでくるのです。この時、プラズマと衝突した大気中の原子や分子が「励起状態」になり、酸素なら緑色や赤色、窒素ならピンク色というように特定の色を帯びながら、「電離層」と呼ばれる地上から100キロメートルぐらいの領域で光るのがオーロラです。オーロラの発生しやすい領域は地球の磁極を囲むように楕円形になっていて、オーロラオーバルと呼ばれています。

地球データはピンポイントデータ

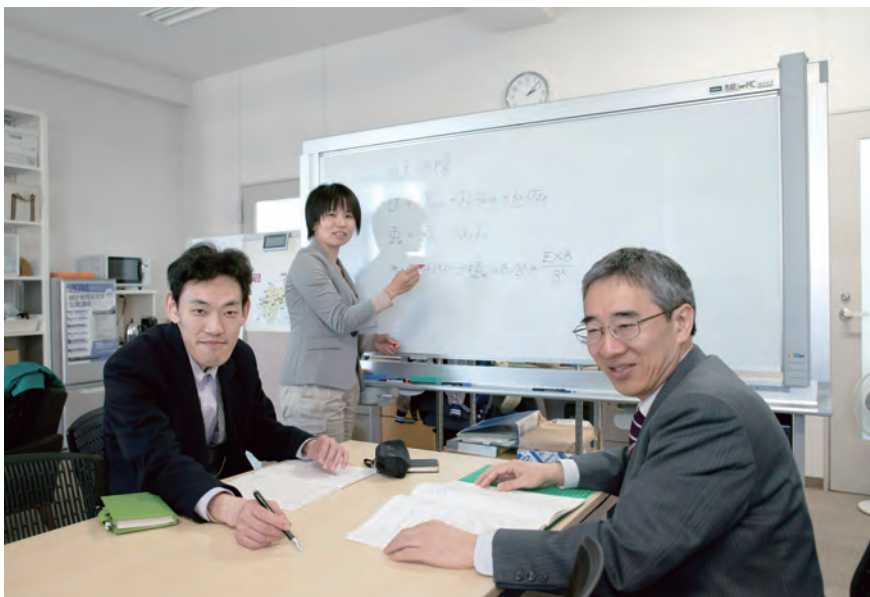
私は統数研に来る前は極地研に所属しており、実は南極で磁気測定のための観測器を設置した経験があります。オーロラを研究するようになったのは2007年頃からですが、この現象には太陽と地球の関係すべてが含まれているといってもよく、オーロラをみることは、宇宙環境の変動をみることでもあります。このような超高層域の大気を対象とした分

野は「超高層物理学」と呼ばれ、物理学的なモデルづくりが行われていますが、必ずしも現実を反映しきれない面がある。というのも、地球物理のデータは場所も時間も、非常に限られたものなんですね。人工衛星が観測できるのは、この広大な宇宙の空間的にも時間的にも、ある1点に過ぎません。そこで統計的な手法を使うと、限りあるデータから全体を考える、大きな手掛かりになるんですね。今までの私の研究では、その時点その場所での現象についてしか言えなかったことが、データのない別の場所で起きている現象についても議論ができるようになるのです。

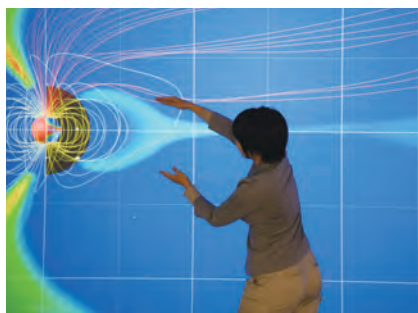
地上の暮らしに役立つシミュレーションを

しかもオーロラの位置は地上から観測できるので、シミュレーションの値を現実に近いに近づけることができます。たとえば2011年には、オーロラが





経度方向に大きく移動する様子をシミュレーションによって再現し、南極の昭和基地とアイスランドで観測されたオーロラの形の違いを説明することができました。現在は特に、磁気圏と電離圏の境界の計算をもっと正確に行うために、電離圏の観測データをシミュレーションに合わせるデータ同化を進めています。ただ、オーロラがどこに出るかがわかっただけではまだまだ、です。せっかくデータ同化を使うならば、磁気圏と電離圏の間でどんなふうに物理量が交わされていて、どんな過程を経て、どんな関係性が成り立つのか、リアルに近づけて描きたい。実際、太陽風は生命に有害であるだけでなく、人工衛星を破壊することもあるため、日常的に使っている通信が途絶えたり、大規模な電力会社の変圧計が故障したりといった悪影響も十分考えられます。どのくらいの強さの電流がどう電離層に流れているのか、もっと定量的に見積もって、影響が論じられるところまで到達したいと考えています。



スクリーン左の赤い半球が、地球の「昼」側。地球の磁場によっていったん遠ざけられた太陽風が、手で示しているあたりで「磁気リコネクション」によって、ふたたび地球の「夜」側へ降り込んでくる。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/a03/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/math/>



Research View

Research View 009

**どんなしくみで「種」が
分かれるのだろう。**

新領域融合研究センター・特任研究員

岡 彩子

Project

遺伝機能(生命)システム

INTRODUCTION



まずマウス系統の 世界分布を俯瞰する

パンダのような柄をした写真のネズミは「JF1(ジャパニーズファンシーマウスの略)」といって、デンマークのペットショップで見つかった個体を起源とする系統です。しかし、そもそもは日本で育てられていたもので、江戸時代には趣味人の愛玩用として『珍翫鼠育草』という書物に記録が残されており、幕末にヨーロッパに伝えられました。現在、世界に分布するマウス亜種には、アフリカからアメリカや西ヨーロッパへ渡った「ドメスティカス」、ユーラシア大陸から東ヨーロッパにかけての広い範囲に分布する「ムスクルス」という大きく2つの亜種があります。この2つの亜種は、今から50～100万年ほど前に共通の祖先から生まれ、長い間に突然変異を蓄積して、数千年前にドイツからギリシア東端へ円弧を描くラインで出会ったといえます。これを「ハイブリットゾーン」といい、この領域では、遺伝的にか

雄のロバと雌のウマの掛け合わせである「ラバ」は、繁殖力を持たないために子孫を残すことはできません。このように2つの種を掛け合わせた時に、生殖や生存に障害が起こる現象を「生殖隔離」といい、長い間、生物学者の関心を集めてきました。どんなしくみで「種」が成立するのか……このような難問に挑むには、系の純度が高く、精密な遺伝解析と遺伝子操作が行える「近交系」というバイオリソースの存在が欠かせません。国立遺伝学研究所ではさまざまな実験用「モデル動物」の系統維持を行っています。なかでも近交系間で染色体を交換した「コンソミック系統」というマウスを用いて、2014年4月、「生殖隔離」という積年の謎に光を当てる研究成果が生まれました。「遺伝機能(生命)システム」プロジェクトから、岡彩子特任研究員(新領域融合研究センター)の研究をご紹介します。

離れてしまった2つの亜種が生殖隔離を起こすことが知られています。ではそれらのゲノムを比べたら、いったい何が起きているのでしょうか？

「ドブジャンスキー・ミュラー」 モデルとは？

この問題に答えるのが、有名な「ドブジャンスキー・ミュラー(Dobzhansky-Muller)モデル」です。仮に「AABB」という遺伝子を持つ共通祖先から分化した集団Aに、新しい遺伝子型「a」が出現して「aaBB」が定着し、一方、集団Bには「AAbb」が定着したとします。その後集団AとBが出会い、遺伝子型aとbが不適合である場合に生殖隔離が起こるのだ……というわけです。これを説明するには、おそらくAABBという遺伝子によってつくられるタンパク質が原因だろうと考えられ、これまでさまざまな研究が行われてきました。しかしながら哺乳類で種分化

を説明する遺伝子は、実はまだほとんど見つかっていません。マウスでは2009年に一例だけ、Prdm9という原因遺伝子が報告されましたが、その具体的な分子メカニズムは解明されていません。私自身もずっと原因遺伝子を探す仕事をしてきたのですが、遺伝子を限定していけばいくほど、生殖隔離という現象が消えてしまう。原因遺伝子が存在するという前提自体が間違っているのかもしれないという疑問から、現在の研究に至ったんです。

実験室で生殖隔離を精密に再現する

ハイブリッドゾーンで起きているような現象は、実験室の中で再現することができます。現在、実験で最も標準的なB6系統は、詳しく言うところ9割がヨーロッパ、1割が日本由来という遺伝的な背景を持っていますが、ドメスティカスとして取り扱うことが可能です。このB6系統に、ムスクリスと遺伝的に近い「モロシヌス」亜種にあたるMSM系統のX染色体を導入した「コンソミック系統」を調べると、精子形成が

異常になることが既に確かめられています。このような異常は、マウス亜種が持っている19対の常染色体や性染色体のY染色体を導入したコンソミック系統では起こりませんでした。そしてX染色体コンソミック系統の精巣の遺伝子の発現を調べたところ、X染色体の遺伝子の15%に発現低下、5%に発現増加が起こり、とてもばらついていましたね。そこで私たちは、遺伝子の近くにあつて発現量をコントロールするシス制御因子に注目してさらに統計解析を行い、発現低下は〈X染色体上にあるMSM系統のシス制御配列〉と〈常染色体上にあるB6系統の転写調節因子〉の遺伝的不適合が原因であり、また発現低下していた遺伝子の多くが生殖に関わる「生殖幹細胞」で機能していることを突き止めました。常に厳しい生存競争にさらされている生物は、生殖関連遺伝子の進化が速いと考えられています。今回の発見は、そのような遺伝子発現をさかのぼって、その起点となる転写調節因子が速く進化したことにより、分化の途中にある亜種間に起こる生殖隔離を説明し得たことに意味があると言えるでしょう。





人間の目だけでなく統計解析で現象を確認する

振り返ると、発現データの分布をグラフで見たときに、もしかしたらこのゆらぎが生殖隔離の根本原因なのではないか、つまり生殖隔離の原因は、これまで想定されてきたような特定のタンパク分子の亜種間差とは別のところにあるのではないかと直感したんですね。というも、人間の目で見ても明らかにX染色体のばらつきは幅広いわけです。では、このばらつきは単に確率的なものではなく、「有意差」といわれる統計的に意味のある違いなのか？そこで、「遺伝機能(生命)システム」プロジェクトの一貫として統計数理研究所の藤澤洋徳先生との共同研究で、この問題を解決していただきました。今回提案したモデルは、統計学的にも難しいと言われる藤澤先生の「分散の有意差検定」が大きな根拠の一つとなっています。私たちは現在、さらに大規模なRNA-seqによる発現データを蓄積しており、今後は国立極地研究所の近藤伸二先生(新領域融合研究センター)との共同研究で、このデータから、いっそう詳細な生殖隔離の分子メカニズムを解明していく予定です。



