

# 特許文献のランキングへの機械学習技術の適用

Application of machine learning technology to ranking of patent documents

特許庁 審査第一部調整課

**後藤 昌夫**

2001年特許庁入庁。特許審査・審判に従事のほか、調整課で特許審査に関するシステムの企画・開発等に携わり、日本貿易振興機構ドバイ事務所駐在を経て、2020年4月より現職。

特許庁 審査第一部調整課

**目黒 光司**

2007年特許庁入庁。特許審査に従事のほか、調整課審査企画室、情報技術統括室を経て、2019年7月より現職。

特許庁 審査第一部調整課

**菅家 裕輔**

2008年特許庁入庁。特許審査に従事のほか、情報技術統括室で庁内インフラ整備に携わり、人事院行政官短期在外研究員（カナダ知的財産庁）を経て、2019年7月より現職。

## 1 はじめに

特許庁では、平成28年度から、人工知能（AI）技術の活用に向けた検討を行い、同年に行った調査研究（「人工知能技術を活用した特許行政事務の高度化・効率化実証的研究事業」）の結果を踏まえ、平成29年4月に、「特許庁における人工知能技術の活用」についてのアクションプランを公表した。そして、平成30年11月及び令和2年7月に当該アクションプランを改定した。

このアクションプランに沿って、特許審査に関連して、平成29年度には、「Fターム等付与支援システム」及び「先行技術調査支援システム（検索式の用語の拡張、ヒット箇所のハイライト表示）」について、平成30年度には、「画像検索技術の特許図面への適用」について、それぞれ調査事業を実施した。そして、平成31年度に

は、「特許文献のランキング及び要約生成への機械学習技術の適用」について調査事業を実施した。本稿では、平成31年度に実施した調査事業における、特許文献のランキングへの機械学習技術の適用について紹介する。

特許庁における特許審査業務は、審査対象の発明の技術的特徴を有する先行技術文献を調査する過程を含むが、特許審査業務において先行技術調査の占める時間的な割合は非常に大きい。調査対象となる文献が年々増加する中、先行技術調査の長期化が課題となっており、より効率的な先行技術調査が望まれている。当該課題に対して、検索結果として得られた複数の文献を、審査対象発明に類似する文献順に並び替えて提示できると、より効率的な先行技術調査につながると考えられる。

一方、昨今の機械学習技術の進展により、検索結果のランキング学習が論文等で報告されている。このような

機械学習技術の活用により、従来の統計的手法やルールベース等の非機械学習手法に比べ、より高精度な検索結果ランキング及び要約生成の実現が期待される。

そこで、平成31年度の調査事業においては、機械学習技術を用いた特許文献のランキング表示の実用性について検証し、特許審査業務への導入可能性を検証することとした。

なお、本稿は、平成31年度に実施した調査事業の結果を踏まえ、著者らの知見に基づいて記載したものであり、特許庁としての意見・見解を表明するものではない。

## 2 特許文献のランキングに用いる機械学習モデルと素性

本調査では、ランキングに適用する機械学習モデルとして、RankSVM (Ranking Support Vector Machine)、及び、LightGBM (Light Gradient Boosting Machine) の2種類を用いた。また、精度比較のベースラインとして、標準的な順位付けの手法であるBM25を用いた。

### (1) Rank SVM

RankSVMは高速・高精度であり、また、ランキングに有用な情報(素性)を特定可能であることを特長としており、ランキングというタスクとの親和性が高いモデルである。図1に示すように、RankSVMはまず、ランキング学習データから正解文献としての重要度が異

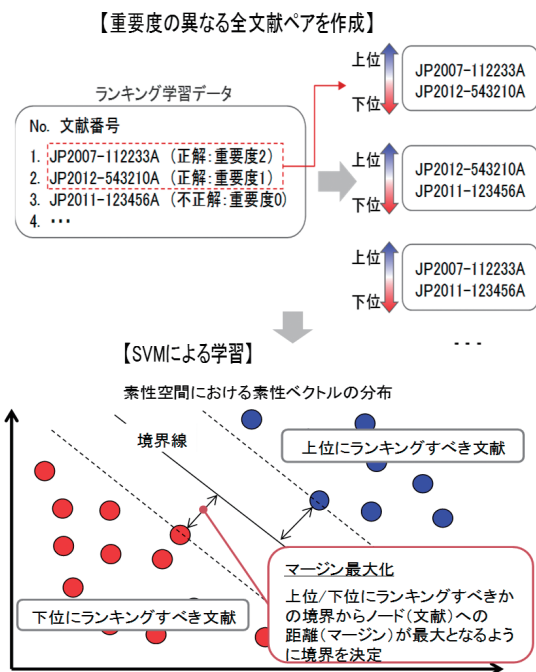


図1 RankSVMの概要

なるすべての文献ペアを作成する。次に、学習データから抽出した素性ベクトルに対し、「マージン最大化」という考え方に基づいて分類の境界を学習し、文献ペアのどちらを上位にランキングすべきかを識別する。

