

「大分け」の精度向上に向けて 2022年版

IPCC's trial to improve the precision of "OWAKE"

一般財団法人工業所有権協力センター（IPCC） 研究所総括研究員

塚田 肇

令和3年10月より現職

✉ tsukada-hajime@ipcc.or.jp ☎ 03-6665-7870

1 「大分け」とは

我が国で為された特許出願には、IPC、ファイルインデックス（FI）、テーマコード、Fタームといった分類が付与されている。

これらの分類を付与するため、機械学習による分類の付与という試みもされている（例えば、特許庁による調査^[1]）が、機械学習による分類の付与は、付与の精度（正解率）が充分でないため、現在でも、人手（技術専門家）によって分類の付与がされている。（なお、IPCとテーマコードはFIから機械的に付与できるため、実際に人手によって分類の付与がされるのはFIとFタームである。）

この人手による分類の付与を行っているのが、一般財団法人工業所有権協力センター（IPCC: Industrial

Property Cooperation Center）である。IPCCは、審査官が新規性・進歩性の判断の際に用いる先行技術文献を検索する先行技術調査事業に加えて、分類付与を行う分類付与事業を2000年度から実施している。

IPCCでは、図1のような流れで分類の付与を実施している。まず、特許庁から、専用の回線を通して特許出願の情報を受け取る（図1の「分類付与の外注」）。次に、その特許出願の技術分野を決め、特許出願の情報を、その技術分野を担当している技術専門家に送る（図1の「自動大分け処理」）。その後、技術専門家が分類の付与を行い（図1の「分類付与」）、付与された分類の指導者による確認（図1の「指導者チェック」）と、付与された分類の審査官による確認（図1の「審査官チェック」）を経て、その特許出願に付与される分類が確定し、

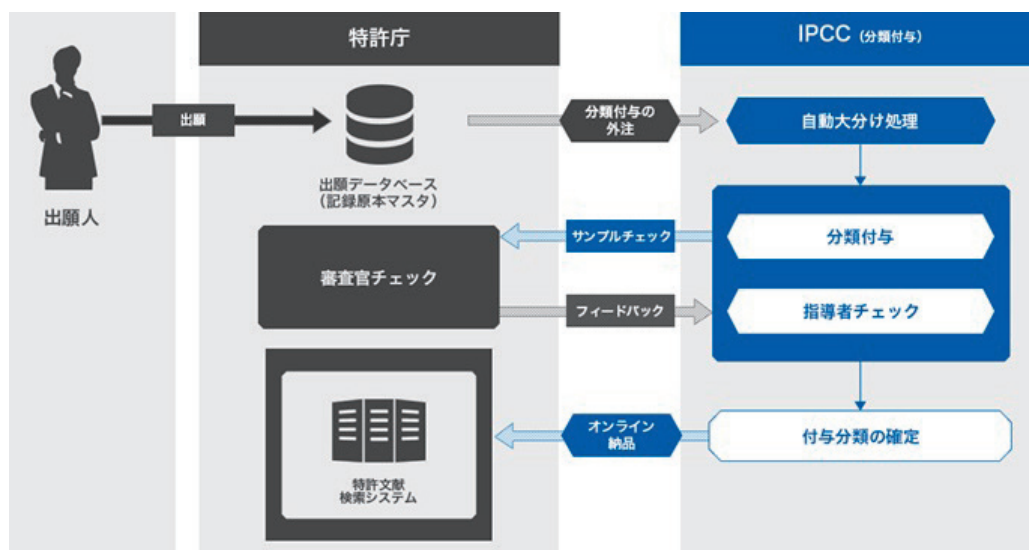


図1 分類付与の流れ

専用の回線を通して特許庁にオンラインで納品される(図1の「オンライン納品」)。

この流れの中で、「特許出願の技術分野を決め」ることを、IPCCでは、「大分け」と呼んでいる。

IPCCにいる技術専門家は約40の技術分野に分けられており、技術専門家はそれぞれが所属する技術分野の分類付与を行うことはできるが、他の技術分野の分類付与を行うことはできない。そのため、ある特許出願の分類の付与を、その特許出願の技術分野の技術専門家に振り分ける「大分け」が必要となる。