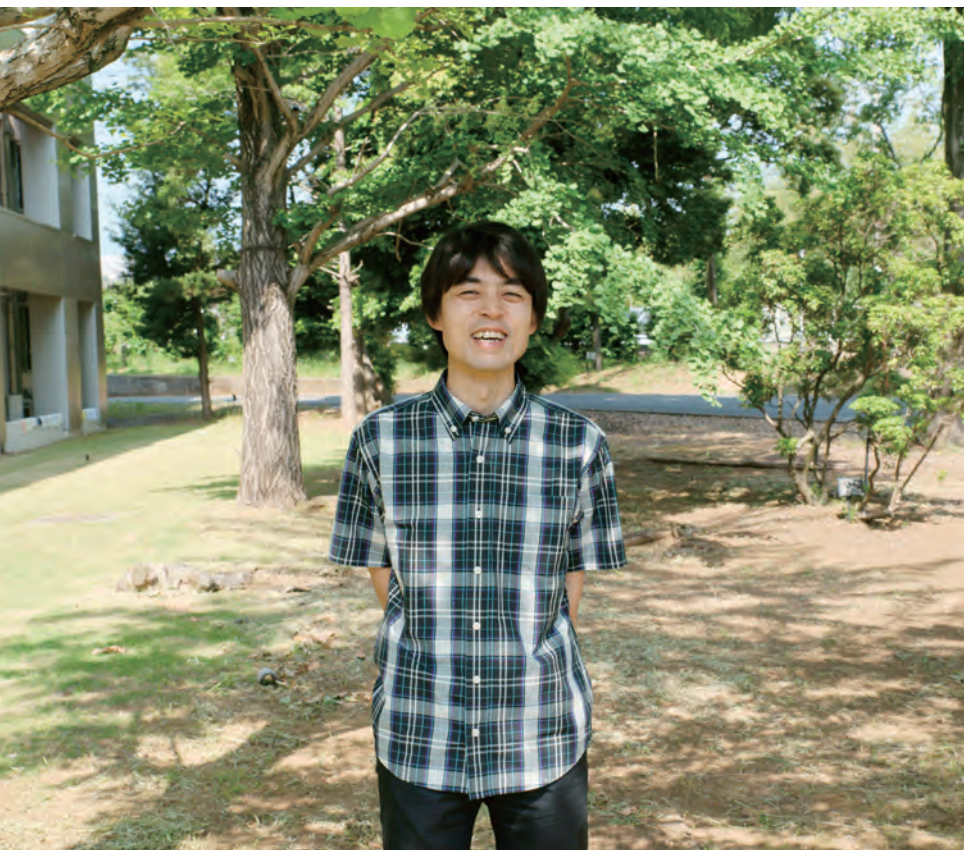
飼育による系統維持の一方で、冷凍保存によってマウス亜種の多系統を維持する、国立遺伝学研究所の一角。人工授精による発生初期の胚細胞を冷凍保存し、使うときは保存された胚を融解してマウスの子宮に移植する。2つの亜種の正式名は「ムス・ムスクルス・ムスクルス」と「ムス・ムスクルス・ドメスティカス」。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/b02/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/life/>



Research View 010

**「建築」という視点で
細胞を解明する。**

国立遺伝学研究所・准教授

木村 暁

Project

データ同化・シミュレーション支援技術

Research View

INTRODUCTION



なぜ細胞生物学が 統計数理とつながるのか？

われわれは細胞の研究をしています。細胞は生命の最小単位であり、生きているものとそうでないものの境目とも考えることができる。生き物の中で一番単純で、しかもいろいろ動きのある様子を研究したいというのが、私の生物学的な関心です。では細胞をどう理解するのか？ もし自らの手で作ることができれば大きな理解につながると思うのですが、作れないので、せめてコンピュータ上で「シミュレーション」として作ろう、という研究をしてきました。細胞の中で核などの細胞内小器官が動いている様子を解析して、そのようなダイナミックな動きを実現するには、そこにどのような物理的な力が働いているのか、建築家が構造計算をするように調べてみよう。シミュレーションモデルをつくるためには、まず顕微鏡撮影をするのですが、そこから三次元的な位置情報、時間的推移といった膨大な量のデータが出てきます。するとコンピュータサイエンスや情報科学の問題にもつながってきて、データが

統計数理が中心となって進めている、リサーチコミュニティ「データ同化・シミュレーション支援技術」に参加する、木村暁准教授(国立遺伝学研究所)。「数理に強いというわけじゃない」という“ラボ育ち”。大学院生時代に、いち早くヒストン修飾の研究を手がけながら、一方で「折りたたまれた染色体の“ここからここまで”をどう区別するのか？ 本当にこれでいいのか？」と、歯がゆい思いも持っていたといいます。「計算できれば正しいかどうかはわかる。間違っていれば次の発見にもつながる」。そういう研究ができる場を選んだ現在の環境で「データ中心科学」を推進し、今、その手法を他の研究室へも展開しています。最も得意とするモデル動物は、写真の線虫です。

らモデルを探っていく「データ中心科学」の方法論に行きつきます。今話題のビッグデータが、細胞生物学の世界でも非常に大事になってきているんですね。

昔ながらの解釈を復活させる 「細胞建築学」

今から約100年前に顕微鏡が発明されて以来、細胞生物学の分野では細胞を押しやり、つぶしたりして、スケッチを描きながら丹念に観察するという時代が続きました。しかし約50年前にいわゆる分子生物学が勃興して、研究者がみんなそっちへ行ってしまった。でも、取り残された分子生物学以前の研究が、今や僕には新鮮というか、もし当時の科学者たちが今のツールを使ったら、たぶんもっとおもしろいことをしているんじゃないかと思うんです(笑)。より小さなスケールで見たいという方向性ではなく、少し大きな塊で見たり、解釈によってちょっと単純化したりすることがすごく大事なのではないか。そこで僕はスピリットはなるべく昔の人のまま、新しいツールを駆使して挑みたい、これを「細

胞建築学」と呼んでいます。しかし、このような中間的な視点の研究はまだ確立されていないため、これが正解という道はありません。登山計画みたいに、どんな道を行くのかという戦略を立て、コースのデザインを十分吟味することで、自分にしかできない研究を目指しています。

データ同化で「細胞質流動」を解剖する

「データ同化・シミュレーション支援技術」プロジェクトでの最近の成果のひとつが、線虫を使った細胞質流動の研究です。細胞が分裂する際、通常は2つとも同じ細胞ができるのですが、稀に性質の異なる細胞がつくられることがあります。線虫のような動物においては、このような分裂の際、細胞質の中で流動が起き、中心に近い部分にある物質がある方向にびゅっと押し出されるような動きが起きます。ではこの動きは、どこに、どれだけの力をかければ再現できるのでしょうか？ 力を直接、

正確に測ることはできないので、力を未知の値として取り扱い、まず細胞内の流れをシミュレーションで再現します。一方、顕微鏡で撮った画像から流動の速度を算出し、シミュレーション結果とつきあわせて、両者が一致するまで力の値を修正します。このような「データ同化」の手法により、画面の1ピクセル=0.1マイクロメートルぐらいのスケールで、生き生きとした細胞内の動きを対象としながら、定量的な確かな土俵に立つことで、細胞質流動という現象をいっそう具体的に解明することができました。

獲得した手法を他大学の遺伝子研究に応用する

このようなリサーチコモンズでの成果を基に、他の生物系の研究室と連携して、データ同化の手法を適用する、いわば布教活動のようなものにも取り組んでいます。たとえばある遺伝子の役割を知るためには、該当する遺伝子を除いた細胞を作製して、正常な細胞との差を

Research View006 p.30

Research View





010

調べるという方法が一般的ですが、この作業は結局、「どこが違うか」を人間がまず見込みを立てるところからスタートしています。でもわれわれが推定するのではなく、まずデータがあって、コンピュータがその現象に内在するメカニズムやパラメータを探る——余計なおせっかいかもしれないけど、データ中心科学で遺伝子研究のやり方を変えるような共同研究を、少しずつ増やしています。実際、きちんとしたデータさえ集まれば、比較的短期間でできる醍醐味もあり、いろんな生物種に幅広く挑戦できる利点もあります。共同研究によって、統計的にも正しくて、実験的にもできるラインを探し出す……そういった研究の進め方も併せて、他の分野の研究者に伝えて行ければいいなと思っています。



「細胞は、環境に応じてある量以上には増えなかったり、“空気を読む”的な部分もあったりして、“足るを知る”ように見えます。線虫はこんなふうにいる——といった理解から、生物学者は、生物の本質を指摘するような役割を担うべきだと考えています。たとえばたくさん分子が集まれば細胞ができるというわけではない……と僕はずっと思っているのですが、今でも答えられない。ただ、心の琴線に触れた疑問は持ち続けて、少しでもヒントらしきものを発見できればと思っています。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/a03/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/math/>



Research View

Research View 011

**未探索の新物質を
デスクトップで発掘しよう。**

国立情報学研究所・准教授

プロジェクトディレクター 佐藤 寛子

Database

データ中心ケミストリ

INTRODUCTION



ビッグデータから 新しい物質を探り出す

私たちの世界にある化合物は、人工化合物も含めて約7,000万種類あり、毎年数10万～100万ぐらいずつ増え続けています。ところが、量子化学計算によって理論的に存在可能な分子を探索していくと、それをはるかに凌駕する数の物質が存在することがだんだん分かってきました。ベンゼン環の部品だけでも何千種類以上もあるのですから、まさにビッグデータになるんですね。化学の分野では、これまで物質の沸点・融点などの基本的な物性や、質量分析、結晶構造解析などの分光学により得られる分子構造に関するデータ、化学反応の情報など、主に実験によって集められた実験データが扱われてきています。今、私たちが進めているのは、もう一つのデータ、つまりコンピュータを使って計算によって得られるデータです。しかも、いったん式を与えればコンピュータが自動的に探索して、どんどんデータを蓄積して

化学でおなじみの分子模型を使って最初につくってみるもののひとつに、炭素6個、水素6個が対称に並ぶ「ベンゼン環(C₆H₆)」があります。「では炭素6個+水素6個で作ることができる分子は、全部でいくつあると思いますか？」——原子間のトポロジカルな結合関係に基づいて分子構造を数え上げると、答えは217種になるそうです。ところが「量子化学計算で探索していくと、安定な構造だけで6,000以上」にもものぼるとのこと。「データ中心ケミストリ」が推進する「埋蔵分子発掘プロジェクト」は、新しい計算方法が可能にした膨大な物質の可能性の中から、目的に合った「新規物質」が効率的に発見できる高精度なデータベースを目指しています。科学的な発見にとどまらず、創薬、バイオ、材料科学にもつながるなど、基礎研究ならではの応用範囲の幅広さも大きな特徴です。研究を率いる佐藤寛子准教授(国立情報学研究所)をご紹介します。

くれるようなシステムを構築しているところです。

化学反応シミュレーションに 不可欠な量子化学計算

分子構造を計算によって求める理論的な方法には大きく2つあり、その1つは、分子や原子の間に働く力学をパラメータ化した「分子力学計算」です。計算量が比較的少ないため、タンパク質などの大きな分子構造の構造解析によく使われています。もう1つは、量子力学をベースに電子の動きまで考慮できる「量子化学計算」です。化学反応は電子の移動によって起こるため、化学反応のシミュレーションには、量子化学計算が不可欠なんです。ただし計算に時間がかかるため、大きな系に適用する際には、一般的に、2つの計算方法を組み合わせるなどの工夫がされることが多いです。量子化学理論に基づいて化学反応経路を探索する具体的なソフトウェアとしては、10年ほど前から当時東北大学の大野公一先生と前田理先生が「GRRM (Global Reaction

Route Map) 法」の開発を始められ、従来法では4原子までしか扱えなかったのが、これによって理論的には無限に大きな系でも計算可能になりました。「データ中心ケミストリ」ではこのGRRM法を用いて、化学反応経路データベースを構築し、ケモインフォマティクスやデータマイニングによる新規物質や化学反応の発見や設計を行うことを目指しています。また可視化や化学反応経路解析ツールの提供によって広く量子化学の研究者・技術者にデータベースを利用可能にする計画です。

化学反応の広大な探索空間の トレイルマップを描き出す

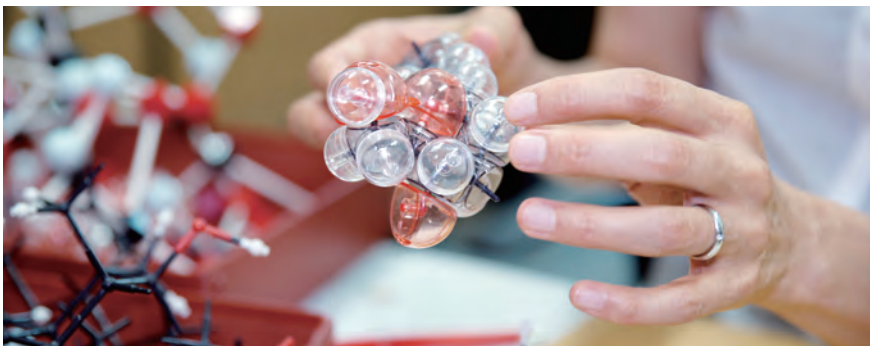
私たちは、分子構造を数え上げるもうひとつの方法「トポロジカル法」と対峙させて、GRRM法を別名「ポテンシャル法」とも呼んでいます。ポテンシャルエネルギーの谷にいるとき分子は安定であり、分子構造が変化するにはエネルギーの壁を越えなければなりません。例えばAという物質からBという物質ができる場合には、このエネルギーの山を超えることによって結合の状態に変化が起きているのです。山は高いほど越えにく

