

## 特許情報テキストマイニングによる企業の研究開発 体制分析

飯野, 由里江  
九州大学大学院システム情報科学府

廣川, 佐千男  
九州大学情報基盤センター

<https://hdl.handle.net/2324/1526120>

---

出版情報：経営情報学会全国研究発表大会要旨集．2008年秋季（B2-3），2009-01-07．The Japan Society  
for Management Information

バージョン：

権利関係：

## 九州大学学術リポジトリ(QIR)

|            |   |
|------------|---|
| Title      | 特許情報テキストマイニングによる企業の研究開発体制分析   |
| Author(s)  | 飯野由里江 廣川佐千男   |
| Citation   | 経営情報学会 2008 年秋季全国研究発表<br>大会要旨集 : B2-3   |
| Issue date | 2009-01-07  |
| URL        | <a href="http://hdl.handle.net/2324/1526120">http://hdl.handle.net/2324/1526120</a> |

# 特許情報テキストマイニングによる企業の研究開発体制分析 Analysis of RD division by Patent Text Mining

飯野由里江\*1

廣川佐千男\*2

\*1 九州大学大学院システム情報科学府

\*2 九州大学情報基盤研究開発センター

**要旨** 一つの企業から出願された特許群を対象として、その企業の研究開発体制を分析する手法を提案し、企業群について比較を行った。具体的には、発明者の共起情報から発明者間の関連をグラフとして可視化する。これにより、主たる研究者と補助的研究者から構成される研究グループが何グループ存在するかが分析できる。さらに研究者間関連度の強弱をパラメータとして変化させたときのグラフの変化率により、多数のグループが相互に関連する研究開発体制か、あるいは、各グループはほとんど独立に運営されているかの違いが見える。

**Abstract** This paper proposes a method to analyze the research and development division of a enterprise based on the patent information of the company. This paper reports results of the method applied to the real companies. The relationships of the inventors of patents are analyzed using co-occurrences and are visualized as a graph. The graph reveals who is the key researcher and how many support researchers are working in a group. By changing the parameter of relation-strength between researchers, a characteristic of each company appears as the change rate of the graph. The rate reveals how each group of researchers influence tightly or sparsely.

## 1. はじめに

特許情報は、権利情報として先願特許調査に用いられる他に、重要な技術情報として、市場（技術開発）動向調査、企業動向調査にも利用される。

特許は、事業発展のための技術開発の結果として出願され公開される。従って、特許情報を特許出願人（企業）の研究開発投資の結果として捕らえることができ、その明確な指標となるのが特許出願件数と発明者数である。一方、発明者情報は発明を行った研究技術開発者の情報であり、その分析により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。

発明者情報を解析する意義には以下のものがあると考えられる。(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、(2) 発明者の研究技術分野別の分布は、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る手がかりとなる、(3) 発明者間の関係は、その技術開発体制を知る手がかりとなる、(4) 発明者の変遷は、技術の流れを知る手がかりとなる、(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る。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。

市川ら[市川 01]によれば、企業業績（一人当たりの売上高）と知的成果物（発明者一人当たりの特許出願件数）との間に高い相関関係がある。また、後藤ら[後藤 06]は、特許庁「技術動向調査」に記載されている重要特許の発明者情報を解析し、重要特許には発明者数が多いという結果から、発明者数が特許の質を測る指標として有効であると結論付けている。重要特許の発明者数が多くなる理由としては、企業が重要と判断した技術には多くの発明者が投入されること、異なる多くの発明者がお互いの技術知識を利用して技術開発を行った結果、価値が高い技術が生まれる可能性があること、などをあげている。

一方、技術イノベーションと研究員の共同活動や組織体制についての研究も行われている。西澤ら[西澤 08]は、論文等から見られる研究組織の協力体制を可視化し、技術分野ごとの関係性を把握しやすくするソフトウェアを作成し報告している。また、松尾ら[松尾 05]は、特定技術分野における研究活動を分析する上で、人間関係ネットワークの抽出が重要な基礎データになることを示唆している。

発明者数は、数量的客観的事実ではあるが、知識ベースの関連を示すものではない。しかし、特許情報における発明者を知識のつながり、技術の群として捕らえることにより、企業における研究技術開発活動を知る上で重要な指標となると考えられる。

本稿では、概念グラフ[廣川 05]を用いて、特許情報における発明者群を解析することにより、研究開発体制を分析する手法を提案する。

## 2. 概念グラフ

廣川ら[廣川 05]は、技術文書群における単語の上位下位関係の定式化として概念グラフを提唱している。たとえば、大学教員の活動概要の文書群に用いた場合、意味のある概念グラフが作成できることを報告している。

