

NII Today

National Institute of Informatics News

特集 人工知能 — 個人の知から社会の知へ —
推論で生命活動の謎解きに挑む

「意味」が情報をつなぎ
Webの世界を変えていく

AIが拓く新しいデータマイニング
人工知能が裁判過程を推論する



人間に代わって、コンピュータが答えを導き出す。

情報学をベースとして、人工知能の研究がNIIでも盛んに行われている。

研究の向かう先にあるのは、さまざまな学問との融合やWebサービスの進化。

豊かな社会づくりに不可欠な「知」として、人工知能研究は今、大きな注目を集めている。

NII Interview

推論で生命活動の謎解きに挑む

人工知能が仮説を立て、 実験し、確かめる

吉田 コンピュータというと、決められた規則に沿って計算するのは得意だけれど、規則外のことはお手上げ、という印象があります。井上教授が研究されているのは、それを越えた仕組み、つまり結論に到達するために必要な規則をコンピュータが推測し「仮説」として考え出すものです。それを現在はシステム生物学という分野に応用されていますが、まずは推論による仮説の発見とはどういうことか、聞かせてください。

井上 有名な三段論法の例としてA「人間は死ぬ」B「ソクラテスは人間である」C「ソクラテスも死ぬ」を考えてみましょう。Aは大前提あるいは規則、Bは小前提、Cは結論です。まず一般的な原理としてAがあり、Bはその条件を満たす個別事実であることから、個別の結論であるCもまた正しいことを導くのが三段論法の論理です。この時、Aの「人間は死ぬ」とCの「ソクラテスは死ぬ」の二つしかない時、これを成り立たせるにはどのような前提を付け加えればいいでしょうか？

吉田 「ソクラテスは人間である」という小前提ですか？

井上 そうです。「ソクラテスは人間である」という「仮説」を推論で立て、それが正しいか確かめればよいのです。これを、アブダクション(※1)と呼んでいます。逆に、B「ソクラテスは人間である」とC「ソクラテスは死ぬ」からAの「人間は死ぬ」という大前提を見つけるものはインダクション(※2)、帰納と呼ばれています。

吉田 何かを発見するのは人間ならではの創造的な思考だと思うのですが、この場合コンピュータがそれまでなかった新たな仮説を創り出すということですか？

井上 論理的に可能であるような、隠された関係を推測するものですが、それが人間にとっては新しい仮説を創造することに相当することも、原理的にはあり得ます。

吉田 そのような推論による仮説発見の分野は新しいのですか？

井上 いえ、計算機分野では20年以上も前から研究されています。しかし、私の研究の新しい点は、知識が欠けている場合にも、うまく働くことです。

海外の研究者に見いだされ、 システム生物学の世界へ

吉田 井上教授ご自身は、推論の世界にどのような経緯で入られたのですか？

井上 学生時代の専門は、アルゴリズムでした。問題を速く効率的に解くために、使う計算の種類や組み合わせ、手順などを考えるものです。一方、ゲームの探索についても興味がありました。こちらは、相手に勝つという最終目的を達成するために、刻々と変わる状況の中でどのような手を展開していけばよいか、膨大な選択肢から選ぶ作業です。このお手本は人間の知能です。

吉田 どちらかという機械的な計算手法の分野ですが、これらが論理の世界へと入るバックグラウンドになっているのですか？

井上 その後、将来には何が起るのか、あるいは過去に何が起こったのかなど、より広く知的な活動に対応できる手段として、知識が不完全な下での推論方式のための理論に興味を持ちました。

吉田 現在につながる研究の流れが生まれたのはいつごろですか？

井上 2001年ごろです。以前発表した仮説発見手法が、英国のポストゲノム研究グループの目に留まり、問い合わせが来しました。ロボットを使って、遺伝子の機能推定に関する仮説生成、その仮説に基づいた実験の計画、実験実施、その結果からのフィードバックまでを自動的に行うという野心的な研究で、その仮説生成の理論背景に私の手法が使えるのだとのことでした。その後、フランスの研究者からも同様の問い合わせがあって共同研究したこともあり、われわれが開発した「SOLAR(※3)」という手法の本格的な応用を始めました。この時に選んだテーマは、代謝経路における酵素の反応に関して、何が反応を促進し、何が阻害要因になっているかを発見することでした。

吉田 計算機を使った学習や仮説発見の手法がさまざまある中で、SOLARの特長は何ですか？

井上 かなり広い問題のクラスにおいて論理的に可能な仮説を列挙できることです。とくに、これまでに既存の理論やデータに欠けている部分を仮説として補完できる機能を実現している仕組みは他にはありません。これを実現するためには、無駄を省き、効率的



井上 克巳
Katsumi Inoue

国立情報学研究所
情報学プリンシプル研究系 教授

に有力な仮説の候補を見つけ出す必要があります。不十分な前提から結論を導き出す時、補完すべき候補は多く考えられます。しかしそれを一つずつ確認していたら膨大な時間がかかります。場合によっては候補は爆発的に増え、実用的な時間では解けないこともあります。それを回避するため、候補の有望性を確率的に評価してランキングし、上位から実験で確かめたりするような仕組みを取り入れたりしています。

吉田 アルゴリズムやゲーム探索といった、計算機科学の技術が総合的に生かされているということですね。こうして生まれたSOLARがシステム生物学にも応用されているとのことですが、システム生物学とはどのような学問なのでしょう。

生物の働きの ミッシングリンクを探す

井上 生命活動を、さまざまな要素が絡み合って構成している一つのシステムとしてとらえ、全体的に理解する学問です。

吉田 これまでの生物学は、ミトコンドリアは細胞の呼吸に関係していたり、リボソームはDNAをコピーして新しいたんぱく質を作り出したりといった機能解明が主というイメージがありますが。

