

特許出願技術動向調査報告書の自動更新に向けて

Towards Automatic Update of Patent Application Technical Trend Survey

難波英嗣
Hidetsugu Nanba

中央大学
Chuo University

The Japan Patent Office carries out the patent application technical trend survey for fields in which the development of technologies is expected, or fields to which social attention is being paid, and more than 250 reports and their lists of research papers and patents have been released up to now. Unfortunately, these reports will become obsolete as time goes on. Therefore, we investigate automatic update of patent application technical trend survey reports automatically. In this paper, we propose a method to update lists of research papers and patents, and experimentally confirm the effectiveness of our method.

1. はじめに

特許庁では、市場創出に関する技術分野、国の政策として推進すべき技術分野を中心に、今後の進展が予想される技術テーマを毎年10件以上選定し、特許出願技術動向調査を実施し、調査報告書を公開している¹。この報告書は、各テーマに関する国内外の論文と特許を分析して作成されており、この調査報告書を読むことで、あるテーマについて注力すべき分野を見ついたり、有力企業や共同研究先を探したりするといった目的に利用することができる。しかし、選定される技術テーマは毎年異なるため、これらの報告書は、公開された直後は最新の技術動向を含んだ内容であっても、時間経過と共にその内容が古くなるという問題がある。そこで、本研究では、過去に人手で作成された調査報告書を自動更新することを目的とする。

特許出願技術動向調査報告書には、報告書本文の他に、その報告書を作成するために収集された論文と特許のリストがある。また、調査報告書を作成する際には、分析軸と呼ばれる技術的な複数の観点が定義されており、リスト中の論文と特許は、それぞれどの分析軸に該当するのかが記載されている。このリストは、テーマごとに1つ以上の Excel ファイルとしてまとめられているが、テーマごとにフォーマットが異なるため、研究目的で利用するためには、新たに人手で整備し直す必要がある。本研究では、現在この整備を行っている。また、その一部を用いて特許出願技術動向調査報告書の自動更新に関する実験を行ったので、その結果について報告する。

2. 関連研究

あるトピックに関する複数の論文や特許の内容を要約し、技術動向マップとしてまとめる研究やプロジェクトがこれまでにいくつか行われている。国立情報学研究所主催の評価ワークショップ NTCIR では、学術論文と特許から要素技術(解決手段)とその効果に関する表現を抽出し、技術動向マップを作成する特許マイニングタスクが実施されている[Nanba 2010]。酒井らは、特許マップの生成を視野に入れ、特許から技術課題情報を抽出

する手法を提案している[酒井 2009]。これらの試みは、いずれも技術動向マップを作成する上で重要であるが、実際に人間が技術動向マップを作成するには、要素技術や技術課題を詳細に分類する必要がある。

複数の論文をまとめるこの他の試みとして、Hashimoto らと Hou らの研究がある。Hashimoto らは、共通タスクのワークショップに着目し、タスク参加者による投稿論文の集合からレビューマトリックス(参加システムの一覧表)を自動生成する手法を提案している[Hashimoto 2017]。Hou らは、学術論文からタスク名、データセット、評価尺度、評価値を抽出し、リーダーボードを自動作成する手法を提案している[Hou 2019]。いずれの研究も、共通タスクのデータを用いた論文を対象としているため、論文は要素技術と評価値という観点で比較しやすいという利点がある一方で、共通タスクのデータを用いない論文については、たとえ論文の中で扱うタスクが類似していても、比較できないという問題がある。

複数の論文や特許をまとめるには、まとめるための観点を見つけることが必要であるが、これは、人間にとっても非常に難しいタスクである。そこで、すでに人間が作成した要約を元に、新たな要約を作成するという試みがある。飯沼らは、版の異なる専門書籍を対象に、古い版の書籍中の章を更新することで新しい版の対応する章の生成を試みている[飯沼 2014]。本研究でも、同様の考え方に基づいた要約作成を試みる。すなわち、過去に人間が作成した特許出願技術動向調査報告書、その元になった論文と特許、さらに分析の観点を入力とし、報告書に含まれていない論文や特許の情報を追加することで、最新の情報を含んだ報告書の自動出力を目指す。

