



# 自然言語処理がつなぐデータと社会

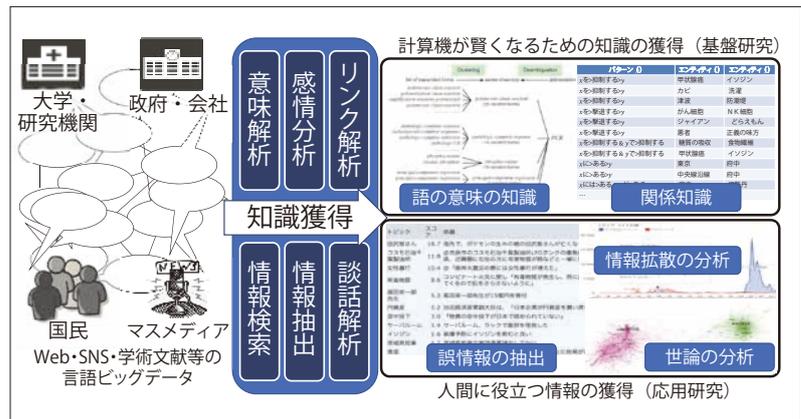
【受賞タイトル】 言語データの意味的解析による知識獲得と社会観測に関する研究

岡崎直観 東北大学大学院情報科学研究科

このたび、マイクロソフト情報学研究賞という栄誉ある賞をいただくことになりまして、とても嬉しく思います。この受賞は、コンピュータが大量の言語データを解析し、コンピュータが賢くなるための知識の獲得（基盤研究）、および人間にとって役に立つ情報の獲得（応用研究）に関するものです。

子供の頃からプログラミングが大好きだった私は、コンピュータが役に立つ場面をもっと増やしたいと思い、学部の卒論配属で人工知能に関する研究室を選びました。その当時はWebの普及が急速に進み、自然言語で書かれた情報が氾濫し始めた頃でした。幸運にも、博士課程在籍中に渡英する機会に恵まれ、生命医学分野の学術論文に対するテキストマイニングの研究・開発を行いました。生命医学分野では、さまざまな分野の研究者が最先端の研究成果を報告するため、曖昧性（同じ表現が異なる実体・概念を指す現象）や多様性（同じ実体・概念が異なる表現で言及される現象）が深刻な問題になっていました。そこで、用語の意味目録を自動的に獲得したり、用語の曖昧性解消（意味の推定）、用語の柔軟なマッチング（類似文字列検索）の研究などに取り組みました。現在は、記号処理の弱点を克服するため、深層ニューラルネットワークに基づく知識獲得の研究を進めています。

応用研究では、ソーシャルメディア上の投稿などの言語データを自然言語処理技術で分析し、社会や世論の動向を「観測」する試みをメディア企業等と共同で進めました。これまでに、震災ときに流通した誤情報の拡散収束過程、東京電力福島第一原子力発電所事故による農作物の風評被害、国政選挙における政策や候補者に対する意見、サッカーW杯における日本チーム



に対する声援など、さまざまなテーマの分析を経験できました。データ駆動型ジャーナリズムなど、自然言語処理の新しい応用を見出す一方で、精度の高い分析を行うには、応用タスクごとに訓練データを作成するという戦略に頼らざるを得ません。基盤研究で世の中の常識的な知識を自動的に獲得・蓄積しておき、応用研究で未知のタスクが与えられたときでも、そのタスクの訓練データに依存することなく、蓄積した知識を活用してタスクを解けるようにして、基盤研究と応用研究の溝を埋めていきたいと考えています。

最後になりましたが、私が研究を楽しく継続できたのも、これまでご指導していただいた先生方、研究室の学生やスタッフ、研究プロジェクトの領域総括・アドバイザーの先生方、共同研究者との出会いに恵まれたからだと考えています。また、大学の研究科や学科、研究室の充実した教育・研究環境、および家族のサポートの賜物でもあります。この場をお借りして、感謝申し上げます。

(2017年5月17日受付)

岡崎直観（正会員）okazaki@ecei.tohoku.ac.jp

2007年東京大学大学院情報理工学系研究科博士課程修了。同大学院の特任研究員を経て、2011年より東北大学大学院情報科学研究科准教授。自然言語処理、テキストマイニング、機械学習の研究に従事。