

概念グラフを用いた発明者解析

飯野由里江¹ 山田泰寛² 廣川佐千男²

Yurie Iino, Yasuhiro Yamada, Sachio Hirokawa

¹九州大学大学院システム情報科学府

¹Graduate School of Information Science and Electrical Engineering, Kyushu University

²九州大学情報基盤研究開発センター

²Research Institute for Information Technology, Kyushu University

要約: 単語の文書頻度による単語間の順序関係に基く概念グラフを使って特許情報における発明者群と単語群の階層構造を構成し、企業や分野ごとの研究開発体制の分析手法を提案する。

1 はじめに

近年、さまざまな情報が Web 上に公開されるようになった。「研究技術」に関する情報は、専門の技術雑誌等に掲載され、その書誌的な事項と専門家による抄録が特定のデータベース [JDream] 上に格納されるのが一般的であったが、特に、大学をはじめとした各研究機関においては、自己の魅力をアピールしたり、複数研究機関によるプロジェクトの研究成果を公にしたりする目的で、発表文献や学内資料などを Web 上で公開するところも増えてきた。

研究技術に関する重要な文献としての「特許」情報についても、例外ではない。特許情報は、従来限られたデータベースから有料で提供されており [PATOLIS]、必ずしも身近な存在ではなかった。しかし、インターネットの普及に伴い、1999年3月に日本特許庁が「特許電子図書館」[IPDL]を Web 上に開設し、誰もが特許情報を無料で検索・ダウンロードできるようになった。また、諸外国においても多くの特許庁が自国の特許情報を無料開放するようになり、ヨーロッパ特許庁のサービスする esp@cenet@[EPO]では 71 カ国の特許データが無料で提供されている。更に、民間企業においても Google Patent[Google]のように特許情報を無償で検索・提供するサービスが現れてきた。

特許情報は、その制度上、一定期間を経ると公開されるだけでなく、出願日、出願人、発明者、発明の名称、要約、特許請求の範囲などを一定の様式で出願することが要求される。また、特定の研究技術を収録する技術雑誌とは異なり、特許情報のデータベースにはあらゆる技術が収録され、更に専門家により共通の分類が付与されている。これらの特性を利用して、特許情報を様々な観点から解析すること

が行われている。

筆者らは、特許情報の特性の活用を目指し、特に発明者情報と、要約文中の特徴用語や技術分野情報に注目することにより、企業における研究技術開発者の関連や技術開発体制についての推測が可能ではないかと考え、筆者らの提言する概念グラフを用いることにより新たな知見を得たので報告する。

2 発明者情報解析の背景

先に述べたように、特許情報は定型化されていること、専門家による技術分類が付与されるなどの特徴があり解析に適した情報であること、特許が（将来の）事業の自由度を左右する「権利を規定する技術情報」であることから、以前よりそれらの情報を戦略的に活用することが求められてきた。その代表的な活用事例として「特許マップ」というものがある。「特許マップ」とは、「膨大な特許情報を特定の利用目的に応じて加工・分析して、ビジュアルに表現したもの（特許庁）」であり、特許庁では「産業界での特許情報活用の一助となることを目的に独自に解析を行い」Web 上で公開している [特許庁]。「特許マップ」作成において、解析しうる情報には様々な項目があるが、筆者らはその中で特に「発明者」情報に注目をした。

出願人（企業）は、事業の発展のために投資をして技術開発を行い、その結果として発明が生まれ、特許として出願される。従って、企業の投資の結果が特許情報として公開されることになるが、特に投資という側面に直結した指標となるのが「特許出願件数（特許を出願するにあたって費用が発生する）」と「発明者数（発明者数 企業の研究員数）」である。また、「発明者」情報はすなわち「発明を行

った研究技術開発者」の情報であり、その発明における「研究技術開発者」のつながりや変遷を知ることができる。そのため、技術論文情報における著者情報とは、その情報のもつ意味がおのずと異なってくる。

著者が、考える「発明者」情報を解析する意義には、以下のものがある。

- 1) 発明者数はその研究技術開発にかかった人員数と推測できること
- 2) 発明者の（研究技術分野別）分布は、ある企業がどの分野の研究技術開発に注力しているかを知手がかりとなりうること
- 3) 発明者間の関係は、その技術開発体制を知手がかりとなりうること
- 4) 発明者の変遷は、技術の流れを知手がかりとなりうること
- 5) 筆頭発明者としての出願が多い発明者は、その技術開発テーマにおけるリーダー的存在でありうること

