

学生フォーラム AI Inter-View

第47回 奥村 学氏インタビュー 「自分の感覚を信じて」

今回の学生フォーラムでは、東京工業大学精密工学研究所の奥村 学准教授にインタビューを行った。言語処理の分野で活躍する氏は、blogWatcherやPRESRIなど、研究成果を積極的にシステムとして公開している。そうした氏の研究室の運営理念や、今後のWeb研究に対するお考えなど、さまざまなお話を伺うことができた。

1. 直感で行き着いた言語処理研究

京都に生まれた奥村氏は、京都大学教員の父親をもち、卒業生の多くが京都大学に進学する高校を卒業した。京都大学への進学が予想される経歴であったが、氏は東京工業大学5類（電気・情報系）へ入学する。「京都にずっといるのも何だなと思って」というのがその理由だった。学部選びの基準も、コンピュータサイエンスは何かおもしろいことができそうだ、という漠然としたものだった。氏は型にはまった考え方が苦手だったのだ。そうして選んだ学部であったが、実習の授業におもしろさこそ感じたものの、ガリガリとプログラムを書いてきっちりしたものをつくる、という作業は性に合っていなかったと氏は振り返る。

そんな折、氏に転機が訪れた。学部3年次に米澤明憲教授（現 東京大学教授）の講義を受けてAIに興味をもち、AIに関する書籍を読んでいく中で言語処理の研究があることを知った。そして、言語を使ったコミュニケーションという曖昧さのある研究こそ、自分に合っていると思えたそうだ。4年次の卒業研究配属時に田中穂積教授（現 中京大学教授）の研究室に入った氏は、ここで言語処理の研究を始めたのである。卒業論文での構文解析に始まり、修士で意味解析、博士で文脈解析と、言語的な曖昧性を解消するためによりセマンティックな情報を入れる必要性を感じながら研究を進めていった。研究室として新世代コンピュータ開発機構（ICOT）の第五世代コンピュータプロジェクトに携わっていたこともあり、企業や他大学の研究者らとともに、最先端の技術に触れながら有意義に研究を進めることができた。

ところが、氏は当時の研究を「博士修了後も課題は多く、現在からすれば古典的な言語処理に過ぎなかった」と振り返る。当時のAIの価値観は、人間と同じことを計算機にやらせようという「ハードAI」であり、言語処理においても人間のもつ言語理解モデルをそのまま計算機に載せたい、という風潮があった。しかし、現在ではそうした手法は不可能と判断され、より現実的なアプローチに移行しているのである。「古典的な手法が不可能であるという判断の一助となっているとすれば、ある

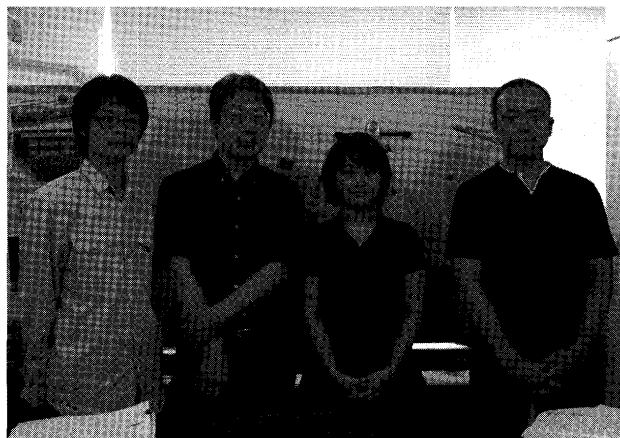


図1 奥村氏を囲んで

意味で現在の言語処理研究に貢献しているといえるのかもしれないのですが」と氏は苦笑する。

2. リアルなコンテンツを対象に

博士課程を修了した奥村氏は、同研究室で3年間の助手を経た後、92年に北陸先端科学技術大学院大学（JAIST）の立上げに携わり、助教授として着任した。学生を指導する立場となってもち始めたのは、「学生も研究者（のタマゴ）である」という考え方であった。もちろん、研究室に入りたての学生は、研究の進め方や技術などの面において研究者として未熟である。しかし、学位を取得する頃にはスキルをもった研究者になってくれることを氏は期待していた。だからこそ、むりやりやらされる研究ではなく、自分でおもしろいことを見つけて頑張ってもらいたいという価値観をもって、氏は研究室を運営していった。学生のアイデアであっても、氏にとってもおもしろいと思えるものは意味のある研究であると判断し、言語処理という立場から逸脱した内容でなければ、研究室として積極的に研究の幅を広げていったのである。

