

No. 46 December 2009

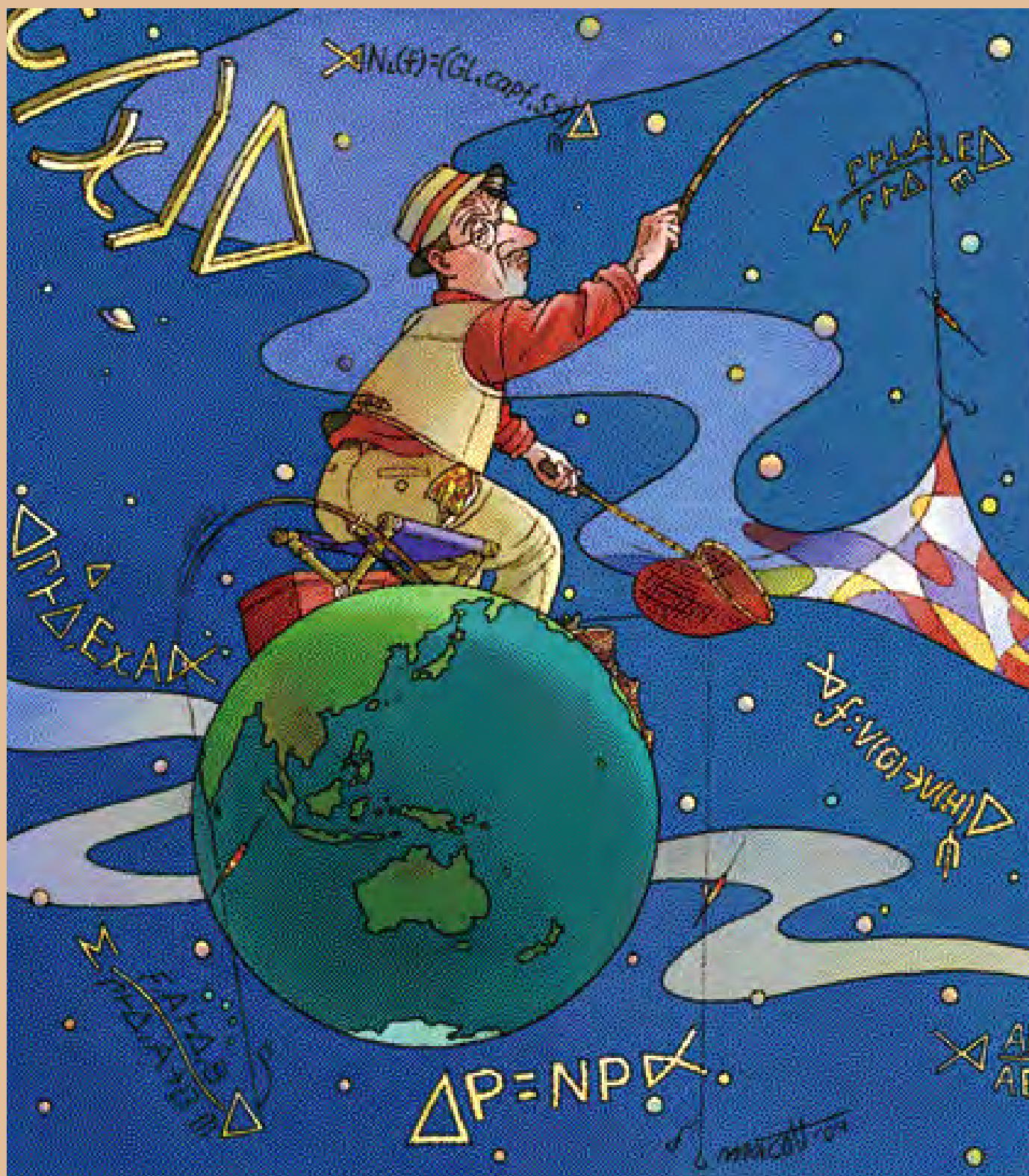
実世界の問題を数学で解く

— 数理・論理とその応用

数理・論理と情報学

アルゴリズムがゲノムの空白を埋める!

NIIでなぜ数学を志したか



Mathematics and Logic as the Cornerstone

河原林健一



かわらばやし・けんいち
国立情報学研究所
情報学プリンシプル研究系教授

数理・論理とその応用 —— 実世界の問題を数学で解く

高橋 ご専門はグラフ理論ということですが、私がこの言葉を最初に聞いたのは1990年に日本の高校生が初めて数学オリンピックに出た時です。よく出る問題のジャンルとして認識しました。

河原林 そう、グラフ理論(*1)とか離散数学(*2)とか組合せ論(*3)はよく出ます。

高橋 でも、グラフ理論とは何ですかと聞いても、先生方も一言では説明してくださらない。「点と辺と線の幾何学だ」とか「一筆書きの算数だよ」とかいいう感じで、なかなかピンと来ません。

河原林 そうでしょうね。ただ、世の中に起こっているいろいろな現象は、例えば道路網などグラフで表わされるものばかりです。僕は学生の頃までは数学的な関心しかありませんでしたが、ドクターをとってからプリンストン大学で1年ほど過ごした時、グラフ理論が自分の知らないところでいろいろ使われているとわかって、計算機の理論の方に入ってきました。実世界で起きている現象というのは、けっこう面白いのです。

高橋 具体的には？

河原林 例えば渋滞メカニズムで、どこかで事故が起こると渋滞は広く薄く大きく波及します。この「広く薄く」をどう解析するかがけっこう難しい問題で、数学の土俵に乗るのです。

高橋 今、渋滞学は流行っていますね。

河原林 渋滞学とはちょっと違う。渋

滞学というのは、渋滞がなぜ発生するかですよね。渋滞が発生した後で、どう影響を与えるかというのは、たぶんまったく違う話だと思います。例えば、カーナビで最短経路を示すとき、事故の発生など時々刻々の状況に応じて、まずまずの精度で速く計算しなければならない。いくら正確でも計算に1時間もかかったら意味がない。どれだけ速くどれだけ正確にやるかというのは、完全に理論の問題です。

