

技術-効果型パテントマップに基づくシーズ知的探索システムの開発

太田貴久¹ 鶴田雅信¹ 野中尋史² 山本悠二¹ 酒井浩之³ 増山 繁³

1. 豊橋技術科学大学 大学院工学研究科

2. 名古屋大学 研究推進室

3. 豊橋技術科学大学 大学院工学研究科 情報・知能工学専攻

1. はじめに

近年、グローバル化の進展と産業構造の急激な転換に伴い、製造業を営む企業にとってオープンイノベーション戦略¹の重要性が増してきている。オープンイノベーション戦略において、研究開発部門や知財部門の人間は、社外の技術開発状況を絶えず確認しなければならない。このような中、公開特許公報は、豊富な技術情報を含み、年間 30 万件以上出願されている重要な情報源となっている。企業は、これらの特許公報から自社の持つ技術に関連する特許を探し出し、その出願動向を調べる特許調査を行う必要がある。しかしながら、従来型の特許文書検索システムは、使用に際し手間と経験を必要とするため、知財の専門家を持たない企業にとって、特許調査は困難なものであった。そこで、本研究では、知財の非専門家でも容易に利用可能な特許探索システムの開発を行った。

2. パテントマップ

オープンイノベーション戦略における特許調査では、「自社・他社の強みや弱みの分析」や「特定の開発目的や、特定の技術に関する出願動向調査」を行う。これらの調査では、各特許がどのような目的で開発され、どのような技術が用いられているかを抽出し、分析を行う。この際、特許調査の現場では、パテントマップと呼ばれる図がしばしば用いられる。パテントマップとは、「特許情報を整理・分析・加工して図面、グラフ、表などで表したもの」[1]である。特に上記調査の場合では、特許に記述されている発明の効果(発明の目的、「リサイクル性向上」など)と、解決手段(発明が使用する特徴的な技術)の2つを軸として、軸項目に対応する特許の件数をバブルで表したバブルチャートによるパテントマップ(以下、このパテントマップを技術-効果型パテントマップと呼ぶ)が役立つ。

技術-効果型パテントマップの例として、特許庁が作成したアルミリサイクルに関する技術-効果型パテントマップ

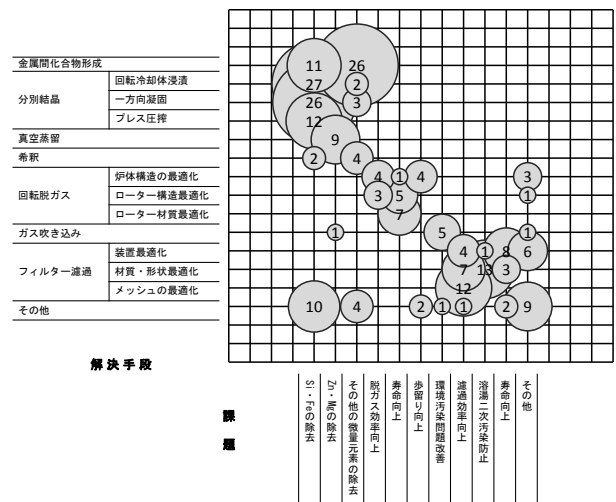


図 1: 技術-効果型パテントマップ

[2]を図 1 にあげる。図1より分かるように、技術-効果型パテントマップは、どのような技術が活発に開発されているかが一目でわかり、さらに、バブルの内容を確認することで、どのような組織がその技術を保有しているのかが容易に分かる。このように、技術-効果型パテントマップは、知財の非専門家でも容易に理解できる非常に有用な図である。また、特許明細書のサポート要件(特許法第 36 条第 6 項第 1 号)により、権利を主張する技術は、発明が解決する課題(効果)を加味しながら審査されるので、技術-効果型パテントマップにより、パテントマップの対象となった分野において、特許がどのように記述されているかを把握することもできる。

しかしながら、現在、このような技術-効果型パテントマップは、知財の専門家によって人手で作成されている。独特の文体で書かれた多数の特許から、適切な効果や技術を抽出することは知財の専門家でも非常に手間のかかる作業である(規模や適用分野にもよるが、1 つのパテントマップを作成するために 2~3 ヶ月かかる場合もある)。そこで、本研究では、このような技術-効果型パテントマッ

