

論文と共著者の関係に基づく研究分野間ネットワークの成長予測

Prediction of the growth of the research fields network based on the relationship between research papers and co-researchers

小野亮輔
Ryosuke Ono片上大輔
Daisuke Katagami新田克己
Katsumi Nitta*1 東京工業大学 大学院 総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

The purpose of this paper is to predict the growth of the research fields network in the near future. The network based on a relationship among the fields that shared technology or a theme by using category data has been built and visualized in order to analyze the changing of research trend. We examine the method in which the relationship between research papers and co-researchers in an academic meeting is focused on by using the data about large-scale joint research projects. To confirm the relationship, we conducted experiments to verify an effect on number of research papers and link structures.

1. はじめに

近年、社会の高度化に伴い学術領域が細分化、複雑化しており、研究者が研究分野動向を把握することが困難となっている。現在までに、研究分野間の関係性に基づいたネットワークを構築することで学術研究の動向を分析する研究が盛んに行われている。先行研究者である奥岡は、同一領域内において文献内容を考慮した研究分野間ネットワークを構築し、年代ごとの分野間の関係性を可視化したモデルについて研究した[奥岡 09]。また安田らは Web マイニングの手法で収集された研究者ネットワークを年代ごとにより、研究者のコミュニティの様子を分析する研究を行っている[安田 04]。篠田らはネットワークの中心性に着目し、ネットワークの形成過程をシミュレートするモデルを提案した。実在するネットワークに近似するネットワーク生成モデルを構築し、ネットワーク分析を行っている[篠田 08]。珍田らは、補正余地が存在するスケールフリーネットワークに対して、Triads の統計情報を用いた補正が有効であることを確認した[珍田 09]。

しかし、従来手法では構築された発表論文に基づくデータのみ利用しているため、様々な外部要因によって影響を受け多様に変化する研究分野間の関係性の動向予測は困難であると思われる。

そこで本研究は、共同研究プロジェクトに基づく論文と共著者の関係の情報を利用し、先行研究である奥岡の研究分野間ネットワークを利用した研究分野動向分析の手法を拡張する。さらに共同研究プロジェクトが将来の研究分野間の関係に与える影響を分析し、共同研究者や研究者の専門分野、プロジェクトの規模やキーワードなどの情報によって予測を試みる。

2. 研究分野間ネットワークとは

研究分野間ネットワークとは、学会の文献データを用いて学会・年代ごとに研究分野同士をそれらの関係によって結び付けた構造のことである。

ネットワークとして研究分野間のつながりを時系列でみたときに視覚的に把握しやすくするために、研究分野間ネットワークを生成し可視化する。そのため先行研究者の奥岡が作成したネットワーク生成システムとネットワーク可視化システムを利用する。このシステムにより生成されるネットワークは 2 種類ある。1 つはカテゴリ情報を用いることで研究分野をノード、各研究分野に属する論文により付加されている文献キーワードを比較することにより関係性を生成した研究分野間ネットワークである。もうひとつはカテゴリ情報を用いず、文献キーワードをノードとし、同一の論文によく同時に付加される文献キーワード間をエッジで結んだ研究分野間ネットワークである。

学会によってカテゴリ情報と文献キーワードのような利用できるデータに違いがあるため、それぞれの学会に対応したネットワークを生成する必要がある。そこで本研究では電子情報通信学会総合大会、日本ファジィ学会学会誌、情報処理学会研究会、人工知能学会全国大会を文献データとして用いる。

カテゴリ情報を持つ電子情報通信学会総合大会、情報処理学会研究会、人工知能学会全国大会においては、文献キーワード、共起キーワードによる分野間関係の生成手法を用いる。

カテゴリ情報を用いた方法では、カテゴリをノードとしたネットワークを生成する。2 つの論文間にエッジを張る条件は、2 つの論文に同じ文献キーワードが付加されているとき、エッジの重みは、共通のキーワードの数とする。次に研究分野間の関係性を生成する。カテゴリ情報は各論文に 1 つ割り当てられている。異なるカテゴリの論文がエッジを張るとき、カテゴリ間でエッジを生成する。エッジの重みは論文同士に張られたエッジの重みの総数をカテゴリに属する論文の文献キーワード数で正規化した値とする。