概念グラフには、2つの大きな特徴がある。

まず第1に、ある文書集合においてユーザーの与えるクエリに対し、文書集合中での特徴語の上位下位関係を次のような手順で抽出する。

まず、ユーザーの与えるクエリ  $q$  によって決まる文書集合  $D(q)$  における単語  $w$  の特徴量  $s(w, D(q))$  を式(1)で定義する。

$$s(w, D(q)) = \frac{df(w, D(q))}{df(w, U)} \quad (\text{式1})$$

この式は、単語  $w$  を含む文書集合  $D(w)$  と  $D(q)$  の交わりの文書集合  $D(w)$  での割合を表している。また、 $D(w)$  と  $D(q)$  の交わりの  $D(q)$  での割合を  $w$  のもうひとつの特徴量と考える。

$$s_2(w, D(q)) = \frac{df(w, D(q))}{|D(q)|} \quad (\text{式2})$$

を関連度閾値とすると、 $s(w, D(q)) >$  かつ  $df(w, U) < |D(q)|$  となる単語  $w$  を  $D(q)$  の下位の特徴語、 $s_2(w, D(q)) >$  かつ  $df(w, U) > |D(q)|$  となる単語  $w$  を  $D(q)$  の上位の特徴語と呼ぶ。

特徴語間の関連抽出には、二つの特徴語  $u, v$  の  $D(q)$  での共起頻度  $df(u * v, D(q))$  を用いて、 $v$  からみた  $u$  の関連度  $r(v, u)$  として定義する。

$$r(u, v) = \frac{df(u * v, D(q))}{df(v, D(q))} \quad (\text{式3})$$

$df(u, D(q)) > df(v, D(q))$  かつ  $r(u, v) >$  であるような、すなわち  $v$  よりも一般的でかつ関連度が  $v$  を越すような  $u$  を  $v$  の上位語と呼び、 $u$  は  $v$  の上位であると言う。特徴語と同様に特徴語間の上位下位関係も着目する文書集合によって変化する。全ての特徴語対の間の上位下位の関連を計算し、下位の語から上位の語へパスを引いた有向グラフを概念グラフと定義する。これにより、全文書集合中における単語の特徴度と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明確となり、その関係も有向グラフとして明示されることになる。そこで、本技術の特許情報に応用すれば、膨大な量の特許情報から切り出される膨大な量の用語において、上位・下位の概念を取り入れることが可能となる。

概念グラフのもう1つの特徴は、技術記載がされている特許明細書だけでなく、(発明者情報等の)書誌事項や分類といった項目の違う情報についても特徴語として同時に処理し、表示・可視化することができることである。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(出願人)間で切り出された用語との関係を示す特許マップもあるが、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、本技術の特許情報に適用し、発明者間の関係と研究開発体制の分析を行う。

## 3. 概念グラフによる発明者群の可視化

本稿ではまず、2008年7月24日現在に公開されていた公開系特許情報のうち、2000年~2002年に出願された化粧品分野のデータを元に出願番号から出願年を切り出してシステムを構築し、実験・解析を行った(対象8865件)。特許明細書には、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の情報が一定の様式で記載されており、IPCやFI、FTといった技術分類コードが割り当てられている。

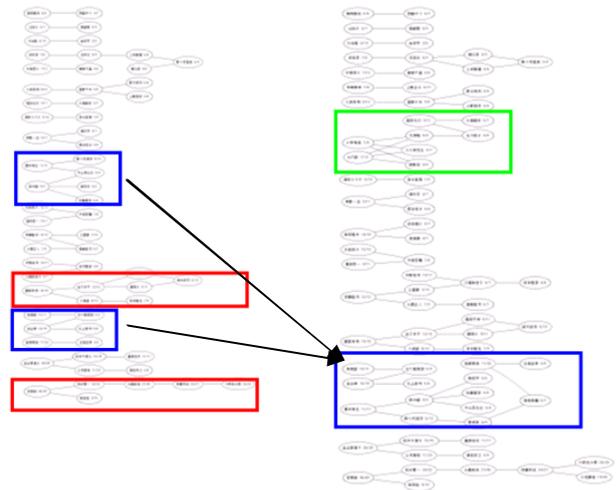
特許明細書群についての概念グラフを構築すれば、検索結果として得られる特許明細群に現れる膨大な量の用語の関連解析が動的に可視化される。また、特許明細書の特定の部分、例えば、課題や解決手段等の項目だけに着目すれば、項目間の違い、項目間の関連を分析することができる。

この特性を利用して、企業名をクエリとして、用語を発明者名として概念グラフを作成したところ、発明者間の関連が明確に表示され群として捕らえることが容易であり、また、発明者群と特定の技術用

語を同時に表示させることにより、特定技術分野における発明者人数の解析が可能であることがわかった。

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 ( a ) 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。各枝の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発明となっている。