い、つまり反応が起こりにくいと言えます。そして実際には2次元ではなく多次元的な広がりを持っており、宇宙にもたとえられる広大な「ポテンシャル曲面」を構成しているのです。この曲面を越えるどんなルートがあるのか、いわば“ハイキングトレイル”を自動的に探索し「化学反応経路マップ」を描き出すのがGRRM法です。さらにこのようなトレイルに特徴づけを行い、幾重にも描くことによって、“道しるべ”になるような、より詳細な化学反応の特徴の発見も目指しています。このような「ケモインフォマティクス」に加え、データからの発掘を目指す「データマイニング」による新規物質等の発見・設計、さらにGRRMの分散化・高速化の研究も進めています。

化学反応経路マップの解析ツール 「RMapView」公開へ

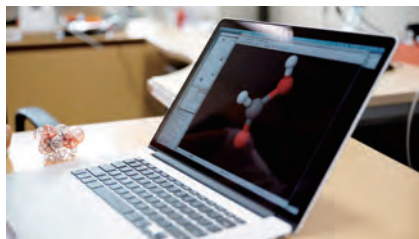
2014年7月にはその最初の成果として化学反応経路マップの解析ツールである「RMapView」をウェブ上に公開しました。これを使って、たとえば「CH₂O₂」という単純な組成式についての化学反応経路マップを選択すると、ネットワーク状のマップ





011

が画面に現れます。マップでは、分子と分子がどのような関係で結ばれているか、それらの分子がどの程度安定か、また、分子間のポテンシャルエネルギーの山の高さなどが可視化されています。最もエネルギーの低い2種の分子を分子モデルで表示すると、ギ酸の回転異性体であることがわかります。仮に2つの分子を選んで、一方を「反応物」、もう一方を「生成物」にしてやると、反応物から生成物に至る、可能な「経路」がすべて算出され、エネルギー的に起こりやすい順に表示されます。この中のひとつの経路を選択して「ムービー」を指定すると、化学反応による分子構造の変化が分子モデルのムービーで表示されます。GRRM法は、これまでに、逆合成解析や反応中間体解析、クラスター構造の解析や触媒機構の解析などに適用されてきていますが、「RMapView」が提供する化学反応経路マップの解析ツールにより、応用研究が加速され、また、応用研究の幅も広がられていくことを期待しています。化学反応経路マップには、理論的に存在可能な、様々な骨格の分子構造が「埋蔵」されていますので、例えば、ここから創薬のデザインのモチーフを探したり、あるいは既存の実験データも参照しながら、欲しい性質を持った新材料を創り出したりと、さまざまな可能性を拓くことが期待されます。



上の写真は「RMapView」の画面。今回のバージョンでは、まだ本格的な解析機能は盛り込まれていないものの、化学反応経路マップがどのようなものを体験することができる。2014年度中に、GRRMからの出力ファイルをRMapViewの入力ファイルに変換するプログラムや、新しい解析機能を搭載したバージョンなど、段階的に公開を行い、2015年春にはオープンソース化まで進める予定という。「化学ではビジュアルや分子模型はとても重要。“電子が移動して反応する”という電子の動きを矢印で記す、教科書にあるような化学反応機構の図と合わせて、理論的に計算された分子構造の変化をムービーで追うことができればやはりインパクトが違う」と、今後は化学反応経路マップから得られる特徴的な化学反応のムービー等を抽出して、ムービーのみをダウンロードできるようにもする予定。「化学の研究者だけでなく、教育的な目的にもぜひ利用して欲しいですね」。

▶ プロジェクト概要

<http://rc.rois.ac.jp/database/d04/>



Research View

Research View 012

極限環境に生きる 「コケ坊主」のすべて。

国立極地研究所・教授

伊村 智

Project

地球・環境システム

INTRODUCTION



昭和基地近くの池の中へ潜ったところ、ちょうど陸の「アリ塚」にも似た、山のかたちをした緑の物体が見つかったのは、およそ20年前のこと。発見者の伊村智教授(国立極地研究所)が名付けたこの「コケ坊主」は、その後の研究から、コケ類を中心に藻類、バクテリア、クマムシなどから成る集合体であり、約1000年かけて60cmぐらいの高さまで成長することなどがわかってきました。コケ坊主が世界的にも注目される大きな理由は、極地という極限の環境下で生きるモデル生物として、まさに理想的な特徴を備えていること。現在「地球・環境システム」プロジェクトの、氷床コアと並ぶもうひとつの柱として、遺伝子解析を含めた“データ中心”によるコケ坊主の全貌解明が進められています。

世界で3カ所しか 棲息していない生命体

コケ坊主を構成しているコケの仲間が正式に報告されているのは、昭和基地周辺の池、南極大陸の半島部の火山噴気孔、そして南米と、世界で計3カ所しかありません。南極大陸はオーストラリア大陸のほぼ2倍に当たる広い面積を持つ大陸で、たくさんの池があり、また30以上の国々が観測基地を持っています、多くの国が湖沼研究を行っています。ところがこの20年来、南極の池に新たにコケ坊主が「見つかった」という報告がない。南極の池は、地上よりむしろ穏やかな環境であり、水温は1年を通じて3〜8度くらい。水面に氷が張っていても光が差し込むので十分光合成が行えるし、天敵となる魚や大きな昆虫もいないので、コケ類にとっては増え放題です。にもかかわらず「とびとび」にしか分布しないのはなぜなのか？ 何らかの大气のかく乱があつてたまたま胞子が飛来したとすれば、半島部と昭和基地周辺を

直接結ぶ大気の流れのようなものを想定する必要があります。一方、南極の池の底を掘削して、地球の古環境を復元しようという研究も広く行われています。この堆積物データから、湖底のコケ類はほんの2000〜3000年前に入ってきたことがわかっていきます。現在とさほど気候も変わらないため、大気の流れをモデリングによって再現し、コケ類の胞子の飛来を解明する方法も模索中です。

コケとクマムシ、 極地に生きる2つの戦略

本山教授の説明

にあるように、私たちの地球はおよそ10万年単位で、寒く長い時期と暖かく短い時期という周期を繰り返していたのでは？ 私たちは今、その暖かい時期のおしまいのほうにいて、コケもこの暖かい期間にどこから侵入してきた。ところが氷期を生き延びてきた生物もいるらしい……地衣類やクマムシな

Research View004 p.21

どの微小動物がそうだと考えられています。氷期には地表も海面も氷河で覆われてしまうため、現在コケ類などが棲息している海岸近くの地表にいたのでは、生き残れるはずがありません。そこで南極大陸は氷期にどこまで氷河に覆われていたかという研究と合わせると、どうも高い山に避難していたのではないかと考えられるのです。最近注目されているのは、なかでも地熱の高い火山です。南極にもいくつか火山があり、実際、その周辺には多くの種類の地衣類やクマムシが豊富に分布しています。すると南極には、コケのように暖かい時期に侵入してきては氷河に流される「一時的な滞在者」と、クマムシのように氷期も世代を継いで生き延びる「定着者」の2タイプの生物がいる。このような生物の多様性やパターンを詳細に解明していくことで、コケ坊主の起源地の特定を狙っています。

極限の環境下では 有性生殖なんか要らない

私の研究テーマは、究極的には、極限環境での生物のあり方だと言えます。これだけ厳しい環境ですから、飛んできた胞子がどうやって定着して、どう世代を回していけるのか、ギリギリのところにいるだろう。たとえばクマムシは、厳しい時期を生き延びるときだけ有性生殖すると考えられる生き物のひとつです。個体数を増やすには、効率の悪い有性生殖よりも、雌だけで効率よくクローンを再生産していく手段をとるわけですね。この例だけでなく、極限を生き抜く彼らのやり方の中にサバイバルの基本原則が見えるのではないかと——この意味で南極は、まさにここにしかない実験場と言えるでしょう。またそのような環境下で、コケ坊主はそれ自体がひとつの森、





あるいは社会のような生態系と考えることができます。どんな生き物が住み込んでいるのか、未知の生命を含めて洗いだらい遺伝子を読み、中にいる生物たちがどんなネットワークを構成しているのかを、国立遺伝学研究所、そして広島大学との融合研究によって分析しています。例えば窒素をアンモニアに固定するバクテリアがあり、別のバクテリアがこれを亜硝酸、そして硝酸に変換し、コケがこれを吸収して利用した後、バクテリアがこれを分解してふたたび窒素へ戻していく循環なども、わかってきました。

データ&シミュレーションが紡ぐ 証拠立て

ところでコケ坊主はそもそもなぜこんな特徴的な形をしているのか？ またなぜ表面と中身が層状の構造をしているのか、考えてみれば不思議ですよね？ そこで今、総合研究大学院大学を中心とした共同研究によって、たとえば採光に最適化しているとしたらどういう形に成長していくのかななどを、コンピューター上のシミュレーションによって解析しています。また、藻類の場合には強すぎる紫外線をブロックする色素を表面に作り、内部で光合成を行います、この2層がどのようなバランスをとることで厚い藻類のマットが成長しているのかななども、計算によって確かめて

います。環境が単純で、余計なパラメータを考慮する必要がない南極では、このようなモデル計算の精度が期待できるし、計算結果と実物を照らし合わせることもできます。ああ、コケ坊主ってこんなシステムになっていたのか、とわかる成果を目指しています。



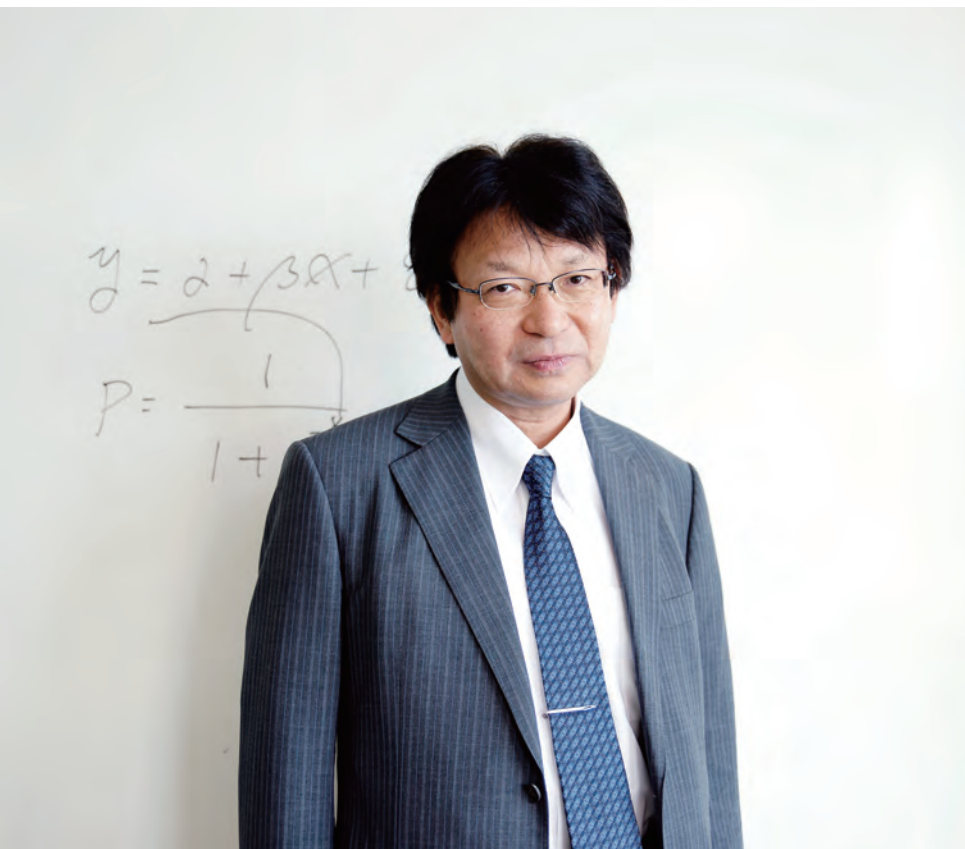
最近は、コケ坊主を広く伝えるためのぬいぐるみ「こけ坊」も完成。研究室では、生命の多様性とパターン解明へ向けて、コケ坊主の全貌解明へ向けた取り組みが進行中だ。コケ坊主を主役に、南極観測で生み出されるさまざまな地球環境データを組み合わせ、さらに構成物の全遺伝子解析というゲノムのビッグデータに挑む。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/b01/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/earth/>