従って、発明者情報は、出願人（企業）の研究技術開発活動を知る上での手がかりとなると共に、場合によっては、優秀な発明者をヘッドハンティングする際の指標にもなりうる重要な情報であると考えられる。

3 関連研究

3.1 大学等の研究機関における研究技術開発者に関する研究

研究技術開発活動における国際競争力の強化が叫ばれる中で、大学等の研究機関においては特に「産官学連携」の重要性が増している[文科省]。産官学連携をスムーズに行うためには、研究技術開発者間の相互交流が重要となってくるが、高度に細分化/専門化した技術分野では、互いの研究開発テーマの関連や技術開発におけるキーパーソンを知ることが容易ではないため、様々な手法で研究技術開発者間のネットワーク情報を表示させる支援技術が提唱されている。

浅田ら[浅田 06]は、Web上のページにおける研究者の名前の共起関係を、検索エンジンを用いて調べ、テキスト処理を組み合わせることで研究者の関係の強さ及びその種類を特定することによって、研究者ネットワークを図式化することを提案している。具体的には、ネットワークを構成するメンバーの氏名及び所属（同姓同名の別人を除去し、メンバーを確実に特定するために所属情報を必須としている）リ

ストをあらかじめ作成した上で、全メンバーの氏名をクエリとしてWeb検索を行い、検索結果における上位一定件数のページを解析し、リスト中にある他の研究者名との共起情報を解析することによって、膨大な量のWeb情報から特定の研究者のネットワークを構成させようとするものである。この手法によれば、あらかじめ限定された研究者間の関係を示す場合であったとしても、大量の情報への検索と解析を繰り返すことになるため、情報源を絞り込み、関係の網羅性を犠牲にするか、関係の網羅性を優先して膨大な量の検索を行うかの選択が必要となる。

また、廣川ら[廣川 05-1]は、所属する大学における教員データにおいて、教育研究活動概要や研究成果一覧などの複数の項目を独立した文書として捉え個別インデックスを構成することにより、項目ごとのクラスタリングを提案している。この手法によれば、課題や研究概要を個別項目として捉えて2軸に設定することにより、共通の研究課題や目的をもつ教員を効率的にクラスタリングし、視覚化することを可能としている。この手法によれば、大量の情報から特定の研究者間の関係や、活動概要を視覚化しようとする試みとして興味深いものであるが、クラスタリング可能な個別インデックスのクラスタ数の設定によって結果が左右される可能性があること、研究者間の関係有無については明示されるものではないことに注意が必要である。

3.2 特許情報の解析事例

近年、特許情報を膨大なテキスト情報として扱い、形態素解析やマイニング技術を応用したさまざまな研究がなされている。代表的な検討は、国立情報学研究所が中心となって行われている[NTCIR]。また、間瀬らによる出願人別使用傾向の分析[間瀬 06]や岩山らによる特許請求項の構造解析[岩山 04]などが報告されている。

「特許マップ」を作成するに当たっては、様々な解析ソフトが市販されており、特許情報の活用を研究するグループにおいて、それら解析ソフトの比較検討が行われている[知財協 07]。その報告では、特許マップソフトの統計処理機能を利用して「発明者相関件数」を解析した後、共同発明者を手作業でまとめ発明者グループを作成し、テーマに対する資源投入を推測する手順について述べられている。

4 検討

4.1 発明者解析の手法

廣川ら[廣川 05-2]は、文書頻度を用いることにより、文書集合に現れる単語の上位下位関係の定式化する概念グラフを提案している。一つの単語に関連する検索結果の文書集合については、その単語に関連する単語群の階層構造が動的に得られる。概念グラフではまず、検索対象全体の文書集合 U を固定する。 U の部分集合 X と単語 u, v について、 $|X|$ は X に含まれる文書の個数、 $df(u, X)$ は u を含む X 中の文書集合の個数 $df(u^*v, X)$ は u と v の両方を含む X 中の文書集合の個数、 $r(v, u) = df(u^*v, X) / df(v, X)$ とする。単語 w と文書集合 X について、 $r(v, u) > 0.5$ となるとき w は X の特徴語であるという。二つの単語 u, v と単語集合 X について、 $df(u, X) > df(v, X)$ かつ > 0.5 であるとき、 X に関して v について上位であるという。この条件の後半の式を本稿では $r(v, u)$ と表記する。概念グラフは、クエリ q に対する文書集合 $D(q)$ の特徴語を節、 $D(q)$ における上位関係を枝とする有向グラフである。具体的な可視化においては、単語 v についてすべての上位語に枝を描くのではなく、その隣接上位だけとすることで枝の数は抑えられる。