RankSVMを特許文献集合のランキングに適用する際には、検索結果文献集合を構成する各文献について、本願、検索式、検索結果文献のデータを解析・照合することによって、ランキングの手掛かりとなる素性データを抽出し、素性データを要素とする素性ベクトルを生成し、検索結果文献集合における重要度の異なるすべての文献ペアの素性ベクトルを抽出し、SVMによる学習を行う。

### (2) Light GBM

LightGBMは、ランク学習のデファクトスタンダードのアルゴリズム LambdaMART に高速化の仕組みを追加する形で Microsoft 社が開発したモデルである。

LambdaMARTは、アンサンブル学習の一種である勾配ブースティング木(MART)(図2(a))と、ランキングの全体最適化を可能にする目的関数(LambdaRank)(図2(b))から構成される。MARTにおいて複数の弱学習器を学習する際に、ランキング上位の文献を含む文献ペアをより重視した(損失に対してより高い重み付けをした)場合のランキング結果が最適になるように学習する。

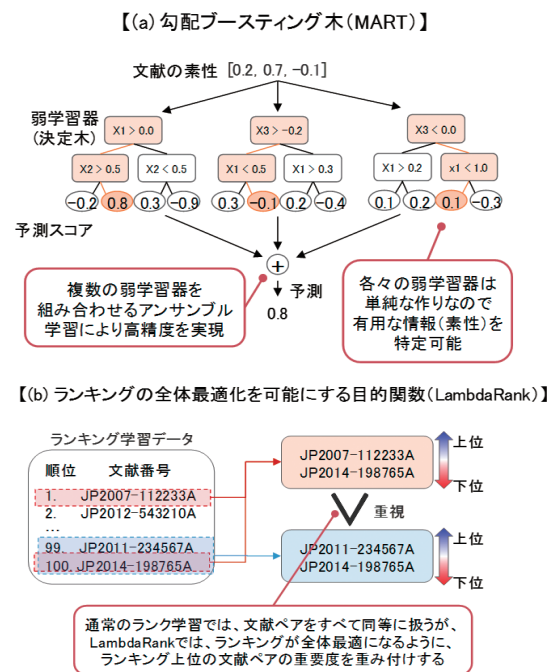


図2 LightGBMの概要



「ある素性の値が閾値より大きいかな否か」というような判定を決定木の上位から順に繰り返して、ある文献のすべての素性データから予測スコアを出力する。学習時には、出力した予測スコアによるランキングが最良となるように、弱学習器（決定木）を学習する。

### (3) 機械学習で使用する素性

本調査におけるランキングでは、本願データ（テキスト、書誌情報）、先行技術調査で作成した検索式、及び、その検索結果文献集合の三つを入力とした。これらの入力データから、表1に示す4種類の観点で、ランキング精度の向上に有用と考えられる24種類の素性を抽出して、機械学習に使用した。

表1 ランキングで使用する素性を選定する際の観点

観点 No.	観点	観点的意義
観点1	本願の属性に類似している文献であるか	・本願の属性（テキスト内容、分類、出願人等）に類似する文献は、先行技術として重要である
観点2	検索式との整合性が高い文献であるか	・検索式は、本願の発明内容を圧縮したものであり、検索式との整合性が高い文献は、先行技術として重要である ・検索式にヒットしたが本願との関連性の低い文献を下位にランキングすることにより、ノイズ文献を除外できる
観点3	検索結果文献自体の特性	・検索結果文献の単語数、請求項数、ファミリー文献数、特許登録の有無など検索結果文献自体の特性を判別できる
観点4	その他(引用・印刷回数等)	・検索結果文献が過去の拒絶理由通知書の中で引用実績がある文献であるかを判別できる ・検索結果文献が過去の先行技術調査においてしばしば印刷・精読される文献であるかを判別できる

## 3 特許文献のランキングに係る精度検証

### (1) 精度検証の対象技術分野と使用するデータ

本検証では、より幅広い技術分野における精度を検証するために、表2のとおり、特許庁の各審査部からそれぞれ1つずつ技術分野（テーマ）を選定した。

表2 本検証で対象とする技術分野（テーマ）の一覧

審査部	テーマ	テーマ名称
1部	2H033	電子写真における定着
2部	4C160	手術用機器
3部	4K029	物理蒸着
4部	5C164	双方向TV、動画配信等

