

特許出願技術動向調査

—機械学習を活用した最新動向推定—

Report on trends of patent applications and technology on specific technology fields

特許庁 総務部企画調査課

立花 啓

機械分野の特許審査、特許文献の機械翻訳、特許情報施策等に携わり、令和元年7月より現職にて出願動向調査や統計調査に従事。

✉ PA0930@jpo.go.jp

☎ 03-3592-2910

1 はじめに

特許出願は公報として広く一般に公開される。特許の公開情報は、企業・大学等における研究開発の成果に係る技術情報や権利情報である。これらの特許情報から、先端技術分野等の特許出願状況や研究開発の方向性を分析することで、企業・大学等における今後の研究開発の方向性を決定する指針となる。そこで、特許庁では、「特許出願技術動向調査」と題し、国の政策として推進すべき技術分野、社会的に注目されている技術分野について、特許出願動向を中心に分析を行っている。

この調査では、特許分類やキーワードを用いた検索式により作成した特許文献の母集団からのノイズ排除（調査対象の技術分野に該当しない文献の排除）、及び技術区分の付与を人手の文献読み込みにより行っている。そのため、調査に係るコスト（費用、作業時間、作業負荷）が膨大となっており、より多くのテーマについて調査を実施したいというニーズに対応できていない。また、一度調査を実施したテーマに対して継続的に調査を行うことも、費用対効果の面で難しい。さらに、IoTのように、複数の技術分野にまたがる調査テーマにおいては、調査対象となる文献数が膨大となるため、人手による作業は困難である。

そこで、特許庁では、調査の際に人手で行っている母集団からのノイズ排除及び技術区分の付与の作業を、機械学習により代替し、効率的に調査を行う手法を検討している。

本稿では、過去に調査を行った技術分野について、機

械学習により最新動向を推定した結果を紹介する。

2 機械学習を活用した最新の特許出願動向の推定

2.1 推定方法の概要

過去の特許出願技術動向調査において、人手で母集団からのノイズ排除と技術区分付与を行った結果を学習データとし、過去の調査と同じ検索式を用いて作成した調査終了後の特許文献の母集団に対し、機械学習によりノイズ排除と技術区分付与を行う。

機械学習モデルとしては、①最新性、②適用実績、③学習コスト、④タスクとの親和性、の4つの観点から、サポートベクターマシン (SVM) とマルチヘッドニューラルアテンションモデル (MH-NAM) を採用する。機械学習の適用イメージや各モデルの詳細は、昨年寄稿¹を参照されたい。

機械学習に使用する特許文献のテキストデータは、日本語に統一することとし、外国語文献については、全文の日本語機械翻訳文を用いる。

2.2 最新動向の推定結果

過去に調査を行った4つの技術テーマについて、2017年出願分までの特許出願動向を推定した。対象の技術テーマを表1に示す。「検索式文献数」は、調査時に検索式により作成された、ノイズを含む母集団の文

1 根生拓弥、特許出願技術動向調査～AI関連技術テーマの調査結果／機械学習を用いた調査手法～、Japio YEAR BOOK 2019, pp.84-91

献数であり、「調査対象文献数」は、ノイズ排除後の文献数である。

これらの技術テーマの機械学習による推定の精度と、これらのうち2テーマの推定結果の一部を紹介する。

2.2.1 推定精度

(1) 精度評価方法

人手解析済みで、ノイズ/非ノイズの情報と技術区分の情報のある文献のうち一部を評価用とし、残りを学習用として使用する。各技術テーマにおける文献数を表2に示す。

(2) 精度評価指標

精度評価指標として、文書検索や文書分類等において最もよく使われている、再現率、適合率、F値を採用する。これらの算出方法を図1に示す。

再現率は、ノイズ排除や技術区分付与の漏れのなさを定量化したものであり、再現率が高いほど、ノイズ排除/技術区分付与の漏れが少なくなる。

適合率は、ノイズ排除や技術区分付与のノイズのなさを定量化したものであり、適合率が高いほど、ノイズ排

除/技術区分付与のノイズが少なくなる。

F値は、再現率と適合率の調和平均（バランスの良さ）である。再現率と適合率の値が高いほど、また、再現率と適合率の値が近い（バランスが良い）ほど、値が高くなる。

(3) 精度評価結果

各技術テーマにおけるノイズ排除の精度を表3に示す。各技術テーマのノイズ排除のF値は、75%~95%程度である。いずれの技術テーマにおいても、MH-NAMの方がSVMよりも高精度であった。