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  $= 0.3$  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 ( b ) )。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、個別の製品開発においては細分化された少数人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、特に重要な技術開発においては、有効な権利を少しでも早く開発するため、複数の異なる部門で共同開発を行う可能性があるなどの理由が考えられる。



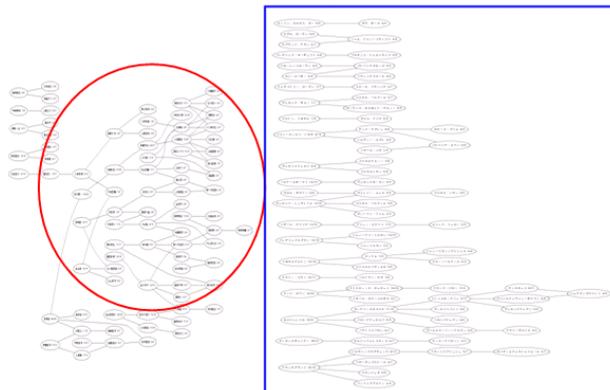
( a )  $=0.5$  ( b )  $=0.3$   
図 1 : A 社における発明者の概念グラフ

#### 4. 関連度閾値による概念グラフの変化率と研究開発研究体制

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、発明者関連度の閾値を 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 )。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。

そこで、発明者関連度を変化させた場合の、 $r(u, v)$  の分布を企業別に表したものが図 3 である。図 3 において、関連度 以上である発明者ペアの全発明者ペアに占める割合は外国企業の方が高く、発明者関連度を変化させても、関連度 以上である発明者ペアの割合に変化が少ない。このことは、外国企業においては、発明者関連度が高い発明者ペアの割合が高く、限られた独立した発明者の下に発明者間の関係がある程度固まっていることを示す。一方、日本企業は、発明者関連度を変化させると、関連する発明者ペアの比率が変化することから、発明者関連度の変化により、発明者ペアのつながりが変化することを意味する。これは、筆者らが提案する概念グラフ ( 図 2 ) において外国企業には変化が小さく、日本企業では変化が大きいことを裏付けている。



( a ) 日本化粧品企業 ( b ) 外国化粧品企業  
図 2 : 企業別発明者の概念グラフ (  $=0.1$  )

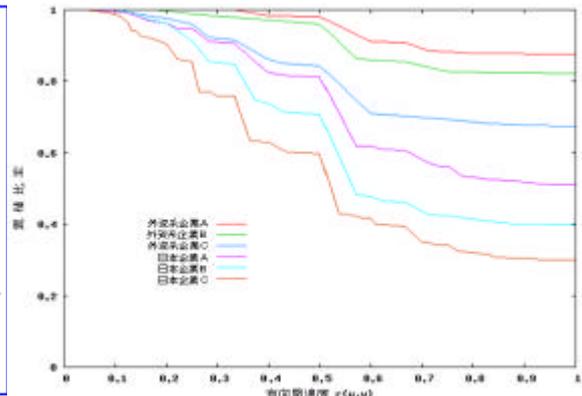


図 3 : 関連度変化と発明者ペアの比率

## 5. まとめと今後の課題

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、日本企業と外国企業を比較することにより、企業によって発明者間の関係が異なることについても示唆され、企業の研究開発体制を知る上で有用なツールとなりうることを示した。

一方、現行の概念グラフには、システム上まだ解決すべき数々の課題がある。

- 1) 関連度閾値を変化させると概念グラフは影響を受けるが、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、年代ごと解析を行う際には、図における特徴語(発明者名)の場所がずれてしまうため、変遷がわかりづらいこと
- 3) 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
- 4) 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と思われる筆頭発明者を中心に解析ができていないこと

今回の検討では、化粧品分野という特定の技術分野における発明者解析にとどまっているが、今後は、他の技術分野における企業発明者とその研究開発体制についての解析を進める予定である。また、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を抽出するために、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。

## 4. 参考文献

- [市川 01] 市川照久,辻 秀一,笠原久美雄,片木孝至 企業内研究所の知的生産性評価法に関する提案と実証,経営情報学会誌,vol. 9, No.4, pp19-32 (2001)
- [後藤 06] 後藤 晃,玄場公規,鈴木 潤,玉田 俊平太 重要特許の判別指標,経済産業研究所,ディスカッションペーパー 06-J-018 (2005).
- [西澤 08] 西澤正己,孫 愛媛,柿沼澄男:日本の論文誌や科研費における研究組織の協力体制や動向の可視化,情報知識学会誌, vol.18, No.2, p.123 (2008)
- [松尾 05] 松尾 豊,友部博教,橋田浩一,中島秀之,石塚 満: Web 上の情報から人間関係ネットワークの抽出,人工知能学会論文誌, Vol.20, No.1, pp. 46-55 (2005)
- [廣川 05] 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)