¹ 自社技術だけでなく他社が持つ技術やアイデアを組み合わせ、革新的な商品やビジネスモデルを生み出すこと。

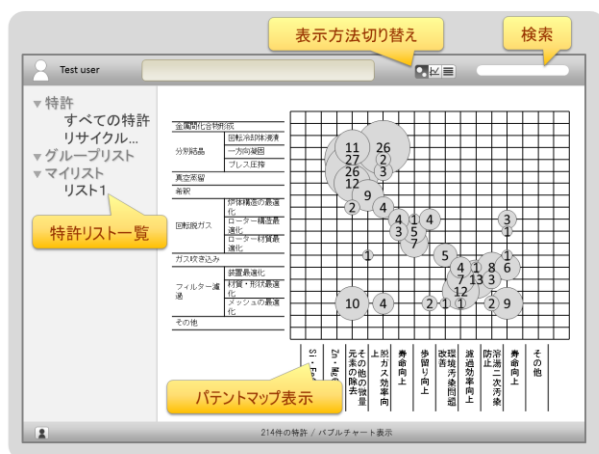


図 2：システム概要

システムを自動的に生成するシステムを構築することで、知財の非専門家でも直感的に特許の出願動向を把握し、技術シーズを探索できるような、高度な特許検索システムを開発した。

3. 従来研究

従来、テキストマイニングを利用したパテントマップ自動作成に関する研究[3]や実用的なシステムとして NRI True Teller[4]等が提案されている。しかしながら、上記の研究をはじめとする従来研究は、tf-idf をはじめとする重みづけを語に行った上で、語を基底とするベクトル(効果や解決手段に分類することは行わない)を生成し、そのベクトルに対して、主成分分析等を施すことで次元数を削減(多くは 2 次元)し可視化するものである。このようなタイプのパテントマップは、次元の削減の際に、情報が大幅に失われており、また、軸の意味の解釈が自明でない等の問題がある。これに対して、本研究で用いる図 1 のような技術-効果型パテントマップは、効果と解決手段により特許を分類しているため、各軸の解釈が明確であり、情報も失われていない。

さらに、直接、図 1 のような技術-効果型パテントマップを自動生成する従来研究は存在しないが、その基礎となる従来研究は、いくつか存在する。その中で、特許文書から効果や特長に相当する表現を抽出する研究として、文献[5,6]がある。これらの手法は、一部の係り受けに使用される表現(「ことにより」や「改善する」など)を辞書に定義し、抽出するものである。そのため、網羅性に欠けるといった問題点がある。

一方、教師あり機械学習を利用した手法として、文献[7,8]の手法などがある。これらの手法では、学習のために「語」などの素性を用いているが、各特許の重要な部分を表す表現は技術分野ごとに異なるため、新たな技術分野

に対応するためには、新しい学習データや素性などが必要になる。ここで、技術分野は明確に分類できるものでなく、また、新しい技術分野も次々と生み出されるため、多くの技術分野に対応することは手間がかかる。

また、これらの基礎的な研究は、技術-効果型パテントマップ生成を意識した研究ではないため、意味のまとめあげを行っていないなど、軸項目に適した表現を抽出するものではない。これに対して、我々は、実用的な技術-効果型パテントマップの生成を意識した手法[9～14]を実装した特許探索システムを開発した。

4. システム概要

本研究で開発したシステムの概要を述べる。本研究では、システムをウェブアプリケーションとして実装する。ウェブアプリケーションとして実装することで、今後の機能追加(アップデート)が容易となり、さらに、OS に依存せずに実行可能なシステムを構築できる。

本システムの動作の概要と、表示される画面のイメージを図 2 に示す。簡単に動作を説明すると、本システムにおいて、ユーザーは、特許の集合である特許リストを中心に作業を行う。ユーザーはログイン後、興味の対象とする特許リストを選択する。その後、選択した特許リストを対象に、技術-効果型パテントマップをはじめとする表示方法の切り替えや、検索を行うことで、特許の効率的な探索を行う。さらに、本システムは、知財の専門家を持たない中小企業での利用も想定しているため、これらの企業が容易に知財コーディネーターへ相談できる機能を実装している。

5. 技術-効果型パテントマップ生成手法

本システムの核である技術-効果型パテントマップの生成手法について説明する。パテントマップ生成は文献[9]