着任から数年間は、それまでの奥村氏の研究同様、言語の意味や文脈の扱いに関する研究などの基礎的な言語処理研究がメインであったが、Webの出現や学生の興味の移り変わりとともに、サービスとしての言語処理研究の扱いや、固有の性質をもったコンテンツへの適用などへと、方向性が徐々に変化していった。そうした中で生まれた実サービスの一つが、PRESRI^{*1}である。これは当時の博士学生であった難波英嗣氏（現 広島市立大

*1 <http://www.presri.com/>

学講師)の提案で生まれた,論文・特許などの学術系コンテンツの検索システムである。当時こそ特に意識したことではなかったが,リアルなコンテンツを扱い,あればうれしい,そんな誰にでも使ってもらえる実用的なシステムをつくっていかうとするきっかけが生まれたのがPRESRIであったと氏は振り返る。

3. Webに散在するコンテンツの編纂

2000年,JAISTの任期を終えた氏は,現在の職場である東京工業大学精密工学研究所へ赴任する。「研究所」という名称が示すとおり,ここでは氏にアカデミックさよりもむしろ実学を志向した研究が求められた。もとより「使えるものをつくりだしていききたい」という価値観のもとで,PRESRIのように学術系コンテンツを題材としたシステムづくりを進めてきた奥村研究室だったが,当時の学生の興味の対象は何よりWebであった。Webを扱った研究が研究室の主流となり,本格化していった。

その最たる例が2003年に開発されたblogWatcher*2である。これも南野朋之氏・藤木稔明氏(ともに現グーグル株式会社)ら博士課程の学生が中心となり,IPAの未踏プロジェクトに応募してつくられたアプリケーションである。そのコンセプトは,大量に発信されるリアルタイム性の強い情報の中から,世の中の人々にとって有用な情報を抽出・提供したいというものであった。日本でブログが話題になり始めたばかりで,RSSフィードを配信するようなブログサービスもCMSもまれであった当時,このシステムは自然言語処理の技術を用いてWeb上に散在する日記を収集し,話題の盛り上がりや商品の評判を分析することに成功した。「ご飯を食べるときにお薦めの店を提示してくれるくらいのものであったが無理だった」と奥村氏は語るが,口コミやトレンドを知りたいという人々の潜在的なニーズを突いたblogWatcherは,新しいアプリケーションとして評価された。プロジェクト審査員を驚かせるようなものをつくらうという強い意気込みも,クローリングやデータベースの扱いなど慣れない開発の苦勞を乗り越える原動力になったという。

blogWatcherは実際にWeb上に公開され,反響を呼んだ。Webは比較的成本をかけずに研究成果を実際のサービスとして提供できる環境であるため,公開することは自然の流れだった。メンテナンスが大変であったり,「内輪での閲覧のために公開しているブログだから情報を使わないで欲しい」といったクレームが来たりするなど苦勞もあったが,ユーザのニーズを拾うことができ,「すごい!」といった声を聞くことができたことは,研究開発を進めるうえで有意義なことであったと氏は語る。

その後もblogWatcherは研究室の中心的なテーマとなり,開発が進められている。例えば評判情報抽出の仕組

みにおいても,情報発信する側の信頼性や,抽出された情報の提示を受ける側の知識の度合いによってその意味合いは全く異なってしまうため,ブログの書き手や利用者のプロファイリングは重要な課題である。また,ブログとマスメディアとの間の相互の影響を見いだすことによって情報の重要度を測ることもできる。このような考えから氏は,人々の関心をより理解し,アプリケーションとしての価値を高めようという努力を続けている。そして現在,blogWatcherは名前をSHOOTI*3と改め,事業化されている。ビジネスにしたことで豊富なニーズの情報を開発の参考にできる利点がある反面,自然と実用性の部分ばかりに気をとられ,研究としての深みが失われてしまうことを氏は危惧している。企業の人と接する際には役割分担を明確にし,研究者としての立ち位置をしっかりと保つことを心懸けているようだ。ただし,学生の場合はその限りではないと氏は付け加えた。未踏プロジェクトやインターンなど,さまざまな人との共同作業は,目に見えない形であれ確実に将来役立つ刺激が得られるので,積極的に参加して欲しいと氏は強調する。