井上 個々の組織や部品を切り分けてその働きを解明する研究はこれまでの生物学の中心であり、今も主要な営みです。一方システム生物学は、細分化したものを再び一つにまとめ上げていく統合を目指しています。

吉田 生物を構成する、臓器や、その中の細胞、神経、DNA、酵素、分泌物質など大小さまざまなレベルのものをまとめて相手にするのですか？

井上 そうですが、全部をまとめることはすぐには実現できません。まずはある機能に限定したとしても、構成要素そのものがどのように働いているかということに加えて、構成要素間の関係がどうなっていて、それらの関係が全体としてどのような機能を実現しているかを探ることが重要です。例えば細胞内では、呼吸などの代謝や変化を起こすため、あるいは反応を抑えるためにシグナルが伝達され、遺伝子の発現や抑制の制御が行われています。これらは、多数の要素が複雑に絡み合ったネットワークとして機能しています。また、一つの要素は他のネットワークにも属して

いたり、ネットワーク同士で影響を与え合ったりするなど複雑な重層構造になっています。生物という一つのシステムの中で各要素が互いにどのように関連して働いているか、モデルを作り、計算機でシミュレーションを行って検証するのがシステム生物学です。

吉田 システム生物学において、計算機はどのように有効に使われるのですか？

井上 現在では、遺伝子の働きなどがどんどん解明されており、膨大なデータが蓄積されています。このようなデータを扱うために計算機は必要不可欠ですが、一方で全体を統合的に把握するためには、まだまだ分かっていない「ミッシングリンク」も多いのです。

吉田 だから計算機による推測と仮説発見が大切なのですね。取り組みの具体例にはどのようなものがありますか？

井上 例えば紫外線による皮膚がんの発生メカニズムがあります。紫外線を浴びると、皮膚のDNAが傷つきます。異常な皮膚細胞が増えようとがんにつながり、p53というがん抑制遺伝子が存在します。主役はp53ですが、実際には数多くの反応が連鎖して働いていて、p53が他のたんぱく質と結合して変異すると逆にがん化を促進することもあります。こうした仕組みは複雑で、制御ネットワークを補完することで新しいがん抑制遺伝子を発見することが期待できます。

吉田 最後に将来の目標についてお聞かせください。

井上 こうした研究の究極の目的である、人類にとって未知で有用な仮説の発見につながればと思っています。そのためには、システムを広く使ってもらうことも必要です。生物学の研究者と話していると、大量の情報処理の必要は感じていても、システム生物学についてはまだ十分に知られていない感じがします。膨大なデータを扱うシステム生物学で、仮説発見システムが隠れたメカニズムを見いだす有用な道具として貢献できると思いますし、応用の機会が増えれば、そこから成果が生まれていくと思います。

※1 アブダクション(abduction): 論理学では仮説的推論という意味で用いられる。

※2 インダクション(induction): 帰納。アブダクションとともに、不完全な知識からの推論形式。

※3 SOLAR: 与えられた知識(論理式)の集合から、結論を効率的に導き出すシステム。これを変形すると、足りない知識(仮説)を導き出すことができる。



吉田典之
Noriyuki Yoshida

読売新聞東京本社
編集局科学部記者

インタビュアーの一言

推論や仮説の発見は人間の専売特許と思っていた。計算機はそれをさらに高速、緻密にこなすようになってきた。このままでは人間の出る幕はなくなっていくのではないか。そんな疑問を井上教授は明確に打ち消した。「もっと上から見たり、違うところからヒントを得たり、人間の順番はずっとある」。それは、生命と機械両方の本質にかかわる部分で研究する立場から生まれた言葉だろう。生命活動とよばれるシステムを解明するなかで、その奥深さや素晴らしさがさらに明らかになっていくのではないか。そんな予感がした。

「意味」が情報をつなぎ Webの世界を変えていく

Web上から収集した情報を選別し、有用なものだけを提供してくれる。
コンピュータが、まるで人間の意思を理解しているかのように振る舞うのだ。
共同のプロジェクトに取り組む3名の研究者が
セマンティックWebの魅力と可能性について語る。

人間のニーズを汲み取る セマンティックWebとは？

「例えばこの取材中に、『おなか減った』とWebに投げかけるとします。そうすると、この取材の終了時間、所在地、直近の私の食事メニューなど、私が伝達していない要素をも考慮して、飲食店の情報をいくつか提示してくれる。そんなSFのようなことが実現するのも、そう遠い未来の話ではないかもしれません」と、セマンティックWebの持つ可能性を語るのはNIIコンテンツ科学研究系の大向一輝准教授だ。

コンピュータが多面的かつ複合的に情報を返すことで、Webで情報を収集して意思決定する際の人間の手間を減らすこと。それを実現しようというのが、セマンティック(※1)Webだ。このような考え方が登場した背景には、誰もが当たり前Webを使って情報収集をするようになったにもかかわらず、膨大に存在するWeb上の情報から、意思決定に必要な情報を取り出す仕組みが確立されていないという現実がある。「ホワイトカラーの方の業務の30%は情報検索だと言われているくらいです。これでは、Webがなく図書館などで情報を探していた時代と、さほど業務効率は変わらないでしょう」

人間がなかなか見つけられない「意思決定に必要な情報」。セマンティックWebの世界で、コンピュータはどのようにしてそれを見つけてくるのだろうか。

情報の関係性を学ぶ 賢いコンピュータ

セマンティックWebを成熟させるアプローチは、大きく2つあるという。1つ目は、機械学習(※2)の手法でコンピュー

タの情報処理能力を向上させるというものだ。「具体的には、言葉が持つ意味の関係性を考え、コンピュータに記憶させるということです」とNII情報学プリンシプル研究系の市瀬龍太郎准教授は語る。「例えば、『源氏物語』と『紫式部』という2つの言葉には、文学作品とその作者という関係性がありますが、これらはイコールで結ばれるものではありません。しかし、これらに関するWeb上の情報を収集し分析したところ、Web上ではほぼ同じような意味合いで使用されているということがわかりました。このような場合、『源氏物語』という語句でWeb検索をかけたユーザに、『紫式部』で検索をかけたときにヒットする情報も合わせて提示してあげることが、そのユーザの意思決定を助けることになります。このような処理は、コンピュータが情報を分析し、その結果から『源氏物語 = 紫式部』ということ記憶することで可能になります」。Web上に存在する何億ページもの情報相手にコンピュータがこのような情報処理をする場合、いかに人手をかけずに抜け漏れなく、かつ効率よく行うかが重要になってくる。その鍵は、もう一方のアプローチが握っているという。