次に、各技術テーマにおける技術区分付与の精度を表4に示す。技術区分付与の精度指標は、マイクロ平均F値とマクロ平均F値を使用する。

マイクロ平均F値は、すべての技術区分を一つの技術区分に統合した場合のF値である。個々の技術区分に対する文献ごとの正解/不正解の結果をすべての技術区分でマージ（加算）し、F値を算出する。付与文献数の多い技術区分の精度に影響されやすい。

マクロ平均F値は、各技術区分に対して算出したF値の単純平均である。付与文献数の少ない技術区分の精度

表1 対象とした技術テーマ

技術テーマ	調査実施年度	調査対象年	技術分野	技術区分数	検索式文献数	調査対象文献数
LTE-Advanced及び5Gに向けた移動体無線通信システム	平成28年度	2010-2014	電気・電子	143区分	20,050件	18,372件
ヒト幹細胞関連技術	平成29年度	2007-2015	化学	368区分	26,733件	11,625件
自動走行システムの運転制御	平成29年度	2010-2015	機械	324区分	38,472件	8,554件
有機EL装置	平成29年度	2010-2015	一般	493区分	36,536件	32,340件

表2 学習データと評価データの文献数

技術テーマ	学習データ			評価データ		
	調査対象文献数	ノイズ文献数	合計	調査対象文献数	ノイズ文献数	合計
LTE-Advanced及び5Gに向けた移動体無線通信システム	14,698件	1,342件	16,040件	3,674件	336件	4,010件
ヒト幹細胞関連技術	9,300件	12,086件	21,386件	2,325件	3,022件	5,347件
自動走行システムの運転制御	6,842件	23,934件	30,776件	1,712件	5,984件	7,696件
有機EL装置	25,872件	3,356件	29,228件	6,468件	840件	7,308件

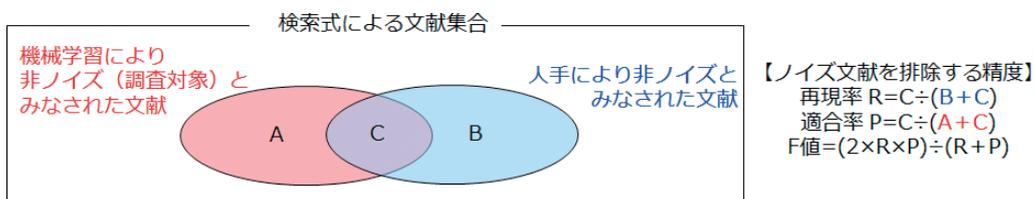


図1 精度評価指標

表3 各技術テーマのノイズ排除の精度

技術テーマ	評価文献数	モデル	再現率	適合率	F値
LTE-Advanced及び5Gに向けた 移動体無線通信システム	4,010件	SVM	93.41%	94.23%	93.82%
		MH-NAM	93.74%	94.33%	94.03%
ヒト幹細胞関連技術	5,347件	SVM	84.26%	82.66%	83.45%
		MH-NAM	86.32%	85.62%	85.97%
自動走行システムの運転制御	7,696件	SVM	73.83%	75.33%	74.57%
		MH-NAM	76.75%	79.30%	78.01%
有機EL装置	7,308件	SVM	94.56%	94.24%	94.40%
		MH-NAM	94.93%	94.87%	94.90%

表4 各技術テーマの技術区分付与の精度

技術テーマ	評価文献数	モデル	マイクロ平均			マクロ平均		
			再現率	適合率	F値	再現率	適合率	F値
LTE-Advanced及び5Gに向けた 移動体無線通信システム	3,674件	SVM	49.10%	48.87%	48.99%	26.22%	25.96%	25.69%
		MH-NAM	48.60%	48.07%	48.33%	23.69%	23.03%	23.11%
ヒト幹細胞関連技術	2,325件	SVM	51.10%	50.93%	51.02%	25.23%	24.86%	24.23%
		MH-NAM	50.15%	45.01%	47.44%	20.42%	19.06%	19.15%
自動走行システムの運転制御	1,712件	SVM	69.96%	69.68%	69.82%	33.16%	33.64%	32.23%
		MH-NAM	66.81%	63.95%	65.35%	26.75%	24.29%	24.21%
有機EL装置	6,468件	SVM	43.55%	42.72%	43.13%	25.51%	25.03%	24.72%
		MH-NAM	38.68%	39.16%	38.92%	18.48%	19.08%	18.41%