高橋 どうやって問題を選ぶのですか？

河原林 僕は海外の理論計算機分野のトップ会議に行って、話を仕入れてきます。20分くらいの講演だと、前半でなぜこういう問題をやるか、実世界ではこうなっている、といった説明をする。向こうの大学院生はそういう説明が特にうまい。理論計算機のトップ会議というのは面白くて、数学のすごい人たちも来るし、実用的なところをやる人もいるし、システムの人も暗号の人たちもいっぱい来るので、非常に刺激になりますね。

高橋 日本ではそういう会議は開かれない。

河原林 そもそも、数学科と計算機学科や情報学科のインタラクションが多くない。それに、現実の問題をモデル化するには実用面をやっている人たちと一緒に考える必要がありますが、そのへんの協力もちょっと欠けている。日本の数学自体の価値観が、純粋数学こそ「由

(*1) グラフ理論とは、数学の一分野で、頂点の集合と辺の集合で構成されるグラフの性質について研究する学問。コンピュータのデータ構造、アルゴリズムなどに広く応用されている。

(*2) 離散数学とは、離散的な(言い換えると連続でない)対象をあつかう数学のことで、組合せ論、グラフ理論が中核をなす分野である。

(*3) 組合せ論とは、特定の条件を満たす(普通は有限の)対象からなる集まりを研究する数学の分野で、整数論、代数学、幾何学、解析などの主要な数学に密接に関連している。

NII Interview Ken'ichi Kawarabayashi+Mariko Takahashi



たかはし・まりこ
朝日新聞東京本社編集局 記者

高橋真理子

緒正しい」といって、応用を軽んずる歴史的な経過があったと思うのです。でも、世界の数学界で今トップにいる人は離散数学の人で、いろいろな応用もやっているしコンピュータサイエンスにも顔が利く。時代は非常に変わってきていると思います。

高橋 日本は、世界の流れからちょっとずれている。

河原林 非常にずれていますね。米国のトップ大学では、応用数学の先生のほうが給料が高いということすらよくある話です。

高橋 世界の流れをつくっているのはやはり米国ですか。

河原林 コンピュータサイエンスで言うと、イスラエルがかなり強いんです。あとは、インドの人たちがすごいですね。私がプリンストン大学にいた時は、学科の3分の1がインド人、3分の1が中国人、残りがその他みたいな感じでした。

高橋 そこで日本人は？

河原林 それはもう、全然見えません。まったく聞えていないですね。

高橋 日本のコンピュータは、最近ハードもあやしくなりましたが、昔はハードはよいけれどソフトはだめだと言われてきました。それは情報科学の理論的な部分、基礎的な部分の弱さとリンクしているのですね。

河原林 それはすごくリンクしている。昔、マイクロソフトリサーチに短期間いた

ことがあるのですが、理論というグループにいる人は義務がほとんどないですよ。自分の研究をしていればよいと。ただし、ほかの人がソフト開発していて変な問題が発生したとき、それが理論の問題だったら集まって皆で解決する。この間AT&T(米国電話電信会社)から



来た人も、同じようにしていると言っていました。そうやって理論の優秀な人たちは数学科の教授よりも明らかにたくさんの給料をもらっている。こういう環境を日本でもつくっていかねばと思います。

高橋 鳩山さんが首相になって、少し数理計画に光が当たるのでは…。

河原林 アメリカが90年代に不況になった時は数学に対して投資を増やした。基本的に理論は人件費だけですから、そんなにお金もかからない。鳩山さんの専門だったオペレーションリサーチの話ですが、例えば大リーグのスケジューリングで、うまくやると飛行機代が何十億ドルも節約できるんですよ。ホームとアウェイでゲームは半分ずつとか、ホーム9連戦はだめだとか、そういういろいろ

ろな条件を加えてコンピュータで計算するわけですが、それをハーバード大学やMITの学生とかが考えて、奨学金をもらったりしています。

高橋 浮いた額の何割かをもらえばよい。

河原林 そうなのです。10%でも、ビルが1棟建ってしまうくらいなので。日本のプロ野球やJリーグでも、きっと何億円か節約できると思いますよ。

高橋 そういう例が出てくると、皆が理論の力を見直しますね。

河原林 それには、優秀な若い人たちがこの分野に入ってくれないと。そのためにも、ここのような研究所がうまくやって、それが広く知られることが大事だと思っています。

◆ インタビュアーの一言

慶応大学の博士課程に1年いただけで博士号をとったと聞けば、天賦の才がわかろうというもの。数学は40代、50代の人がずっと君臨している世界、情報科学は時代の進み方が速いので30代が主役でないといけない世界、とこのころ。なるほど日本が情報で世界に伍していくのは大変そうだが、34歳で教授になった河原林さんには大きな期待がかかる。ガラパゴス的な日本数学界を変え、社会の数学者に対する見方を変え、企業で仕事をする数学者を増やす。その実現に日本の将来がかかる。

Mathematics and Logic as the Cornerstone

特集

「礎としての数理・論理」

数理・論理と情報学

宇宙の星の謎を探るにしろ、台風の進路予測をするにしろ、ロボットを動かすにしろ、その礎に数学がなければ問題を解決することはできない。しかし、数学とはそもそも人間が知りたいと思う欲求、好奇心から始まった学問であり、芸術に似て、それ自身が人間を感動させる「美」をもつと数学者たちは言う。純粋数学の意義から、世の中に貢献する数学、とりわけ情報学に応用される数学まで、数理・論理の研究の魅力について話を聞いた。



佐藤 健
Ken Satoh
情報学プリンシプル研究系教授

実社会を支える数理論理と数理工学

佐藤 まず、龍田先生の研究分野についてお聞かせください。

龍田 私の研究分野は数理論理です。これは、人間が数学を考えたときの思考方法を、その思考手順を厳密に記述し、対象化することにより分析するという学問。ギリシャ時代の古から人間が探求してきた非常に長い歴史をもつ学問ですが、数十年前からコンピュータへ応用されるようになり、より脚光を浴びるようになりました。とりわけ現在では、ハードウェアの回路、ハードウェアやソフトウェアの検証、ソフトウェアの仕様記述、プログラミング言語の設計、人工知能での推論などの分野において、必要不可欠な学問となっています。