ファジィ学会の場合は共起キーワードネットワークを生成する。これは、文献キーワード間の共起度を計算し、文献キーワードをノードとしたネットワークである。その際キーワードの共起度の計算には Simpson 係数を使用する。

生成されたネットワークを可視化した研究分野間ネットワークにおいて、ノードの大きさは分野に投稿された論文数が多いほど大きく表示され、エッジの太さは分野間の関係性が強いほど太く表示される。

3. 共同研究プロジェクトと研究分野間ネットワーク

研究分野が発展するには共同研究プロジェクトのように複数の専門家・研究者が協力し、予算を獲得して研究することが必要である。現に共同研究の数は増加の一途をたどっている。

そこで今回は科学研究費補助金を獲得している共同研究プロジェクトに着目し、研究分野の発展との関係性を調べる。

3.1 共同研究プロジェクトの論文数に与える影響

共同研究プロジェクトが実際に研究分野間ネットワークに影響を及ぼし、ひいては研究分野間ネットワークの成長予測に利用できるかどうかを確認する。そこで共同研究プロジェクトが行われることで、その研究に関する論文の数に影響を与えているかどうか調査した。

今回の調査の対象として、科学研究費補助金に採択された共同研究プロジェクトである「情報洪水時代におけるアクティブマイニングの実現(2001年9月～2004年3月)」を選んだ。この研究を選んだ理由は、特定領域研究である大型プロジェクトであったこと、「アクティブマイニング」という専門的なキーワードを用いていることで、今後の論文数への影響の調査が容易になると考えたことが挙げられる。

科学研究費補助金データによると、この共同研究プロジェクトに付加されているキーワードは、データマイニング、データ収集、ユーザインタラクション、機械学習、知識獲得、知識発見、視覚化である。また、共同研究者に付加されているキーワードは、データマイニング、機械学習、エキスパートシステム、人工知能、学習といったように、それぞれの研究者の専門分野といえる分野となっている。このようなキーワードを確認することで、どのような専門分野を持った研究者が、どのような研究プロジェクトに参加しているか把握できる。

3.2 調査方法と結果

共同研究プロジェクトの与える影響を調べるため、まず論文数の変化を分析する。そこでアクティブマイニングに関する論文数の変化を追うために論文検索として広く用いられている Google Scholar を利用した。

1998年から2009年にかけて、「アクティブマイニング」についてのキーワードを含む論文を検索し、その数をアクティブマイニングに関する論文数とした。その際に検索ワードとして、以下の2つのパターンで調べた。

(ア) 「"アクティブマイニング" OR "active mining"と記述したもの。

(イ) 「"アクティブマイニング"と「研究プロジェクト参加研究者の氏名」を記述したもの。

(ア)の場合は、アクティブマイニングまたは active mining というキーワードをタイトルや本文に含む論文すべてを検索する。一方で(イ)はアクティブマイニングに関する論文でかつ研究プロジェクトに参加した研究者が著者としてもしくは被引用者としてかかわっている論文を検索する。ここで(ア)の数-(イ)の数を

(ウ)の数とすると、(ウ)の数は、アクティブマイニングに関する論文でかつ、研究プロジェクトに参加した研究者が著者としても被引用者としてもかかわっていない論文数を表している。この論文数は、共同研究プロジェクトが論文数に与えた影響の大きさを表していると考えられる。この論文数の推移を研究プロジェクト

