

分析者の意図を反映した 対話的な特許マップの作成技術

試行錯誤を要する特許マップ作成作業の支援技術

株式会社日立製作所 中央研究所

佐藤 祐介

PROFILE

2004年（株）日立製作所入社。特許文献を対象とした文献重要性評価、検索支援、半自動分類など、マイニング技術の研究に従事。

✉ yusuke.sato.pr@hitachi.com ☎ 042-323-1111

株式会社日立製作所 中央研究所
東京工業大学精密工学研究所客員教授

岩山 真

PROFILE

1992年（株）日立製作所入社。以来、文書検索、文書分類、自然言語処理等の研究に従事。また、NTCIRにおいて特許検索用テストコレクションの作成に携わる。2009年度特許版産業日本語委員会委員。

✉ makoto.iwayama.nw@hitachi.com ☎ 042-323-1111

1 はじめに

注目する技術分野に関する特許文献を収集し分析することは、経営戦略の方向性の決定や新しい技術の発掘に不可欠である。FIやFターム等の特許分類やキーワード検索により収集した大量の特許文献を、出願人や出願年、課題や解決手段の違いにより分類してその技術動向を分析することで、自社・他社の技術的な強み、弱みを把握し、次に行うべき研究開発の指針を立てることが可能となる。

このような技術動向調査では、収集した特許文献集合に対しさまざまな統計手法を用いて分析・集計を行い、特許マップを作成する場合が多い。特許マップにはさまざまなバリエーションがあるが、例えば以下のようなものが挙げられる。

- ・ 出願人の年ごとの出願件数を折れ線グラフに表すことでその分野への注力度合いを把握
- ・ 特許文献間の引用関係を可視化することで技術の派生関係を俯瞰

本稿では、こうした特許マップの中でも、要素技術や課題、解決手段の違いごとに分類するマップを作成する技術に関する筆者らの研究を紹介する。試行錯誤が必要となる特許文献の技術分け作業を、簡単な操作と少ない

作業量で行うことが可能なツールを実現した。なお、本稿ではこのような技術の違いによって分類されたマップのことを技術分類マップと呼ぶこととする。

2 技術分類マップの特徴

技術分類マップの例を図1に示す。技術分類マップでは、特許文献を複数の観点で分類して配置する。図1は、「プラズマディスプレイパネル」に関する特許文献を「課題」と「解決手段」という2つの観点で分類している。観点には、「課題」と「解決手段」が一般的であるが、この他にも「要素技術」や「発明の効果」、「産業上の利用分野」等を用いることもある。得られた技術分類マップ内の特許文献の分布から、競合他社が注力している技術や、また逆に未着手の技術を見つけることが可能となる。

		← 解決手段 →		
		手順	駆動回路・システムの構成	信号処理・制御
↑ 課題 ↓	表示品質の改善	特開平3-1 特開平3-2		特開平1-1
	低消費電力化		特開平2-1	
	高信頼性化		特開平2-2 特開平2-3	特開平1-2

図1 「プラズマディスプレイパネル」の技術分類マップの例

3

技術分類マップ作成
技術の研究の背景

技術動向調査は膨大な件数の特許文献を対象とする場合が多い。例えば新規事業に参入する際には、場合によっては数千件から数万件の先行技術に関する特許文献を収集し、要素技術ごとの優劣を分析する必要がある。しかしながら、新たな分野にどういった要素技術が含まれており、どういった観点で技術分類マップを作成すべきかの判断は、当該分野の専門家以外には難しい。また、FIやFターム等の特許分類が1つの基準となり得るが、分析者が意図する体系と必ずしも合致していない場合も少なくない。こうしたことから、新たな技術分野を対象とした技術分類マップの作成作業は試行錯誤を伴うことは想像に難くない。未知の分野の膨大な特許文献に対して、図1にある「表示品質の改善」や「手順」などの技術項目を分析の初期段階から設定してマップを作成していくのは困難を極める。実際には、ある程度の件数の特許文献を読んだ後にどういった技術項目が適切であるかを吟味し、さらに他の特許文献を読み進めていく過程で技術項目の取捨選択や特許文献の分類しなおし等を行いながら技術分類マップを作成することになるのが一般的である(図2)。

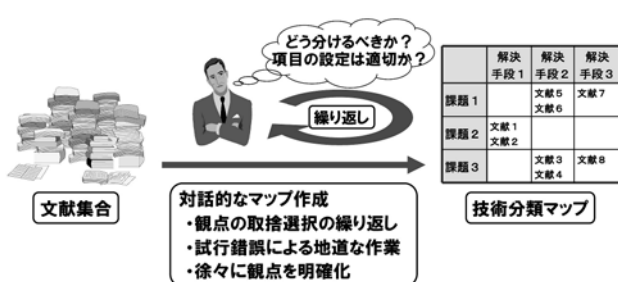


図2 技術分類マップ作成のプロセス

このような作業には、大量の人手と時間が必要であるため、自動化できるツールが望まれている。図1の例で言えば、「課題」、「解決手段」それぞれの項目名を文献集合の特徴から見つけ出し、その項目に従って特許文献を該当箇所に仕分ける作業を自動化できることが望

