

「AI×商標：イメージサーチコンペティション」の概要とその成果について

Outline and results of the AI x Trademark: Image Search Competition

特許庁 審査業務部一般役務

神 亜耶人

平成 24 年入庁。令和 2 年から令和 4 年まで商標審査推進室にて商標審査の効率化や商標審査における AI 技術の導入を担当。令和 4 年 4 月より現職。

特許庁 審査業務部商標課商標審査機械化企画調整室

綿貫 音哉

平成 21 年入庁。審査業務課法条約担当、審査業務部機械を経た後、令和 2 年 10 月より現職。

1 はじめに

人工知能（AI）技術の発展は、審査業務の高度化・効率化に資する可能性を秘めており、商標審査分野においても、より効率的かつ品質の高い審査を実現すべく、審査の一部に AI 技術を導入する取組みを進めている。

その一環として、特許庁では、先行図形商標の検索において、画像認識技術を利用した図形商標検索システム（イメージサーチツール）をアジャイル開発方式にて開発し、審査において試用しているが、当該イメージサーチツールは、一部の精度において課題があった。

そこで、特許庁では、当該イメージサーチツールの精度向上を目指し、特許庁として初となる機械学習コンペティション「AI×商標：イメージサーチコンペティション（以下「本コンペ」という。）」を開催した。

本コンペのイメージは、以下の図 1 のとおりである。

本稿では、本コンペを開催するに至った背景、本コンペの概要、本コンペ開催に当たっての課題、本コンペの成果等について言及する。

なお、本稿は個人の見解を示したものであり、特許庁や経済産業省の見解を表明するものではない。

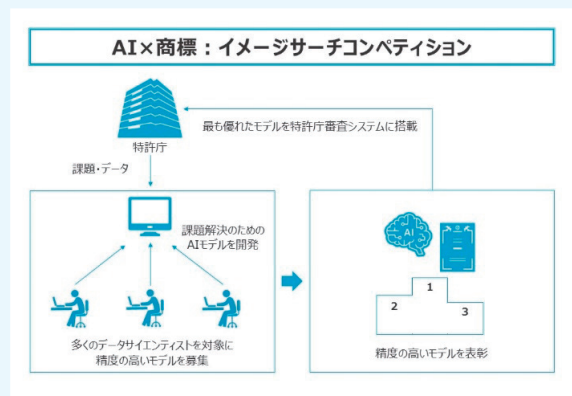


図 1

2 本コンペを開催するに至った背景

本コンペ開催の背景として、従来から使用している図形分類を利用した検索システムの概要とその課題、イメージサーチツールの概要とその課題について説明する。

(1) 図形分類を利用した検索

特許庁では、先行図形商標の検索において、商標の図形要素の分類として国際的に広く採用されているウィーン分類を基に、独自に図形分類を作成し、それらの分類を基準に検索を実施する、図形検索システムをメインで

利用している。

この図形分類を利用した検索は、出願された図形商標に対して人手で図形分類を付与し、付与された分類を基に検索式を作成して検索を行い、類似する図形商標の絞り込み調査を行う。

図形分類による検索は、ある商標に付与されている図形分類と同じ分類が付与された先行図形商標を漏れなく検索できるメリットがあるが、一つの商標にはその特徴に応じて複数の図形分類が付与されるため、検索結果として抽出される商標が莫大な数となり、結果多くのノイズが生じるデメリットがある。

また、この検索手法は、出願商標に付与された図形分類を有しない先行商標は検索結果として抽出されず、この検索手法のみだと、付与された図形分類が異なる類似商標が、審査官の目に触れない可能性がある。

(2) イメージサーチツールの導入

上記のとおり、図形分類のみに基づく検索手法は一部課題があったため、特許庁では、図形の全体的な特徴を機械的に抽出し、それを基に検索を実施する、イメージサーチツールを開発し、審査の支援ツールとして、2019年から試験的に利用している。

イメージサーチツールの導入により、極めて簡素ではあるものの、検索ノイズを極力排除した検索結果の参照が可能となり、かつ、先述の図形分類による検索で検出困難であった事例も、一部検出可能となった。

