

リンクマイニングによる研究者ネットワークの抽出： 成長プロセスと国内外からの見え方

Social Network Extraction among Researchers by Link Mining:
Its Growth and Characteristics from International and Domestic Views

安田 雪^{*1}
Yuki Yasuda

松尾 豊^{*2}
Yutaka Matsuo

武田 英明^{*3*1}
Hideaki Takeda

^{*1} 東京大学
University of Tokyo

^{*2} 産業技術総合研究所
AIST

^{*3} 国立情報学研究所
National Institute of Informatics

This paper describes analyses on a researcher network extracted from the Web: we extract a network of researchers using a search engine. By putting two names of researchers (in Japanese and in English), we can get two different kinds of networks among researchers. One is a network which represents the collaborations among researchers in domestic, and another is a network of the collaborations internationally. We compare the two networks and find that there are several researchers who work as a bridge between the domestic and international communities. We discuss the roles of researchers as well as the other findings from the analyses.

1. はじめに

本稿の目的は、人工知能分野における研究者について、Web マイニング技術によって得られる研究者ネットワークの構造を明らかにすることである。具体的には、Web の上の単語の共起関係から単語間の関係構造を抽出する、松尾らの方法[松尾 05]を用いて、研究者のネットワークを抽出する。

本報告は、2003 年から継続的に行われている「イベント空間情報支援プロジェクト」の一環として、研究者の創造的な関係形成支援を目的とする研究のひとつと位置づけられる。我々は、研究者同士の結びつきは、学問の発展に貢献し、地理・世代・研究組織など、多様な次元で隔たりをもつ研究者をつなぎ、その関係維持を支援するシステムは、当該領域の学問の発展のための重要なプラットフォームになると考える。

局所的には当然だと見なされている知識も、異なる場にもたらされれば貴重な新知識でありうる。異なる場に存在する局所的知識と融合することで、一部では既知の一般的な知識も、新たな知識を創発させる可能性がある。この可能性は、学会、研究機関や国といった制度的枠組みを超えて、知識が連結することの重要性を示唆している。

このような知識連鎖を支えるのが、学術誌などのメディアとともに研究者の人的交流である。研究者同士の知人関係や協働関係に関する情報は、一部の共著関係や師弟関係をのぞき、原則的には他人が入手・理解することが難しい。また、研究者自身も自分の共著関係や師弟関係を把握していても、その共著相手の共著関係や、自分の師弟の師弟関係を完全に把握しているわけではない。研究者の関係情報は推移性をもたない情報であり、知人・友人関係などと同様、研究者も有効活用すべき関係情報を、認知制約のため活用できない可能性が高い。上記の松尾らの技術は、人間の関係の認知制約に挑むものと見なされる。

以前の報告[安田 04]では、研究者の集団が「見えざる大学」(invisible colleges)[Burt 82]を形成していること、Web マイニング

で得られたネットワークが研究者のパフォーマンス(論文数や引用数)と相関があることを確認した。さらに、[安田 05]では、先端的な研究を担うコアの研究者グループ内では紐帯の数の偏りが年々大きくなり競争が起こっていること、また媒介中心性が次年度以降の固有ベクトル中心性と相関が高く、複数の島をつなぐ研究者が中心的な役割を担うようになる傾向があることを述べた。昨年の報告[安田 06]では、ネットワークの時系列変化を調べ、中心的な人物は年々異なり、高い中心性を維持することは難しいこと、コミュニティを形成するトピックが移り変わっていること、それが文脈の共有が進むことに原因があり、知識の体系化、構造化につながっていくことを述べた。

このような研究者同士の共著関係や協働関係を抽出し、その連鎖の状態を研究者集団の共有知識にすることは、研究者の情報交換のみならず、研究者集団の知識の共有化にも貢献すると、我々は考えている。そのためには、現状の研究者間ネットワークを把握すること、そして、関の欠損部分を架橋する関係の形成・支援に役立つツールの構築が重要であろう。