ましい。

文献集合の自動的な分類には、テキストマイニング技術の1つであるクラスタリングが広く用いられている。クラスタリングは、単語の頻度等の情報により内容の似た文書どうしをグループ化する方式であるが、分析者の意図を反映した分類がなかなか行えないという問題があった。この問題に対して、クラスタリング技術に改良を加えることで、分析者の分類の意図を試行錯誤的に反映可能な技術分類マップ作成技術を開発した。

4

技術分類マップ作成
ツール

図3に著者らが開発を進めている技術分類マップ作成ツールのイメージを示す。ツールは2つの画面から構成されている。左上が特許文献の分類結果の分布を示す件数表示画面であり、右下が各観点に分類された特許文献の内訳を示す詳細表示画面である。

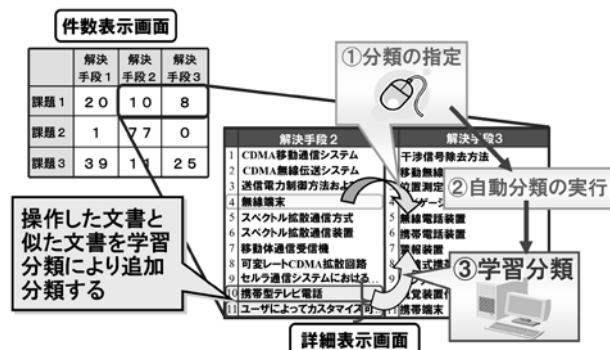


図3 技術分類マップ作成ツールのイメージ

分析者は、ツールが自動分類した結果の概要を件数表示画面で俯瞰し、その詳細を詳細表示画面で確認する。ここで、従来は分析者の意図どおりの分類結果が得られなかった場合は、分類のための初期設定からやり直す必要があった。例えば、各文献や表中の各項目の特徴を表す単語の選び直しなどである。しかし、単語には「プリンタ」と「印字装置」のような数多くの同義語が存在するため、それら全てを網羅するのは容易ではな



い。また、こうした単語の取捨選択により分類結果が必ずしも改善するかどうかはわからないという問題があった。

本ツールでは詳細表示画面中の特許文献の分類先を直接指定することにより、ツールの分類結果を改善する方式とした。必要な操作は以下の2つである。

操作①：特許文献の分類先の指定

操作②：自動分類の実行

分析者はツールが自動で分類した特許文献のうち、「A. 正しく分類された特許文献」と「B. 誤って分類されている特許文献」であることが容易に判断のつくものから数件を選ぶ。Aについては、正解であることをツールに指示し（チェックボックスにチェックを入れる）、Bについては正しい分類先に修正する（操作①）。ある程度の件数の文献に対する指定が終了した時点で、ツールによる自動分類を実行する（操作②）。ツールは操作①により指定した文献の傾向から、分析者が意図している各技術項目の特徴を学習して文献集合全体を分類し直す。その結果、ツールの自動（学習）分類により、未着手の文献のうち分類指定を行った文献と類似の文献をツールが自動で分類する（図3の③）ため、その分のマップ作成作業量を減らすことが可能となる。

このように、判断が容易な数件の文献の分類先の指定とツールによる自動分類を段階的に繰り返していくことで、全件の文献を精査するよりも少ない作業量で技術分類マップを完成させることができる。

5 技術分類マップ作成ツールの性能評価

本ツールによりどの程度の特許文献の分類先を指定すれば十分な分類精度が得られるのかを評価した結果を述べる。

国際特許分類（IPC）におけるA61B（診断；手術；個人識別技術）に属する特許文献のうち、図4に示した4つのメイングループをもつ各500件程度を抽出した計2268件の文献集合を作成した。この集合に対して、

ランダムに選択した数文献の分類先の指定と自動分類を繰り返した場合の分類指定件数と分類精度の関係を示したグラフが図4である。グラフ中の各点は、1つ前の（点での）分類結果の文献に指定操作を追加し、再度自動分類を行った際の分類精度を示す。