そのため、イメージサーチツールは、審査品質の向上に寄与するものといえる。

(3) イメージサーチツールの課題

イメージサーチツールは、同一又は酷似の図形商標の検索は、比較的高精度で可能である一方で、以下の図2のとおり、図形と文字等を組み合わせた結合商標における、商標の部分的な一致に係る検索の精度に課題があった。

また、色彩の濃淡に差異がある場合、図形の向きが異なる場合も同様に、抽出が難しい事例が存在した。









同一又は酷似	商標登録第 4276290 号 (商願平 08-108461) 	商標登録第 4188623 号 (商願平 04-143308) 
部分一致	商標登録第 4200612 号 (商願平 08-127888) 	商標登録第 3280480 号 (商願平 04-273292) 
色彩相違	商願 2008-42337 	商標登録第 5141404 号 (商願 2007-114590) 
向き相違	国際登録第 948760 号 	商標登録第 4550159 号 (商願 2001-046444) 

図2

(4) 機械学習コンペティションの導入

イメージサーチツールについては、上記の課題についての精度についての向上が求められていたが、大規模なシステム開発を行うためには開発コストがかかる。

また、内製開発を進めるためには、AI 技術、特に画像認識技術に特化した人材を確保しなければならないが、優秀な AI 専門家は常に人材不足の状態であった。

そのため、イメージサーチツールを導入した2019年以降は、根幹となる検索エンジンについて、大規模な改修はできずにいた。

他方、精度向上の方法についての模索は続けており、その一環として、AI 技術について知見を有する企業との意見交換を行った際、機械学習コンペティションという枠組みを知ることができた。

そして、調査の結果、事業全体の仕様設計やモデルの実装調整は困難ではあるもの、機械学習コンペティションの活用により、イメージサーチツールの精度が向上する見込みが高いことが判明した。