特許明細書群についての概念グラフを構築すれば、検索結果として得られる特許明細書群に現れる膨大な量の用語の関連解析が動的に行うことができる。また、特許明細書の特定の部分、例えば、課題や解決手段等の項目だけに着目すれば、項目間の違い、項目間の関連を可視化して分析することができる。現在知られている「特許マップ作成ソフト」には、異なった項目を同時に表示する可視化機能をもつものはない。発明者（あるいは出願人）ごとに使われる用語の関係を示す特許マップはあるが、個別の用語と特定の発明者（出願人）ごとの関係の違いが明確になる一方で、用語間の関係は明示されないため、技術としての把握をすることはできない。本稿では、発明者の氏名も単語と考え概念グラフを適用することで、発明者と（技術的）特徴語の関係、発明者間の関係、リーダー的発明者の情報の抽出が可能であることを、いくつかの実例を通じて示す。

4.2 実験と考察

4.2.1 発明者群の解析

本稿ではまず、2007年9月23日現在に公開されていた公開特許情報のうち、2000年～2005年に出願された化粧品分野のデータを元にシステムを構築し、そのうちの2000年～2002年に出願された特許

(9191件)に関して、実験・解析を行った。

大手化粧品企業 A 社を対象に、発明者名を特徴語として概念グラフを作成したところ、図 1、左図のような結果が得られた。この概念グラフでは、出現頻度の高い発明者名を抽出し、更に発明者名間の共起頻度の上位・下位解析を行うことにより、出願件数の多い発明者とその関係者（共同発明者）を効果的に可視化できることがわかる。図 1、左図の結果より、A 社において複数の開発グループが存在することが容易に把握することができる。なおこの解析においては、右側に示された発明者の出願のうちの過半数が左側に示された発明者との共同、すなわち関連度、 $r(v, u) > 0.5$ としている。

ところで、発明者間の関連度を変化させた場合には、概念グラフの形状が変化する。A 社のグラフにおいて、関連度、 $r(v, u) > 0.3$ とした場合には、新たな発明者と関連線が表れたり、異なる 2 つの少数グループのように見えた群が 1 つにまとまったりすることがわかった（図 1、右図）。この理由としては、企業における研究開発では、複数人が共通の技術開発をしながら、末端の製品の開発には細分化された少人数でフレキシブルにグループを形成して研究開発を進めることがあるためと考えられる。

4.2.2 研究体制と発明者群解析

企業により研究開発体制が異なることは容易に類推できるが、発明者群の違いとして表示可能であるか解析を試みたところ、以下の知見を得た。

大手化粧品企業 A 社の場合は、発明者関連度 $r(v, u)$ を 0.5 0.3 0.1 と変化させた場合には、関連線が大きく変化して、複雑な発明者群を形成するのに対して、外国企業 B 社の場合は、新たな発明者は表示されてくるものの、発明者群としての基本的な形状には変化が無い（図 2）。すなわち、日本企業の A 社の場合は、フレキシブルな発明者群を形成しているのに対し、外国企業の B 社の場合は、発明者群がある程度固定されていることが予測できる。

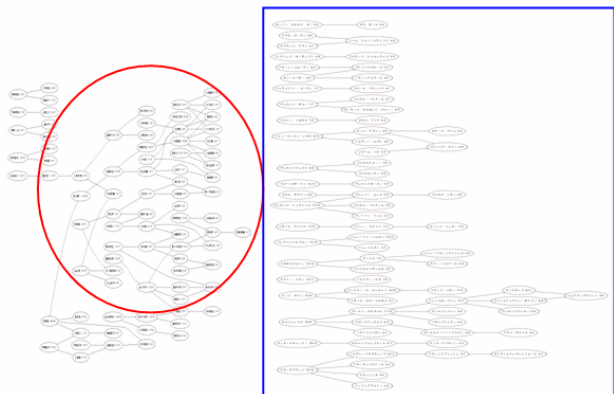


図 2：企業別発明者の概念グラフ（関連度 0.1 の場合）
（左図：日本化粧品企業、右図：外国化粧品企業）

4.2.3 発明者群と技術内容の解析

概念グラフでは、複数種の特徴語を同時に解析することが可能であるため、特許情報のうち、特許に記載されている発明者情報と要約 / 請求項から切り出される技術用語と同時に表示させることにより、各発明者群の研究内容についても類推することが可能である。そして、大まかではあるが、企業の注力技術が明示されることになる。図 6 は A 社の解析に

