

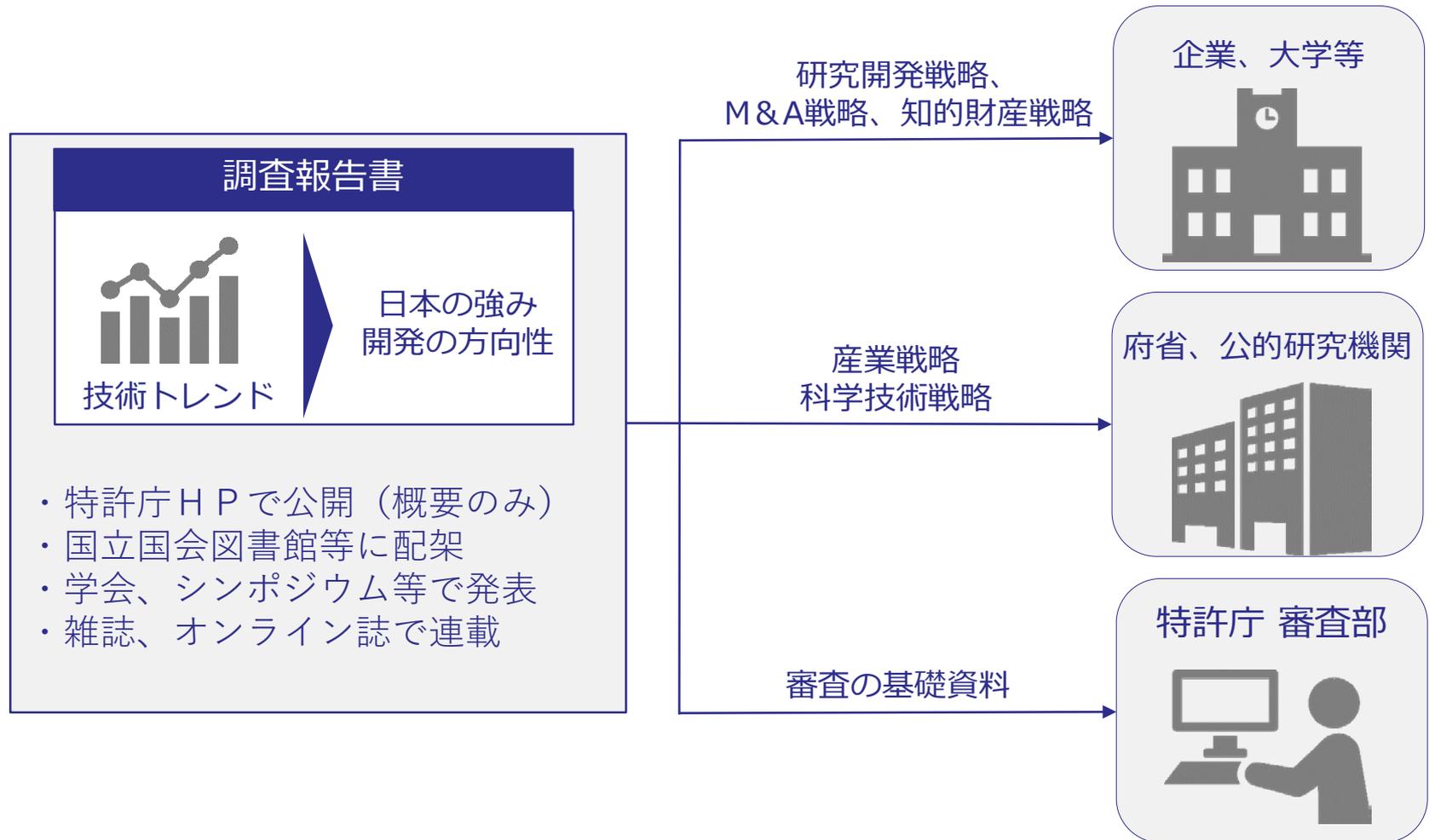
# 機械学習を活用した特許出願動向の 推定の概要

令和2年4月

特許庁

# 特許出願技術動向調査とは

- 「特許出願技術動向調査」は、注目度の高い技術テーマを対象に、その出願動向等を調査して技術トレンドをつかみ、日本の研究開発の方向性を見定めるもの。
- 調査結果は報告書としてまとめられ、迅速・的確な審査の基礎資料として、また、企業・大学・研究機関等が研究開発戦略等を策定するための基礎資料として活用される。

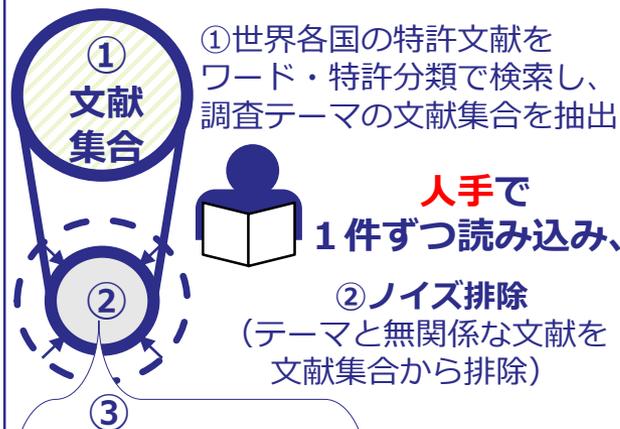


# 機械学習を活用した出願動向調査

○過去に「特許出願技術動向調査」を実施した技術テーマについて、機械学習を用いて、その後の特許出願動向を分析する。

○「特許出願技術動向調査」では特許文献を人手で読み込み解析するが、機械学習を活用した調査では、AIが文献を読み込んで学習し、最新の出願動向を推定するため、低コストで継続的な調査が可能となる。