本検証で使用するテーマごとの学習データ、validation データおよび評価データの数を、表3に示す。これらのデータは、本願、検索式、検索結果文献集合の三つ組から構成されるものであり、1件の本願が複数の検索式・検索結果文献集合を持つことがあるため、本願の件数はデータ数よりも少ない。

表3 学習データと評価データの件数

テーマ	学習データ	validation データ*	評価データ
2H033	11,180 件	1,628 件	1,445 件
4C160	8,524 件	1,204 件	1,679 件
4K029	6,104 件	851 件	645 件
5C164	12,936 件	1,689 件	897 件

(※) validation データは、学習データの一部であり、パラメータ調整等を行う際に精度評価用として使用されるデータである。

本検証で使用する評価データには、表4の4種類があり、各評価データ数は表5のとおりである。

表4 本検証で使用する4種類の評価データ

データ名	定義
評価データ1	(条件1) 正解文献=X 文献のみ (条件2) 検索結果文献集合 ≥100 件 (条件3) 起案日 =2010/4 ~ 2014/8
評価データ2	(条件1) 正解文献=X 文献のみ (条件2) 検索結果文献集合 ≥100 件 (条件3') 起案日 =2010/4 以降
評価データ3	(条件1) 正解文献=X 文献のみ (条件2') 検索結果文献集合 ≥30 件 (条件3) 起案日 =2010/4 ~ 2014/8
評価データ4	(条件1') 正解文献 =X 文献または Y 文献 (条件2) 検索結果文献集合 ≥100 件 (条件3') 起案日 = 制限なし

表5 各評価データの件数

テーマ	評価データ1	評価データ2	評価データ3	評価データ4
2H033	58 件	134 件	168 件	1,085 件
4C160	99 件	142 件	168 件	1,270 件
4K029	19 件	45 件	56 件	525 件
5C164	39 件	47 件	81 件	730 件

なお、学習データおよび validation データは、上記の評価データのいずれにも含まれないものであり、正解文献=X 文献または Y 文献としている。

### (2) 精度検証における評価指標

本検証における精度評価指標として、平均精度 (MAP) および再現率を用いた。それぞれの概要は、表6のとおりである。

表6 本検証で採用する精度評価指標

評価指標	概要	長所・短所
平均精度 (MAP)	正解文献の出力順位における適合率（当該順位以上の順位の文献集合における正解文献の割合）の平均を評価データごとに算出し、全評価データで平均した値	・ ランキングの精度を相対的に比較する際に有用 ・ ランキングの精度を絶対的に示す指標ではない
再現率	ランキング結果の上位 N 件または上位 M% において、正解文献が含まれる割合を表す	・ 最も一般的な指標の一つ ・ N 及び M の値の設定について検討必要

(3) 精度検証の結果と考察

① モデル別の精度

精度を検証した結果、表7および表8に示すとおり、どのテーマにおいても、概ね LightGBM、RankSVM、ベースラインの順で、平均精度 (MAP) および再現率

が高くなった。LightGBM と RankSVM は、ベースラインで使用している文献テキストの類似度だけでなく、様々なデータを素性として活用して機械学習を適用することにより、ランキング精度が大きく向上している。

LightGBM の方が RankSVM よりも平均精度 (MAP) が高くなった原因の一つとしては、例えば素性「本願と検索結果文献のテキスト類似度」においては、「類似度が大きいほど正解文献のランキングが上位になる」という線形の傾向にあるが、素性「出願日の差」においては、一定の値 (2 ~ 4 年) においてピークが見られるという非線形の傾向にある。RankSVM では、カーネル関数として線形カーネルを採用しているため、「出願日の差」のような非線形の傾向にある素性を学習結果に適切に反映できていない可能性が高い。

表7 モデル別の平均精度 (MAP) の比較

テーマ	モデル	平均精度 (MAP)			
		評価データ 1	評価データ 2	評価データ 3	評価データ 4
2H033	ベースライン	0.1805	0.1228	0.2243	0.1138
	RankSVM	0.2571	0.2375	0.2574	0.2255
	LightGBM	<u>0.3294</u>	<u>0.2725</u>	<u>0.3516</u>	<u>0.2632</u>
4C160	ベースライン	0.1285	0.1728	0.1794	0.1371
	RankSVM	0.1913	0.2097	0.2473	0.1953
	LightGBM	<u>0.2310</u>	<u>0.2712</u>	<u>0.3256</u>	<u>0.2519</u>
4K029	ベースライン	0.1989	0.1335	0.2512	0.1132
	RankSVM	0.1843	0.1083	0.3763	0.2245
	LightGBM	<u>0.2701</u>	<u>0.1813</u>	<u>0.4374</u>	<u>0.2556</u>
5C164	ベースライン	0.1262	0.1785	0.1987	0.1198
	RankSVM	0.1969	0.1982	0.2644	0.2120
	LightGBM	<u>0.2365</u>	<u>0.2430</u>	<u>0.3540</u>	<u>0.2707</u>