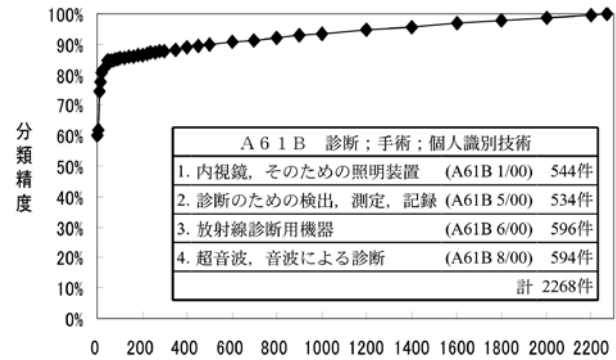


図4 文献の分類先の指定と自動分類を交互に繰り返した場合の、分類指定件数と分類精度の関係

全件が2268件であるのに対して、約200件程度の文献の分類先の指定を行うことで、90%の分類精度を得ることができた。これはつまりは、全件の10%程度を読んで仕分ければ、ツールによる自動分類により残り80%を正確に分類できることが可能であることを意味しており、技術分類マップ作成作業量の大幅な削減が期待できる。

図4に示したA61Bに関する文献集合では大きな効果を得ることができたが、文献集合の性質によっては性能に差が出ることがある。他の文献集合により評価した結果、以下のような傾向があることがわかった。

■対象とする技術分野により性能が異なる

技術分類マップを作成する分野により、必要な分類指定件数が異なる傾向があった。図5は、年間1万件以上の出願のあるIPCのサブクラス下の4つのメイングループから抽出した約2000件（例えば、B41Jは「B41J 2/00:495件」、「B41J 5/00:492件」、「B41J 11/00:522件」、「B41J 29/00:594件」）を本ツールで分類した際の、90%の分類精

度が得られるのに必要な分類指定件数の割合を示した棒グラフである。対象とする技術分野によって必要となる件数に大きく差が出ていることがわかる。同義語や新語など、語彙の多い分野ほど必要な件数も多くなる傾向が見られた。

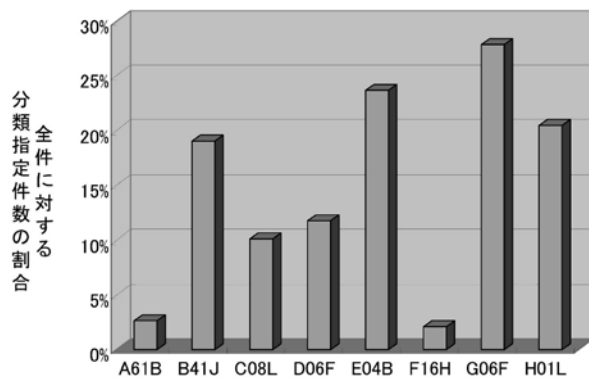


図5 90%の分類精度に必要な分類指定件数の割合の技術分類ごとの違い

■詳細な技術进行分类する程、必要な件数も多くなる

特許分類における展開記号や分冊識別記号といったように、技術が詳細化するにつれて十分な精度を得るのに必要な件数が多くなる傾向があった。図6は、IPCのH01L 21/00下のドット数ごとに3つのサブグループを抽出した文献集合300件（例えば、表1）进行分类した際の90%の分類精度が得られるのに必要な件数の割合を示した棒グラフである。

ドット数 4	ドット数 6
H01L 21/20	H01L 21/8242
H01L 21/22	H01L 21/8244
H01L 21/26	H01L 21/8246

表1 ドット数ごとに抽出した例

ドット数が増加する、つまりは技術の詳細度が増すほど、十分な精度を得るのに必要な文献数が多くなることがわかる。

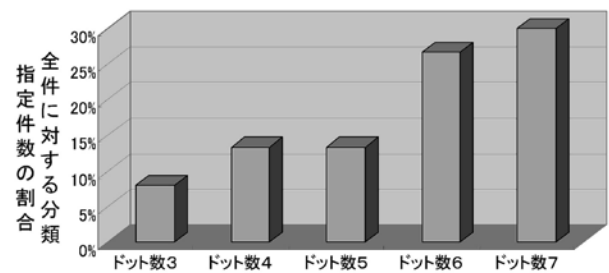


図6 90%の分類精度に必要な分類指定件数の割合の技術分類の詳細度（ドット数）ごとの違い

6 おわりに

本稿では、大量の文献集合に対して分析者の意図どおりの技術分類マップを作成するためのツールの開発について紹介した。ある一定量の文献を本ツールによる自動分類と併用して分類していくことにより、全件を読んでマップを作成していた従来よりも大幅に作業量を減らせる可能性があることがわかった。評価の結果、最も良い場合で全件の20%程度の文献を仕分けることで、90%の文献进行分类するのと同じ効果が得られることがわかった。一方で、対象とする技術の分野や詳細度合いによって、必要な分類指定件数に差があることもわかった。

今後は、より少ない文献数で十分な精度が得られるための改良や自動分類結果の理解支援等の周辺機能の開発を進めていく予定である。

7 参考文献

- [1] 佐藤祐介, 岩山真. 半教師有りクラスタリングを適用した対話型文書分類技術の提案. 情報処理学会 FI 研究会, Vol.2009-FI-95, 2009.
- [2] Yusuke Sato, Makoto Iwayama. Interactive Constrained Clustering for Patent Document Set. Proceedings of PalR09, 2009.