## 特許出願技術動向調査

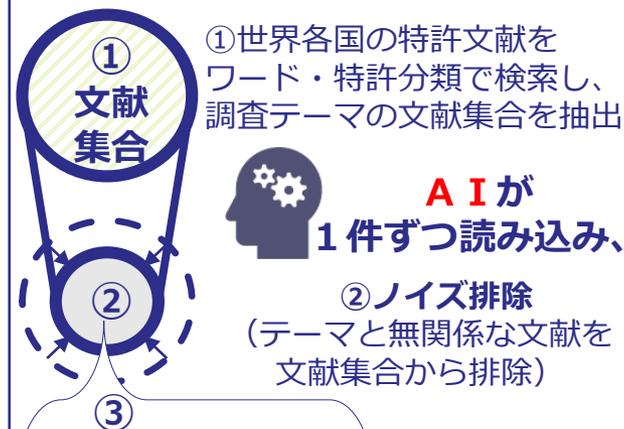


	要素技術
文献1	A、C、D...
文献2	B、C、E...
⋮	



④ 文献数を集計後、マップにして解析

## 機械学習を活用した調査



	要素技術
文献1	A、C、D...
文献2	B、C、E...
⋮	



④ 最新の出願動向を推定

## ＜採用した機械学習モデル＞

機械学習モデル	観点① 最新性	観点② 適用実績	観点③ 学習コスト	観点④ 親和性	モデルの概要
サポートベクターマシン (SVM) (ベースライン)	×	○	○	○	<ul style="list-style-type: none"> <li>・非ニューラルネットワーク系のモデルとして最も広く普及しているモデル</li> <li>・マージン最大化という考え方により、精度が高い分類付与を実現</li> </ul>
マルチヘッドニューラル アテンションモデル (MH-NAM)	○	△	△	◎	<ul style="list-style-type: none"> <li>・特許文献内のどこの記載箇所に着目して出力するかという注目度合を学習できるアテンション機能を持つディープラーニングモデル</li> <li>・一般に、SVMよりも高精度の分類付与が可能</li> </ul>

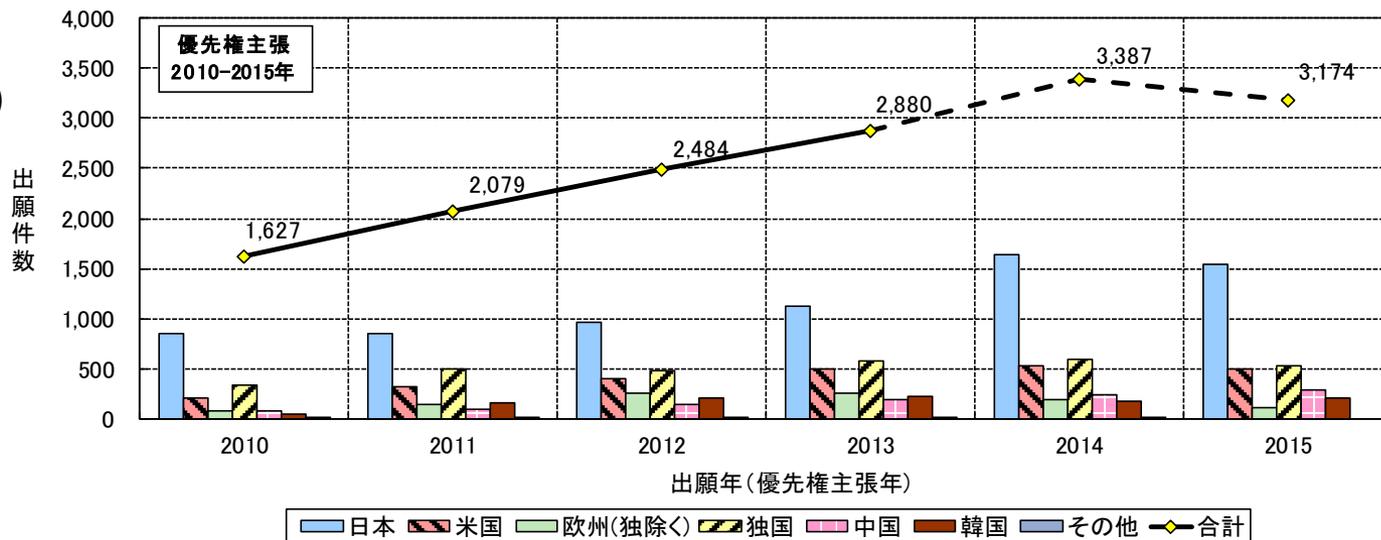
## ＜対象とした技術テーマ＞

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	2010-2014	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