Research View 013

オープンデータ

×

金融のチャレンジ

同志社大学・教授

津田 博史

Project

社会コミュニケーション

Research View

INTRODUCTION



金融ほど数字でできた世界はない。

銀行の窓口では人と人が対面していますが、それ以外はすべてと言っていいほど「数字の世界」なんです(笑)。金融機関というのはデータの固まりであり、実際にお金の流れは全部数字でできている。しかも、数字がちょっとでも狂うと大変なことになりますね。だからこのようなデータを解析することによって、さまざまな評価へとつなげることができます。たとえば銀行が企業に資金を融資する時には、金融工学や統計ファイナンスによるデータ解析に基づいて倒産リスクを推計していますし、日本の公的年金運用では、2014年の4月から従来の株式指数よりもより高いリターンやより低いリスクなどの何らかの付加価値をもった“スマート”な株式指数をベースとした運用を積極的に取り入れる姿勢が示され、株式市場でスマートベータ運用の動向が注目されています。しかし、このような日本の金融機関の動きは、実は世界的に見るとやや立ち後れています。

金融工学のひとつの最先端であるウォールストリートでは、トレーディングの主役はとくにロボット(プログラム)だと言われます。人が売買する取引とは違ってアルゴリズム取引と呼ばれるプログラムによる取引は、100万分の1秒を競って行き交う世界。「買い」か「売り」か——その意思決定を行うプログラムの背後にあるのが、ビッグデータです。世の中に生起する現象をどう数値化し、どのように組み合わせで解析したら、経済を先読みすることのできる情報へと結びつくのか。リサーチcommonsの「社会コミュニケーション」プロジェクトに加わり、「金融」という視点から観光というテーマに取り組む、同志社大学の津田博史教授に聞きました。

一方、観光は産業としてこれからますます重要になります。これまで主流だった団体旅行に対して、インターネット経由で個人旅行(Free Independent Travel)が伸びているのですが、この実態はまだよく把握されていません。そこで、ウェブ上にあるデータを日々大規模に集めて個人旅行者の嗜好やニーズからどんなホテルプランが選ばれているかを把握し、この行動が見えるようにしよう——この世界的にも新たな試みとして曾根原先生、Research View005 p.24 研究員の藤さんがデータベース化した情報を、使わせていただき、研究してきました。

宿泊プランの人気度への影響を分析する「ロジットモデル」

データベースは京都市にある約200の宿泊施設のウェブ予約システムから取得されたもので、時間と

ともにだんだん空室が減っていく様子を見ることが
できます。ところが、このようなシステムでは、たと
えば実際は100しかない部屋数に対して宿泊プラン
が提示する部屋数の合計が200であるように提示さ
れていることがほとんどです。そこで同じ部屋に対
してどのプランが重なっているかを見破って、実際
の稼働率を推定していきます。次にプランに付いて
いるさまざまなオプションをカテゴリー分けし、「ロジ
ットモデル」を使ってどの要因が販売スピードに
影響を与えているのかを調べます。「ロジットモデル」
は、金融工学では会社の倒産リスク、医学の世界で
は薬効の判別などによく用いられるモデルで、生起
確率を導いてくれるため、事象を把握しやすいのが
大きな特徴です。たとえばツインの部屋では「ドリ
ンク付き」が効いていたり、売れたプランの中でも
特に「女性向けお土産」に人気があったりします。こ
のモデルを用いて人気の理由を個々の要因まで掘
り下げて分析できます。

公開されたIR情報を使って 客室稼働率から売上高推定へ

またデータベースは3年分蓄積されているため、稼

働率の季節変動や、曜日効果などのデイリーな動き
も把握できます。また客室規模や駅からの距離な
ど、いろいろ切り口で稼働率の平均値を推定するこ
ともできます。さらにこのデータを、京都市が発表
する主なホテル稼働率の月別集計と照合し、「空室
あり」の場合の実際の空き部屋数の推定などにも役
立てています。さてこのようにして精度アップした
稼働率に、素泊まり／朝食付きプランの価格をベー
スに算出した基準価格を掛けると、売上高が推定で
きますね？——すると、これは実際に銀行がホテル
経営のリスク管理に、あるいは投資会社がホテルの
収益還元価値評価に使えるようなデータになって
きます。そこで今度は、ホテルだけに投資している
一部上場リート（会社型投資信託会社）が投資家向
けに公開しているIR情報と、組み合わせてみましょう。
リートが公開しているホテルの稼働率・売上高・収益
率などの実績値とわれわれの推定とを比較して、そ
の誤差を小さくすれば、実際の決算に先んじて決算
情報を知ることができるわけです。これは二重の意
味で画期的なことであり、まずホテルやリートが発
表するより前に、決算状況を知ることができるとい
うこと。もう一つは、以前なら機関投資家だけが知
り得るような高品質な情報でありながら、使ってい





るのは誰にでもアクセスできるオープンデータです。

新たに出現する巨大データを いかに金融の分野に活かすか

ホテルの稼働率の推定にウェブデータを利用したのと同様に、ウェブ上の不動産情報は、商業用不動産の収益還元価値評価モデルや、今や銀行の主力商品の1つであるアパートローンの信用リスクの評価などにも応用可能です。実際に現在、沖縄の那覇を中心に地域の賃貸マンション、アパートのような商業用不動産全体の家賃収入等の推計を試験的に進めているところです。また京都だけでなく日本全国のホテルについて同様に稼働率や売上高の推定を行えば、日本の景気動向の指標の1つとして、より大きな経済動向を知ることができるでしょう。ところで、人工衛星が日本のほぼ上空を通る準天頂衛星システム「みちびき」の4機体制の運用が2018年に始まり、cm級測位の精度の測定が可能となる見込みです。このような高精細で新しいデータが利用可能な将来には、高精度の土地の隆起・沈下の度合いの把握が可能になるなど地震や地滑りなどの自然災害の事前の把握と、また、被害の経済損失のシミュレーションシステムの実現にも大きな可能性を感じ

ています。そして、トレーディングの主役がプログラムにとって代わったように、今は企業訪問による調査など人海戦術で行っている企業の収益予想などの分析業務も、データ収集から推定まで一貫して自動化できる可能性があります。結局のところ、それがいろんなサービス向上につながっていくはずだ、と考えています。



所属する同志社大学がちょうど京都にあることも、プロジェクト参加の理由のひとつだったという。2014年8月に立ち上がった京都市の「未来交通イノベーション研究機構」でもデータ解析を担当する予定。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/b03/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/human/>



Research View 014

バイオハッカソン2014、 データ統合への道のり。

DBCLS・特任助教

片山 俊明

DBCLS・特任准教授

山口 敦子

Database

ライフサイエンスデータ (DBCLS)

INTRODUCTION



バイオハッカソン だからできた7年間

そもそもは海外で開催されたあるハッカソンに招待されて、すごく生産性が高い方法であることを知ったのがきっかけなんです。研究は人間の営みですから、人が会って一緒に作業するのが一番効率がいいわけですね。今や、研究発表や講演のようなものならネットでも視聴できる。学会でも、むしろ質疑応答やコーヒープレーク時のディスカッションのような部分が有意義なんですね。ならば、その良いところだけ集中してやろうよというのが、ハッカソン。世界中の主要なデータベースを作っている人たちに集まってもらって、まずは問題を洗い出すところから始めます。それらの課題に対して、僕らがちょっとプログラムを書くことで、面倒な作業が1クリックで済むようになったり、検索やデータ整備が100倍早く進むようになったりする。そのよ

“ハッカソン”とは、一言でいえば気鋭のプログラマーたちが集まって、短期間ながら凝集性の高い交流を通じて開発に取り組み、課題解決を目指すコーディング・イベント。バイオハッカソンは2008年にこの形式を採り入れ、以来ライフサイエンス分野のデータベースを扱う世界トップレベルの研究者や技術者を集めて、毎年開催しています(NBDC、DBCLS主催)。近年は、Web 3.0時代のセマンティック・ウェブにおいて標準とされるデータ形式「RDF(Resource Description Framework)」化を推進しており、世界的にも、データの統合化と標準化を担う重要な機会として定着してきました。第7回目の2014年は11月9日～14日、東北大学 東北メディカル・メガバンク機構との共催により仙台・松島で開催され、10ヶ国から78名が参加。今回はこのイベントについて、オーガナイザーの片山俊明特任助教(DBCLS)、山口敦子特任准教授(DBCLS)をご紹介します。

うな解決の1つ1つによって、バイオインフォマティクスを非常に効率よく進展させることができるんですね。さらにこれを毎年継続していく中から、海外との共同研究もたくさん生まれています。(片山)

日本人のゲノム情報を医療に 役立てるために

今回は、地域の人々の世代間にわたるゲノム情報を集めている東北メディカル・メガバンク機構と連携して、これまであまり扱ってこなかった個人ゲノム情報や医療データにもフォーカスしています。というのも、個人ゲノムのようなデータはプライバシーの問題等もあって、セマンティック・ウェブが担うオープンデータの思想になかなか馴染みません。そこでこれらのデータをライフサイエンスデータに統合していくための仕様策定や、医療系アプリケーション等の登場に備えたデータの整備といった技

術開発を、今回の課題のひとつに挙げました。ちなみにデータの「標準化」はバイオハッカソンの大事なミッションの1つであり、過去にもすでに複数の国際標準を提案しています。遺伝子の特徴と日本人の病気傾向をつなぐデータが公共的に蓄積され、これを医療の現場で一人一人のゲノム情報と照合して、診断や治療に役立てる——そんな将来的なビジョンを共有しながら、いま出来る開発を進めています。(片山)

RDFを実用的に 使いこなすための「縁の下」

このような標準化・統合化を進めるためには、データベースの基盤的な技術開発も不可欠です。そのひとつが、分散して置かれたオープンなデータベースと、手元にあるデータをシームレスにつなぎ、欲しい情報が取得できるようにするための技術開発です。これにはデータがRDF化されていることが前提となります。RDFは、データの持つ意味を機械が取得

することを可能にし、多くのデータを意味的に連関を持たせてつなぐことができる形式として、現在最も代表的なものです。ところが分散化されたさまざまなRDFデータを対象に処理を行おうとすると、現状では現実的な時間内で処理できるほどには技術が成熟していません。そこでどこにどのようなデータがあるかという目次を標準化したり、メタデータを自動生成したり、個々のサーバにあまり負担をかけることなくデータを自動収集したり……といった基礎技術を、今回もさらにブラッシュアップしています。(山口)

データベースは統合された 知識を編み出す仕事

ところで、生物学のラボでは多くの人材が実験を担っていますが、かつて行っていたような個別研究から、次世代シーケンサーの登場によって一気に、大量にデータが生み出せるようになりました。実験手法においても、細胞内の量的変動を計測する等のさまざま

Research View003 p.17

Research View





まな技術革新があり、このこともデータの増加に寄与しています。このような大量データを分類して再利用できるようにして、さらにその中からエッセンスをまとめることができれば、生物学として全体を成すような知識を編み出すことができるはず。データベースというのは、本質的に図書館にとても近い事業なんですね。ライフサイエンスのすべての分野を俯瞰できて、しかも詳細が分かるようなデータベースを作れば、それは教科書のようなものになるはずだというのがデータ統合のひとつの考え方です。しかもそのようなモデルをデータから「発見」するのは、人間ではなくコンピュータかもしれない——データ統合の道りは、きっとそんな未来にもつながっています。(片山)