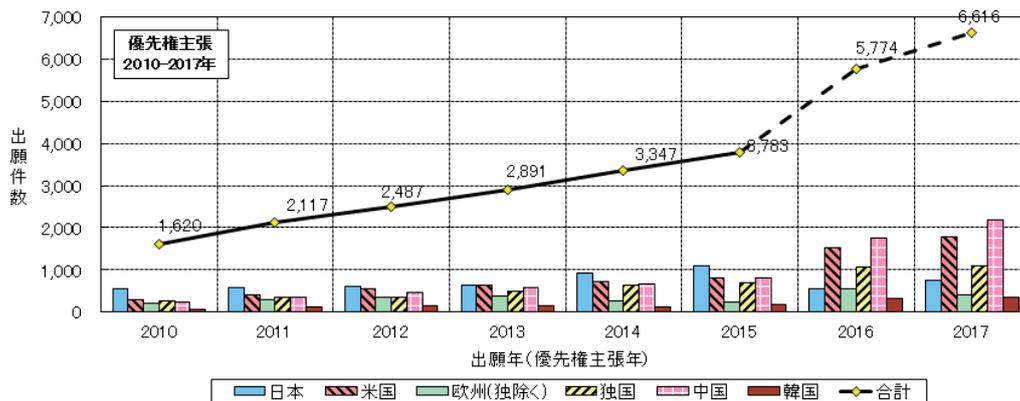


図2 出願先国別の出願件数推移 (日米欧独中韓への出願)

に影響されやすい。

各技術テーマの技術区分付与のマイクロ平均F値は39%~70%、マクロ平均F値は18%~32%であった。総じてSVMの方がMH-NAMよりも高精度であった。

2.2.2 推定結果の概要

4つの技術テーマのうち2テーマについて、最新の出願動向を推定した結果の概要を紹介する。なお、ここではMH-NAMによる推定結果を紹介する。また、出願件数推移の各グラフにおいて、直近2年程度は、データベース収録の遅れ等により、全データを反映していない可能性がある。

(1) 自動走行システムの運転制御

平成29年度に調査した「自動走行システムの運転制御」では、2010-2015年の出願を対象としたが、機械学習により2017年の出願までの出願動向を推定した。

全体動向(ノイズ排除の結果)として、図2に出願先国別の出願件数推移、図3に出願人国籍別の出願件数推移を示す。出願先国別にみると、日本を除く各国で近年出願が増加しているが、出願人国籍別にみると、日本を含め各国で増加傾向にある。

技術区分別動向として、図4に運転支援システム(運転負荷軽減システム)の出願件数推移を示す。運転負荷



図3 出願人国籍別の出願件数推移 (日米欧独中韓への出願)