コンピュータが 情報を読み取り、つなげていく

2つ目のアプローチは、コンピュータが読み取りやすい形式の情報を、できるだけ多くWeb上に流通させるというものだ。具体的には、XML(※3)などで、情報の種別となるメタデータ(※3)を付加するということだ。Webに一般的に流通



Ikki Ohmukai

大向一輝

コンテンツ科学研究系
准教授

Ryutaro Ichise

市瀬龍太郎

情報学プリンシプル研究系
准教授

Hideaki Takeda

武田英明

学術コンテンツサービス
研究開発センター長 教授

している情報の記述形式、ハイパーリンクは、人間にとっては読み取りやすい情報だが、コンピュータにとって読み取りやすいものではない。「コンピュータが読み取りやすい形式の情報」がWeb上に増えてくれば、情報を収集・分析し、情報の関係性を記憶するという一連の作業をコンピュータが自動で行うことも可能になるだろう。

現在、Web上に情報をアップロードする際、XMLなどでメタデータを付加することは義務化されている訳ではない。「コンピュータが読み取りやすい形式の情報」を増やしていくためには何らかのルールが必要になる、と考えたくなるどころだが、一概にそうとも言えないようだ。

「ソーシャルタギングという言葉をご存知ですか？簡単に言うと、他のユーザと共有することを前提として、Web上にアップロードする情報にタグをつけることです。写真共有サイトのflickrやソーシャルブックマークサイトのdeliciousなどで、かなり大きな動きになっています。それぞれのユーザは、セマンティックWebを意識している訳ではありませんが、タグがついた情報のかたまりを俯瞰的に見ると、それらが体系化されていることがわかります。このような、複数のユーザの集めた情報のかたまりを集合知と呼びます」と語るのはNII

学術コンテンツサービス研究開発センター長の武田英明教授だ。「人工知能の研究というのは、元々は個人の知能を対象としたものでした。それが、Webの登場により対象が社会の知能となった。扱うデータは膨大で、常に変化し続けるため、研究者にとってはすごくチャレンジングです。集合知というのは、まさに社会の知能を象徴するものだと言えますね」

「今、ソーシャルメディアの急速な普及により、かつてないスピード感でさまざまなスタイルの人間関係がWeb上で構築されてきています。この人間どうしのつながりと膨大な情報とが絡み合うことで、集合知の研究もますます面白いものとなっていこうと期待しています」と大向准教授は笑顔を見せる。

NIIにおける

セマンティックWebの取り組み

「今まで隠れていた新しい知識を発掘する手段としてはもちろん有効なのですが、長い時間とたくさんの人手をかけて蓄積した情報をセマンティックWebの文脈に置くことも、意味のあることだと考えています。NIIが所持している膨大な