たとえば、ソフトウェアの検証というのは、そのソフトウェアが間違いなく動くかどうかを数学的に証明して確かめるというもの。というのも、ソフトウェアは、いまだに人が手作業で書いて生産するしか方法がなく、必ず間違えるものだからです。一方で、ソフトウェアが間違っていると、パソコンが動かなくなるとか、銀行のオンラインシステムが止まってしまうとか、飛行機や原発などの社会インフラの制御ができなくなるといった具合に、私たちの生活に重大な支障をきたします。これらはすべて、プログラマーが意識しないで間違えて書いてしまったプログラムの部分、「バグ」によって引き起こされるものです。

そこで通常は、テストプログラムを動かして、だい

たい動くようであれば、そのソフトウェアを走らせるという方法がとられています。しかし、テストには漏れがあり、バグを見逃すことで重大な事故が起こることもある。これはソフトウェアの危機とも呼ばれています。そうしたことから、完全にバグのないソフトウェアを生産するために、ソフトウェアを数学的に検証し、どんな場合でも動くという保証が求められているのです。

一方、現状では、この手法はあまり採用されていません。なぜなら、ソフトウェアの数学的な検証には、大変な手間と費用がかかるからです。とはいえ、ソフトウェアは社会インフラを支えるうえで非常に重要な役割を担うものですから、将来的にはすべてのソフトウェアについて数学的に検証されるようになるでしょう。プログラム理論は、新しい種類の数学であり、ソフトウェアに内在する計算の本質を解き明かすことができる深みのある学問であり、興味は尽きることがありません。

佐藤 たとえば、現在、車にはたくさんのコンピュータが搭載されていますが、それが正しく動くかどうかを検証することも、大きな課題になっています。しかも車の場合はリアルタイムに、瞬時に判断し制御する必要がある。そういった場面でも数学論理の役割というのは非常に大きなものがありそうですね。

龍田 そうですね。さらに、数学論理はコンピュータ教育にも役立っている。たとえば、ソフトウェア技術者の考え方の基礎を養う役割を担っています。

TLCA List of Open Problems

<http://tka.di.unito.it/opitca/>

Updated July 11, 2008

Problem # 20 [SOLVED]

Submitted by Mariangela Dezani-Ciancaglini

Date: 2006

Statement. Type theoretic characterisation of hereditary permutations

According to [Bergstra and Klop, 1980, Barendregt, 1984] (Definition 21.2.9) a closed lambda-term M is a *hereditary permutation* if the Böhm tree $BT(M)$ has the following properties:

- the label \perp does not occur in $BT(M)$;
- each variable occurs exactly once;
- the head variable is the first abstracted variable;
- all other variables occur at one level lower than their abstractions.

The problem is to find a type assignment system in which all and only the hereditary permutations get types of a fixed shape. Many other sets of λ -terms have been characterised in this way, see for example [Kurata, 2002, Dezani-Ciancaglini et al., 2005] and the references there.

Solution: Makoto Tatsuta proved [Tatsuta, 2008] that the set of all hereditary permutations is not recursively enumerable, and therefore cannot be characterized by typability in a finitary type-assignment system. On the other hand, there is a system such that a term can be assigned a certain countably infinite family of types in that system if and only if it is a hereditary permutation.

References

[Barendregt, 1984] Barendregt, H. (1984). *The Lambda Calculus. Its Syntax and Semantics*. North-Holland, second edition.

[Bergstra and Klop, 1980] Bergstra, J. A. and Klop, J. W. (1980). Invertible terms in the lambda calculus. *Theoretical Computer Science*, 11:19–37.

[Dezani-Ciancaglini et al., 2005] Dezani-Ciancaglini, M., Honsell, F., and Motokawa, Y. (2005). Compositional characterization of λ -terms using intersection types. *Theoretical Computer Science*, 340(3):459–495.

[Kurata, 2002] Kurata, T. (2002). Intersection and singleton type assignment characterizing finite Böhm-trees. *Information and Computation*, 178(1):1–11.

[Tatsuta, 2008] Tatsuta, M. (2008). Types for hereditary permutations. In *Proceedings of the Twenty-Third Annual IEEE Symposium on Logic in Computer Science, LICS 2008, 24–27 June 2008, Pittsburgh, PA, USA*. <http://research.nii.ac.jp/TechReports/07-010E.html>.

1

図1 解かれた 20 番目の問題

プログラム理論の国際学会は、数学分野の未解決問題として 22 の問題を挙げている。龍田教授はその 20 番目の「全単射に相当するラムダ式を特徴づける型理論を与えよ」という問題に挑戦し、図のように見事に解決した。

ソフトウェアというのは、論理的に書かれ、論理的に実行されるものであり、数学論理を学ぶことで、技術者に必要な基礎的な素養を身につけることができるのです。

佐藤 なるほど。では、次に速水先生、お願いします。

速水 私の研究分野は数理工学です。一言でいえば、実際に解かなければならない問題があった場合、まず数理の言葉に置き換えて、問題を解決したり、解決するための方法論を編み出したりする学問です。なかでも私の専門は、数値計算という分野。自然現象、工学現象、社会現象などを、コンピュータを使ってシミュレートするための基礎技術の開発を行っています。

たとえば、台風などの気象予測や、飛行機の設計のためのシミュレーションには、流体現象を表す方程式を解く必要があります。流体の数学モデル、数理モデルというものです。次にこの方程式を解くわけですが、複雑な現象や複雑な形を扱う場合、手計算では解けないので、コンピュータを使って〈近似的〉に解きます。そこで必要なのが「離散化」というプロセス。離散化とは、連続またはアナログな方程式を、コンピュータが扱えるようなデジタルな式に変換、または近似することです。さらに、離散化された方程式の近似の精度を上げるためには、たとえば、3次元を扱う場合、空間を網目状に膨大なメッシュで切り分ける必要があり、その分、コンピュータで扱う変数の数も膨大になります。したがって、コンピュータの限られたメモリーと計算速度で答えを