開催初日のシンポジウムの風景から。「地球には何百万通りもの多様な環境があり、そのなだらかなランドスケープのそれぞれに適応した生物がいる。そして地球上のすべての生命が、他の生物と共通の遺伝子を持って生きている」と片山特任助教。「データベースから得た知識と理論で、工学的に生物を合成し、そこに何かの刺激を与えたら、それが生命としてポツと生き始める可能性もあるわけです。またこのことは、生と死のスイッチを理解することにもつながりますね」

▶ データベース概要

<http://rc.rois.ac.jp/database/d02/>

▶ DBCLS ウェブサイト

<http://rc.rois.ac.jp/database/d02/>

▶ BioHackathon ウェブサイト

<http://www.biohackathon.org/>



Research View

Research View 015

南極上空の風を アーカイブする。

国立極地研究所副所長・教授

部門責任者 中村 卓司

Database

地球環境データ

INTRODUCTION



レーダーで大気の三次元的な構造を捉える

身近なレーダーといえば、スピード違反の取締に使用されるスピードガンかもしれません。電波を発信すると、車に当たって跳ね返ってきます。動いている物体に当たった電波は周波数が少し変化しているため、ドップラー効果により車の速度が検出できるという仕組みです。PANSYも原理的には同じで、電波を発信し、大気に当たって返ってきた電波を受信します。自分から出した電波ですから、近くほど早く、遠くほど遅く返ってくる。さらに電波を出す向きを少し振ることで、異なる角度から測定して、大気の三次元的な構造を知ることできます。電波望遠鏡や光の望遠鏡と違って、奥行き方向に高精細に観測できる点が、レーダーによる計測の大きな特徴です。PANSYの場合は47メガヘルツというVHF(超短波)の電波を使って、南極の上空を最大約75～150メー

たくさんのアンテナを並べて一括制御する技術「アクティブ・フェーズドアレイ方式」による「PANSYー南極昭和基地大型大気レーダー(Program of the Antarctic Syowa MST/IS radar)」は、2012年から本格観測を行っています。フル稼働時には1,045本のアンテナを使って、南極上空の真上方向の風の成分を高精度に捉えるという、他にはない大きな特徴を持つPANSY。ところがレーダーから得られるのは一般に電磁波の受信記録であるため、生データそのものから温度や速度のような意味ある物理量を知ることはできません。そこでリサーチコモンズ「地球環境データ」部門では、東京大学、京都大学との連携で、PANSYの観測データを適切に解析処理し、5年、10年という長期にわたってアーカイブする仕事に取り組んでいます。地球温暖化をはじめ、特に地球全体の気候変動予測の研究に欠かせない貴重な観測データとして、いっそうの共同利用の推進が期待されるデータベースです。

トルごとの細密さで見ることができます。高度500キロ(メートル)の超高層大気までをカバーしますが、特に気候変動などと直結している地表から100キロ以下の大気の動きを見るために最適化された大気レーダーと言えるでしょう。

地球の大気は薄っぺらい層状になっている

地球の直径は約1万2,700キロ、大気的主要な部分はせいぜい数百キロの薄い層です。そこではほんの1キロ、2キロ上へ行くだけで、刻々と温度が変化し、環境が変わっていきます。つまり、地球の上空にある大気はすごく薄い層の重なりでできているんですね。そこでPANSYも、地上から上空まですっと連続的に詳しく見ていく、言わば観測ロケットを数分ごとに一日何百発も上げるような観測を行っています。太陽のエネルギーが地球に届くと、大気の一部

が温まって、大気の動きが起こりますね？——これが風となります。大気は下の方ほど重く、上の方ほど軽くなっているため、風は基本的には重力に逆らわず、地表に対して水平に移動する場合がほとんどです。ところが積乱雲、雷雲、前線の活動などによって上昇気流が起こったり、ヒマラヤ山脈のような高い山に沿って空気が持ち上げられたりする現象も至るところに発生しています。実はそれらが、気象の変動や大気の循環の解明の鍵を握っています。このような鉛直方向の動きを測る手段は、レーダーを

大気の中の小さな波 「大気波動」を捉える

このような上昇気流の中から出て来るとても小さな波「大気波動」が、地球全体の大気の流れを駆動する働きがあることがおよそ30年前にわかり、注目を集めてきました。たとえば赤道付近では積乱雲を伴うような上昇気流で大気波動が発生して遥か上方に伝わり、上空で潰れて大気の流れを作っていますし、南極でもこの波がたいへん強いことがわかっていま

す。地球全体の大気の流れは、何千、何万キロスケールの構造を持ちますが、大気波動は10キロ程度の大きさであつたり、上下方向の波長が3キロ程度だったりというように、天気図などに比べると比較的小さな現象にあたります。こういったものは、衛星からでは観察できないし、気候予測モデルが扱う空気の塊よりもかなり小さくて、分解能が足りません。PANSYは、そういった大気の細かい構造を、南極上空で捉える役割も担っているんです。

地上から流れ星や宇宙線を キャッチする

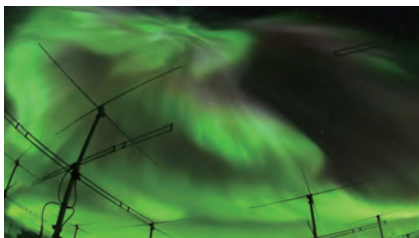
レーダーによる観測では、原理的にはいくらでも詳しく奥行き方向を調べることができますが、一方で、そのデータ量は莫大になってしまいます。つまりレーダーは、常にその時代その時代の記憶容量との戦いなんです。たとえばPANSYが、大気層の幅に合わせておよそ1マイクロ秒ごとに違う値を取得したとすると、1日は8万秒ですから、1日のデータ数は80ギガ点にものぼります。そこでデータ中心科学の時代にふさわしく、このようなビッグデータから、価値あ





る情報を掘り起こす「マイニング」にも挑戦しています。たとえば人工衛星の追跡や、地球へ降ってくる宇宙線の観測など、大気レーダーの建設当初には想定していなかった新しい使い方もいろいろ見つかっています。われわれはこのようなデータ処理や解析手法の開発に加え、衛星回線の限られた容量を駆使して送られてくる南極からの観測データを遅延なく配信するシステムを継続的に改良しています。現在の公開データに加え、対象データの拡大や、時間的によりきめ細かいデータ等の公開へ向けて、地道な準備を進めています。

2014年3月から東京・立川にある国立極地研究所で一般公開されている「リアルタイム観測データ表示システム」のモニター。地球温暖化と関連して、大気の温度の自然による変動と人為的な変動の区別や、オゾン層の変化などを説明する数値モデルに、決め手となるような観測データを提供する。また国立極地研究所では、現在北欧で建設が進められている、より上空のプラズマを観測する新しいレーダー「EISCAT_3D」のプロジェクトにも参加。将来的には、EISCAT_3Dとのデータ連携により、グローバルな気候変動の様子が、いっそう詳しく解明されるはずだ。



▶ データベース概要

<http://rc.rois.ac.jp/database/d01/>

▶ プロジェクトウェブサイト

http://tric.rois.ac.jp/math/geo_pansy/



Research View

Research View 016

統計データを社会に生かす 未来像。

統計数理研究所 副所長・教授

樺 広計

Database

人間・社会データ

INTRODUCTION



1947(昭和22)年に公布された統計法が、2009(平成21)年4月、およそ60年ぶりに大改正されました。「統計」を意味する「Statistics」は、まさに国「State」の状態、すなわち国勢を知ることによって国を動かす指針にする「国勢学」がその語源です。リサーチコモンズの研究データベース「人間・社会データ」は、その応用研究としての「社会コミュニケーション」と一体化しながら、研究コミュニティに役立つデータ基盤を整備・提供し、さらに社会科学等における統計手法の共有・発展を支えていくためのプロジェクト。サブプロジェクトのリーダーを務める椿広計教授(統計数理研究所)をご紹介します。

科学による政策と、 公的統計データの「2次利用」

統計学は16～17世紀、「歴史は動く統計である」とも言われ、その重要性が認識されてドイツを中心に広まりました。データを集めて科学として利用する、今日のようなエビデンスに基づく政策決定が始まるのは、19世紀末頃からです。日本では1881(明治14)年、大隈重信によって統計の整備が進められ、大正時代に国勢調査が始まりました。このように国が収集するデータは、これまででもっぱら政策という「1次利用」に限られてきましたが、データは国民全体の知的財産であるという考え方から「2次利用」を推進して、もっと研究に活用したり、その研究成果を国民にフィードバックしたりしようというのが、統計法改正のポイントです。そこでわれわれは、研究利用を含む2次利用を統括する独立行政法人統計センターと連携協定を結び、一橋大学など他の大学とともに、実際の研究利用をお手伝いしています。

研究利用の基盤をつくり、 モデリングを支援する

なかでも情報・システム研究機構は、データ提供だけでなく、社会科学等の研究者に解析・モデリングを支援する役目があると考えています。たとえば社会科学的分析に用いられるモデリング手法がほんとうに妥当なのか、われわれのところにログなどのかたちで記録が残るようにすれば、社会科学で今こんな手法が使われているという生きた情報をアーカイブしていくことができます。そしてその中から、どんなものが問題解決や社会への提言につながるのか、よいもの・ダメなものを選ぶというプロセスも重要です。さらに、このように草の根的に集めた「モデリングの知」からある程度汎用的なモデルを生成したり、研究者が共有・再利用したりできるようにすることを目指しています。

公的統計がサイバー空間の 情報と結びつく強み

私の専門は応用統計なので、統計の新しい応用分野を拓くのが役目であり、それによってもう一方の柱

である数理統計も伸ばしていければと考えています。「社会コミュニケーション」プロジェクトで取り組んでいる自殺対策、食品安全というテーマは、まさに融合研究から生まれたものであり、今や主力分野に成長したと言ってよいでしょう。また私が専門とする統計データはどちらかというと“クラシックな”統計であり、データの中から対象が持つ特徴を探り、モデル化する作業が中心になります。一方、世の中には収集するだけで価値を生み出せるデータも多く、プロジェクトディレクターの曽根原登教授・一藤裕研究員(国立情報学研究所)が取り組むサイバースペース・データはその代表例です。2つのデータはまさに一長一短で、たとえば伝統的な人口統計である国勢調査は5年に1回ですが、今や携帯電話を使えばリアルタイムに人の移動を把握できます。人口調査と移動端末数とはまったく別物ですが、もしある

Research View005 p.25

時点における国勢調査と携帯電話の情報をうまく連携させてモデリングすることができれば、国勢調査が行われている間の人口移動を推定できます。このように一方の統計が持つ偏りを他方が補正するしくみを工夫していくと、すごい力になると思いますね。

ビッグデータからの発見が 問題解決を加速する

統計の考え方の基本に「Plan・Do・Check」というサイクルがあります。これに「Action」が加わったものを「PDCA」と言い、1960～1990年代前半の日本の自動車工場がこれを品質管理手法として採用して、世界をリードしたことで知られています。サイクルのうち、最近の統計が活躍するのはチェックのと



公的統計データをセキュアな環境で利用するため、監視カメラを備えるなど、高い審査基準をクリアした統計数理研究所の「オンサイト室」。2010年から匿名化された公的データの提供を開始し、2次利用が進められている。同研究所の宿泊施設「赤池ゲストハウス」を活用して、遠方からの研究者を受け入れる準備もある。一方、個人情報保護に対応した匿名データは情報量が少なく、社会科学にとっては大規模なモデリングが行えない。そこで総務

省2次利用研究会で検討を重ね、応用統計学会を通じて学術会議マスタープランに提言したところ、2013年に採択された。全国の大学などが公的統計データを活用できるこの有効な方法は、内閣府統計委員会の基本計画に盛り込まれ、2014年3月閣議決定された。また統計センターには、将来的には総務省だけでなく全省庁のデータが集まる方針も固まり、今後いっそうの大学共同利用が期待される。