併せて、実務利用しているシステムへのモデルの実装を想定した機械学習コンペティションは極めて珍しく、

特許庁にとって新しい取り組みにチャレンジすることは、検証の意味でも有益であると考えられたため、本コンペを開催する運びとなった。

3 本コンペの概要

本コンペにおいて、特許庁が参加者に提供したデータとその役割、入賞者選考方法、結果の概要について説明する。

(1) 提供データの概要

特許庁が参加者に提供したデータは、審査官が実際に審査で用いるデータを基に作成している。

データは、学習用データ、評価用データ、全図形商標データに大別される。

①学習用データ

本コンペ参加者が機械学習に使用するデータが学習用データである。

学習用データは、出願商標と引用商標が1つずつのペアとなっている。

商標登録出願され、審査の対象となる商標が「出願商標」、審査官が出願商標と類似すると判断する可能性の高い商標が「引用商標」である。

また、学習用データの数は2,311ペアである。

本コンペ参加者は、学習用データを用いることにより、商標審査における「類似」を機械学習する。よって、学習用データは本コンペの結果を大きく左右するデータといえる。

また、審査で使用するツールへの実装を考えれば、学習用データは実務に即したデータにする必要がある。

これらのことから、学習用データは、過去の審査例の中から審査官が目視で選定した。

②評価用データ

本コンペの検索精度を評価するために用いるデータが評価用データである。

データ自体は、学習用データと同様、出願商標と引用商標のペアからなるものの、参加者には、出願商標に該当するデータのみ配布する。

本コンペ参加者は、出願商標をクエリ（検索対象商標）として、後述する全図形商標データの中から正解となる引用商標（正解データ）を検索する。例えるならば、出

願商標が試験問題で、引用商標がその解答という関係にある。

つまり、参加者に出願商標に該当するデータのみ配布するのは、解答（引用商標）を共有してしまえば、試験にならないためである。

なお、評価用データの数は、1,542件ペアである。

学習用データと同じく、評価用データも本コンペの結果を大きく左右するデータであるため、審査官が目視で選定した。

③全図形商標データ

本コンペ参加者が検索を行った際、類似商標を探し出してくる対象となるデータが全図形商標データである。

特許庁に出願・登録された図形要素を含む商標を用いており、約80万件の商標データとなっている。

この中に、評価用データの正解データが含まれている。

(2) 入賞者選考方法

本コンペでは、以下のとおり、入賞候補者を選出後、モデルの検証を行った上で、審査委員会を開催し、最終的な入賞者を決定した。

①入賞候補者の選出

検索結果上位20件以内に、問題に対する解答が含まれているかという指標を基準に、入賞候補者の選出を実施した。

問題の数は、上述の評価用データに準ずるので、1,542問である。

採点は、人手で実施するのは不可能に近いので、評価関数を用い、システム側で自動算出する。

②予測モデルの検証

入賞候補者の決定後は、候補者から予測モデルを入手し、利用しているツール等が本コンペのルールに従っているか、指定した環境下で、スコア等が再現できるかといった点について、検証を実施した。

③審査委員会の開催

予測モデルの検証が終了した後、審査委員会を開催し、最終的な入賞者3名を決定した。

同委員会を構成する委員は、機械学習又はデータ分析に関して専門知識を有し、当該分野において優れた功績を有する者3名である。

同委員会では、入賞候補者の予測モデルのソースコードについて、機能的なエラーの有無や実装に適した汎用

的な記載になっているか等の専門的観点からの検証と、委託事業者により実施された予測モデルの検証が公正に実施されたか等の公正性の観点からの検証を行った。

(3) 本コンペの参加者と入賞者

本コンペは、参加者数 637、投稿件数 1,453 という結果であった。

また、本コンペの入賞者は、以下のとおりである。

第 1 位：ヤフー株式会社：tmsbir チーム

第 2 位：穴井晃太氏

第 3 位：NRI デジタル株式会社：チーム TDX

4 本コンペ開催に当たって検討した課題

本コンペにおける検討課題は、大別して以下の 3 点である。

(1) 提供データの準備

本コンペを開催するに当たっては、提供データの準備が課題であった。

本コンペでは、審査官が過去に類似と判断した審査例を基に提供データを用意したが、実際の審査例は、見た目の類似性以外の理由で類似商標と判断された事例も含んでいる。

そこで、本コンペを開催するに先立ち、過去の審査例の中から、見た目の類似性の高い商標を目視で選定し、それらのペアを、学習用データ・評価用データとして提供した。

本コンペ参加者は、学習用データを機械学習に利用することで、予測モデルの精度を向上させ、かつ、評価用データは、問題とその答え合わせに用いられる。

そのため、学習用データ・評価用データの品質は、予測モデルの精度に直結するものであり、イメージサーチツールの精度向上のための鍵といえる。

(2) 多数の応募者の獲得

多くのデータサイエンティストが本コンペに参加することで、精度の高い予測モデルが作成される可能性が上がる。

他方、本コンペは特許庁として初の取組みであったため、データサイエンティストにおける本コンペの認知度

を向上させ、多数の応募者を獲得することが課題であった。

この点、本コンペ開催に先立つ 2021 年 9 月、デジタル庁が発足し、2021 年 10 月 10 日及び 11 日を「2021 年デジタルの日」と定め、デジタル関連の技術又はサービスを利用した祝祭を実施する日としていた。

そこで、本コンペの開催告知を「デジタルの日」に合わせて行う等、デジタル庁とのコラボレーションを実施し、その広告力もお借りすることで、本コンペの認知度向上を目指した。

(3) 本コンペで得られたモデルの実装

本コンペで得られたモデルは、特許庁の商標アジャイルシステムに実装する想定で検討を進めていたが、そのためには、受入先であるシステム側との連携が重要となる。

そのため、本コンペ開催前には、提出モデルの動作環境を、特許庁側の動作環境と合わせるよう設定した。

また、モデル搭載後の運用・保守の費用負担を最小限とするため、本コンペにおけるモデル作成に利用できるソフトウェアは、オープンかつ無料なソフトウェアのみとした。

さらに、本コンペ開催前、開催期間中及び開催後の実装段階を通して、庁内のアジャイルシステム開発担当者、本コンペ運営事業者、我々担当者を中心に、週一程度を目処に、綿密な打合せを実施することで、円滑な実装を目指した。