導き出すためには、効率的な計算法(アルゴリズム)を開発する必要があるのです。

また、開発した数値計算法が一般にどう振る舞うのか、どんな場合にちゃんと機能するのか、答えの精度はどうか、答えに近づく(収束する)速さはどうかを解析し、保証するためにも数学が必要になります。

このように、現象のモデル化、離散化、アルゴリズムの開発と解析、すべてにおいて数学(数理)が必要になります。つまり、計算機を使っていろんな現象をシミュレートすることを情報学の一つのミッションとするなら、そのあらゆる局面で数理が必要だということがおわかりいただけるでしょう。

佐藤 現象のモデル化と離散化、アルゴリズムの開発と解析——そのいずれも、実社会の現象を、いかに数学に置き換えて、わかるように提示するか、ということですね。

速水 そうですね。その使命は現実の世界にできるだけ近づくことにあり、より精度を高めるために、つねに、モデルも計算法も計算機も進化が求められているのです。

数学的な難問を解決する喜び

佐藤 それぞれ、具体的にはどのような研究に取り組んでいらっしゃるのですか？

龍田 私が手がけているのは、プログラム理論と数理論理学ですが、前者は先述のソフトウェア検証などのための数学理論、後者がギリシャ時代から



龍田 真
Makoto Tatsuta
情報学プリンシプル研究系教授



速水 謙
Ken Hayami
情報学プリンシプル研究系教授

Mathematics and Logic as the Cornerstone

続く学問といえます。最近の私の研究成果としては、プログラム理論の国際学会が未解決問題として挙げた 22 の問題のうち、20 番目の「全単射に相当するラムダ式を特徴づける型理論を与えよ」という問題を、2 年前に解決したことです(図1)。言い換えると、これは、「1対1対応を表す抽象的プログラムを、データの種類の分析で特徴づけよ」という問いですが、結果としては「それは存在しない」ということを証明しました。22問の未解決問題のうちで最初に解決した問題であり、ソフトウェア理論の進展に貢献するだけでなく、真理の探究という意味においても大きな成果といえます。

速水 私のほうは、大規模な連立一次方程式や最小二乗問題を計算機で解くための計算法の研究を行っています。大規模な連立一次方程式というのは、先述の流体の方程式を離散化した際に現れるものです。たとえば、空間を各方向から網目で 100 に切ると、空間全部で 100 万の接点に分解されます。となると、100 万元の連立方程式を立てることになる。それをいかに効率的に解くか、というのが課題です。

もう一つの最小二乗法とは、もともと 19 世紀の数学者ガウスが、天体の軌道の予測や測量のために開発した手法です(国立情報学研究所平成 19 年度市民講座「社会に生きる数学」参照)。応用と

してはほかに、実験などで生じる大規模なデータの統計解析、画像処理、信号処理、制御など多岐にわたります。

大規模な最小二乗問題を計算機で解こうとすると、繰り返し計算して解を改良していくという、「反復法」が重要になります。私たちは現在、大規模で解きにくい、たちの悪い問題でも高速に解けるような計算法を開発したり、その方法の振る舞いの数学的な解析を行ったりしています。大学院生が新しいアイデアを次々に考え出してきてくれるので、一緒に楽しみながら研究に取り組んでいるところです。

佐藤 いずれも一般の人にはほとんどなじみのない難しいテーマではありますが、科学技術の発展に、数理、理論が不可欠だということがよくわかりました。

数学の意義は好奇心と感動にある

佐藤 一方で、数学というのは、具体的に社会の役に立つことばかりではありません。とりわけ純粋数学では、真理の探究にこそ重きが置かれています。

龍田 数学の歴史を振り返ってみると、真理の探求として意義のある定理や法則は、数十年、あるいは数百年後に、結果として、実社会で役立ってきた例がいくつもあります。逆に言えば、最初から役に立てることを念頭に研究をしているわけではない、ということなんですね。

私は、数理、論理の意義には三つあると考えていますが、最も大切なのは好奇心だと考えます。人間が知りたいと思う気持ちに答えて、心を豊かにすることにこそ数学の意義がある。たとえば天文学も、あの美しい星のかなたはどうなっているのかを知りたい、という気持ちを原動力にして始まった学問です。要するに数理や論理は、文学と同様の意義をもつということです。

たとえば、幾多の天才数学者が長年にわたって挑み続けた「フェルマーの定理」($x^n + y^n = z^n$ となる整数 $n > 2$, $x, y, z > 0$ は存在しない)も、どのような地図も 4 色で塗り分けられるという「四色定理」も、あるいは、大きな素数についての疑問も、すべてのプログラムを模倣できるようなプログラム、いわゆる

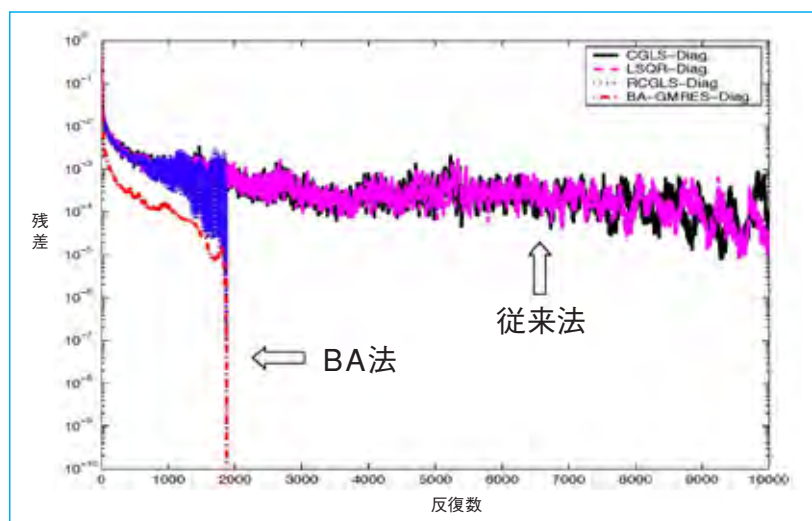


図2 BA法と従来法の比較
BA法という大規模な最小二乗問題に対する新しい反復解法を開発した。BA法を用いると、従来法に比べてはるかに少ない反復数で最小二乗解を求めることができる。