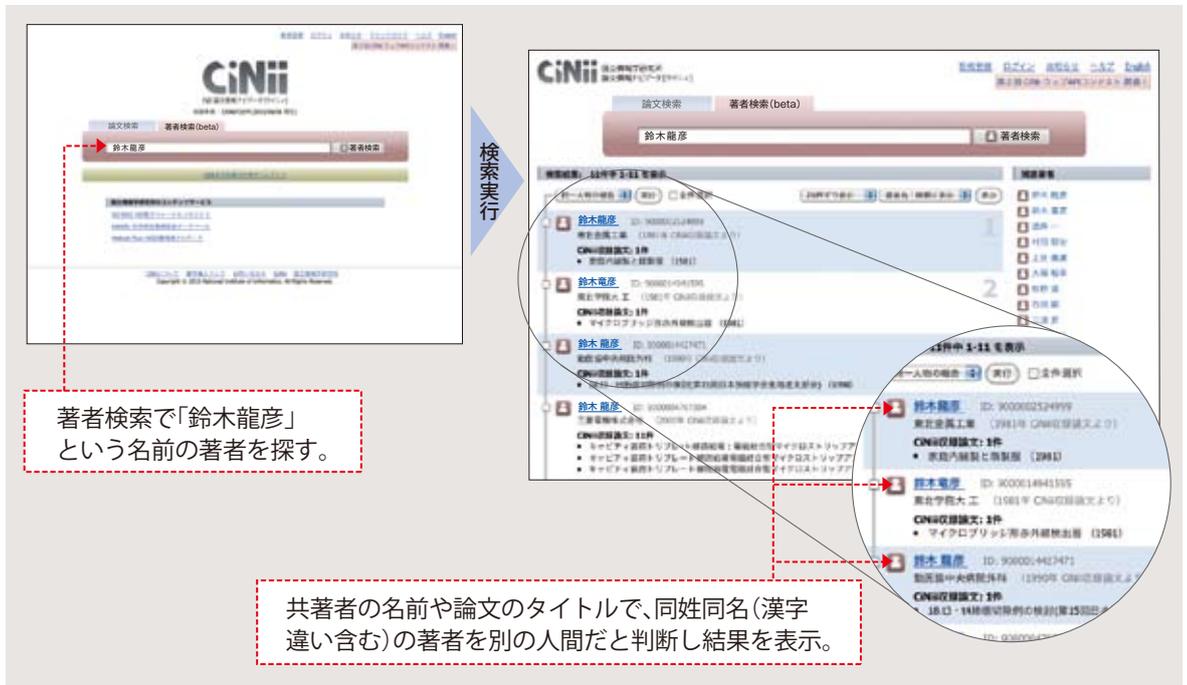


図 論文検索サービス「CiNii」

学術情報を、コンピュータが読み取りやすい形式で流通させることで、今までと違ったかたちでそれらが活用されるのではないかと考えています」と語る大向准教授は、NIIが展開する論文検索サービス「CiNii(サイニイ)」(図)のセマンティックWeb対応を進めている。その第一歩として2010年の4月に、同姓同名の人を区別して検索する機能を公開した。本来名前というのは単なる文字列であるため、同姓同名の人同士を区別することはできないが、論文ならではの特徴を生かしてこの機能を開発したのだという。「論文には、複数の人で1本の論文を書き上げることが多い、タイトルが研究テーマに左右されやすい、という一般の書籍にはない特徴があります。『共著者の名前』『論文のタイトル』という情報と『著者の名前』の関係性をコンピュータに記憶させることで、文字列としては全く同一である同姓同名の方々の名前を区別しているのです」

既存のWebで有用であった膨大な情報を、セマンティックWebの世界でも有効利用するために、前述の2つのアプローチをうまく使った先進的な取り組みと言えるだろう。

また、武田、市瀬、大向の3名全員が関わる「LOD.AC(Linked Open Data for Academia)」も、NIIの代表的なプロ

ジェクトの1つだ。これは、コンピュータが読み取りやすい形式で情報を流通させ、さまざまなデータのつながりをつくることで、検索という行為を「ページを探す」から、「モノやコトを探す」へとステップアップさせようとする取り組みである。例えば、ある画家の作品が所蔵されている施設を網羅的に知りたいと思ったとき、今まではハイパーリンクをひとつひとつ人手でたどっていく必要があったが、「LOD.AC」がうまく機能すると、その画家の作品の所蔵施設の情報を一度に入手することができるのだ。

このように、あるものをキーとした情報のつながりが無数に存在するのが、セマンティックWebの世界の大きな特徴と言える。

言語の壁を飛び越えて 情報をつないでいく

WebやITという言葉聞いて、真っ先にアメリカを思い浮かべる人は少なくないだろう。しかし、セマンティックWebの研究が盛んに行われているのは意外にもヨーロッパだという。「彼らが生活していくためには、いくつもの国が集まってできたEUという共同体を動かしていく必要が

あります。うまく動かしていくためには、言語の壁を越えた情報共有が欠かせないのです。情報の裏にXMLなどで付加された『意味』によって情報のつながりをつくることのできるセマンティックWebは、彼らのニーズにぴったりと合った情報共有の手段なんですね」と市瀬准教授は話す。「日本のような共通の言語でコミュニケーションできる国にいと、『意味』を意識しなくても情報共有が可能なので、『意味』とは何なのか、ピンとはこないと思います。でも、だからこそセマンティックWebの研究に力を入れていく必要があると思います」と話すのは武田教授だ。

政治や経済など、さまざまな分野で今後グローバル化の流れは加速していくだろう。その中で、日本が世界的潮流の中で孤立しないためにも、いつでも最新の情報を受発信できる態勢でいること。それが大前提として必要だ。セマンティックWebの研究は、日本の国際競争力にも大きくかかわっているのである。

広がる セマンティックWebの可能性

情報収集という行為に留まらず、さまざまな分野で活躍する可能性を秘めているセマンティックWeb。最後にそれぞれが考える、セマンティックWebの今後について語ってもらった。武田教授は、生活そのものをWeb上に記録していく「ライフログ」とセマンティックWebの関係に注目しているという。「人間の生活を記録していくデバイスの方は、確実に進歩してきていて、行動を逐一記録していくことは可能です。しかし、何故そのような行動をとったか、というその行動の『意味』もいっしょに記録できてはじめてライフログは有用な情報になると思うのです」。ライフログが、行動の意味といっしょに記録できれば、それはとても画期的なことだ。普通に生活するだけで、日記のような形式の記録が残り、それをWeb上にアップロードすれば、遠くに住んでいる友人がどんな想いを持って何をしているかをリアルタイムに共有することが可能になる。また、遠隔で定期的に高齢者の状態をチェックする方法や犯罪抑止策としても期待できるので

はないだろうか。

市瀬准教授は、コンピュータが自動的にWeb上の情報をつなぎ合わせ、新しい知識を発見するようになると考える。「セマンティックWebの世界では、性質の違うデータがつながるようになってきています。例えば同一の化学物質について書いてある、製薬会社の実験データと、学術機関のデータがあるとします。それら2つのデータが同じことをテーマとしていると判断し、中身を分析し、新薬開発の方法を提示する、という一連の作業をコンピュータが自動的に行うことが可能になると考えています」。カメラロボットと連携して、ロボットが撮影した映像をWeb上で共有できれば、新種の生物を自動的に探索する仕組みをつくることも可能になるかもしれない。

大向准教授は、セマンティックWebの広がりに伴い、人間の側にも求められることがあると考えている。「セマンティックWebが普及するということは、コンピュータが自動的に情報を取ってきて、必要だと判断した部分だけを提示してくれるようになるということです。ただし、それが本当に正しい情報かどうかは、コンピュータは判断してくれません。それは私たち一人一人が判断しなければいけないことなのです。情報収集の仕組みの変化に合わせて、私たち人間が変わっていくことなしには、セマンティックWebの本当の意味での発展は望めないと思います」

「コンピュータの能力」と「情報の記述形式」双方が整うことがセマンティックWebの普及には必要だ。その結果として普及したセマンティックWebが、社会にとってプラスに作用するためには、同様に、人間の成長が必要なのである。

(取材・構成 磯貝 里子)