016

ころ——実際のデータ(実測)とモデルによって期待されるデータ(予測)を見て、どのずれが解消されなければならないかという問題を発見するのが最近の統計の役割です。たとえば現在、日本全体としては自殺数が減ってきていますが、年齢や地域によっては依然として増えているところがある。このような地域によってばらつきが存在することが問題だと考えるわけです。そして原因と思われるデータと結果のデータを集め、モデリングを行います。これを使って効いている原因をコントロールしたり、環境因子がどう影響を与えているかを評価したりして、さらに分析を進めていきます。この辺は検証的で古典的な統計モデリングが依然として重要です。ただ、このような問題解決プロセスの中で近年大きく進歩したのは、ビッグデータから要因を発見するような統計だと申し上げたい。機械学習、データマイニング、人工知能など、問題のありかや原因候補を探索するさまざまな方法が登場し、統計のサイクルを活性化させていると言えるでしょう。



「原因と結果の相関グラフを描くだけでも重要なトレーニング。統計のサイクルをきっちり回せれば、小学生だっていろんなことができる」という。データサイエンティストの育成にも関連して、小・中学生を対象とした“統計による問題解決”の教育・普及啓蒙にも奔走する。

▶ データベース概要

<http://rc.rois.ac.jp/database/d03/>

▶ オンサイト利用

<http://www.rois.ac.jp/tric/micro/>



Research View 017

多様な野外生物たちの 戦略と進化。

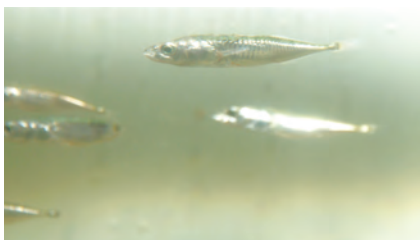
国立遺伝学研究所・特任准教授

北野 潤

Project

システムズ・レジリエンス

INTRODUCTION



2011(平成23)年の東日本大震災では、巨大な津波が東北地方を襲いました。以前から北海道から東北にかけての冷たい水域に住む魚「イトヨ(写真左)」を研究してきた北野潤特任准教授は、この時、自分の研究する魚の生息地が津波で襲われたことで、魚にどのような変化があったのか、実際のフィールドへの強い関心を持って「システムズ・レジリエンス」プロジェクトに参加したといいます。野外生物たちの持つ形態や行動の多様さに魅せられ、その進化の過程を次世代シーケンサーを用いたゲノム解析などによって明らかにしようという研究の現場から、その最新の成果をご紹介します。

ここにヘンテコリンな生き物がいる！

世の中には、ちょっと信じられないような派手な形をした生き物や、ものすごく精巧にできた生き物などがいて、その形態や行動は実に多様です。例えばアズマヤドリという鳥は、木の枝などを集めてきて大きなオブジェを作るのですが、巣として使うのではなく、単にメスを呼び込むためだけのものなんです。ところが立派に作ると本当にもてる(笑)。なんでこんなにヘンテコリンな生き物がいるのか、というのが僕の最初の疑問で、それがどのように進化してきたのかを知りたいと考えています。そこでトゲウオ科に属するイトヨという魚を対象に選び、行動や形態の違いを生み出しているゲノム構造を調べることにしました。イトヨ属は、ゲノム塩基数がヒトの約7分の1程度とかなり少ないことも、イトヨを選んだ理由のひとつです。

津波という環境変化とイトヨの進化

釜石の北、海岸沿いに大槌という、井上ひさしの長編小説『吉里吉里人』と関連して知られる地域があり、大槌川と小槌川というふたつの川が流れています。この大槌川を遡ったところにイトヨが生息していたのですが、津波でものすごい瓦礫に覆われた。絶滅したのではないかと思われたんですが、結局、生きていたんですね。ボランティアの人たちや自衛隊によって比較的早く瓦礫が取り除かれたことに加えて、湧き水の出る地域であるため、湧水口に入り込んで急場をしのいだのではないかと考えられます。また、湧き水によって新しい池ができたところがあって、そこにはイトヨが侵入してきていて、これの遺伝子を調べると、上流にいる回遊しないイトヨと海にいる回遊するイトヨの雑種が形成されていることもわかりました。生物が実際に移動して起こる進化の過程を、リアル

タイムで見ることができたわけです。

「種」の違いは連続的である

ところで、極めて多様な生物がいるなかで、そのどこからどこまでがひとつの種なのかというと、その境は連続的であろうと僕は考えています。一般的には、ランダムに交配する集団を一つの種といい、ランダム交配する似た集団とは生殖隔離によって隔てられているというのが種の定義です。ところが生殖隔離は必ずしも100%ではなく、完全に隔離されているものもあれば、ごく少数交配するものもあって、明確な線を引くことができません。日本海のイトヨと太平洋のイトヨは、一見しただけではほとんど違いがわからないのですが、DNAがまったく違います。日本海のメスと太平洋のオスを掛け合わせると、オスが精子を作れなくなる生殖隔離が起こりますが、その他の組み合わせでは問題が起こらない。つまり、別種として分岐するかしらないかの、ちょうど中間に

いる魚たちなんです。したがって、種が分岐する過程を研究するために格好のモデル系であると考えています。

インドネシアのメダカに見る性染色体

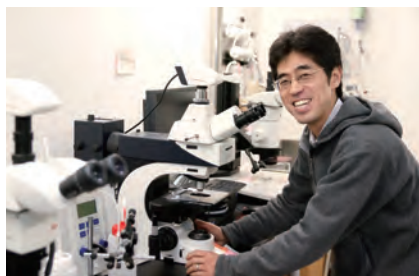
イトヨのゲノムを解析してわかったことは、日本海のイトヨだけに、オスのY染色体の一部と9番の常染色体がくっついて性染色体の一部になった「ネオY染色体」が見られることです。さらに常染色体9番の残りの部分も、性に関わる「ネオX染色体」として進化していました。ちなみに性染色体はメスがXX、オスがXYですから、Y染色体にはオスに有利な遺伝子が溜まっていくという理論があるんです。これらを考え合わせると、性染色体の進化が種分化に関わっているのではないかという仮説が生まれます。そこで今度はこの仮説を検証するために、メダカの研究を始めました(写真右)。哺乳類では性染色体に





017

のっている遺伝子は、ヒトでもチンパンジーでも基本的にはほぼ同じですが、メダカやイトヨでは、オスという性を決めるY染色体の組成が、種によってまったく違います。インドネシアのメダカの場合、性染色体はほぼすべてXX、XYという構造ですが、そこにのっている遺伝子がまったく違うのです。性染色体にのっている、特にオスの形態の派手さのような性差の進化が、種分化を生み出しているのではないかと——今、この疑問を解き明かそうと、フィールドと実験室内での実験の双方を駆使して検証しているところです。



「地球上で、生命はいかに生まれたか？——それは神ではないと言ったのは進化論ですが、真実はわからないという状況は今も変わっていません」。しかしながら「答えはあるはず」だという。「サイエンスとは、ある疑問に対する答えの与え方の一文化。ある仮説の立て方、証明のしかたをずっと積み上げて、一つの知識体系として綿々とながっている」。一方、科学が弱い点もある。「科学としては1回きりのイベントを証明するのはかなり難しいんです。たとえば震災や原発等の環境変動によって突然変異率が高まったという仮説があっても、偶然的誤差よりも確実に高いということを実証するのはかなりハードルが高い。レジリエンス解明の難しさのひとつは、そこにあります」。

▶ プロジェクト概要

<http://rc.rois.ac.jp/research/b04/>

▶ プロジェクトウェブサイト

<http://tric.rois.ac.jp/resilience/>

プロジェクト紹介

 イメージデータ解析プロジェクトディレクター
松井知子〔統計数理研究所〕

生物では、近年の細胞・分子イメージング技術の発展を受け、バイオインフォマティクスの新しい分野として、バイオイメージ・インフォマティクスの研究が急速に活発化しています。——医療では、医用画像処理はたいへんニーズの高い技術でありながら、統計数理の手法が十分に浸透していない現状があります。——人間・社会に目を向ければ、近年、社会の個人化・孤立化が進み、経済不況が続くなか、一人暮らしの老人やストレスを抱える人々など、

ケアを必要とする人が急増しています。また、温暖化などに伴う環境変動が深刻な問題となっています。本プロジェクトでは、医療・生物や人間・社会に関する4つの分野で実際に生み出されたイメージデータを取り扱い、これらを高度解析する技術を研究開発します。さらにその成果を、さまざまな問題に適用できる汎用的なイメージデータ解析ツールとして公開し、各研究分野の学術的発展にも貢献していきます。

1. バイオイメージング・インフォマティクス(医療・生物関連)
2. オプティカルイメージングデータを用いた脳の動作原理の解明(医療・生物関連)
3. 医用画像処理における統計的手法に関する研究(医療・生物関連)
4. 人間・社会の視覚情報データ解析(人間・社会関連)

 メタ知識構造解析プロジェクトディレクター
宮尾祐介〔国立情報学研究所〕

現在、社会活動の様々な成果がテキストデータという形で蓄積されています。例えば、学術分野では、多様な研究の成果が学術論文や特許として、日々大量に公開されています。しかしながら、これらの成果や波及効果を客観的かつ定量的に把握・活用する方法は、今のところ存在しません。そこで、テキストの中で明示的・非明示的に表現された因果関係、理由、目的といったメタ知識構造を自動認識し、それに基づき構造化された知識を自然言語テキ

ストから自動抽出する手法について研究を行います。当面は、メタ知識構造が明確な学術論文や特許文書などのテキストを対象とし、将来的にはより一般的なテキストデータを対象とすることを検討しています。

- ・ メタ知識構造の定式化
- ・ 学習・評価データとしてアノテーションコーパスの構築
- ・ 言語的手がかりに基づく自動抽出手法の開発
- ・ 統計的特徴に基づく自動抽出手法の開発
- ・ 言語的・統計的手法を統合した自動抽出手法の開発

データ同化・シミュレーション支援技術

プロジェクトディレクター
中野純司〔統計数理研究所〕

「ビッグデータ」時代においては、データ量の単なる増加のみならず、まったく異なるタイプのデータを扱わなくてはならない点も特徴的です。本プロジェクトでは、シミュレーションとデータ解析の作業を一体化する手法の高度化と一般化を目標に、地球科学と生命科学における具体的な問題解決に向けた共同研究を通して、データ同化手法のさまざまな分野への適用拡大を図ります。地球科学においては、地球の磁気圏と呼ばれる領域において、シ

ミュレーションモデルと観測データの同化実験を行い、オーロラの再現に挑戦します。南極大型大気レーダー(略称:PANSY)プロジェクトのデータ処理を題材として、観測現場でリアルタイムに得られる高速、高分解能、多次元データの問題について研究します。生命科学においては、データ同化手法を発生細胞生物学分野の問題(ほ乳類の体節形成、精子形成および動物細胞の分裂、細胞質流動)に適用し、同分野における有力な手法として確立する

ことを目指します。本プロジェクトは以下の3つのチーム編成により、相互に連携し合いながら、研究を推進します。

- ・ 数理・計算チーム
- ・ モデリングチーム
- ・ データデザインチーム

e-サイエンス基盤技術

プロジェクトディレクター
新井紀子〔国立情報学研究所〕

爆発的な勢いで増加しつつある研究資源および研究情報を研究者が理解し、分野固有の文化や空間的制約に妨げられることなく分野横断的な研究が促進されるよう機械が支援する「サイエンス3.0基盤」構築を目指します。より具体的には研究資源・研究情報を機械可読化するために、大規模データマイニング技術・検索技術、既存学術データベースと連携するためのセマンティックウェブ技術、異種データベースをつなぐ情報リン