リンクマイニングの技術により、日本語の氏名の共起関係を用いて複数の時点における我が国の AI 研究者の連結構造が抽出された現時点では、この AI 研究者ネットワークが国際的にはいかなる様相を呈しているのかについて、探索的な分析が必要であると考えられる。

本研究では、海外、ことに英語を使用言語とする研究者集団からは、我が国の AI 研究者群がどのように見えているのか、すなわち WWW 上で発信される英語情報に限定すると、日本の AI 研究者ネットワークはいかに構成されているのか、そして海外の AI 研究者と国内の AI 研究者はどのような連鎖の構造を持っているのかを明らかにしたい。学術の世界においても情報流通、コミュニケーションといった面でグローバル化が進む中で、国内外の関係とそのあり方を検討することは重要な課題であろう。

研究者は、日本あるいは海外の大学院で研究キャリアを開始し、その地をホームポジションとして、当地の学会への参加と報告を始め、その後、海外の学会への参加、報告へと段階的に活動領域を広げていく。

研究業績の蓄積にともない、大多数は、国内外の大学や研究機関に就職をし、そこを新たな研究拠点とする。就職後は、

連絡先: 安田 雪、東京大学ものづくり経営研究センター、東京都文京区本郷 3-34-3 本郷第一ビル 8F, 03-5842-5501, yasuda@mmerc.e.u-tokyo.ac.jp

研究以外にも、大学であれば教育活動、企業組織であれば管理・運営業務が発生するが、この時期には、研究予算が獲得しやすくなるなど、経済的に安定する。

研究者のキャリアにおいて、研究拠点である所属機関の所在地において、研究グループや学会の運営活動を担うことが要請されるのは、この段階である。一部の研究者は、これ以降、学会関係者により多大な期待を抱かれ、任務を課されることになり、研究時間の多くを国内の教育・学会活動に捧げることになる。このような研究者は、研究とともに、当地の教育、組織運営、学会活動の基幹人物となっていく。

粗手に対して、一部の研究者は、研究キャリアの一時期に活動の拠点を海外の大学や研究機関に移し、国際的な研究活動を展開する。この地理的移動は、海外の研究者とのネットワーク形成、拡大の機縁となる。人的ネットワークのみならず、研究テーマ、分析手法、研究動向なども含めて、海外の研究者との接触は、多様な知的ネットワークの拡大をもたらす。これらの知見は、海外と国内とを行き来する者によって、わが国を研究拠点にする人々にも豊かな刺激をもたらす。我々はこのような、異なる研究者集団間で、知識を仲介する人々の重要性を指摘したい。

もちろん、日本以外の国の大学や研究期間に就職し、数十年にわたる長期的所属状態を保ちつつ、教育・研究の最先端を進めていく者もいる。彼らは国際的な知識の仲介者として研究活動を行っているというよりも、海外で研究を行っているという解釈すべきであり、我々は、むしろ国内外の研究活動のインターフェースとなり、研究者間の空隙をつなぐ国際的な研究活動者に焦点をあてたいと考えている。

以下では、英語の氏名情報と用いてマイニングした、わが国の AI 研究者のネットワークと、海外の AI 研究者のネットワーク構造を分析し、双方がいかなる特徴や連鎖の構造をもつのか、また、本領域において国内外の知識の連鎖を担うネットワークが持つ問題点を検討したい。

2. 研究方法と抽出データ

近年、さまざまな種類のデータ(特に関係データ)が利用可能になっており、特に、関係性をもったデータに関するマイニング技術である、リンクマイニング[Getoor06]の研究が活発に行われている。Web マイニングによるネットワーク抽出は、関係データを取得する技術であるが、こういったリンクマイニングの技術のひとつであると考えられる。関係情報は基本的に不可視であり、その抽出には困難が伴う。とりわけ人的ネットワークの関係抽出では、ネットワークを構成する個々に多大な負担をかけざるを得ず、個人情報保護やプライバシー問題など、法的・倫理的課題が立ちだかるため、データ抽出の困難が指摘されている。そのため、電子的記録媒体からの関係抽出など、リンクマイニング技術には大きな期待が寄せられている[安田 07]。