5 本コンペの成果

本コンペにおける成果としては、以下の 3 点が挙げられる。

(1) イメージサーチツールの精度向上

本コンペの評価用データの中から 100 件程度サンプルを抽出し、本コンペの入賞モデルの最終スコアと、イメージサーチツールとを比較した。

その結果、イメージサーチツールは、上位 20 位以内に正解データが表示される割合が約 28% であったが、本コンペの入賞モデルの最終スコアは、評価用データに対して上位 20 位以内に正解データが表示される割合が

約73%であった。

両者は、サンプル比較の結果ではあるが、その精度の差は2倍以上となっている。

そのため、本コンペは、イメージサーチツールの精度向上に、大きく寄与したといえる。

(2) 機械学習コンペティションの有効性確認

本コンペの成果物の一部は、現状、特許庁の商標アジャイルシステムへの仮実装が完了し、審査官へ試行提供中である。

また、試行利用段階ではあるものの、審査への有効性が確認できた事例も存在する。

上記により、機械学習コンペティションで作成されたモデルは、実務利用しているシステムへの実装が可能であることと、機械学習コンペティションの成果物は、実務課題の解決に有効となり得ることが、本コンペの取り組みにより明らかになったといえる。

(3) イノベーションの促進

機械学習コンペティションの特徴は、政府や企業が特定の課題に関連したデータを公開し、広く一般に課題解決のための予測モデル等の開発を募る点にある。

これにより、個人で活動するデータサイエンティストが、機械学習コンペティションを通じて政府や企業におけるシステム開発に参画する機会となる。

本コンペにおいては、600者を超える参加者が参加し、かつ、参加者からも、「本コンペで培ったスキルを自社での業務に役立てたい」、「画像認識が専門ではないが、本コンペで画像認識分野のスキルを身につけたい」といった声が寄せられた。

上記により、本コンペの開催は、画像認識分野におけるイノベーションの促進に資する成果があったと考える。

6 おわりに

本コンペは、特許庁として初めての企画ではあったものの、上述の成果を得ることができた。

この一番の要因は、何よりも、本コンペ実施にあたりお世話になった皆様のお力添えが大きかったと考える。

具体的には、本コンペ開催にあたっては、コンペの設計・運営等、多岐に渡る業務をご対応いただいた

Nishika 株式会社の皆様をはじめ、多くの企業の皆様にご尽力いただいた。

また、本コンペ事業の仕様検討や広告、成果物の実装調整にあたっては、特許庁総務部総務課情報技術統括室の皆様をはじめ、庁内及び庁外の多岐にわたる部署の皆様にご協力いただいた。

併せて、本コンペ開催期間中は、多くの方がコンペに参加し企画を盛り上げてくださるとともに、素晴らしいモデルを数多く提出いただいた。

上述の皆様には、この場を借りて、心から感謝を申し上げます。

本コンペで誕生した成果物は、現状、一部を審査官に試行提供している段階にあるが、その精度が高くなるとともに、商標審査の品質向上に寄与する度合いも高まると考える。

その審査の品質向上が、商標制度を利用される皆様の満足度の向上へとつながることを、本コンペに携わった一担当者として、心から願っている。

参考文献

1. 一般財団法人経済産業調査会「特許ニュース」No.15636（令和4年4月18日発行）に掲載された拙著「特許庁初の機械学習コンペティションー「AI×商標：イメージサーチコンペティション」の開催ー」
2. 経済産業省「「AI×商標：イメージサーチコンペティション」の入賞モデルを決定しました」2022年3月14日 (<https://www.meti.go.jp/press/2021/03/20220314001/20220314001.html>)
3. Nishika 株式会社「AI×商標：イメージサーチコンペティション（類似商標画像の検出）」(<https://www.nishika.com/competitions/22/summary>)