※1 セマンティック (semantic): 「意味上の」「意味に関する」「語義の」という意味を表す英語。

※2 機械学習: 人間の学習行為と同様の機能をコンピュータで実現させるための技術・手法のことである。データの集合を解析し、規則性を導き出すことが基礎となるため、統計学との関連が非常に深い。

※3 XML、メタデータ: 内容や作成日時、作成者、形式など、そのデータに関する情報をメタデータという。XMLは、メタデータを記述するための言語の1つである。

AI データマイニング が拓く新しい

山田誠二

Seiji Yamada

国立情報学研究所
コンテンツ科学研究系
教授



岡部正幸

Masayuki Okabe

豊橋技術科学大学
情報メディア基盤センター
助教



Web上の情報を検索により収集することは、いまや当たり前だが、一般ユーザにとって、日々増え続けるデータの中から目的に合った情報を取り出すのは容易なことではない。

そこで現在、NIIが中心となって進めているのが、制約クラスタリングという手法を使った

インタラクティブ情報収集・データマイニングの研究だ。

AIの基盤技術を活用することで、情報収集を画期的に刷新する技術とは？

クラスタリングに制約を与え、 欲しい情報を手に入れる

現在、NIIコンテンツ科学研究系の山田誠二教授が中心となって進めているのが、「最小ユーザフィードバック(※)によるインタラクティブ情報収集・データマイニングの枠組み」の研究である。山田教授はこれまで機械にやさしい人をテーマに、ヒューマンインタフェースやヒューマンエージェントインタラクション(NII Today No.44 参照)の研究を手掛けてきたが、自身の研究の中で一貫して主張してきたのが、従来のようにシステム単独の機能向上だけでは限界があるという点だ。

「システムを使う人間に、簡単で有効な支援をしてもらい、人間とシステムが協調作業を行うことではじめて大きなブレークスルーが生まれるだろうと考えています。つまり、ユーザフィードバックをシステムに与えることで、再クラスタリングを行い、よりよい結果を再度ユーザに提示するというわけです。ただし、人の手をできるだけ煩わせずに最大の効果を生み出すというのが本研究の課題です。単に直感的に使い勝手がいいというだけでなく、情報学の理論に裏打ちされたシステムをつくるのが狙いなのです」と山田教授は言う。

ここで鍵を握るのが、「クラスタリング」という技術だ。クラスター(cluster)といえど同種のもの集まりを意味するが、クラスタリングは大量データから知識を取り出すデータマイニングなどに使われている一般的な技術

のことを指す。クラスタリングを使えば、例えば数千枚もあるデジカメで撮った写真の中から、山の風景、人物ポートレートといった具合に、共通の特徴をもつ写真を抽出し、自動的にいくつかのグループに分けることが可能になる。画像に限らず、テキストでも音声でも動画でも、データをベクトルの表現に置き換えることさえできれば、クラスタリングの手法が適用できるという。例えば文章なら、どんな単語がいくつ含まれているかを調べることで、高次元ベクトルとして表現できるというわけだ。

「本研究で採用したのは、このクラスタリングに制約をつけた“制約クラスタリング”です。ただ単にクラスタリングをするのではなく、人が方向付けをしてやることで、クラスタリングの精度を上げるのです(図)。例えば、山と川の写真が別々のグループとして分けられているけれど、自然風景として一つにまとめたい、あるいは犬と猫の写真が動物のグループとして一緒になっているけれど、こちらはそれぞれ分けたいという場合には、それを人間が指定してやる。このときに、人間のフィードバックが最小になるように制約クラスタリングをいかに効率化するか、またどのようなGUI(グラフィカル・ユーザ・インタフェース)であれば人間が制約を与えやすいか、クラスタリングを担当する小野田領域リーダー、GUIを担当する高間准教授、制約クラスタリングのアルゴリズムを担当する岡部助教とチームを組んで、研究を進めています」と山田教授は語る。



小野田 崇

Takashi Onoda

財団法人電力中央研究所
システム技術研究所
領域リーダー



高間康史

Yasufumi Takama

首都大学東京
システムデザイン学部
情報通信システムコース
准教授

制約クラスタリングの 精度を上げる取り組み

そもそもクラスタリングとは、人工知能の機械学習の分野から発展した技術である。機械学習では、アウトプットするものの情報をあらかじめ人間がコンピュータに与えておく「教師あり学習」と、事前にアウトプットすべきものの情報を与えなくても結果を導き出すことができる「教師なし学習」があるが、自動的に似た種類ものを抽出しグループ分けするというクラスタリングとは、すなわち、教師なし学習の基盤技術の一つなのである。ただし、教師なし学習による結果が必ずしも、人間の求める結果にはならない場合があるため、人が制約を加えて軌道修正しようというのが制約クラスタリングであり、いうなれば「半教師あり学習」といえる。

では、具体的にどのようにすれば制約クラスタリング自体の精度を上げることが可能なのだろう。研究チームの一員としてクラスタリングの開発を手掛ける電力中央研究所の小野田崇領域リーダーは言う。

「計算機が自動的に情報を分けると、いつも同じ答えをはじき出します。ところが人が情報を分ける場合は、人によって結果はそれぞれでしょう。そこで人間が制約を与えやすいように、ただ似ているものを見せるのではなく、似ていて、なおかつ離れているものを比較させる。そうすることで、できるだけ少ない負荷で、人間が情報をどう分けたいのかを機械にフィードバックさせるのです。明確な答えがないものを扱うだけに大変難しいですね。今後はもっと人間臭いというか、なぜ機械がこのように分けたのか、人間が直感的にわかるようなクラスタリングの手法を追究していけたらと考えています」

