

論文の内容の要旨

論文題目 引用ダイナミクスを用いた萌芽的論文の
早期特定に関する研究
(Early Detection of Emerging Research Fronts Using Citation
Dynamics)

氏 名 岩見 紫乃

本論文は、論文の引用ネットワークに関する時系列変化の情報を用いて、早期に萌芽的論文を特定する基礎的な手法を提供するための研究である。

本論文は 7 章から成り、第 1 章に、本研究が必要である背景と研究目的を述べ、第 2 章に提案手法を説明し、第 3 章と第 4 章に、時系列分析から得られた結果と考察を記述し、第 5 章に、時系列分析を観察した上で、研究目的を達成するための設計を検証し、第 6 章に、提案手法を実装したシステムを紹介し、第 7 章に結論としてまとめている。

まず、第 1 章では、研究の背景、研究目的等について述べている。今日、情報量が急速に増加して、専門家でも把握しきれないほど科学技術の知識が増えるといった知識の爆発現象が起こっている。先端技術企業は萌芽的技術シーズの探索に力を入れていることが多いが、この現象の中で、探索が次第に難しくなっている。一方で、技術シーズが製品化・産業化され、市場において利益をあげ得る期間(ライフサイクル)は短くなっており、競合者に対し有利な立場を得るには、早期に有望な技術シーズを見つけることがますます重要になっている。こうしたことを背景として、先行研究では、論文の引用その他の書誌情報を利用して、萌芽的技術を探索する手法を開拓することが行われている。具体的には、例えば、論文の引用先は、論文の著者が当該被引用論文を自著と内容的な関連が深いとして選定したものであるという性質を利用して、引用ネットワークを構築し、クラスタリング手法を

用いて論文をグループ化することで、学術知識や産業技術知識をボトムアップに構造化する手法が実用的なものとして完成されている。このような引用ネットワークまたは各クラスターのネットワークの中心に位置する論文は、当該分野において、多くの専門家から深い関心を持たれ、または、支持されているものと評価することができ、それら論文を精査することで、萌芽的技術シーズを特定することも可能であるとの議論がなされている。ネットワーク中心性を測定するにあたって適切な指標としては、複数のものが候補として挙げられている。論文は学術分野の成果物であるが、近年、サイエンス・リンケージから産業応用に繋げることに注目が集まっており、特許やインターネット上のデータもある中から、本研究では論文のみ注力している。また、論文のデータは、特許などのデータに比べるとデータの質が高く、手法の確立にあたってはより正確な結果が導出されることが期待される。一方、こうした先行研究に基づく手法は、十分に論文が被引用数を得た後でなければ、重要性を推測できないものであることや、参考とする被引用数が直近の数年のみである等といった制約があることから、被引用数は少ないが将来大きな分野を開拓することにつながる萌芽的論文を早い段階で見つけ出すことが難しいという問題点がある。そこで、本論文では、各年の論文の引用ネットワークからネットワーク構造に基づく指標群とそれらに関する特徴量を計算し、分野の登場初期からの指標の中長期的な変化を可視化、定量化することで、萌芽的論文を早期に特定する基礎的な手法を提案する。萌芽的論文とは、複数の重要論文の定義があるうち、特に、分野を開拓した重要論文を指している。

第 2 章では、本論文の提案手法について述べている。本論文で着目した時系列分析は、検知すべきパターンが定まっていないものを特定するにあたって、時間の前後のデータを利用することにより、質の安定した恒常的に供給されるデータを利用することができるというメリットがある。提案手法は、次のようなステップで行うものである。まず、論文の書誌情報を論文データベースサービスである **Web of Science** から収集し、論文の引用関係から引用ネットワークを形成する。引用ネットワークのうち、最大連結成分に含まれる各論文に対して、**in-Degree Centrality** を始めとする複数のネットワーク中心性指標を計算する。指標は各年まで論文を集めたデータセットごとに計算する。1つの論文に対して各年の指標が算出されると、横軸を時間軸、縦軸を指標の値とし、各論文を 1 本の線として指標の値を繋げ、時系列で指標の変化を見ることを可能とする。これら時系列分析から萌芽的論文の特徴を抽出し、未知の分野でも萌芽的論文を特定できる特徴量として数値化する。