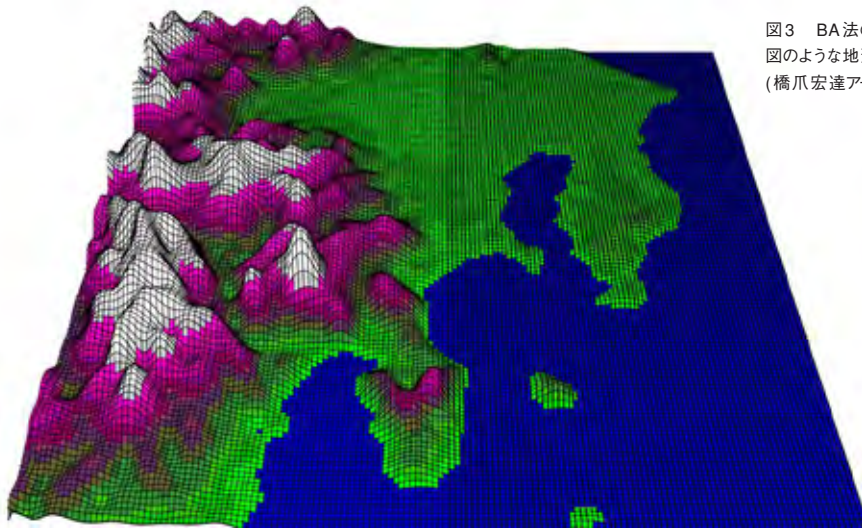


図3 BA法の最小二乗問題への適用例
図のような地形データの補間に用いることができる。
(橋爪宏達アーキテクチャ科学研究系教授作成)

る「万能チューリングマシン」は存在しない、という難問も、すべての源泉は好奇心にあります。また、文学や芸術は、人間が価値を感じて感動するから存在しているのだと思いますが、それは数学や論理も同じです。先述の難問の解決というのは、まさに感動的です。

佐藤 確かに、いい証明を見ると、いい仕事をしているなと思うし、美しいと感じますよね。それは、科学者や数学者がもつ共通の感覚であり、難問と呼ばれるものは、数学者が純粋にそこに面白みや感動を感じるから、全身全霊をかけて取り組むことができるのでしょね。

速水 興味深いことに、好奇心や感動を与えるような、数学者が直感的に美しいと感じるような理論であればあるほど、結果的に社会に役立つ場合が多いのも事実です。

龍田 ただ残念なことに、数学の意義として好奇心や感動というのは一般の人にはわかりにくいいため、現在は、社会の役に立つという意義が強調されすぎているくらいがあります。実際に、科学技術の発展に、数理、論理は不可欠に違いありません。ところが数理、論理の研究は非常に息が長いもので、フェルマーの定理が解かれたのは、360年後の1994年ですし、四色定理も120年後の1976年でした。さらにそれらが、具体的に役に立つかどうかわかるのは、ずっと先になってからのことが多いのです。ギリシャ時代から研究されている論理はいまやコンピュータに不可欠ですし、四色定理はいろいろな計算の最適化に、そして、大きな素数は暗号の分野で役立っていますが、そうなるには、数十年、数百年という時間が必要でした。

佐藤 数学に限らず、物理でも同様のことがいえ

ますね。ニュートンの万有引力も、我々の生活の何かに役立たせるつもりで発見されたものではありませんが、そこからアインシュタインの相対性理論が生まれ原子力エネルギーの開発に進んでいったと考えられます。

速水 もちろん、何かを解決したいという欲求から理論が生み出されることもあるわけですから、科学技術の発展に役立つということを否定的に捉える必要はないでしょう。

ちなみに、ガウスという人は、若いうちは数論の研究をしていたのですが、その後天文学へ転向し、最小二乗法を考案、彗星の出現を予言して一躍有名になりました。さらに最小二乗法を測量や地図の作成に用いています。また、測量の問題に取り組む中から微分幾何学のきっかけを創ります。この微分幾何学は後世、アインシュタインの一般相対性理論の構築に用いられることになります。このように学問というのは、理論が世に役立ち、さらに好奇心を刺激して、新たな理論を生むという連鎖構造をもっているんですね。

龍田 さらに付け加えるなら、数学者の真理に対する直感と良心だけが、数理、論理の正しい研究方向を与えることができるということではないでしょうか……。

佐藤 役に立つだけだと、問題が解決したらおしまいです。数学のような深い学問では、一般性があるので、さまざまな応用ができる可能性がある。それがまた新たな理論を生み出し、思いがけない分野で爆発的に役に立つということにつながるということでしょうね。

本日はお忙しいところありがとうございました。

(取材・構成 田井中麻都佳)

アルゴリズムがゲノムの空白を埋める！

生物の全遺伝子配列を次々に明らかにしてきたゲノムプロジェクト。

多くの生物のゲノムが解明されたといわれているが、実はその中にはまだ正確に解明されていない領域が点在している。この不明な空白を新しいアルゴリズムが埋めようとしている。

さらに、このアルゴリズムによって生物学研究に新しいアプローチが生まれつつある。



宇野毅明
Takeaki Uno
情報学プリンシプル研究系 准教授

2006年12月、情報・システム研究機構は新領域の融合研究を生み出そうと、静岡県伊東で“若手クロストーク”を開催した。NII情報学プリンシプル研究系の宇野毅明准教授は、自身が作成した類似性解析のプログラムの応用分野を求めて、ヒトとマウスのゲノム配列から似ている部分を非常に短時間で見つけ出せることを発表した。一方、国立遺伝学研究所の梅森十三特任研究員は、自分が研究しているマウスのゲノム配列の解析方法はないかと参加していた。「これだ！」と思った梅森氏が、宇野氏に声をかけて意気投合。一晩中プログラムの応用について話し合ったという。