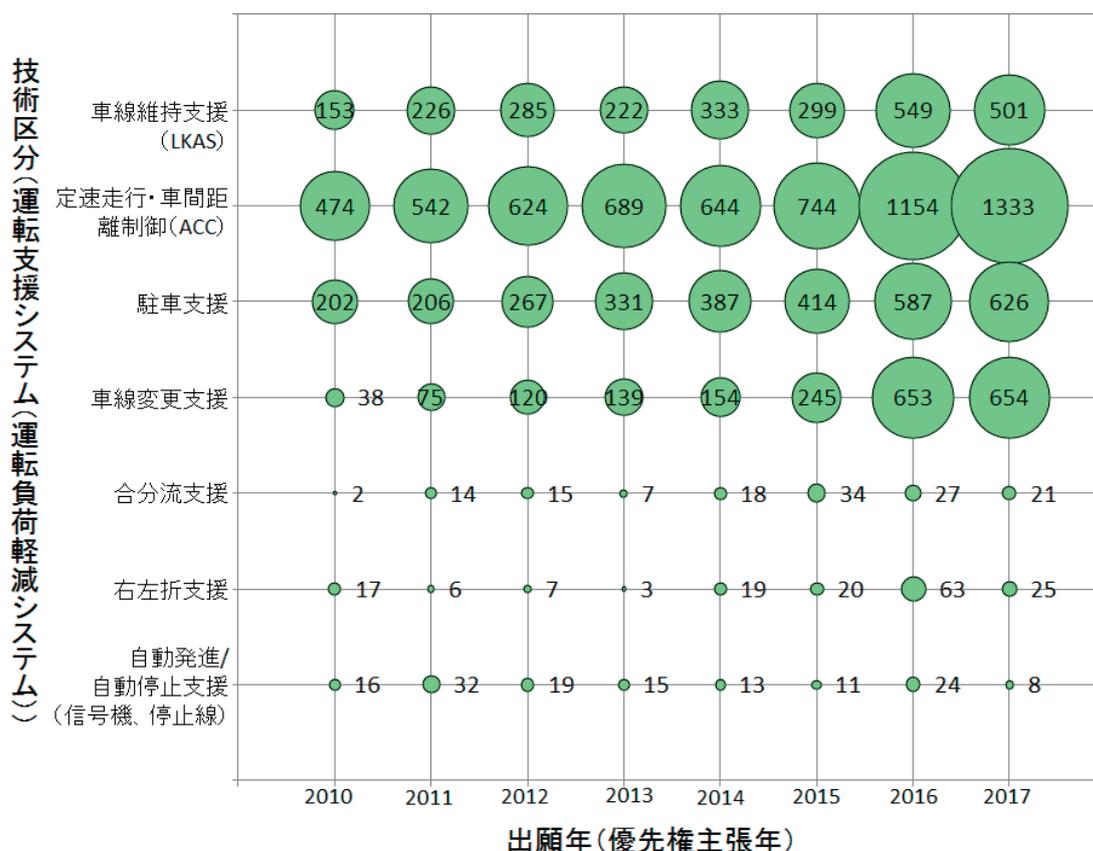


図4 技術区分(運転支援システム(運転負荷軽減システム))別出願件数推移 (日米欧独中韓への出願)

軽減システムの出願件数は全般的に増加傾向にある。

の各観点の出願件数推移を示す。

(2) LTE-Advanced 及び 5G に向けた移動体無線通信システム

平成 28 年度に調査した「LTE-Advanced 及び 5G に向けた移動体無線通信システム」では、2010-2014 年の出願を対象としたが、機械学習により 2017 年までの出願動向を推定した。