第 3 章では、この提案手法に対して、ノーベル賞受賞分野や ACM チューリング賞から 7 分野を選定し、受賞対象となった論文(非常に重要な萌芽的論文)に関して本提案手法を適用することで、萌芽的論文の特定の観点から、指標の有効性を具体的に検証している。ノーベル賞は、分野でもっとも重要な発見、発明、改良をした人物に贈る意図で創始されており、分野の始まりに貢献した研究者に贈られるため、本研究における萌芽的論文の特徴を抽出するのに適した素材であると考えた。いくつかのノーベル賞受賞分野では、分野が開拓された当初、その研究成果が顧みられなかった逸話が残っている分野もあり、そのよう

な分野での萌芽的論文の特徴を抽出することで、未知の分野における萌芽的論文を特定することに繋がると考えられる。12 の指標群で時系列分析を行い、比較検討を行った結果、in-Degree Centrality、Communicability Centrality、PageRank の 3 指標が萌芽的論文の特定に有効であることが明らかとなった。これらの指標においては、萌芽的論文が分野の登場初期に高い値を持つ傾向が明確に見られた。

第 4 章では、7 分野を横断した考察を行っている。重要論文の特定に有効な指標どうしを比較すると、精度と早期特定には、in-Degree Centrality が最も適切だが、データセットによっては PageRank の方が有用な場合があることが示された。また、手法を活用するに当たっての制限事項として、学術分野の登場初期において引用があまりに疎であるために引用ネットワークが形成できない場合があること、同じく初期に重要論文が 2 番目以降の連結成分に含まれてしまい、重要論文の特定が遅れる場合があることも明らかとなった。更に、Web of Science に収録されていない論文や書籍が萌芽となる重要な文献に該当した場合には、データセットの性質上、特定ができないことも示された。

以上の第 3 章、第 4 章では、萌芽的論文の早期特定に適した指標を見つけるために探索的研究であり、提案手法による時系列の可視化から有効な指標群を見出すことができた。しかしながら、可視化した時系列を観察して発行年の異なる論文を比較することは容易ではない。そこで、第 5 章においては、分野の成長初期において高い精度で特定可能となるように、萌芽的論文を客観的な数値で特定するための特徴量群の開発を試みている。最初に、第 3 章、第 4 章の分析結果から時系列中における各論文の指標の最大値(Height)、各論文の指標の増加率の最大値(Slope)、指標の増加率が最大値に達するまでの増加率の平均値(SlopeAve)等の候補特徴量を設定した。これら候補特徴量群について、第 3 章と同様の 7 つのデータセットを用いて実験を行った結果、in-Degree Centrality および PageRank の時系列変化のデータについて、各論文の指標の最大値(Height)と毎年の変化量のうちの最大値(Slope)が、萌芽的論文を特定する特徴量として有効であることが明らかになった。また、それら特徴量間の相関が高いことも示された。一部のデータセットでは、対象と設定した分野ではなく、不可分に関連する分野の萌芽的論文を特定するという現象が見られたが、関連領域も含めて萌芽的論文を特定する特徴量として有用であることが判明した。さらに、専門家に確認いただけた分野では、未受賞の萌芽的論文が、特徴量で挙げられる萌芽的論文候補に特定されているという評価が得られた。これらの特徴量の成果は、将来、萌芽的論文を自動的に抽出するにあたって、利用可能であると期待される。

第 6 章では、開発した指標、特徴量のシステム化について述べている。本システムは、論文データベースと引用ネットワークの形成については既存のシステムを利用しているが、それ以外の部分の分析に関しては新規に作成したものである。

第 7 章では、結論を述べている。先述したように、政策や経営の場において投資すべき技術分野を決定せねばならない意思決定者や研究テーマを模索する研究者にとって、萌芽的な技術シーズの早期特定が重要な技術経営上の課題となっている。本論文では、過去の計

量書誌学による技術経営研究の成果も踏まえつつ、重要な萌芽的論文の早期特定に役立つ独自の手法を提案し、ケース研究により、その有効性を確認した。本研究の新規性は、ネットワーク指標を時系列で表した時に萌芽的論文ならば初期に高い値を持つことから、学術分野の中で萌芽的論文の特定する特徴量を設定したことである。この特徴量は、異なる出版年で複数の萌芽的論文があっても、時代を横断して一元的に評価できる点で優れている。また、先行研究では、**Betweenness Centrality** で特定できる萌芽的論文に限られていたのを、それ以外の萌芽的論文にも適用できるようになった。本研究の研究成果は、先端技術を巡る市場化競争が激化する今日のイノベーション環境において、意思決定者を効果的に支援する技術経営上の基礎的な手法を提供するものであると確信する。