3. 特許出願技術動向調査報告書の自動更新

3.1 特許出願技術動向調査報告書とは

特許庁は、平成16年以降250件以上の技術テーマについて特許出願技術動向調査を実施し、報告書を作成・公開している。その詳細について、平成29年度の技術テーマ「自動走行システムの運転制御」を例に説明する。このテーマでは、国内外の特許8,832件および論文1,826件から調査報告書が作成されている。また、1節でも述べたとおり、技術分析軸と呼ばれる技

連絡先: 難波英嗣, 中央大学理工学部, 〒112-8551 東京都文
京区春日 1-13-27, 03-3817-1883, nanba@kc.chuo-u.ac.jp

¹ <https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/index.html>

術的な観点で定義され、各特許と論文がどの観点に属するの
かを一覧表としてまとめている。「自動走行システムの運転制御」
の場合、表 1 からわかるとおり、技術分析軸の一つに、運転支
援システムという大きな技術課題があり、その中に運転負荷軽
減システム、さらにその中に車線維持支援や、駐車支援など、
より詳細な技術がある。このトピックには、分析軸が 243 個定義
されている。

このテーマでは、分析軸には技術課題に関するものと解決手
段に関するものがある。技術課題に関するものは、表 1 に示す
ようにある技術を用いることで、どのような改善が行われるか
についてまとめたものである。一方、解決手段に関するものは、自
動走行システムを実現するための要素技術についてまとめたも
のである。

表 1 自動走行システムの運転制御に関する技術分析軸の一部

運 転 支 援 シ ス テ ム	運 転 負 荷 軽 減 シ ス テ ム	車線維持支援 (LKAS)		D11
		定速走行・車間距離 制御(ACC)	先行車軌跡 追従	D121
			その他	D12X
		駐車支援	自動パー キング	D131
			その他	D13X
		車線変更支援		D14
		合分流支援		D15
		右左折支援		D16
自動発進/自動停止 支援(信号機、 停止線)		D17		

文献リストには、特許の出願番号や論文の書誌情報に加え、
どの技術分析軸の属するのかの情報も記載されている。表 2 は
その一部を示したものである。表において、JP から始まるコード
は日本国特許を表している。また、JP-A-2011-141747 には D16
の技術分析軸が該当している。これは、表 1 と合わせると、この
特許が右左折シーンにおける支援機能に関するものであること
がわかる。同様に、JP-A-2011-141802 は、先行車両の移動軌跡
に追従して走行する技術や、信号や停止線といった交通法規
に従っての発進・停止を行うための支援機能に関するものであ
ることがわかる(表 2)。

表 2 技術動向調査報告書の例

文献(特許)	D121	D16	D17
JP-A-2011-141747		1	
JP-A-2011-141802	1		1

これらの文献を元に、709 ページからなる報告書が作成され
ている。なお、Web 上には、80 ページの概要のみが公開されて
いる。報告書本文および文献リストを入手するには、特許庁に
問い合わせる必要がある。

3.2 特許出願技術動向調査報告書の自動更新

3.1 節では、テーマ「自動走行システムの運転制御」を例に特
許出願技術動向調査報告書および文献リストについて説明し
た。このような報告書を現在の人工知能技術を用いて完全自動
で作成するのは不可能であるが、平成 29 年に公開された報告
書から、平成 30 年以降に出願された特許や発表された論文の

情報を加えることで報告書を自動更新することは、決して実現
不可能ではないと考えている。ただし、自動更新の実現には、
以下に示すいくつかの課題がある。

(課題 1) 分析軸ごとに関連論文・特許を自動収集

最新の情報を含んだ報告書を作成する第一歩として、分析
軸ごとに関連論文や特許を収集する。これは、すでに分類済の
論文や特許を訓練用データとし、機械学習手法を用いて実現
することができる。ただし、分析軸によっては、特許や論文が数
件しか存在しないものもあるため、このような分析軸については、
別途検討する必要があると考えられる。

(課題 2) 特許・論文集合から観点の候補を自動的に検出

報告書が作成されて時間が経過するにつれ、定義されていた
分析軸には当てはまらない特許や論文が出現する可能性がある。
そこで、そのテーマに関する特許や論文から、新しい分析
軸を検出する手法が必要となる。