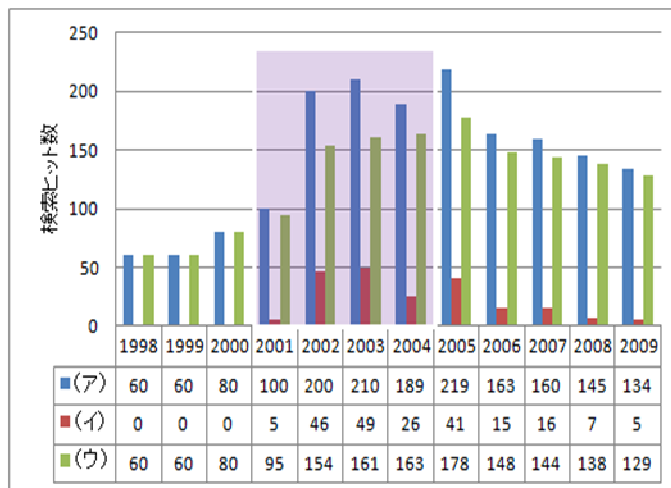


図1. プロジェクトが論文数に与える影響

の実施期間中・前後で分析することで共同研究プロジェクトと論文数の関係性を考察する。

調査の結果を図示したものが図1である。1998年から2009年までの12年間の(ア)、(イ)、(ウ)の数を示している。

このグラフより、2002年から急激に(イ)の数が増えていることがわかる。さらに2003年まで上昇を続け、その後徐々に減少している。これは共同研究プロジェクトの開始翌年から、共同研究プロジェクトの参加研究者自身の論文が多く発表されたことを示している。実際に発表された文献は、研究実績報告書によると、2001年に14件(人工知能学会論文誌4件、電子情報通信学会論文誌1件)、2002年に27件(人工知能学会論文誌5件、日本ファジィ学会誌1件、情報処理学会論文誌2件)、2003年に28件(人工知能学会論文誌2件、日本ファジィ学会誌2件、電子情報通信学会論文誌1件、情報処理学会論文誌1件)、2004年に21件(人工知能学会誌2件、電子情報通信学会論文誌1件、情報通信学会論文誌2件)であった。

(イ)の数と発表文献の数に相違がみられるのは2つの原因が考えられる。1つは発表論文の中にアクティブマイニングというキーワードを用いていないことから、(イ)でヒットしなかったことが挙げられる。2つ目として、(イ)は、引用や発表論文に記載されていない研究会の発表も含まれていることが挙げられる。2001年は発表文献が(イ)の数を超えていたが、2002年からは逆転し(イ)の数が発表論文を大幅に超えている。これは2001年には少なかつた引用の影響が2002年から大きくなったことが考えられる。一方(ウ)の数は、2002年に急激に伸び、2005年まで単調に増加し、その後は緩やかに減少している。これは、共同研究プロジェクトの開始時の2001年から、共同研究プロジェクトが論文数に影響を与え始め、共同研究プロジェクト終了後の2005年に最も影響を与え、その後プロジェクトが終わった後も影響は残り、高い論文数の水準を維持することにつながったと考えられる。つまりこの共同研究プロジェクトは、プロジェクト開始時から、論文数を押し上げる要因として機能していた。

以上のことから共同研究プロジェクトは論文数に影響を及ぼしその結果、論文を構成要素として含む研究分野間ネットワークにも影響を与えると考えられる。

4. 研究分野間ネットワークの成長予測

ネットワークの成長予測とはネットワークの特徴量(密度, クラスタリング係数, 次数, 中心性)のような定量的データの将来の値を予測し, その予測結果から研究動向を定性的に分析することを意味する.

4.1 予測に用いるデータ

科学研究費補助金を獲得している共同研究プロジェクトの情報を使用する. その際, 扱う共同研究プロジェクトは, 特別推進研究, 特定領域研究, 新学術領域研究(研究領域提案型), 基盤研究(S)に属するものを用いる. 実際のデータはウェブ上の科学研究費補助金データベースを利用して収集し, 共同研究プロジェクトに付加されるキーワード, 研究者に付加される参画分野・キーワードの情報をそれぞれ抽出する. 共同研究プロジェクトに付加されるキーワードは, 申請の際に研究者らが自ら決定するものであり, それらは研究者自身のキーワードとしても追加される. キーワードの並ぶ順番は参加したプロジェクトのキーワードが多い順である. それぞれの研究者は各々の専門分野に関する共同研究プロジェクトにより多く参加していると考えられるため, より前方のキーワードのほど研究者の専門分野であることができる.