表8 モデル別の再現率 (上位 20%) の比較

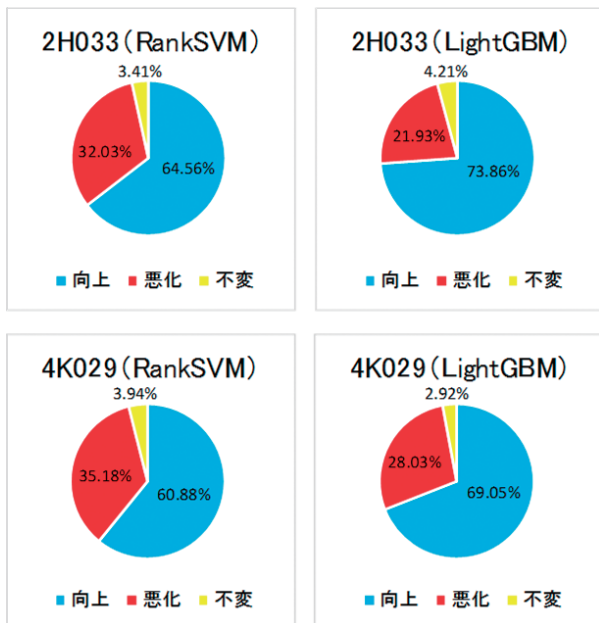
テーマ	モデル	再現率 (上位 20%)			
		評価データ 1	評価データ 2	評価データ 3	評価データ 4
2H033	ベースライン	75.86%	64.93%	58.33%	49.48%
	RankSVM	77.59%	76.12%	66.07%	68.76%
	LightGBM	<u>91.38%</u>	<u>84.33%</u>	<u>77.38%</u>	<u>77.08%</u>
4C160	ベースライン	55.56%	57.97%	57.74%	51.35%
	RankSVM	71.72%	73.19%	70.83%	65.81%
	LightGBM	<u>80.81%</u>	<u>80.43%</u>	<u>82.74%</u>	<u>72.78%</u>
4K029	ベースライン	57.89%	53.33%	62.50%	50.30%
	RankSVM	73.68%	<u>60.00%</u>	83.93%	65.75%
	LightGBM	<u>84.21%</u>	57.78%	85.71%	<u>70.87%</u>
5C164	ベースライン	56.41%	57.45%	50.62%	47.20%
	RankSVM	<u>66.67%</u>	<u>63.83%</u>	70.37%	71.41%
	LightGBM	61.54%	<u>63.83%</u>	71.60%	<u>77.26%</u>



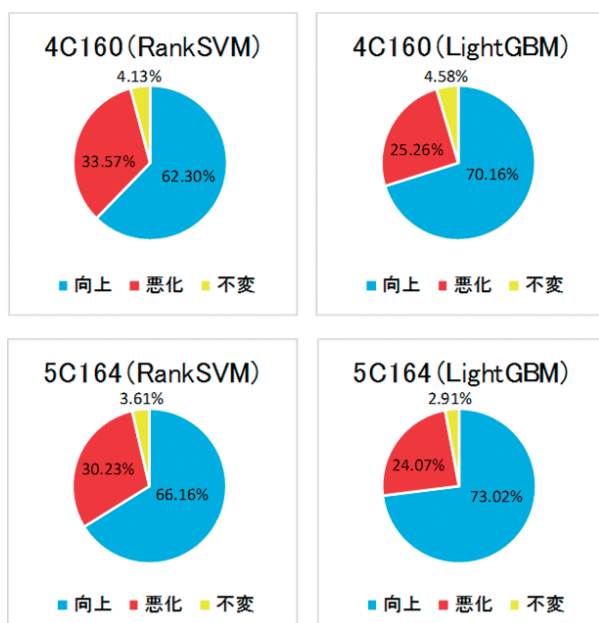
② ランキングの順位が向上／悪化した正解文献数の割合

ベースラインによる正解文献の順位に対して、RankSVM 及び LightGBM による順位が向上／悪化した正解文献数の割合を、テーマごと及び評価データごとに比較した。