(課題 3) 文書生成技術を用いた報告書の自動更新

課題 1 や課題 2 は、いずれも報告書を作成するための特許
や論文を収集するためのものであるが、こうして収集された特許
や論文を元に、文書生成技術を用いて報告書の自動更新を行
う。これには、元となる報告書を解析し、追加すべき特許や論文
を報告書のどの個所に個所に追加するのかを決定し、それに
基づいて元の報告書の文を言い換えたり、場合によっては新た
な文を追加したりする必要がある。

これらの課題のうち、本稿では、課題 1 について実験を行った。
すなわち、すでに分類済の論文や特許を訓練用データとし、
fastText[Joulin 2017]を用いて、どの程度、自動分類が可能であ
るか調査した。fastText は、単語埋め込みの学習やテキストをあら
かじめ決められたカテゴリに分類できる入力層、隠れ層、出力
層の 3 層からなる DNN(Deep Neural Network)をベースとした分
類器である。課題 1 でも述べたとおり、分析軸によって、属する
文献が大きくなる。そこで、分析軸に属する件数と特許の
分類精度について調査した。次節では、この調査について報告
する。

4. 実験

本研究では、前節で述べた課題 1 について調査を行った。

実験用データ

本研究では特許庁で公開されている技術動向調査報告書
を用いて実験を行った。実験には、「自動車」、「ロボット」、「水処
理」、「情報セキュリティ技術」、「熱電変換」、「ビッグデータ分析
技術」、「次世代無線 LAN」の 5 つのトピックに関する特許を用
いた。データの数を表 3 に示す。

表 3 実験用データ

トピック	特許件数	分析軸数
水処理	1611	304
ビッグデータ分析技術	4120	104
自動走行システム	16016	243
次世代無線 LAN	3286	176
熱電変換	1296	220

評価尺度

分析軸ごとに何件の特許が含まれているのかを調べ、件数ご
とに分類精度(=再現率)を集計した。

実験結果

分析軸ごとに含まれる特許件数を調べた。結果を図 1 に示す。この図から分かる通り、1 分析軸あたり 25~29 件の特許を含む事例が突出して多い。一方で、分析軸にたった 1 件だけ特許が含まれるケースが 105 軸(全体の約 5%)あり、多いもので 1 軸あたり 1481 件の特許が含まれるものも存在していた。

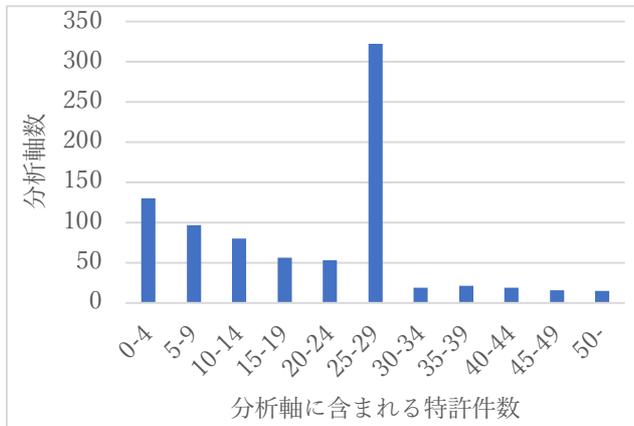


図 1 分析軸ごとに含まれる特許の件数

次に、分析軸の大きさ(軸に含まれる特許数)と分類精度の関係を図 2 に示す。図から、分類軸に含まれる特許数が 5~9 件の時に分類精度が 0.739 と最も高く、分類軸に含まれる特許数が増えるにつれ、分類精度が下がっていることが分かる。一般的には分析軸に含まれる特許件数が増えるほど訓練用データが増えるため、分類精度が増加すると予測されるが、この予測と正反対の結果が得られている。この原因について考察する。表 1 に示すとおり、分析軸は技術課題や解決手段であることが多い。特許を分析軸ごとに分類する上で、判断に必要な個所は特許中のごく一部に過ぎないが、今回は分類する歳、特許全文を利用している。1 分析軸あたりの特許数がそれほど多くないことから、分析軸への分類の根拠となる単語を正しく認識したというよりも、分類対象となる特許と良く似た特許(例えば同一発明者のもの)が訓練データの中に入っているかどうかが強影響したのではないかと推測される。今後は、例えば酒井ら[酒井 2009]の手法を用いて特許から技術課題について記載されている個所を抽出したり、NTCIR 特許マイニングタスク[Nanba 2010]のように要素技術(解決手段)を抽出したりした上で訓練することで、より適切な分類が可能となるのではないかと考えている。