4. コミュニケーションの支援に向けて

広大なWebには,多くの人々によってつくられた玉石混交のコンテンツが散在する。自然言語処理というツールを用いてこれらを編纂し,信頼性や価値のあるコンテンツを再生産するというのが,blogWatcherの開発以降,氏の研究の中心になってきた。そんな氏が今後の研究対象として興味をもっているものの一つが,「ケータイ小説」のようなUGC(User Generated Contents)だ。近年はWikipediaに代表されるように,Web上の情報もそれ自身で整理されるしくみをもつものが増えてきた。そんな中であって,どこにあるかわからない,そもそも存在しないかもしれないおもしろい情報を見つけることにこそ興味があると氏は語る。

もう一つの興味として氏は,対話に代わるチャットのようなコミュニケーションをあげた。blog上で行われているようなユーザ間での価値のある情報のやり取りを,より即時的なシステム上で実現できればおもしろいのではないかと氏は考えている。その際に,コミュニケーションをスムーズに実現できるようにしくみを構築することが氏の目標だ。「情報の価値や信憑性というものはある種の評価。同様にコミュニケーションの仕方を評価する研究があってもよいと思う」と氏は語る。「炎上」と呼ばれるコミュニケーションの暴走の原因として,文字情報しかないことがあげられるが,文字情報だけのやり取りでも,やり方を工夫すれば炎上は防げると氏は考えている。想定しているのは「こういうコミュニケーションは危険だ」と,教示するようなシステムである。その際には,やはり自然言語処理を用いて情報の価値や信頼

*2 <http://blogwatcher.pi.titech.ac.jp/>

*3 <http://shooti.jp/>

性などに尺度付けし、コミュニケーションの仕方を評価するといった手法が有効であると氏は考える。要約システムなど、利用の有効性という側面から評価できる従来の評価とは異なり、今まで評価されなかったものを評価するという新たな取り組みだ。「対話に関する研究にこそ、言語処理の本質があるんじゃないかな」と氏は語る。

5. 若手研究者へのメッセージ

最後に、奥村氏より若手研究者へのメッセージをいただいた。氏はまず、自分がやりたいことを頑張ってやってほしい、とおっしゃった。自分のやりたいことを見つけることは大変である。しかし、自分がおもしろいと思えるものをずっとやっていけるのは研究者として幸せなことだ、と氏は考えている。自身が学生だった頃と比べ、最近の学生は物事を合理的に考えるあまり、ときに慎重になりすぎているように氏には感じられるそう。選択肢の中でどれが一番なのかを悩んでしまうかもしれないが、おもしろいものは一人一人に多くあり、その中から直感を信じて素直に選べば、それぞれにおもしろみがあるということだ。氏の場合はそれが言語処理であり、研究の軸となってきた。

また、合理的という点でもう一つ、時間効率についても、そればかりを追い求めないことも大事だと氏は語られた。長い人生の中ではどんな体験も必ず自分の役に立つので、多くのことにチャレンジすべきだということである。例えば未踏プロジェクトや海外でのインターンなどは、それ自体が研究成果に直結しないかもしれないが、積極的に挑戦してほしいとおっしゃった。氏自身、ICOT のプロジェクトに始まり、JAIST 所属時にはトロント大学のサバティカル研修を経験し、また指導者として **blogWatcher** を開発した未踏プロジェクトなどに携わってきた。これらは、システムの開発能力や人脈を得られるほか、異なる価値観をもった人々との交流を通じてより良い研究スタイルを発見できる機会となるなど、貴重な経験であったと述懐された。短期的には回り道のように見えることも、いつかきっと自分の身になってくるものなのだ。

奥村氏が感じているように、我々学生は研究というものに対して神経質になりすぎているかもしれない。失敗を恐れずにさまざまな経験を積み、それを活かして息の長い研究をやっていきいたいと思う。

〔古川 忠延, 亀田 堯宙 (東京大学)〕