さらに、できるだけ少ない制約で最大の効果を生むためのアルゴリズムの開発を担当するのが、豊橋技術科学大学の岡部正幸助教だ。

「人間の手による制約の数を減らすために、クラスタリング精度を最も向上させることが

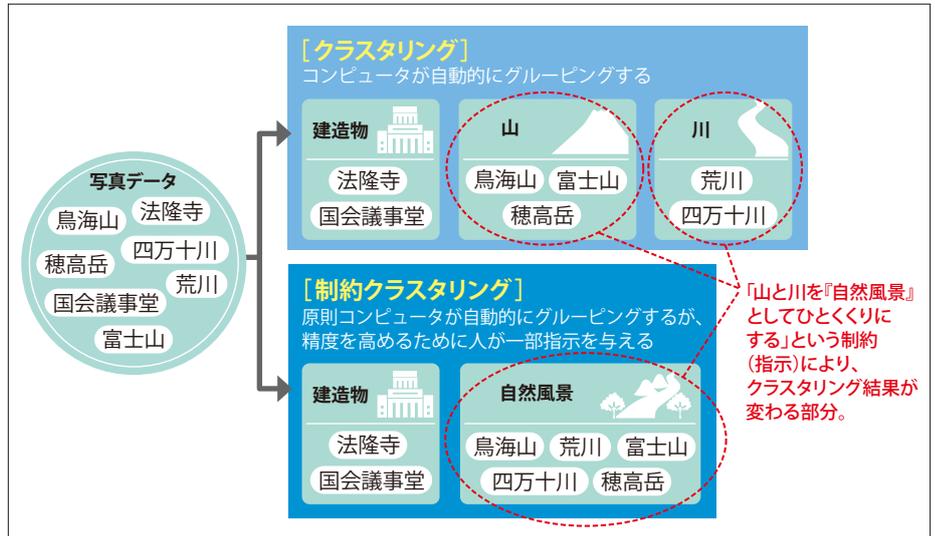


図 クラスタリングと制約クラスタリングの違い

期待できる制約を選択するアルゴリズムや、擬似的に制約を拡張させるアルゴリズムを開発するなど、最小の制約で最大の効果を生むための目的関数を導き出すのが私の役割であり、このアルゴリズムが性能的にどれだけ優れているかを実験的に示すことで定量的な評価も行っています」

人間が比較・判断しやすい GUIを開発する

本研究におけるもう一つの重要な柱が、人間にやさしいGUIの研究だ。GUIを担当した首都大学東京の高間康史准教授は言う。

「クラスタリング結果へ制約を与えるには、人間にコンピュータ画面上に二つのペアを見せて比較してもらい、これを同じグループにしたいか、別のグループにしたいかを選んで入力してもらうわけですが、その際にできるだけユーザの負荷を減らし、なおかつ正しく選べるようなGUIの開発に取り組んでいます。例えば文章を比較してもらう場合、あまり長いテキストだと判断に時間がかかってしまいますが、一方で、割愛しすぎると判断に必要な情報が足りなくなってしまう可能性がある。また、正しく分けようとするれば、いくつものペア

を見てもらう必要があり、負荷が大きくなってしまいます。そこで、どれくらいの情報を見せれば、最適に判断できるのか、実験を通じて検証しています。例えば、長い文章、キーワード、あるいはスニペット(数行の要約文)によって判断時間にどれくらいの違いが出るか、あるいは視線追跡調査を取り入れ、判断の際に、ユーザが何を見ているのかを検証します。その結果、同じ話題に関する文章を比較する場合は二つの文章に共通するキーワードを長く見て、違う話題の文章の場合は一方の文章にしか含まれないキーワードを長く見るとか、通常の検索エンジンで使われているスニペットが我々の実験でも有効である、といった面白い実験結果が始まっています。どうすれば人の感性に届くGUIを開発できるのか、得られた知見をさらに体系化していきたいですね」

今後はインタラクションのコストを抑え、より汎用的なシステムとして開発をしていきたいという山田教授。膨大な情報の海から、求めている情報を瞬時に手に入れるシステムは、AIの基盤技術や応用数学など、情報学の英知の結集によってのみ現実のものとなるのだ。(取材・構成＝田井中麻都佳)

※ユーザフィードバック: 情報を扱う人間(ユーザ)が、コンピュータに対して与えるフィードバック。

人工知能が 裁判過程を推論する

現在、司法の世界では、裁判の迅速化が求められている。

そんな中、民事訴訟のプロセスの一部に論理プログラミングを
応用することで、裁判の効率化を図ろうという研究が行われている。

裁判や法律文書の記述では、生身の人間が意思疎通に用いる文章や言葉が使われ、
人工知能や論理プログラミングとは縁遠いという印象を受けるが、
実は多くの共通点があるという。

研究の現状の成果、そして今後の可能性について、
情報学プリンシプル研究系の佐藤健教授に話を聞いた。

自然言語ゆえの曖昧さを、 論理プログラミングによって回避

論理プログラミングを簡単に説明すると、
論理を数学によって研究する「数理論理学」を
コンピュータ上に持ち込んだもので、人工知
能の記述言語として用いられている。NII情報
学プリンシプル研究系の佐藤健教授は、「これ
まで法律や裁判の世界では、自然言語だけ
ですべての事象に対応してきました。これにシ
ンボリックな記号や数式などを用いた論理学

的な視点を導入し、論理プログラ
ミングを応用することで、新たな
知見が得られ、実際の裁判にも寄
与できるのではないかと考えま
した」と話す。

法律学に論理学的な解析を持
ち込む利点は、大きく3つ挙げられ
る。1点目は、法律文書の中に埋も
れている暗黙の仮定を明らかに
することができる

点だ。論理学的な考えでは、すべての仮定を
明らかにしないと結果が出てこない。しかし、
法律は自然言語で書かれているため、機械的
に導かれない事柄がある。これが明示できる。
2点目は、法律文書を論理学的に解析し、論理
式のような厳密な形で表現できるようになると、
法律概念自体をコンピュータ上で扱うこ