ケージ技術、オントロジー技術等の研究開発を推進します。本研究の実施にあたっては、プロジェクトを以下の3つのサブテーマに分け、これらの要素技術を統合した上で、データ構造の可視化・探索支援技術を盛り込み、研究者にサービスするための基盤の研究開発を行って、実際に大学・国内主要研究グループ・学会等に提供しながら、実証的に研究を推進します。なお本研究の実施に先立って、日本全国の研究者を対象とした研究者向けサイエンス2.0基盤サービス

「researchmap」を公開しており、平成23年度にはReaDとの統合を果たし、24万人の研究者が参加する世界最大級の研究者コミュニティへと飛躍的に成長しました。

サブテーマ紹介

1. 研究資源に関する情報推薦基盤の構築
2. 学術リソースのためのオープン・ソーシャル・セマンティックWeb基盤の構築
3. 融合研究を加速するための情報共有クラウドサービスの確立

地球・環境システム

プロジェクトディレクター
 本山秀明〔国立極地研究所〕

氷床コアからの古環境復元は、地球の過去・現在・未来を知るための鍵であると言われています。氷床コアから復元可能な歴史の長さは約100万年。ちょうど直立猿人から現代までをカバーするスケールに当たることから、氷床コアに含まれる微生物の変遷を読み取ることで、人類の時代における地球生命システムを理解する道を拓くことを目指すプロジェクトです。環境変動史に生物変遷史を加えることで、現在の環境を正確に理解し、さらに

は将来の地球環境がどうなっているのかを予測します。機構4研究所のほかに、北海道大学、筑波大学、千葉大学、東京大学、日本海洋技術研究機構(JAMSTEC)、東京工業大学、玉川大学、京都大学、京都府立大学、広島大学等と連携して研究を行っています。

サブプロジェクト紹介:

1. 氷河、氷床コアに見る地球環境の変遷と生物の変動、人間圏との関わり
2. 極限環境における生物多様性とそのパターン
3. 極限生物の環境適応メカニズムと進化

遺伝機能(生命)システム

プロジェクトディレクター
 倉田のり〔国立遺伝学研究所〕

ヒトゲノム解読計画以来、生命科学の研究スタイルは、大量のゲノム情報を基盤とするデータ駆動型へとパラダイムシフトし、近年の超並列大規模シーケンサの登場は世界的にもその勢いを加速させています。そこで本プロジェクトでは、最新のゲノムテクノロジーを駆使し、系統的にゲノム・遺伝子関連情報の大規模生産を行う一方、多元的な表現型データを系統的に収集し、それぞれの情報を統計学的に解析する手法の開発を目指します。

さらに、これらのデータを統合し、遺伝的相関構造の描出を行うための統計学的手法を開発し、モデル生物の大量のゲノム情報と表現型データに適用してブラッシュアップし、汎用的にゲノム機能と遺伝的ネットワークの抽出を行います。これにより、多数の遺伝因子の高次関連から形成される生物多様性を、システムとして理解することを目指します。3つのサブテーマが相互にその情報を交換しながら生命現象の新たな方法論を見つけ、データ中

心科学ならではの新規な原理と解釈を提案できるよう研究を推進します。

1. 次世代シーケンサによるゲノム関連情報の大規模生産とその情報解析手法の開発
2. 大量ゲノム関連データと多元的な生物表現型多様性データの統合による遺伝的相関構造抽出のための統計手法の開発と最適化
3. 大量で多元的なデータの情報・統計手法を適用したゲノム機能と遺伝的ネットワーク抽出

社会コミュニケーション

プロジェクトディレクター

曽根原登〔国立情報学研究所〕

ビッグデータ時代を迎え、これまでとは桁違いに多様で大規模なデータ、テキスト、音声、画像、映像などのWeb情報を集積することができるようになりました。その一方で今世界は、地球環境の保護(CO2削減)やウイルスへの脅威、医療・健康、持続的発展可能な社会経済、食料・エネルギー、“安全・安心な”社会など、さまざまな問題に直面しています。そこで本プロジェクトは、これらの複雑なシステムの問題を見据え、データを活かす情報システム技術と、人間・社会科学を

融合させた、新しい学問領域の科学的研究手法の創成を目指しています。現実の物理的社会(Real)と爆発的に増加するWeb情報空間(Cyber)が融合した社会において、現実社会のありとあらゆるモノが投影されたWeb空間から、現実社会を制御する「Cyber-Real-Control技術」を研究開発します。さらに、現実社会をWeb空間でシミュレーションするだけでなく、Web空間から人やモノの操作を行い、現実社会にフィードバックすることで、今までにない新たな価値

を生み出だす人間・社会の知の循環基盤を実現します。なお、データベース「人間・社会データ」との有機的連携により推進します。

サブテーマ紹介:

1. Web/SNSデータ駆動の観光・防災政策科学の研究
2. データ駆動型学習支援:個に最適化した学習支援サイバー学習空間の研究
3. 意思決定・コミュニケーションプロセスに関わる情報循環の高度化と標準化
4. 自殺予防対策等に関する独自統計整備と予防政策支援
5. 安全規準など策定のための数理的方法整備とその食品安全分野などへの実装

システムズ・レジリエンス

プロジェクトディレクター

丸山 宏〔統計数理研究所〕

レジリエンスとは、環境の大きな変化に対して、一時的に機能を失ったとしても柔軟に回復できる能力を指す言葉です。私たちの社会が持続可能なものとなるためには、様々な外界の事象に柔軟に対応していかなければなりません。その際、システムが壊れないことに注力するだけでなく、壊れた後いかに素早く回復するかも重要な考えとなってきます。レジリエンスなシステムにはさまざまな種類があります。地球の長い歴史の中で、生物は多く

の事象に対応してきました。また生物に限らず、インターネットなどの人工物、経済や金融などの社会システム、あるいは企業・組織なども、外界の様々な擾乱に柔軟に対応し、その活動を維持しています。これらレジリエントなシステムに共通な性質はあるのでしょうか。そのような性質、あるいは科学的な原理から、レジリエントなシステムを設計し運用するための知識の体系を構築することができるとは。私たちは、多くの

分野にまたがる研究者のネットワークを通して、これらの問いに答えていきます。

サブテーマ紹介:

1. レジリエンスの統合戦略
2. 生物におけるレジリエンス
3. 社会システムにおけるレジリエンス
4. レジリエンスの計算モデル

地球環境データ

データ基盤整備・部門責任者

中村卓司〔国立極地研究所〕

南極昭和基地上空の運動とプラズマパラメータを地上から高度500kmまで高精度で観測する、南極昭和基地大型大気レーダー（PANSY, Program of the Antarctic Syowa MST/IS radar）。ここから送られてくる、極地域ならではの貴重な観測データの解析・配信を担うプロジェクトです。国立極地研究所・東京大学・京都大学はじめ大学や研究機関が共同で推進しています（代表：東京大学・佐藤薫教授）。PANSYレーダーは、衛星などからは得られな

い大気の実際の状態を、高度域を含む広いエリアにわたって観測することができます。気候現象をはじめ全球大気のモデル形成に大きな決め手となることから、そのデータの解析・整備は、内外の学術コミュニティに待望されているデータといえるでしょう。2014年3月から東京・立川にある国立極地研究所で「リアルタイム観測データ表示システム」を一般公開中。現在の公開データに加え、対象データの拡大や、時間的によりきめ細かいデータ等の公

開へ向けて、地道な準備を進めています。

南極昭和基地大型大気レーダー（PANSY）の主な特徴：

- ・ 1,045基のアクティブ・フェーズド・アレイ・アンテナ
- ・ 対流圏・成層圏・中間層の風、電離圏のプラズマ精密観測
- ・ 現象の3次元構造を捕捉可能グローバルなエネルギー収支の定量化や大気結合の解明
- ・ 内外学術コミュニティとの連携による気候変動・温暖化の将来予測モデルの高精度

ライフサイエンス統合データベースセンター（DBCLS）

DBCLS センター長

小原雄治〔DBCLS・国立遺伝学研究所〕

全国の大学、研究機関などに存在するさまざまなデータベースを一元的に利活用可能にする、ライフサイエンスデータベースの統合化の研究開発を推進しています。「統合データベース」に関する将来的なビジョンを国際的な研究コミュニティと共有し、その標準化に深く関わりながら、基盤技術の構築を継続的に行っています。科学技術振興機構バイオサイエンスデータベースセンター（NBDC）、理化学研究所、産業技術総合研究所、京都大学化学研究

所、東京大学、との連携により、統合化に必要な基盤技術を中心とした研究開発を行っています。

ライフサイエンス統合データベースセンター（DBCLS）の主な事業内容：

- ・ データベース統合化のための基盤技術開発（RDF化の推進）
- ・ インターネットを活用した高度な検索を可能にする技術の開発
- ・ ハイスループットシーケンスを中心とする大規模データの利用技術開発
- ・ アノテーション、キュレーションの支援技術開発
- ・ 開発者向けの国際開発者会議「BioHackathon」の開催
- ・ ライフサイエンス分野の研究者を対象にした日本語による情報提供

人間・社会データ

プロジェクトディレクター

曾根原 登〔国立情報学研究所〕

本データベースは、人間・社会における合理的な意思決定や判断をデータに基づいて支援するため、急速に普及するスマートフォンやSNS、多様なセンサから収集される多種多量なビッグデータの収集、保管、共有、分析・合成を可能とする人間・社会データ基盤の構築を目指しています。プロジェクト「社会コミュニケーション」との有機的連携により推進します。

研究開発の概要:

1. 人間・社会データプライバシー保護利活用基盤
2. 人間・社会コミュニケーションデータ収集基盤
3. 人間・社会データ分析サービス合成基盤
4. 政府公的統計のオンサイト拠点形成と全国展開計画の整備、オンサイト拠点データのアジアデータへの拡張
5. 産官も活用可能なデータベース構築とその利用を前提にした共同研究の企画
6. 産業環境情報のアジアでの収集と還元プラットフォームの実効化

データ中心ケミストリ

プロジェクトディレクター

佐藤寛子〔国立情報学研究所〕

現在までに存在が確認されている化学物質は約7000万種類であり、年間数10万～100万種オーダーで増え続けています。しかし理論的には存在し得るものの、いまだ人類が手にしていない化学物質種数は、これをはるかに凌駕することが明らかになりつつあります。そこで本プロジェクトでは、まずこの革新をもたらす「埋蔵分子」を理論的に探索・発掘し、これらを提供する化学反応経路を、エネルギーや電子状態に関する物理化学パラメータとともにデータ

ベース化します。そして研究・教育機関や一般から利用可能なウェブシステムを構築します。

データケミストリの視点から、自己発展型の異分野融合研究資源を創出し、物質科学に関する種々の問題解決のための情報資源を整備したいと考えています。

共同研究者の所属研究機関:

国立情報学研究所、統計数理研究所、東北大学、量子化学探索研究所、京都大学、東京大学、京都産業大学、スイス連邦工科大学

編集後記

「データ中心科学リサーチコモンズ」とは何か？——この大きな問いをシリーズの全体テーマとして、2014年の4月1日～2015年3月16日にかけて「リサーチコモンズ ウェブサイト (<http://rc.rois.ac.jp/>)」に掲載してきた計17本のインタビュー記事を、手軽に読める一冊のブックレットにまとめました。

このウェブサイトは、「データ中心科学リサーチコモンズ」という研究事業の推進を“広報的”に支援する目的で運営しています。しかし、分野を越えた「融合研究」に先駆的に取り組み、「データ中心科学」という今日的な学術課題を担ったプロジェクトの中身をわかりやすく紹介し、その意義を伝えるのは、なかなか容易ではありません。

そこで、文字通りお忙しい研究の合間を縫って、プロジェクトを推進する多くの先生方にご協力いただき、その研究がどのような「融合」であり、どのような「データ」を生み出し、それが大学・研究機関にいかに関与しているのかを、ひとつひとつおうかがいすることにしました。写真を信頼できる水谷充氏にお願いし、記事は、お話を池谷が原稿のかたち書き起こし、これを先生方にチェックいただく共著のスタイルで、まとめていきました。