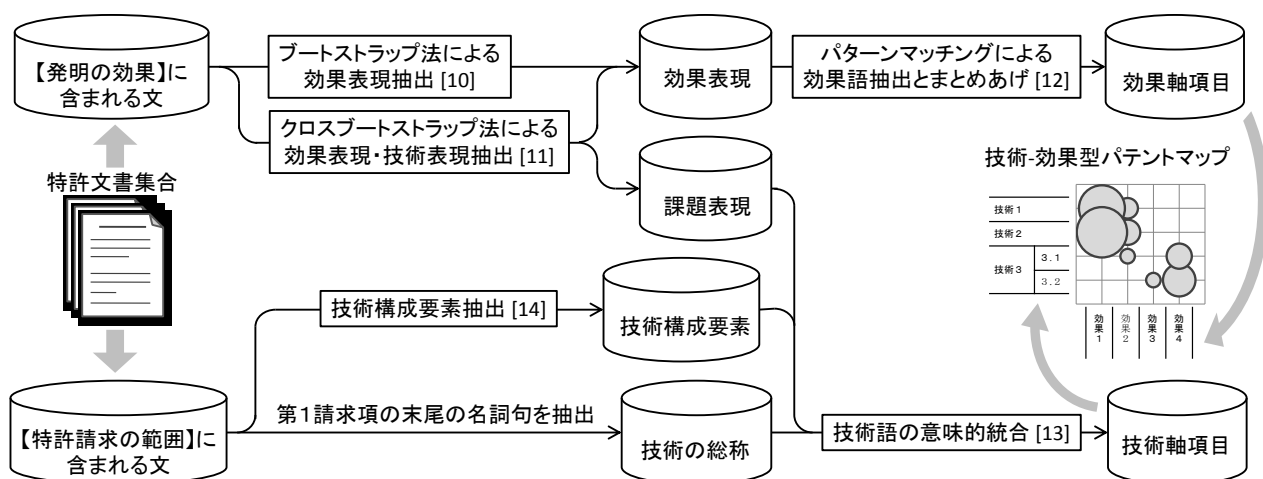


図 3：技術-効果型パテントマップ生成手法の概要

に基づいて行われる。特許文書集合からパテントマップを生成するまでの手順の概要を図 3 に示す。なお、図 3 において、特許文書中の文から最初に抽出される「効果表現」、「課題表現」、「技術の総称」と「技術構成要素」は、以下のように定義する。

- 効果表現
直接的なユーザーの利点に相当する表現(例:「メンテナンスを最小限にすることができる」)
- 課題表現
効果表現の根拠となる技術上のメリットに関する表現(例:「粘着性物質の付着を防止」)
- 技術の総称
その発明の内容を表す技術的な意味を持つ語(例:「熱可塑性樹脂」)
- 技術構成要素
権利範囲を定める発明の構成要素(例:「生体情報を計測する計測手段」)

以下では、図 3 に示した各手法[10～14]について簡単に説明する。

5.1. ブートストラップ法による効果表現抽出

先にも述べたように、特許文書の重要な表現は技術分野ごとに大きく異なるため、文献[7,8]のような手法は、技術分野ごとに学習データを作成する必要がある、その作成には手間がかかる。これに対して、本手法は、特許文書において、一般的に用いられる表現のみを利用し、多くの技術分野の特許に対応可能な手法である[10]。簡単に説明すると、本手法では、手がかり表現と呼ばれる効果を表す語の後に出現しやすい表現を 1 つ(「ができる。」)与え、そこから効果表現を取得し、さらに、取得した効果表現から新たな手がかり表現を取得する、といった処理を繰り返す手法である。実験では、精度 0.780、再現率

0.776 で効果表現を取得できることを確認した。

5.2. クロスブートストラップ法による効果表現・課題表現抽出

本手法は、技術上のメリットに相当する課題表現と効果表現を同時に考慮して抽出する手法である[11]。本手法は、特許明細書の【発明の効果】では、課題表現と効果表現が同じ文に出現することが多いことを利用した手法で、先の手法[10]のように手がかり表現を 2 種類用いることで、効果と課題の両表現を同時に取得する。簡単に説明すると、最初に「課題表現の手がかり表現」を用いて「課題表現」を取得し、そこから、「効果表現の手がかり表現」を取得する、といったように、一方の表現から他方の手がかり表現を取得する手法である。実験では、本手法により、精度 0.960、再現率 0.790 で効果表現・課題表現を取得することができることを確認した。

5.3. パターンマッチによる効果語抽出とまとめあげ

上記の 2 手法[10,11]によって取得した効果表現は、そのままでは長すぎるため、パテントマップの軸項目として適切ではない。そこで、本手法により、効果表現中からノイズを削除し、意味的なまとめ上げを行う[12]。本手法は、はじめに、抽出した効果表現に対して係り受け関係を考慮した 3 つのパターンを適用することで効果語(効果軸項目)と技術語(技術軸項目)を取得する。さらに、軸項目を構成する形態素を素性としたクラスタリングを行うことで意味のまとめ上げを行う。実験では、効果語と技術語の抽出については精度 0.85 を、意味のまとめ上げについては、精度 0.9、再現率 0.9 を達成した。

5.4. 技術語の意味的統合

特許において、権利を請求する範囲は可能な限り広くすることが一般的であるため、技術語は効果語より抽象的な表記がされることが多い。このような技術語の意味のま

とめあげを行うためには、形態素の一致をみるだけでは不十分である。そこで、本手法では、日本語 WordNet を利用し、はじめに特許内で技術語間の上位・下位関係を推定し、その後、異なる特許間で語を比較することで統合を行う[13]。これは、同一特許内で、【特許請求の範囲】に書かれた技術語は、【発明の詳細な説明】に書かれた技術語の上位語であることが多いという特許文書の特徴を利用した手法である。本手法に対する実験では、特許内の語の対応は、精度 0.830、再現率 0.723 を、特許間の語の対応は、精度 1.0、再現率 0.576 を達成した。