とができるようになる。これにより、さらに複
雑な法的問題の解決や、法律における矛盾の
検出が実現できるようになる。3点目は、論理
学的な問題と法学的な問題を区別できると、
法律の専門家は法的な問題に集中できるよう
になることだ。

つまり、法律学に論理学的手法を持ち込む
と、これまで不明瞭だった箇所がはっきりと
浮き彫りにされ、かつ煩雑だった作業の効率
化などが期待できる。

主張を証明するプロセス 「要件事実論」

そして現在佐藤教授が研究しているのが、
民事訴訟への論理プログラミングの展開だ。

民事訴訟では、裁判官は原告と被告の主張
に対し、民法や商法など「実体法」と呼ばれる
法律を参照して判決を下している。実体法の
条文では、ある事実に対して特定の法律効果
が生じるといった規定をしているが、誰がど
のような事実を立証すれば「特定の法律効果
がある・なし」が言えるかまでは規定してい
ない。例えば、品物の購入に際し代金が未払い
の場合、代金を支払って欲しい人は売買契約が
あったことを積極的に証明しようとするが、
誰のどのような主張を通せば証明が成立す
るか、ということまでは明文化されていない
のだ。このような紛争に対して、原告と被告の
どちらが証明すべきなのか、どのように裁判
を結論づけていくのかなど、裁判官がガイド
ラインとして利用しているのが要件事実論で
ある。佐藤教授は、「この要件事実論のプロセ
スが、これまで私が研究してきた非単調推論
(※)に酷似した点があることに気づき、論理
プログラミングの開発意義を認識しました」と
いう。



佐藤 健

Ken Satoh

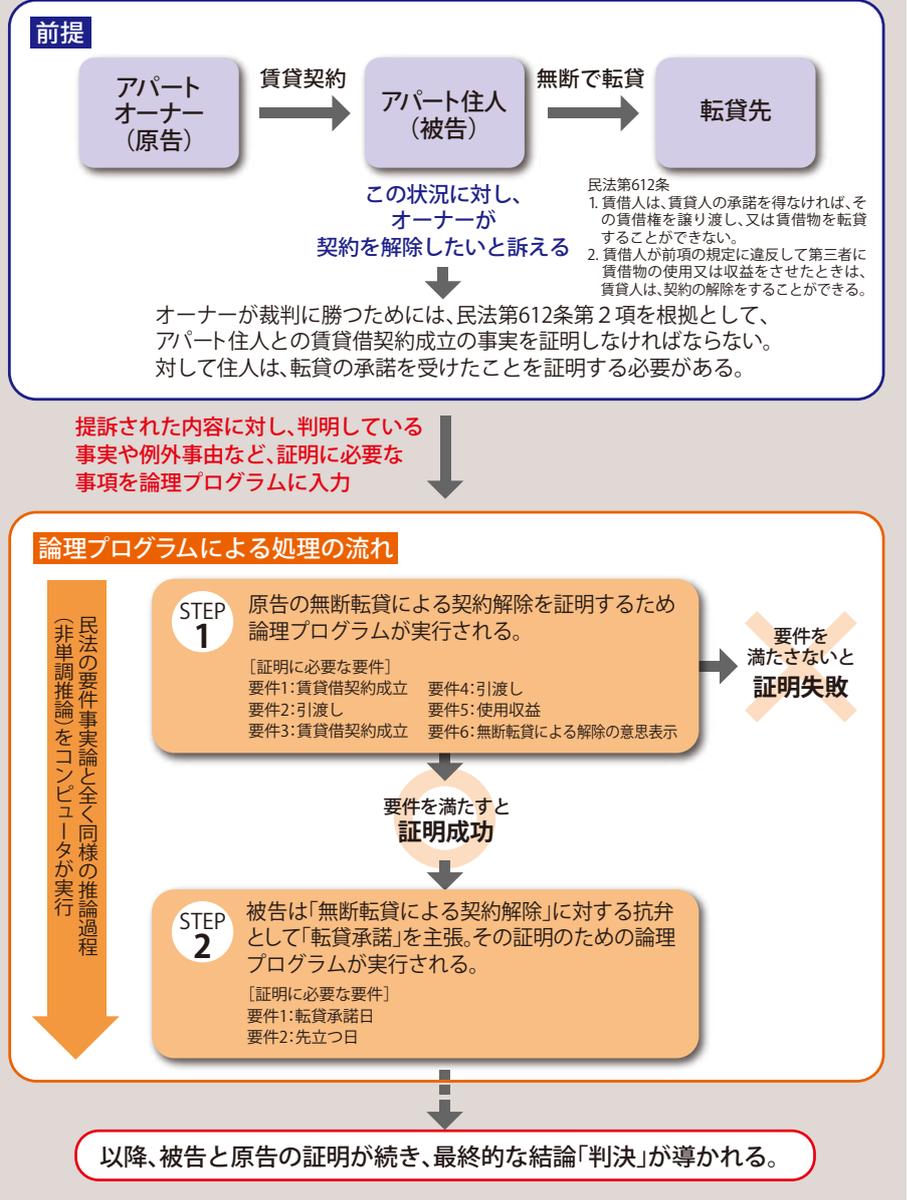
情報学プリンシプル研究系
教授

裁判の不確定な事柄を人工知能がシミュレート

論理学の世界でいう“非単調性”とは情報が付け加わると前の結論が撤回される可能性があることをいう。多くの形式論理はこれとは逆の“単調性”であり、新たな情報が追加されても既にある情報が減ることはない。一方、非単調性は、不完全な情報下での物事の仮定を設定しておく。例えば、“自動車は道路を走る”という仮定があったとする。これに何も追加されない場合、“自動車は道路を走る”ものとして推測されるが、後でこの自動車が“F1などのレース専用車”と判明した場合は「道路を走る」という結論は撤回される、といった具合だ。このように、不完全な情報しかない時にどのような推論を展開するか、理論的に解説するのが非単調推論である。