なかでもウェブサイト公開当初にご登場いただいた統計数理研究所 丸山宏先生(システムズ・レジリエンス)、国立遺伝学研究所 倉田のり先生(遺伝機能

システム)、DBCLS 小原雄治先生(ライフサイエンスデータ)、国立極地研究所 本山秀明先生(地球・環境システム)、国立情報学研究所 曾根原登先生(社会コミュニケーション)には、まだウェブサイトも存在しないなか、この「Research View」という連載についてご理解いただき、多大なるご協力をいただきました。また、北川源四郎機構長、ならびに城石俊彦先生、樋口知之先生、そして各プロジェクトディレクターの先生方には、インタビュー先のご紹介をはじめさまざまな面でご教示いただいたことを、併せて、この場を借りて御礼申し上げます。

取材を通じ、研究そのものへの情熱はもちろんのこと、プロジェクト内の若手研究者の育成に対する先生方の熱意が、とりわけ印象に残っています。読者の方々には、本書を通じていま改めて、さらに一歩先の「融合研究」とは何だろうか、「データ中心科学」はどこへ向かっているのか、と問いかけてくるプロジェクトの現在を、感じとっていただければ幸いです。

2015年3月

大学共同利用機関法人
情報・システム研究機構 URA(広報)
池谷瑠絵

 研究者紹介 2015年3月現在


Research View 001

丸山 宏

統計数理研究所副所長、モデリング研究系／サービス科学研究センター長、教授。1981年東京工業大学理学部情報科学科卒、博士(工学、京都大学)。専門分野は自然言語処理、XML、Webサービス、情報セキュリティ。1983年日本アイ・ビー・エム入社、2009年までの26年間に在籍し、東京基礎研究所長や執行役員を務める。巨大なグローバル企業で、コンピューター言語XMLの研究・開発と標準化など最先端の研究・開発に携わった。1997～2000年、東京工業大学 情報理工学研究科客員教授等を経て、2011年、統計数理研究所教授。



Research View 002

倉田 のり

国立遺伝学研究所系統生物研究センター長、教授。九州大学理学部より農学研究科へ(農学博士)。専門は、植物生殖機構、ゲノム多様性、植物育種遺伝学。三菱生命科学研究所PD、藤田保健衛生大学助手・講師を経て、1991年より農水省生物資源研究所主任研究官、1996年より国立遺伝学研究所助教授、2003年より教授、2009～2013年副所長。イネの染色体研究から始まり、酵母、ショウジョウバエ、ヒト、マウスなど様々な種を対象に分子・細胞生物学研究を展開。1991年からイネゲノム研究に戻り、以後多様性、進化、発生の複合的研究を展開。野生イネ遺伝資源を用いた、Oryza属23種の解明に挑む。



Research View 003

小原 雄治

2004～2012年、国立遺伝学研究所第8代所長、現在DBCLSセンター長、国立遺伝学研究所特任教授。専門はゲノム科学。1951年京都生まれ。1974年京都大学理学部生物物理卒業後、大学院は名古屋大学理学部分子生物・岡崎令治教授の元へ進学した。1980年名古屋大学理学部分子生物助手、1985年頃から大腸菌ゲノムの物理地図作成に取り組む。1988年から2年間に、英国ケンブリッジMRC分子生物学研究所で客員研究員として、モデル生物「線虫(C.elegans)」に親しむ。国立遺伝学研究所遺伝情報研究センター助教授を経て、1996年教授。



Research View 004

本山 秀明

国立極地研究所教授、気水圏研究グループ長、先進プロジェクト研究グループ長。理学博士(北海道大学1988)。第31次日本南極地域観測隊員以来、南極観測隊への参加は11回に及ぶ。1998年同研究所助教授、2006年教授。氷床コアを軸にした地球環境変動研究や雪氷圏での水循環などに独自の研究成果がある。南極大陸ドームふじ基地では降雪や積雪の通年観測を実施したり3035m深の氷床コア掘削に成功した。北極での氷河調査も多い。これらを基に、現在から過去数十万年までの気候・環境変動の復元や、その変動機構についての研究を進めている。



Research View 005

曽根原 登

国立情報学研究所情報社会相関研究系 教授。工学博士(信州大学1994)。専門は情報流通システム工学、デジタルメディア工学。特に近年はサイバー空間とリアル空間が融合する、ビッグデータ時代のライフログに注目した研究開発を展開する。1978年日本電信電話公社入社。国際電気通信基礎技術研究所、ATR視聴覚機構研究所、NTTヒューマンインタフェース研究所、NTTサイバーソリューション研究所を経て、2002～2004年東京工業大学大学院理工学研究科 連携講座教授、2004年より現職。



Research View 006

中野 純司

統計数理研究所モデリング研究系複雑構造モデリンググループ教授、統計科学技術センター センター長。理学博士(東京工業大学1988)。専門は計算機統計学、時系列解析。特にインターネット、並列計算、対話的グラフィックスなどを利用する統計解析システム、大規模データ解析に取り組む。1978年東京大学工学部計数工学科卒、1980年同大学修士課程修了。1980年徳島大学工業短期大学部生産機械工学科助手、1989年埼玉大学大学院政策科学研究科講師、1992年一橋大学経済学部助教授を経て、1998年統計数理研究所教授。2007年より現職。



Research View 007

齋藤 正也

統計数理研究所データ同化センター 特任助教。博士(理学)(総合研究大学院大学2005年3月)。専門は力学系およびデータ同化を用いた感染症数理モデリング。2005年総合研究大学院大学数物科学研究科天文科学専攻博士後期課程修了。同年産業技術総合研究所システム検証研究センターテクニカルスタッフ。2006年同センター特別研究員。2008年科学技術振興機構CREST研究員。2009年統計数理研究所データ同化センター特任研究員。2012年より現職。



Research View 008

才田 聡子

北九州工業高等専門学校 准教授、理学博士(九州大学2005)。専門は超高層物理学。特に近年は「データ同化・シミュレーション支援技術」プロジェクトのもとで地球磁気圏と呼ばれる地球のまわりの宇宙の電磁場を数値シミュレーションで再現し、観測データと比較することで、現在のシミュレーションに使われている地球磁気圏のモデルの検証と改良に取り組んでいる。2005年国立極地研究所プロジェクト研究員、2008年新領域融合研究センター融合プロジェクト特任研究員を経て、2013年情報システム研究機構特任研究員(掲載時)、2015年より現職。



Research View 009

岡 彩子

新領域融合研究センター 融合プロジェクト特任研究員。博士(理学)。専門はマウス遺伝学、生殖隔離、ゲノム多様性・進化。国立遺伝学研究所系統生物研究センターで維持されている様々な野生マウス系統のゲノム情報などを利用しながら、ゲノム多型によって遺伝子発現制御の違いが生み出され最終的に表現型の違いが形成されるまでの遺伝的メカニズムの解明に取り組んでいる。



Research View 010

木村 暁

国立遺伝学研究所細胞建築研究室 准教授。生命の最小単位である「細胞」が構築されているしくみの理解を目指して、細胞の顕微鏡観察と数理モデル化を組み合わせた研究を展開している。1997年東京大学理学部卒、2002年同大学院博士課程修了。博士(理学)。2002年ERATO堀越プロジェクト研究員、同年慶應義塾大学大学院特別研究助手、2006年日本学術振興会特別研究員を経て、2006年より現職。



Research View 011

佐藤 寛子

国立情報学研究所情報学プリンシプル研究系 准教授。総合研究大学院大学複合科学研究科 准教授。博士(理学, お茶の水女子大学)。基礎学問としての化学へ情報学的にアプローチするケモインフォマティクス、計算化学、データケミストリが専門。1996～1998年理化学研究所基礎科学特別研究員、1998～2001年JSTさきがけ研究21「情報と知」、2000年国立情報学研究所助手、2002年より現職。2007～2008年スイス連邦工科大学(ETH)客員教授、2010年よりチューリッヒ大学 客員教授、1998年より理化学研究所 客員研究員、2013年より量子化学探索研究所理事を兼務。



Research View 012

伊村 智

国立極地研究所生物圏研究グループ教授。広島大学卒、博士(理学、広島大学)。専門は南極陸上生態学、南極湖沼生態学。南極湖沼中の大規模なコケ群落である「コケ坊主」をはじめ、蘚苔類(コケ)を扱う。1994年同研究所助手、2002年同研究所助教授、2011年同研究所教授。第36次越冬隊、42次夏隊、45次越冬隊、49次夏隊(総隊長)、イタリア隊、アメリカ隊、ベルギー隊に参加。豊富な越冬経験を活かした南極・北極観測の普及啓蒙活動にも務める。



Research View 013

津田 博史

同志社大学理工学部数理システム学科教授。1983年京都大学工学部卒業、1985年東京大学大学院工学研究科修士課程修了、1999年総合研究大学院大学数物科学研究科統計科学専攻博士課程修了。博士(学術)。専門は、金融工学、コーポレート・ファイナンス、統計科学、時系列解析。1980年代に米国 Wall Street で金融工学の研究に着手し、以後約30年にわたり金融商品・企業のリスク評価、価値評価、資産価格変動の研究に従事。最近は大規模データを利用した観光科学の研究に取り組む。情報・システム研究機構統計数理研究所客員教授を兼任。



Research View 014

片山 俊明

ライフサイエンス統合データベースセンター(DBCLS)特任助教。京都大学大学院理学研究科卒業。専門分野はバイオインフォマティクス。生命科学データベースの相互運用性を向上させるため、情報処理推進機構のIPA未踏プロジェクトにも採択されたBioRubyなどオープンソース・ソフトウェアの開発を進めるとともに、2008年より国際開発者会議BioHackathonを主催し、セマンティック・ウェブによる次世代のデータ統合とアプリケーション開発を国際協力の元に推進している。



Research View 014

山口 敦子

ライフサイエンス統合データベースセンター(DBCLS)特任准教授。博士(情報学、京都大学)。ライフサイエンスのデータにグラフ理論やアルゴリズムを適用することを専門とし、近年はセマンティックウェブ技術を用いたデータベース統合のための技術開発を進めている。1993年(株)日立製作所基礎研究所入社。(株)生物分子工学研究所、京都大学化学研究所等を経て、2007年より現職。

Research View 015

中村 卓司

国立極地研究所副所長、宙空圏研究グループ教授。1984年京都大学工学部卒、博士(工学、京都大学1992)。南極昭和基地大型大気レーダー(PANSY)をはじめとするレーダーを用いた観測を専門とし、極域中層・超高層大気の構造・変動および諸現象の解明を推進する一方、レーダーから得られた貴重な極域大気データの解析・共同利用を推進する。1989～1998年京都大学超高層電波研究センター助手、2003年米国コロラド大学航空宇宙工学科客員教授等を経て、2009年より同研究所教授、2012年より現職。



Research View 016

計 画 策 略

統計数理研究所副所長、データ科学研究系教授。工学博士(東京大学)。科学の文法としての統計的方法を、品質管理、医薬品許可、環境計測、ビジネス科学、公的統計などの諸領域に適用する応用統計を専門とし、その普及啓発にも努める。1982年東京大学工学部計数工学科助手、1987年慶應義塾大学理工学部数理科学科専任講師、1997年筑波大学社会工学系助教授、2000年同大学教授等を経て、2005年統計数理研究所リスク解析戦略研究センター長、2007年より統計数理研究所教授。名古屋大学客員教授、筑波大学客員教授(名誉教授)を兼任。



Research View 017

北野 潤

国立遺伝学研究所特任准教授。
1996年京都大学医学部卒、医学博士(京都大学大学院医学研究科2002)。特に野外生物における表現型多様化の分子機構に注目した生物多様性・分類を専門とする。
2000～2003年京都大学大学院生命科学研究所 研修員・助手、2003～2009年フレッドハッチンソン癌研究所(ケイティー・パイクル研究室)研究員、2009～2011年東北大学大学院生命科学研究所 助教を経て、2011年より現職。目下、トゲウオ科やメダカ科の魚類を用いて、種分化の遺伝機構解明に取り組む。