5.5. 技術構成要素抽出

何を 1 つの発明とするかは出願人に依存するため、発明の一部分である構成要素が重要な場合がある。そこで、本手法は、エントロピーに基づき【特許請求の範囲】において構成要素を区切る表現(例えば、「と、」)を抽出し、その区切り表現(文献[14]内では「区切り文字」)をもとに請求項から技術構成要素を取得する[14]。実験では、本手法は、精度 0.76、再現率 0.66 を達成した。

6. おわりに

本研究では、特許文書の特徴を考慮した様々な手法を組み合わせることで、知財の非専門家でも容易に利用可能な特許探索システム(シーズ知的探索システム)を開発した。今後、効果語と技術語の抽出性能をさらに高めるとともに、知財の専門家と非専門家の両方を対象に、実際にシステムが特許調査に有効か否かを検証する必要がある。

謝辞

本研究は、総務省戦略的情報通信研究開発推進制度(SCOPE)地域 ICT 振興型研究開発、日本学術振興会科学研究費基盤(C)、及び、文部科学省グローバル COE の支援に基づき行われた。

参考文献

[1] 特許庁, "特許マップとは - 技術分野別特許マップについて(機械 6 焼却炉技術)", http://www.jpo.go.jp/shiryou/s_sonota/map/kikai06/map/map.htm
 [2] 工業所有権情報・研修館, "資源循環型社会の構築に向けて", <http://www.inpit.go.jp/blob/katsuyo/pdf/chart/fippan08.pdf>
 [3] Hideyuki Uchida, Atsushi Mano and Takashi Yukawa, "Patent Map Generation Using Concept-Based Vector Space Model", Proceedings of the Fourth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, 2004.
 [4] NRI True Teller, <http://www.trueteller.net/>
 [5] 石川大介, 石塚英弘, 宇陀則彦, 藤原譲, "特許文献における因果関係の抽出と統合", 情報知識学会誌,

Vol.14, No.4, pp.105-118, 2004.

[6] 西山莉紗, 竹内広宣, 渡辺日出雄, 那須川哲哉, 武田浩一, "技術文書マイニングのための特長表現抽出", 第 22 回人工知能学会全国大会, pp.3K-2, 2008.
 [7] Hidetsugu Nanba, Tomoki Kondo, and Toshiyuki Takezawa, "Hiroshima City University at NTCIR-8 Patent Mining Task", Proceedings of the Eighth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, pp.354-358, 2010.
 [8] Risa Nishiyama, Yuta Tsuboi, Yuya Unno and Hironori Takeuchi and Toshiyuki Takezawa, "Feature-Rich Information Extraction for the Technical Trend-Map Creation", Proceedings of the Eighth NTCIR Workshop on Research in Information Access Technologies Information Retrieval, pp.318-324, 2010.
 [9] 増山繁, 野中尋史, 坂地泰紀, 小林暁雄, 鈴木佑輔, 太田貴久, 酒井浩之, "自然言語処理を利用した効果-技術型特許マップの自動生成手法の開発", 第 1 回特許情報シンポジウム, 2010.
 [10] 酒井浩之, 野中尋史, 増山繁, "特許明細書からの技術課題情報の抽出", 人工知能学会論文誌, Vol.24, No.6, pp.531-540, 2009.
 [11] 坂地泰紀, 野中尋史, 酒井浩之, 増山繁, "Cross-Bootstrapping: 特許文書からの課題・効果表現対の自動抽出手法", 電子情報通信学会論文誌, D, Vol.J93-D, No.6, pp.742-755, 2010.
 [12] Hirofumi Nonaka, Akio Kobayashi, Hiroki Sakaji, Yusuke Suzuki, Hiroyuki Sakai and Shigeru Masuyama, "Extraction of the Effect and the Technology Terms from a Patent Document", IEEE The 40th International Conference on Computers & Industrial Engineering (CIE40), cie214jp-1, 2010.
 [13] Akio Kobayashi, Hirofumi Nonaka, Hiroyuki Sakai and Shigeru Masuyama, "An Automatic Thesaurus Construction Method for Technical Terms in Patent Maps", IEEE The 40th International Conference on Computers & Industrial Engineering (CIE40), cie248jp-1, 2010.
 [14] Yusuke Suzuki, Hirofumi Nonaka, Akio Kobayashi, Hiroyuki Sakai and Shigeru Masuyama, "Extraction of Technology Terms from Patent Specifications for Technology-effect Type Patent Map Generation, ITC-CSCC 2010, pp.725-728, 2010.