「大分け」は、日本に出願される約30万件という膨大な数の特許出願を、ありとあらゆる技術を含む約40の技術分野に分ける作業であり、人手で行うとなると膨大な人件費が掛かるため、システム化し、自動で行っている。

大分けシステムの構築における重要な観点は、言うまでもないが、費用と、それから、大分けシステムによって推定される技術分野の精度(正解率)である。(ここでいう「正解」は、大分けシステムによって推定された技術分野に属する分類が、その技術分野の技術専門家によって付与された場合のことをいう。)

もし、大分けの精度が低いと、大分けによって決められた技術分野の技術専門家が、特許出願の明細書等を読

んだ上で、分類を付与しないと判断し、正しい技術分野を決め、正しい技術分野の技術専門家に分類の付与を依頼する、といった無駄なコストが多く掛かってしまうため、大分けの精度は非常に重要である。

大分けシステムの精度向上は継続して行われており、その継続して行われてきた精度向上の取り組みについて、櫻井^[2]が2020年に紹介した。その後、大分けシステムを刷新したので、本稿では、その大分けシステムの刷新について紹介したい。

2 旧大分けシステムについて

刷新前の旧大分けシステムについては、櫻井^[2]が紹介しているが、本稿でも概要を示す。

旧大分けシステムは図2のように構築されていた。特徴は、特許出願の技術分野を推定するために、特許出願の明細書等から技術分野を推定する多数のアルゴリズムを組み合わせていたところにある。旧大分けシステムでは、ある特許出願について、多数のアルゴリズムから出力された、単語の出現頻度による技術分野の推定結果、最大エントロピー法による技術分野の推定結果、…、ルールベースによる技術分野の推定結果を、多数決等によってマージし、最終的な技術分野の推定結果(大分け結果)