本稿では、まず、JSAI2007 の著者・共著者約 750 人についての Web 上の氏名の共起関係から得たネットワークを抽出する。具体的には、2人の氏名を AND でつなぎ、検索エンジンのクエリーとしてヒット件数を測定する。このとき 750 人ではネットワークの視覚化がやや難しいため、750 人のうちから、「氏名 所属」で検索エンジンのヒット件数を調べ、その上位 100 名を選んだ。共起関係の強さは、Simpson 係数[Manning02]を用いる。

図1がそのネットワーク図である。[安田 06]で述べたように、ネットワークは時間的に変化しており、JSAI2007 で中心的な人物、クリークは以前のものとは変化している。例えば、この 100 名のうち 27 名は、JSAI2003-JSAI2006 の 4年間に参加したコアの 90 名[安田 05]に含まれていない。また、コアの 90 名のうち、29 名

は JSAI2007 の著者・共著者に含まれていない。したがって、JSAI2007 でもネットワークが時間的な変化をしていることが確認できる。

次に、研究者の国際的な場での協働関係を見るために、英語名での共起関係を調べたのが図2である。このとき、「松尾豊」と「武田英明」のつながりを調べるのであれば、「Yutaka Matsuo Hideaki Takeda」と AND でつないだものをクエリーとして、ヒット件数を求める。図2は、図1のネットワークと異なっている。図1では比較的多くのノードがコンポネントに含まれているのに対し、図2では、中心的なノードにエッジが集中している。(つまり centralization が高い。)図1で中心的な人物と図2で中心的な人物は一部、異なる。日本語名でのネットワークにおいて中心的な研究者は、研究コミュニティをまとめ、方向付けていく上で重要な役割を果たしている研究者である。また、英語名でのネットワークにおいて中心的な人物は、日本の研究を国際的にアピールする上で、中心的な役割を果たしている研究者であると考えられる。

しかし、日本人研究者の名前を英語名にしてネットワークを抽出しただけでは、実際に国際的なコミュニティで活躍しているかどうかを知ることは難しい。英語でのパブリケーションを行うことも重要であるが、海外の研究者と何らかの協働関係を築きながら研究を行っている様子を観察したい。そこで、上記の日本人研究者 100 人に加え、過去の IJCAI および AAI(どちらも AI に関する主要な国際会議である)の参加者の中からヒット件数の多い人上位 100 名を抽出した。この合計の 200 名に対してネットワークを抽出したものが図3である。

図3では、ネットワークが大きく2つのクリークに分かれており、それぞれ日本人コミュニティ(上側のクリーク)、国際コミュニティ(下側のクリーク)に相当している。Simpson 係数の性質のために、エッジを持たず孤立している海外の研究者が多く存在し、国際コミュニティのクリークがやや小さくなっている。具体的にみると、日本人コミュニティの中に入っている外国人は、Tom Mitchel(スタンフォード大)、Judea Pearl(UCLA、確率推論等)、Daphne Koller(スタンフォード大、AI)、Katia Sycara(CMU、ロボット)など、西海岸を中心とする有名な AI 研究者が多い。また、Tim Finin(メリーランド大)や James Hendler(メリーランド大)、Stefan Staab(Koblenz-Landau 大)、Frank van Harmelen(アムステルダム自由大学)などの Semantic Web の研究者は、日本人コミュニティとしっかりと協働関係を作り上げていることが分かる。



図1: 日本語の氏名によるネットワーク図(100名)

逆に、国際コミュニティのクリークには、例えば Rodney Brooks (MIT) や Stuart Russel (スタンフォード大) らの著名な研究者が含まれている。そういった国際コミュニティに対する強力なハブとなっているのが、例えば、横尾真 (九州大学)、石田亨 (京都大学)、元田浩 (米国国防総省) らの研究者である。これらの人物は、図3のネットワークにおいて媒介中心性が高い。また、図2による英語名のネットワークで中心性の高い研究者でもある。