4.2 予測手法

研究分野間ネットワークに科学研究費補助金データベースの共同研究プロジェクトの情報(研究者の専門分野, プロジェクトの予算, プロジェクトのキーワード)を適用することで研究分野間ネットワークの予測を行う. ノードの拡大とエッジの強化を予測するにあたり, 次の3つを仮定する.

(1) 研究分野間ネットワークは共同研究プロジェクトの影響を受け成長し, 共同研究プロジェクトに参画する共同研究者のそれぞれの専門分野が今後の研究分野間ネットワークにおいて関係を強める. もしくは, 共同研究プロジェクトのキーワード同士の関係が, 今後の研究分野間ネットワークにおいて強まる.

(2) より多く研究された分野が発展する. 研究者の専門分野もしくは共同研究プロジェクトのキーワードが今後の研究分野間ネットワークにおいて成長する.

(3) 多くの予算の下で遂行された共同研究プロジェクトのほうに, 研究分野の成長により大きな影響を与える. 科学研究費補助金データには共同研究プロジェクトに対して補助された金額も記載されていることを利用し, 補助金の金額による重みづけを行う.

予測するにあたり, 共同研究プロジェクトに参加する研究者の専門分野もしくは共同研究プロジェクトのキーワードを, 研究分野間ネットワークを構成するノードに対応付ける必要がある. 共同研究プロジェクトのキーワードや研究者の専門分野は, 研究者が独自に考えるため, ネットワークのノードを大幅に超える数が存在する. そこで概念辞書を作成し相互の対応関係を構築する.

次に共同研究者のもつ専門分野が今後の研究分野間ネットワークに影響を与える場合(1)と, 共同研究プロジェクトのキーワードが今後の研究分野間ネットワークに影響を与える場合(2)の2つを考える.

(1) 研究者 A が専門分野 α をもっていて研究者 B が専門分野 β をもっているとき, α と β のノードの重みと, エッジの重みを増やす.

(2) 共同研究プロジェクトのキーワードが α, β, γ の時, α, β, γ のノードの重みを増やし, α と β, β と γ, γ と α のエッジの重みをそれぞれ増やす. -

さらに補助金の金額に応じた数を掛けることで, ノードの拡大とエッジの強化の際の重みに研究プロジェクトの規模を考慮する. この重みを, 実際の研究分野間ネットワーク可視化システムのノードの大きさ, エッジの強さと比較し共同研究プロジェクトと研究分野間ネットワークの関連性を発見することを試みる.

5. おわりに

本研究は, 研究分野間のネットワークの予測をする手法の構築と評価を目的とした.

従来までの論文情報に基づくネットワーク特徴量の時系列変化によるネットワークの成長予測だけでなく, 共同研究プロジェクトのような情報を利用し, プロジェクトの特徴や研究者の専門性に着目したネットワークの成長予測の手法を構築することで, ネットワークの構成要素の属性をも考慮に入れたモデルを提案した.

現在までのグラフ理論的なアプローチとして, ネットワークの構造に着目し, エッジの欠損やネットワークの成長を仮定したエッジ予測の適用によって研究分野間ネットワークを予測する手法も研究されていることから, この手法を, 共著者関係を用いた研究分野間ネットワークの予測手法と組み合わせることで予測精度の向上が可能であるか検討する.

参考文献

- [松尾 06] 松尾, 篠田, 中島: 中心性に着目した合理エージェントのネットワーク形成, 人工知能学会誌, Vol21, No.1, pp.122-132 2006.
- [奥岡 09] 奥岡, 片上, 新田: 研究分野ネットワークによる学術分野動向の解析と評価, 人工知能学会研究会資料, SIG-KBS-A803, pp.21-26 2009.
- [珍田 09] 珍田, 大沢: Triad 構造を用いた不完全な複雑ネットワークの確率的な構造補正, JAWS2009.
- [篠田 08] 篠田, 松尾, 中島: 複数の中心性尺度によるネットワーク生成モデル, 日本知能情報フェジィ学会誌, Vol20, No.3, pp.410-422 2008.
- [安田 04] 安田: 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, 人工知能学会全国大会, 2004.