その意味で、裁判とは原則的に「非単調」である。もし、裁判が単調性ならば、一度勝訴したらずっと結果は変わらない。しかし実際の裁判では、第一審で得られた証拠を提出して勝訴したとしても、控訴審で新たに有力な証拠が出てきたら敗訴する可能性があり、結論は非単調に変化する。佐藤教授は、さきほどの要件事実論が不完全情報下における合理的な推論の定式化であり、非単調推論が応用できると考え、図のような要件事実論の論理プログラムを作成した。

図は、無断転貸解除の裁判についてのシミュレーションである。あるアパートのオーナーが、「賃貸契約をしている住人が、無断で第三者に部屋を転貸しているので、契約を解除したい」と求めている。この時、オーナーは正規の手続きに則って住人に部屋を引き渡したことや住人が第三者に部屋を貸したことを証明しなければならないが、転貸は認めていないことは証明する必要がなく、住人の方が転貸の承諾をオーナーから受けたことについて証明する責任がある。この要件事実論と同じ推論過程をコンピュータに行わせるのが、非単調推論に基づく論理プログラムである。これまで判明している事実や過去の判例、例外などの必要事項を記述すると、結論つまり契約解除できるか否かが導き出される仕組みになっている。



“Juris-informatics”という新しい学問分野の創成を目指して

佐藤教授は法律学と情報学を組み合わせた“Juris-informatics”という新たな学問分野を創成することを目標に掲げている。その一方で、「法律や裁判のすべてに論理プログラミングが適用できるわけではありません」と指摘する。例えば刑事訴訟では、殺意について、どのようなかたちで事件が起こったのか、凶器はどのように使用されたのか、事件後、被告人はどのように行動したか、などを総合的に判断して認定し、判決を導く必要がある。こうした事実認定、さらには人間の常識による解釈などが入ってくると、論理プログラミングだけでは最適な答えを出すことは不可能であり、人間の洞察や判断に頼らざるをえない。

つまり裁判の過程には、どうしても人間にしか判断できない部分が存在する。それを踏まえると、現時点におけるJuris-informaticsの研究は、「裁判のすべてをゆだねられる人工知能を創る」ことではなく、むしろ、「自然言語では難しいロジカルな作業への応用」にベクトルを合わせているといえる。人間とコンピュータの棲み分けをはっきりさせることによって、法律における論理プログラミングおよび人工知能は、研究の明確なゴールを得ることができるのである。
 (取材・構成 森本淳一)

※非単調推論: 推論を追加することでこれまでの結論や結果が変わる論理。例えば、裁判では新たな証拠が見つかった場合、一審と二審で判決内容が変わってくる。これに対し、単調性は新たな論理式を追加しても結論や結果には変化はない。

感動の共有—プライバシーも共有？

越前 功

(国立情報学研究所 コンテンツ科学研究系 准教授)



久しぶりに休暇を取って家族で旅行に出かけた。ホテルはオーシャンビューで眺めは最高。この感動を皆に伝えたい。よし、Twitterでつぶやこう。ホテルの名前も教えちゃえ…。

Web上の旅行記が「留守」のサインに

携帯端末とネットワークの普及により、このような感動の共有がリアルタイムで実現可能な便利な社会になった。しかし一方で、TwitterやSNSの利用者の位置情報や移動経路などのプライバシー情報が第三者によって不当に把握されることで、利用者が不利益を被ることが社会問題となつている。米国では、休暇を取って旅行に出ていることをTwitterで公開したところ、自宅に置いてあった数千ドルを盗まれた事例がある。また、位置情報の解析を通じて個人情報推測できる可能性も報告されている。Rutgers UniversityのHohlfeldは、65人のドライバーの車に装着したGPSを被験者に1週間観測させたところ、85%のドライバーについて、自宅らしき場所を発見できたという。さらに、TwitterやSNSの利用者のプライバシーに対する意識が低いことも、この問題に拍車をかけている。USA TODAYの記事によると、長期休暇で旅行に出た米国人の多くは、旅行中に自宅への泥棒の侵入を恐れて、留守電話の設定を意図的に解除したり、新聞や郵便物の送付を停止したりしているが、旅行中のSNSの利用に対しては警戒心が薄

く、自宅から何百マイルも離れた旅行先の位置情報や写真をリアルタイムで掲載したり、『1週間に1週間滞在の予定です』などといった情報を安易に投稿してしまう問題があると専門家が指摘している。

安心して「感動を共有」できる技術とは？

先に述べた問題の対策として、プライバシー保護技術が注目されている。プライバシー保護技術の研究では、情報提供者のプライバシー情報を第三者に過度に取得されないようにすることが重要である。社会からの要請もあり、1980年代後半頃から情報セキュリティやデータ工学などの様々な分野の研究者が分野の枠を超えて取り組んでいる。プライバシー保護技術の研究の多くは、情報の内容を「あいまい」にすることや、偽の「情報を加えること」で、第三者によるプライバシー情報の取得を困難にすることに主眼が置かれている。一方で、ほかされた情報の品質がどのように劣化し、受け手にどのような影響を与えるかについての研究は未だ少ないようである。

TwitterやSNSにより不特定多数と感動や話題を共有し、コミュニケーションを図ることが当たり前になった今日、プライバシーを気にするあまり、感動や話題を共有できなくなってしまうのは本末転倒である。情報の品質を維持しながら一定のプライバシー保護を実現する仕組みが求められている。

今月の表紙イラスト：知育玩具で遊ぶ幼児と知能ロボット。知能ロボットも、自分と実世界(社会)と関わりながら、徐々に学び、成長する。このような自立した存在としてのロボットが人と共生する社会はいつ訪れるであろうか。

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49号 平成22年9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：谷口弘幸 制作：株式会社 商業デザインセンター

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:03-4212-2131 FAX:03-4212-2150 e-mail:kouhou@nii.ac.jp