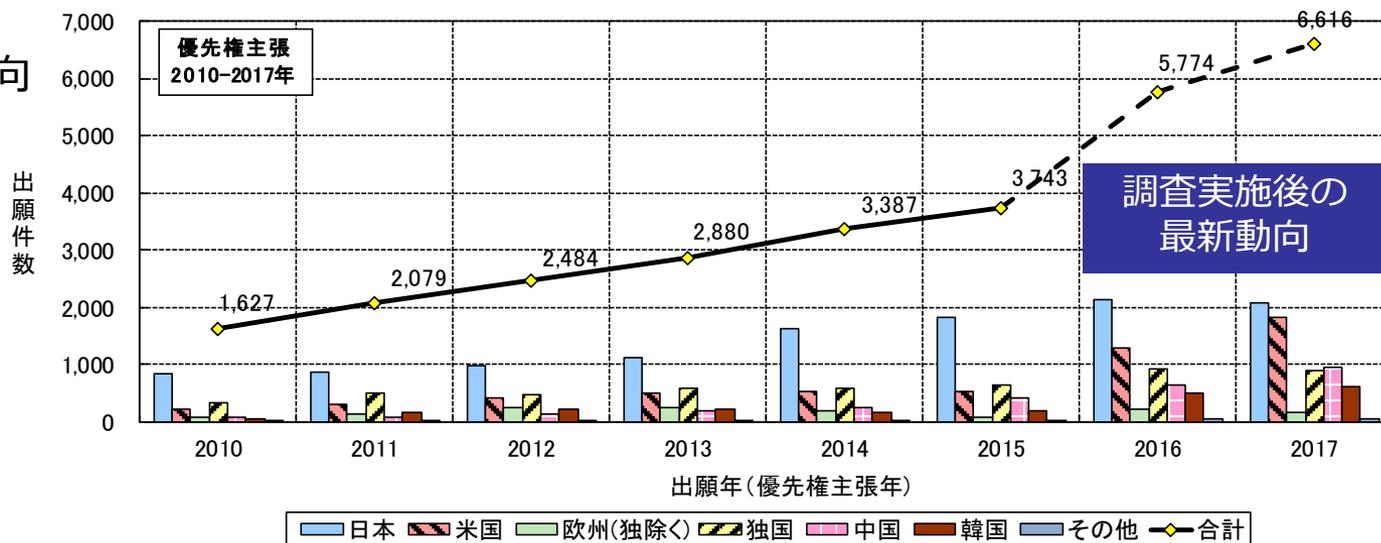
# 機械学習による最新動向推定の例

平成29年度「自動走行システムの運転制御」における、  
出願人国籍別の出願件数推移（日米欧独中韓への出願）

調査実施時の結果  
(調査報告書のグラフ)



機械学習による最新動向  
推定の結果



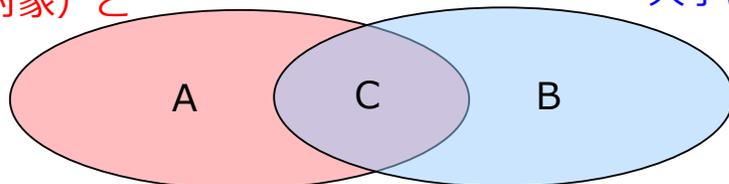
# 機械学習の精度指標（ノイズ排除）

○人手によるノイズ排除の結果に対する再現率（漏れのなさ）と適合率（ノイズのなさ）との調和平均により「F値」を算出し、ノイズ排除の精度の指標とする。

○対象テーマの「F値」は、75%~95%程度。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。

検索式による文献集合

機械学習により  
非ノイズ（調査対象）と  
みなされた文献



人手により非ノイズと  
みなされた文献

【ノイズ文献を排除する精度】

$$\begin{aligned} \text{再現率 } R &= C \div (B + C) \\ \text{適合率 } P &= C \div (A + C) \\ \text{F値} &= (2 \times R \times P) \div (R + P) \end{aligned}$$

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%	<u>94.03%</u>
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%	<u>85.97%</u>
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%	<u>78.01%</u>
有機EL装置	7,308件	SVM	94.56%	94.24%	94.40%
		MH-NAM	94.93%	94.87%	<u>94.90%</u>

# 機械学習の精度指標（技術区分付与）

- 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。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。付与文献数の少ない技術区分の精度に影響されやすい。
- 対象テーマでは、マイクロ平均は39%～70%、マクロ平均は18%～32%。総じてSVMの方がMH-NAMよりも高精度。

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%	<u>48.99%</u>	26.22%	25.96%	<u>25.69%</u>
		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%	<u>51.02%</u>	25.23%	24.86%	<u>24.23%</u>
		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%	<u>69.82%</u>	33.16%	33.64%	<u>32.23%</u>
		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%	<u>43.13%</u>	25.51%	25.03%	<u>24.72%</u>
		MH-NAM	38.68%	39.16%	38.92%	18.48%	19.08%	18.41%