その将来性から、2人の共同研究は新領域融合研究センターのプロジェクトとして採用され、3年間の融合研究がスタートした。

コンピュータにできることを探す

まず取り組んだのは、マウスの第13番染色体の配列の正確な解明だ。この染色体には、*Genic1*という領域がある。梅森氏は、ある系統のマウスを掛け合わせたときに起こる神経異常(写真)の原因が、*Genic1*にあることを突き止めて以来、この領域(図)の研究を続けてきた。しかし、繰り返し配列が多い難解析領

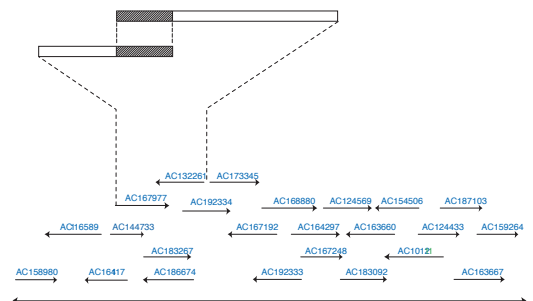
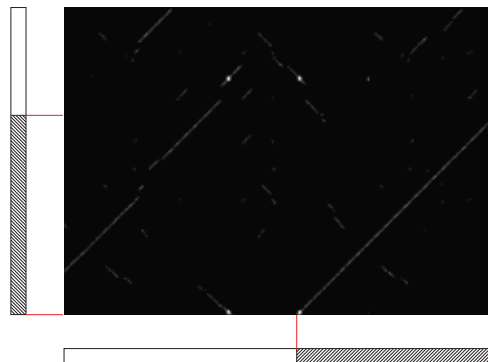
域であったため、ゲノム配列の解明が進まず研究の妨げになっていた。

「コンピュータが活躍できる場所を見つけるのが、私の仕事の1つです」と宇野氏。ゲノム配列決定のどこにコンピュータが活躍できる余地があるだろうか。ゲノムは、AGCTという4種類の塩基が連なってできている。現在の技術では、ぶつ切りにして1000塩基ずつ読むという方法でしか配列を読むことはできない。そこで、大量の断片の配列を読み、それらがどのように重なり合うかを調べることで、元の配列を解明する。まるで、パズルのピースを組むようだ。この際、大量の断片がどのように重なるかを調べるコンピュータプログラムが必要になる。このプログラムの性能を決める計算方法をデザインする“アルゴリズム研究”が専門の宇野氏は、新たに「高速相同性検索アルゴリズム」を提案した。

宇野氏のプログラムの新しさの1つは、30塩基の類似性を計算できる点だ(図の左)。これまでのプログラムの多くは、10塩基程度が正確に一致しているところを足がかりに重なる部分を見つけている。これは4種類の塩基10文字の並び方が 4^{10} =約100万通り存在するため、重なる部分の候補を100万分の1に

図 マウスのゲノム断片の配列比較(左)とマウス第13番染色体の塩基配列

「高速相同性検索アルゴリズム」を用いて2つのゲノム断片の類似性を計算した結果は左図のようになる。図の右側のきれいな白線の部分に対応する2つのゲノム断片の斜線部分が重なることがわかる。こうしてゲノム断片の並び方を決めていくと、マウスの第13番染色体の塩基配列もわかることになる。





梅森十三
Juzoh Umemori
国立遺伝学研究所 特任研究員



小出 剛
Tsuyoshi Koide
国立遺伝学研究所 准教授



写真 正常マウス(右)と遺伝的不適合マウス(左)
韓国のマウスとブルガリアのマウスを交配させると、遺伝的不適合が生じ震えを起こすマウスが現れる。その原因が*Genic1*領域にあることを梅森氏は突き止めた。

絞り込み計算速度を上げることができるからだ。しかし、100億塩基を超えるゲノム配列を扱うときは、100万分の1に絞り込んでも一致する個所は少なくとも1万以上。これでは、事実上計算できない。対して宇野氏のプログラムを使うと、配列の種類は 4^{30} 通りになり、同じでない配列が同じとみなされてしまう可能性はほとんどなくなる。

もう1つ重要なことは、ゲノムの断片配列が、読み取り処理で起こるエラーを含むという特徴を配慮した点だ。こうして考え出された“30塩基のうち1~2個の不一致があっても同じ配列とみなしていい”というアルゴリズムを用いることで、*Genic1*の正確な配列は明らかになった(図の右)。

ゲノムが違ったものに見えてくる

融合研究は、現在、プログラムの側から生物学に研究のヒントを提供するという新たな段階に入っている。「高速相同性検索アルゴリズム」を使えば、ゲノムの中にどのような繰り返し配列がどの程度の頻度で存在しているかを簡単に調べることができる。これまでは見つけ出すだけでも難しいとされてきた繰り返し配列について、新たな知見が得られることになるだろう。

「コンピュータは繰り返し配列の候補を見つけることはできても、最終的に判断するのは人間です。だから、生物の知識をもつ梅森さんはすごいんですよ」と宇野氏はあくまでも道具を提供しただけという。一方、梅森氏は「これらの繰り返し配列は、ゲノムの構造上重要な意味をもっているかもしれないと考えようになりました。ゲノムを単純に塩基の並びと見

る宇野さんの発想から影響を受けた結果です」と話す。異なる分野と融合することで、まったく新しい見方が生まれたという。

それぞれの目標のために協力する

「すれ違いがあるのは当たり前。だから、理解できるまでしつこく話し合わなくてはなりません」と宇野氏は融合研究を進める上での心構えを話す。そもそも、梅森氏は「高速相同性検索アルゴリズム」を使って得られる生物学的知識に興味があるのに対して、宇野氏は開発したプログラムの応用範囲がどこまで広げられるかに興味がある。専門も目標も違う研究者が協力するのは大変だ。しかし、だからこそお互いに足りないところを補い合って、これまででない研究成果が得られるのだ。