おいて発明者と技術用語を同時にあらわしたものである。各発明者群に特徴的な用語を見出すことができる。

5 むすび

今回の検討により、筆者らの提案する概念グラフが、発明者間のかかわりを一目で表すことができ、また、企業のタイプによる発明者群の違いを明確にすることができ、研究開発体制についても示唆され、企業評価に重要な有用なツールとなりうることを示した。更に、技術用語を同時に表示させることによって、発明者群の開発技術を容易に類推することができ、企業における研究技術開発にかかった人員数（投資）を技術ごとに推測することが可能と考えられる。

今回の検討に相当する発明者情報の解析を、市販の特許マップ作成ソフトで行う場合には、発明者間のマトリクスマップを作成した上で、発明者ごとのランキングマップや技術用語解析などを組み合わせるなど、手間と時間がかかることが予想される。また、発明者等の書誌事項と技術用語を組み合わせる視覚化するものは提案されていない。

一方、現在の筆者らの解析は、定性的解析であり



図 3：発明者と技術用語の概念グラフ（A 社）
印の箇所には、発明者名の間に「染色」「色調」といった用語が表示され「染毛剤」の開発者と予測可能
定量的解析ができておらず、今後の課題と考えている。

なお、現行の概念グラフは、システム上まだ解決すべき数々の課題がある。

- 1) 関連度の数値により解析される結果が大きく影響を受けるが、適切な関連度を選択するにはある程度の予測と経験が必要であること。
- 2) 発明者の動きは発明の変遷を表すものであるが、年代ごと解析時には図における特徴語の場所がずれてしまうため、変遷がわかりづらいこと
- 3) ユーザーの与えたクエリによって決まった文書集合における単語の特徴度が、全文書集合中における単語の特徴度とは必ずしも一致しないため、クエリによって表れる特徴語が変わってしまう可能性があること
- 4) 元情報である特許情報には、様々な情報の表記上のゆれがあるが、それらを統合する機能を備えていないこと
- 5) 出願頻度の多い情報を元に、関連線を形成するため、出願件数が多い発明者を中心に解析されてしまい、必ずしも重要な発明者（筆頭発明者）を中心に解析ができていないこと

今後は、上記の課題に対して、特許情報からより有用な情報の抽出するために、他の手法（従来型の統計ソフトやマトリクスマップなど）との併用についても検討する必要がある。

参考文献

- [浅田 06] 浅田洋平, 松尾豊, 石塚満: Web からの研究者ネットワーク抽出の大規模化, 人工知能学会論文誌, Vol. 20, No.6, pp370-378 (2006)
- [岩山 04] 新森昭宏, 奥村 学, 丸川雄三, 岩山 真: 手がかり句を用いた特許請求項の構造解析, 情報処理学会論文誌, Vol.45, No.3, pp. 891-905 (2004)
- [知財協 07] 日本知的財産協会特許情報検索委員会, 特許マップの比較研究,
http://www.jipa.or.jp/content/jyohou_hasin/sympo/temp/06sym_kensaku.pdf
- [特許庁] 特許庁, 特許出願技術動向調査報告
http://www.jpo.go.jp/shiryous/s_sonota/tokumap.htm#1
- [廣川 05-1] 廣川佐千男, 関隆宏, 安元裕司, 山田泰寛: 教員データに対する多面的検索システム, 情報処理学会研究

報告 Vol.105, No.173, pp.665-672 (2005)

[廣川 05-2] 廣川佐千男, 下司義寛, 和多太樹: 文書群からの概念グラフの構成, 情報処理学会研究報告. 情報学基礎研究会報告, Vol.2005, No.94, pp. 79-84 (2005)

[文科省] 文部科学省「国立大学法人等における産学官連携」

http://www.mext.go.jp/a_menu/shinkou/sangaku/main7_a5.htm

[間瀬 06] 間瀬久雄, 大西 昇: 特許文書中のタームの出願人別使用傾向の分析と類似特許文書検索精度への影響評価, 情報処理学会研究報告, Vol.2006, No.33, pp. 77-84 (2006)

[JDream 独立行政法人科学技術振興機構, Jdream, <http://pr.jst.go.jp/Jdream2/>

[PATOLIS] 株式会社パトリス, PATLIS, <http://www.patolis.co.jp/>

[EPO]ヨーロッパ特許庁, http://ep.espacenet.com/?locale=jp_EP

[Google] Google 社, Google Patent Search, <http://www.google.com/patents>

[IPDL] 工業所有権情報・研修館, 特許電子図書館 <http://www.ipdl.inpit.go.jp/homepg.ipdl>

[NTCIR] NTCIR プロジェクト <http://research.nii.ac.jp/ntcir/outline/prop-ja.html>