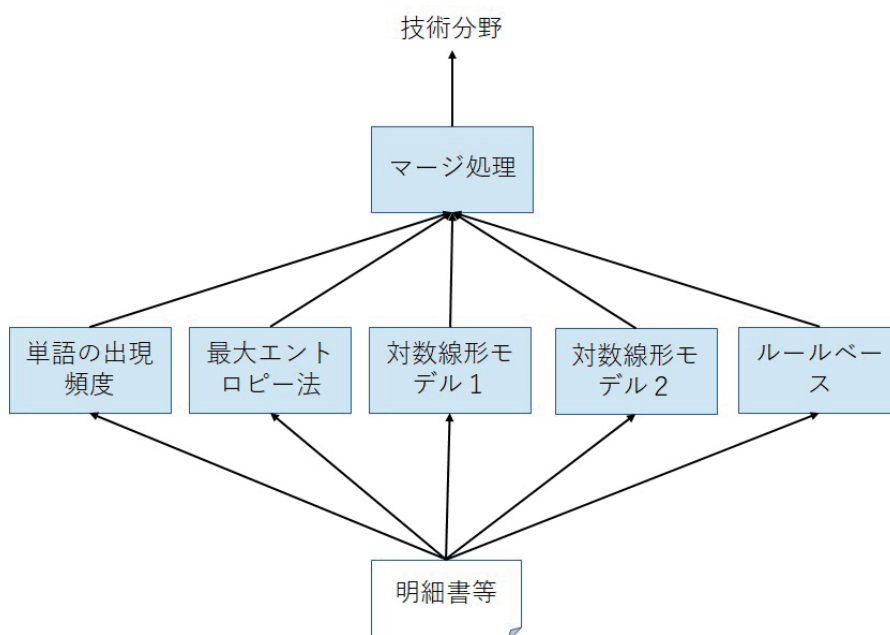


図2 旧大分けシステム

を出力していた。

このような仕組みにより、旧大分けシステムは、特許出願の技術分野を非常に高い精度での推定を実現していた。

一方で、複雑な仕組みのため、システムが巨大化し、システムの運用コストが高くなってしまいうという問題が発生していた。

そこで、多少の技術分野の推定の精度の低下があっても、システムをシンプルにすることによって、システムの運用コストを下げる刷新を行うこととした。

3 新大分けシステムについて

刷新後の新大分けシステムは図3のように構築されている。特許出願の技術分野を推定するアルゴリズムについて、深層学習に一本化することで、システムをシンプルにする一方、アルゴリズムをシンプルにしたことによる技術分野の推定の精度の低下を抑えるため、以下のような仕組みを導入した。

(1) 深層学習のモデルを特許文献に特化したモデルとした。

(2) 特許庁^[1]にあるように、特許出願の分類の推定には、時間の経過と共に推定の精度が低下していくという特徴があることから、一定期間ごとに、過去数年分の特許文献で学習し直すようにした。

(3) 技術分野を推定するためのアルゴリズムの部分の独立性を高め、アルゴリズムの修正や変更を行いやすくするようにした。

その結果、運用コストを刷新前と比べて下げることができた。さらに、事前の検証ではわずかに下がると見込まれていた技術分野の推定の精度も刷新前と比べて高くすることができた。

しかしながら、技術分野によって、アルゴリズムの変更が推定の精度に与えた効果はかなり異なり、非常に推定の精度が高くなった技術分野もあれば、そうではない技術分野もあった。現在は、大分けシステムによる技術分野の推定が、誤推定である可能性が極めて高いと大分

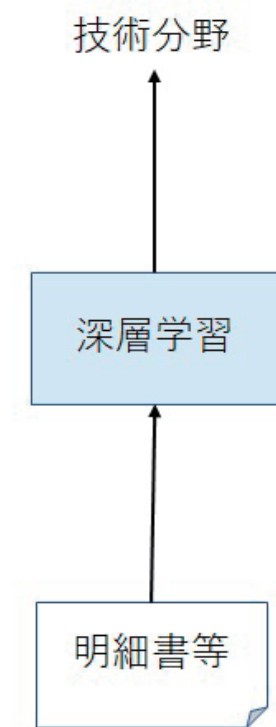


図3 新大分けシステム

けシステムによって判定された極一部の特許出願については、大分けシステムによって技術専門家に分類の付与が割り振られる前に、人手で技術分野を決定(大分け)し、人手で決定した技術分野の技術専門家に分類の付与を割り振るといった対処を行っているが、将来的には、(1) 深層学習のモデルを修正する、(2) 誤推定の可能性が高い場合は、ルールベースで技術分野を推定するようにシステムを改造する、といったシステムでの対応も考えている。

4 今後について

現在、システムで実現できているのは大分けであり、分類の付与は人手で行っているが、将来的には、分類の付与もシステムで実現したいと考えており、IPCC では長年に渡ってシステムによる分類の付与の研究を行っている。しかしながら、分類の付与が正しくされていることは、先行技術調査、ひいては、特許権の品質にも影響することから、分類の付与には 100%に近い正解率が求められているところ、数十の技術分野の中から特許出願の技術分野を推定する大分けであっても 100%に近い正解率は実現できておらず、ましてや、数万の分類 (FI) の中から特許出願の分類を推定する分類の付与をシステムで実現できる目途は立っていない。そこで、現在は、推定した分類を、付与する分類の候補として技術専門家に提示することで、分類の付与のサポートを行う研究を行っている。

その研究の結果については、来年度以降紹介したい。

参考文献

- [1] 特許庁, 令和 2 年度 機械学習を活用した特許動向分析の調査, https://www.jpo.go.jp/resources/report/gidou-houkoku/tokkyo/document/index/2020_gido_machine_learning.pdf
- [2] 櫻井 健太, 大分けの精度向上に向けて, Japio YEAR BOOK 2020