「プログラムを開発する場合、複雑な現実の問題は避けて、計算しやすい想定問題に逃げてしまいがちです。そうした情報研究に対して、この融合研究では実際に役立つ“地に足の付いた研究”ができたと思っています」と宇野氏。この点が高く評価され、データマイニングの国際会議で最優秀賞を受賞した。2人の研究に協力してきた国立遺伝学研究所の小出剛准教授は「情報という新しい手法をつかって、これまで解けなかった生物学の問題を解こうとしたことは、私たちにとってもチャレンジングでした」と振り返る。

この融合研究プロジェクトは、2010年の3月で期限を迎える。しかし、宇野氏も梅森氏も「まだまだやりたいことがあります」と声をそろえる。2人がデータを前にディスカッションする日々はこれからも続く。

(取材・構成 池田亜希子)

NIIでなぜ数学を志したか

科学の基礎、とりわけ情報学の基礎をなす数学を学ぶ醍醐味はどこにあるのか。数学を目指した動機、その面白さなどについて総合研究大学院大学複合科学研究科情報学専攻修士課程2年生の山下翔さん、保國恵一さん、同博士課程2年に在籍するFaisalさんの3人に語ってもらった。

——現在の専攻を目指したきっかけを教えてください。

山下 私は元々は文科系で、人のコミュニケーションに関心がありました。が、数学にもひかれるところがあり、両方を一緒に学びたいと思っていたところ、数理言語学という分野に行き当たりました。大学時代(千葉大学文学部行動科学科)から金沢先生に指導を受けていて、学問の内容に対する興味が膨らみ、継続して教えるを受けたいと思うようになりここで学ぶようになりました。人間の用いる様々な言語現象を模倣できるような数学的なモデルを構築する研究をしています。

保國 学部時は音響分野が専門でした(九州大学芸術工学部音響設計学科)。卒業研究では、音の環境を数学的に予測して設計するため、数値計算を用いた研究をしていました。それがきっかけで、数値計算に対する関心が高まったので、現在の指導教員を探し出し、本専攻を志しました。

Faisal バングラデシュ出身で、修士課程は母国のDarul Ihsan大学でコンピュータサイエンスを学びました。今は龍田先生の下で、プログラムの設計に誤りがないかを検証する手法や、その基礎となるラムダ計算の研究をしています。情報学は、人工物、そして自然のシステムのすべての基礎をなすものだと思います。今日、コンピュータがどんどん社会に入り込み、あらゆる場面で生活を支えています。多種多様なプログラムが稼働しており、その設計のよりどころとして、信

頼のおけるアルゴリズムや管理手法が求められています。

母国では情報学やコンピュータサイエンスの研究はまだ萌芽期にあり、今後力を入れるべき分野です。こうした理由から、研究をしたいと考えるようになりました。

論理の裏付けには数学が必要

——研究内容を説明していただけますか。

山下 人が話す自然言語には、単語を配列して文を構成するための規則がありますが、それを数学的に表現したものを「形式文法」といいます。例えば、自然言語では、「カツオが、<サザエが料理した>魚を食べた」というように、文の中に文が埋め込まれているよ

うな構造があります。このような言語現象一般をある形式文法によってモデル化できるかというのが、基本的な問いです。その問いは形式文法の数学的性質によって明らかになります。このような数学的性質を調べるのが私の研究です。

また、プログラミング言語では、プログラムをコンピュータが実行可能な機械語に変換するコンパイラというソフトウェアがありますが、そこでは形式文法での結果が使われており、そうした場面でも研究成果が応用できます。

保國 私の研究は、中学校で習う連立一次方程式と深い関係があります。中学で習う連立方程式では、未知数と式

の数が同じでせいぜい2~3個のものを扱い、未知数を消去していき、残る未知数を割り出していくという形で解が得られます。

現在、私が研究対象としているのは、未知数と式の数が一致しない連立方程式で、最小二乗問題といわれるものの解法です。数万個もある方程式の場合、未知数を消していく方法では計算量が



膨大になってしまいますが、あるアルゴリズムを作ってそれを繰り返すことによって徐々に求めるべき未知数に近づけていけば、より少ない計算量で解が求められるようになります。それには経験則ではなく、正しい数学的な裏付けのある方法でなくてはなりません。実用面では、観測した雪の状態から雪崩を予測したり、観測したデータから本来の画像を復元するような画像処理などに応用されているようです。

Faisal ソフトウェアを開発している最中に、確かに目指すゴールに向かっていくかどうかを検証することは非常に困難です。私の研究は、構造化された論理



保國 恵一
Keiichi Morikuni
速水 謙 教授研究室



Mahmudul Faisal
Al Ameen
マハムドゥル・ファイサル・
アル・アミーン
龍田 真 教授研究室



山下 翔
Sho Yamashita
金沢 誠 准教授研究室

体系を駆使して、部分的にプログラムされたものが最終的に1つのソフトウェアとして構築された時に、期待した通りに正しく動作するかを、事前に検証しようというものです。今のところ、検証するプログラムの大きさには限界がありますが、もっと大規模で複雑なシステムにも適用できるような方法を目指しています。論理体系を作り上げることができても、そ

大きな挑戦でもあります。

保國 私の場合、研究をしていて最も面白いのは、アイデアが湧き、それを実際に試してみてもうまくいった時ですね。世の中には、数学を学んでもいったい何の役に立つのか分からないといった議論もあるようですが、いろいろな研究の基礎となり、実際に役立っていることが実感できるので楽しいと感じます。

Faisal 数学では、ある性質を証明するということは、すなわち新しい定理を作り出すということです。それがゲームのようで、わくわくします。

——研究成果を発表したり、対外的な活動をする場はありましたか。

保國 国内外の学会や研究会での口頭発表、他研究所

の研究者の先生の訪問や、そこでの研究や発表をする機会に恵まれました。今までの手法より短い計算時間で正しい結果が得られたので、それを発表し、その集大成として修士論文をまとめていくつもりです。

山下 まだ本格的な学会発表はなく、内輪のセミナーで発表する程度です。修士論文が、きちんとした成果をまとめる最初の機会になります。

学会ではありませんが、この夏はフランスのサマースクールに参加しました。ヨーロッパは形式文法の研究が比較的盛んなので、大きな刺激を受けてきました。英語力を磨く必要性