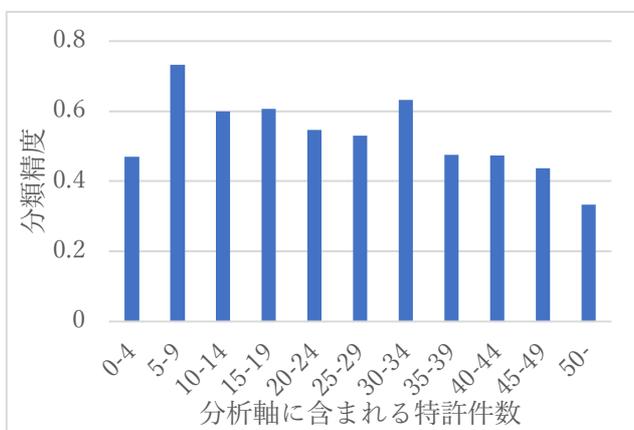


図 2 分析軸の大きさと分類精度の関係

5. おわりに

本研究では、特許庁の技術動向調査報告書の特許リストを自動更新する手法を提案し、文書分類技術を用いて、技術分析軸から特許を自動抽出する実験を行った。また、分析軸によっては文書分類を行うのに十分な数の特許が存在しない場合があることを考慮して、文書分類に利用できるデータ数に応じて分類精度がどの程度影響を受けるのかを調査した。データセットには特許庁の技術動向調査報告書の中から、「自動車」、「水処理」、「熱電変換」、「ビッグデータ分析技術」、「次世代無線 LAN」の 5 つのトピックに関する特許を用いた。調査の結果、分析軸に含まれる特許が 5~9 件の時、分類精度 0.739 が得られたが、分析軸に含まれる特許数が増えるにつれ、分類精度が低下することが分かった。

なお、筆者らは、現在「知財工学会」を設立中である。これは、知財活動に関する方法論を情報学等の工学的観点から議論できる場を提供することを目的とした学会である。本稿で述べたデータは、知財工学会を通して公開する予定である。

謝辞

本研究で用いた技術動向調査報告書は特許庁の許可を得て利用させていただいた。

参考文献

- [Hashimoto 2017] Hayato Hashimoto, Kazutoshi Shinoda, Hikaru Yokono, Akiko Aizawa, Automatic Generation of Review Matrices as Multi-document Summarization of Scientific Papers, Proceedings of BIRNDL'17: Bibliometric-enhanced IR and NLP for Digital Libraries, Workshop at SIGIR 2017, 2017.
- [Hou 2019] Yufang Hou, Charles Jochim, Martin Gleize, Francesca Bonin, and Debasis Ganguly, Identification of Tasks, Datasets, Evaluation Metrics, and Numeric Scores for Scientific Leaderboards Construction, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 5203-5213, 2019.
- [Joulin 2017] Armand Joulin, Edouard Grave, Piotr Bojanowski, Tomas Mikolov, Bag of Tricks for Efficient Text Classification, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, pp. 427-431, 2017.
- [Nanba 2010] Hidetsugu Nanba, Atsushi Fujii, Makoto Iwayama, and Taiichi Hashimoto, Overview of the Patent Mining Task at the NTCIR-8 Workshop, Proceedings of the 8th NTCIR Workshop Meeting on Evaluation of Information Access Technologies: Information Retrieval, Question Answering and Cross-lingual Information Access, pp. 293-302, 2010.
- [飯沼 2014] 飯沼俊平, 難波英嗣, 竹澤寿幸, 新情報の追加によるサーベイ論文の作成支援, 言語処理学会 第 20 回年次大会, pp. 408-411, 2014.
- [酒井 2009] 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.