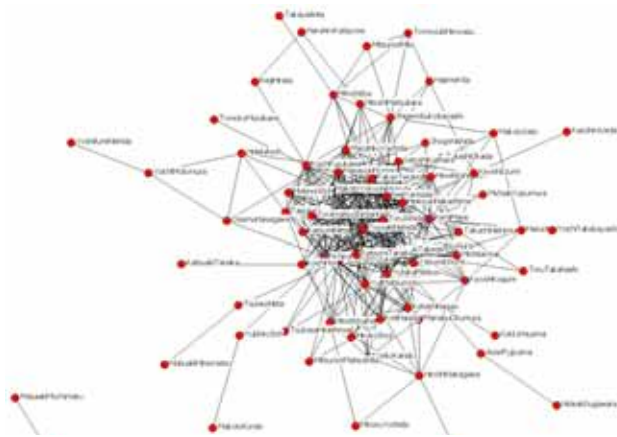


図 2: 英語名によるネットワーク図 (100 名)



図 3: 日本と海外の人工知能コミュニティ (200 名)

3. 考察とまとめ

詳細の結果は当日の報告に譲るが、ここでは、その概要を述べ、まとめとする。まず、日本語圏内で知りうる AI 研究者像と、英語圏から見た研究者群には大きな異なりがある。我が国の AI 研究者に対する認知は国内外で異なっている。国内で活躍する者が必ずしも海外で活躍をしているとはかぎらない。海外ネットワーク、国内ネットワークの双方の上位に含まれている者は、決して多くない。人工知能の分野は、直感的には、文系の学問領域での海外発信数の少なさに比して高いと推察しうるが、それでも国内で研究コミュニティを方向付け、さらに国際的にもアピールしている研究者は一握りである。

認知の異なりを考えたときに、次の2つの異なりがある。ひとつは、英語圏から見た有名な日本人研究者と、日本語圏から見た有名な日本人研究者は異なる。(しばしば海外で認められるが故に国内でも著名になる研究者がいる。)さらに、英語圏から見た有名な海外の研究者と、日本語圏から見た有名な海外の研究者も異なる。(日本で有名な海外の研究者が、それほど著

名でない場合もある。)これは、ネットワークにおける認知の性質を考えるとごく自然である。

しかし、AI の研究の国際的な顕現性を上げ、日本の研究をさらに積極的に海外に発信する上では、海外から見て日本人コミュニティの研究者がどう見えているかを意識すること、また、日本人に良く知られていないが著名な海外の研究者の存在を知ることの両面が重要であろう。

研究者は、必ずしも日本の国内と海外の両方で活躍すればよいというわけではない。日本の研究コミュニティを方向付け、組織化し、また後進を指導していくことによって、研究コミュニティの力を上げることは重要な役割である。また、一部の研究者が積極的に海外に情報を発信していくことも重要である。こういった研究者の役割分担とその効果を示すことができれば、今後の AI 分野の国際的な発展につながっていくのではないだろうか。

参考文献

- [Burt 82] Burt, Ronald: Toward a Structural Theory of Action, Academic Press, 1982 .
- [Getoor06] L. Getoor and C. P. Diehl: "Link Mining: A Survey", SIGKDD Exploration, no.7, vo.2, 2005
- [Manning02] C. Manning and H. Schutze, Foundations of statistical natural language processing, The MIT Press, London, 2002
- [安田 04] 安田雪: イベント空間情報支援システムが抽出するヒューマン・ネットワークの構造、第 18 回人工知能学会全国大会, 2004
- [安田 05] 安田雪、松尾豊: 人工知能学会における研究者ネットワークの分析、第 19 回人工知能学会全国大会, 2005
- [安田 06] 安田雪、松尾豊、武田英明: 人工知能学会におけるネットワーク構造と変化、第 20 回人工知能学会全国大会, 2006
- [安田 07] 安田雪、鳥山正博、"電子メールログからの企業内コミュニケーション構造の抽出" 組織科学, Vol40(3)pp.18-32.