も痛感しました。

Faisal NII テクニカル・レポートで研究成果を発表しました。SEFM(Software Engineering and Formal Methods)という国際学会にもアクセプトされています。筆頭著者は龍田先生ですが、もし私に機会が与えられれば、それが最初の発表の場となります。

研究者として貢献したい

——今後の抱負をお聞かせください。

山下 私のやっている形式文法はマイナーな分野ですが、博士課程を終えた後は研究者を目指して勉強するつもりです。脳の中で言葉がどのように実在しているかを知りたいというのが、最初の大きな動機でした。その究極の真理を究めるのは、自分が生きている間には難しいでしょうが……。

保國 さらに勉強を続けていき、プログラムを活用しつつ、新しい分析を絶えず加えていきたいと思っています。研究成果が、多くの人に実用の場で受け入れてもらえるようなものになることをも望んでいます。

Faisal 私は博士号が得られたら、帰国して大学の先生になり、後進を育てたいと思います。欧米や日本に比べて、バングラデシュでは研究のインフラストラクチャが不十分です。より先進的な研究環境を整備して、この分野で母国に貢献することが夢です。

——本日はありがとうございました。

(取材・構成 塚崎朝子)



れが整合性のある体系であるという証明をすることが必要になり、そこがチャレンジングです。

アイデアが形になる面白さ

——どんなところに、研究の面白さを感じますか。

山下 今までは人文科学と呼びながらも、言語学には科学でない部分が多くありました。数学を用いることで、まだ手つかずだったところを探っていけることが、研究の醍醐味です。とはいえ、半世紀以上前からこうした研究の枠組みが徐々にできてきたわけですが、そこから遅々として研究が進んでいないのが実情で、

ゲームに秘められた情報処理の仕組み

定兼邦彦（国立情報学研究所 情報学ブリッジ研究系准教授）

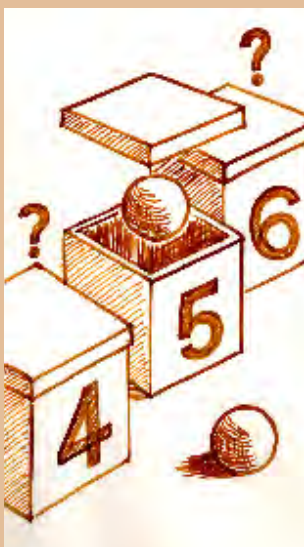
次のようなゲームを考える。部屋に、番号のついた箱とボールがたくさん置いてある。それぞれの箱には一つのボールを入れることができる。ゲーム開始時には、全ての箱は空である。m人のプレイヤーは、別々に待機し、ランダムな順番で部屋に呼ばれる。プレイヤーは、一つの箱のふたを開け、中のボールがあれば取り出せ、空ならばボールを入れることができるし、何もしなくてもいい。しかし、最後に必ずふたを閉める必要がある。この動作をt回繰り返す。ただし、全プレイヤーは、あらかじめ相談し、どのようにボールを入れるかを決めておくことができる。

さて、勝敗はどのように決まるのか？ プレイヤーは、自分が最後に呼ばれた人かどうかを当てれば「勝ち」に、一人でも間違えれば「負け」になる。一見勝ち目のなさそうなゲームだが、ボールの入れ方をうまく決めておけば、必ず勝てる方法があるのだ。その際、箱を開ける回数tをできるだけ少なくするにはどうすればいいか。

必ず勝つにはどうするか

自分が何番目のプレイヤーかが分かれば、必勝法の一つになる。これには、箱とボールを使ってカウン

ターを表現すればよい。すなわち、ボールの入っている箱は1、入っていない箱は0とすれば、箱とボールで2進数のカウンターを表現できる。こうしておけば、各プレイヤーは現在のカウンターの値を見て、それを1増やすことができる。カウンターの値を1増やすには、2進数の一番下の桁に対応する箱を開け、ボールがなければボールを入れ、入っていれば取り出



す。入っていた場合は上の桁に繰り上がりがあるのだ、一つ上の桁で同じことを繰り返せばいい。カウンターの値がmになれば自分は最後の人であることがわかる。カウンターの値は1からmなので、2進数は、 $\log_2 m$ ビットで表現できるから、 $t = \log_2 m$ 回箱を開ければ、必ず勝てるわけだ。

1からmまでの数を表現しようと思うと絶対に $\log_2 m$ ビット必要だから、この方法は一見最適に見

える。しかし、もっと小さいtの必勝法は存在しないだろうか。よく考えてみると、このゲームでは、各プレイヤーは自分が最後かどうかだけわかればいい。つまり、何番目かを厳密に知る必要はないのである。この点で、カウンターを用いる方法は冗長であり、より最適な方法の存在が示唆される。

このゲームでは各プレイヤーが開けることができる箱の数は制限されているが、箱を何個使うかは関係ない。こう考えれば、実は、tを小さくできるので、ボールの入れ方は複雑になるが、 $t = \log_2 \log_2 m$ の必勝法が存在する。

情報のすばやい利用が勝負を決める

このゲームにおいて、箱は、ボールが入っている状態が1、入っていない状態が0を表す計算機のメモリとみなせば、箱の数は、ある情報（自分が最後かどうか）を表現するために必要なメモリの量（ビット数）に対応する。また、箱を開ける回数tは、情報を得るために必要なメモリアクセス回数（時間）に対応する。情報をすばやく取り出せるようにすることが、より良い必勝法につながる。単なるゲームのように見えるが、奥が深い。

情報から知を紡ぎだす。

NII

国立情報学研究所 ニュース (NII Today) 第46号 平成21年12月

発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 <http://www.nii.ac.jp/>

〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

編集長：東倉洋一 表紙画：小森 誠 写真撮影：由利修一 デザイン：鈴木光太郎 制作：サイテック・コミュニケーションズ

本誌についてのお問合せ：企画推進本部広報普及チーム TEL:03-4212-2135 FAX:03-4212-2150 e-mail: kouhou@nii.ac.jp