図3に、評価データ4についての比較結果を示す。向上した正解文献数の割合は悪化の割合を概ね大きく上回っている。また、RankSVM よりも LightGBM の方が、向上した正解文献数の割合が高くなっている。



ベースラインの順位に対して順位が向上／悪化した正解文献数の割合 (2H033, 4K029)



ベースラインの順位に対して順位が向上／悪化した正解文献数の割合 (4C160, 5C164)

図3

③ 一定の再現率を実現するために読むべき文数

ここでは、ランキング結果の再現率を固定値 (90%) とし、この再現率を実現するために必要な上位文献数 (絶対値・相対値) を算出し、モデルごとに比較する。読むべき文献数及び読むべき文献割合 % は、一定の再現率 (90%) を実現するために読むことが必要となる文献数の絶対値と相対値を表しており、当該値が少ないほど、スクリーニングの効率が向上する。評価データ4についての比較結果を表9に示す。

表9 再現率 (90%) を実現するために読むべき文献数 (評価データ4)

テーマ	モデル	読むべき文献数	読むべき文献割合 %
2H033	ベースライン	172	71%
	RankSVM	104	49%
	LightGBM	92	39%
4C160	ベースライン	162	70%
	RankSVM	121	53%
	LightGBM	108	42%
4K029	ベースライン	105	63%
	RankSVM	84	53%
	LightGBM	75	42%
5C164	ベースライン	120	71%
	RankSVM	76	47%
	LightGBM	68	39%

この比較結果を見ると、RankSVM 及び LightGBM ではベースラインに比べて読むべき文献数が概ね 50% ~ 80% 程度に減少している。特に、LightGBM の方が、減少割合が大きい。

④ 特許文献のランキングに有用な素性

機械学習した結果に基づいて、特許文献のランキングに有用な素性を分析した。具体的には、RankSVM では、予測時に特定の素性をシャッフルすることによる精度低下の割合から、素性の重要度を分析した。また、LightGBM では、決定木において使用頻度の高い素性が重要であるとして分析を行った。

分析したところ、ランキングに有用な素性は、RankSVM と LightGBM の間で傾向が異なった。RankSVM で有用とみなされた素性の 83% は、上記表1の観点1「本願の属性に類似している特許文献」に係る素性であった。

これに対して、LightGBM で有用とみなされた素性

には比較的広い観点の素性が含まれていた。具体的には、下記の素性が有用であると分析された。

- 素性 1 : BM25 による、本願との間の類似度
- 素性 2 : 本願と検索結果文献のテキスト類似度 1 (本願テキストと検索結果文献テキストの類似度 (単語ベクトル))
- 素性 3 : 本願と検索結果文献のテキスト類似度 2 (本願テキストと検索結果文献テキストの類似度 (単語分散表現ベクトル))
- 素性 6 : 付与分類の類似度 1 (本願及び検索結果文献に付与された特許分類ベクトルの類似度)
- 素性 8 : 出願日の差
- 素性 10 : 検索キーワードのヒット数 1 (検索結果文献のテキストにおける検索キーワードのヒット数)
- 素性 17 : 関連出願の数 (検索結果文献のファミリー文献数、公開公報等の優先権主張国数)
- 素性 20 : 本願における検索結果文献の提示の有無
- 素性 23 : 本願に類似する文献の引用文献

## 4 おわりに

本検証により、機械学習モデルの LightGBM、RankSVM を適用することで、ベースラインの標準的な BM25 に比べてランキング精度が大きく向上した。特に LightGBM では、「ランキングの順位が向上した正解文献の割合」、「再現率 90% を実現するために読むべき文献数」等の精度が大きく向上しており、特許審査業務に導入することで、スクリーニング作業効率の向上等の効果が期待される。

### 参考文献：

- [1] 特許文献のランキング及び要約生成への機械学習技術の適用に関する実証的研究事業 調査報告書、2020 年 3 月
- [2] 目黒光司、笹野遼平、榊原隆文、菊池悠太、高村大也、奥村学：F ターム概念ベクトルを用いた特許検索システムの改良、言語処理学会第 21 回年次大会 発表論文集、pp.768-771、2015.
- [3] Rashmi Korlakai Vinayak and Ran Gilad-Bachrach. 2015. DART: Dropouts meet

Multiple Additive Regression Trees. In Guy Lebanon and S. V. N. Vishwanathan, editors, Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, volume 38, pages 489-497, San Diego, California, USA, May. PMLR.

- [4] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 3146-3154. Curran Associates, Inc.