全体動向(ノイズ排除の結果)として、図5に出願先国別の出願件数推移、図6に出願人国籍別の出願件数推移を示す。また、技術区分別動向として、図7に技術軸

3 おわりに

本稿では、機械学習を活用した特許出願の最新動向の推定について紹介した。

4つの技術テーマの推定結果は、特許庁ウェブサイトにて閲覧が可能である。

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

機械学習による出願動向の推定は、精度向上の余地は

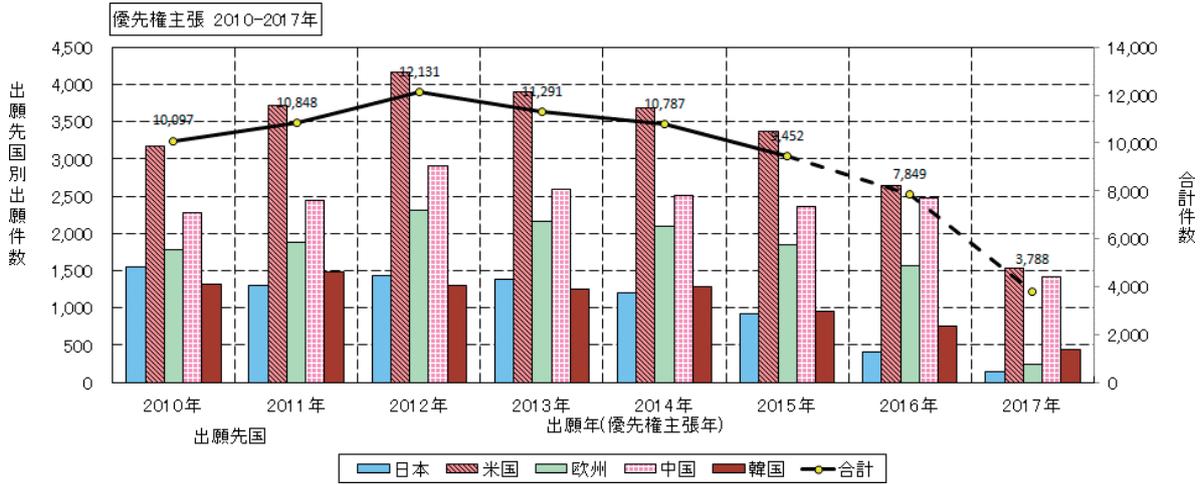


図5 出願先国別出願件数推移（出願先：日米欧中韓）

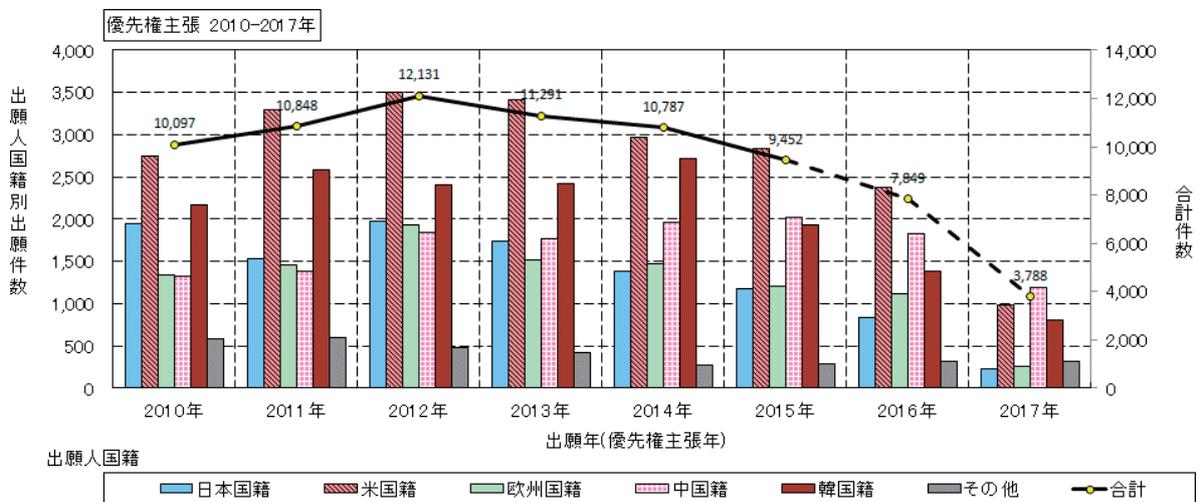


図6 出願人国籍別出願件数推移（出願先：日米欧中韓）

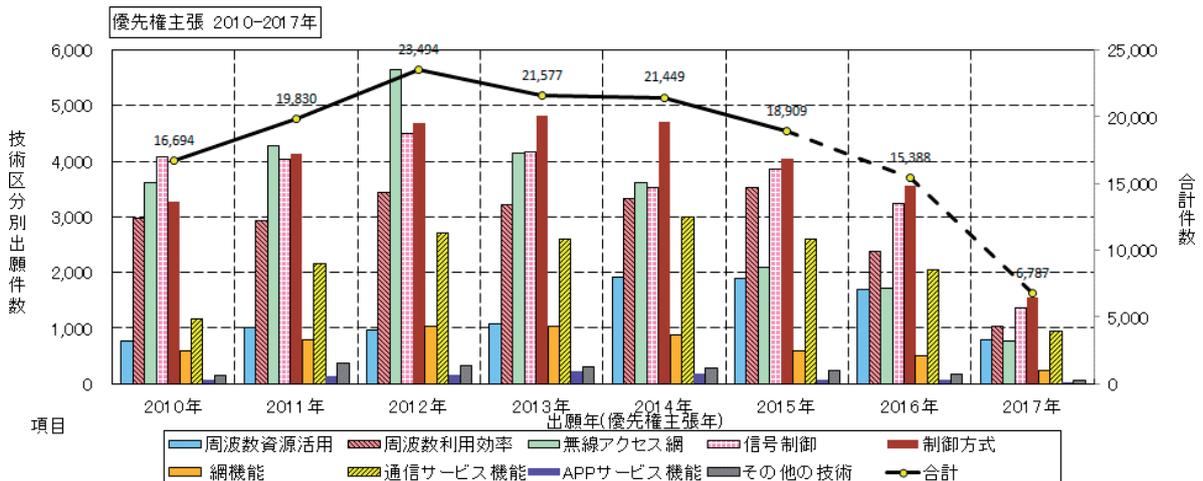


図7 技術区分別出願件数推移（出願先：日米欧中韓）

あるものの、人手での作業と違い、統一した基準に沿った均質な作業を行うことができるという利点を有しており、今後、技術動向調査へのさらなる活用の可能性を有するものと考えられる。

最後に、我が国の企業、大学等が、研究開発戦略策定の際、特許出願技術動向調査の調査結果を有効に活用することで、効率的な技術開発を進め、我が国の国際競争力強化につながれば幸甚である。



1 特許情報施策および事業

