

「大分け」の精度向上に向けて

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 研究所総括研究員

櫻井 健太

令和元年10月より現職

✉ sakurai-kenta@ipcc.or.jp

☎ 03-6665-7870

1 「大分け」とは

特許情報の有用性について、本誌では説明するまでもないが、これをより有効に活用していくためには、必要な情報に迅速にアクセスできること、つまり検索が容易であることが重要である。そして、特許情報の検索を行うにあたり、不可欠なのが特許分類である。

我が国で発行される特許公報には、国際特許分類 (IPC)、ファイルインデックス (FI)、テーマ、Fターム等の特許分類の記号が掲載されている。特許の分類については、計算機による自動付与の研究も活発になされているが、特許公報に掲載されている特許分類の記号は、基本的には明細書を人が読んで理解することで、付与されている (ただし、IPC とテーマは、人手で付与された FI に基づいて、対応するものが機械的に導出される。)

一般財団法人工業所有権協力センター (IPCC: Industrial Property Cooperation Center) は、特許の検索業務の他、特許庁の発注を受けて、FI と Fタームを同時に (一元的に) 付与する一元付与業務を、2000 年度から全技術分野の全特許出願に対し実施している。この FI 及び Fタームの付与は、上述の通り、基本的には人手で行なっており、具体的には、39 のいずれかの技術区分に属する技術専門家が、それぞれ自身の担当技術に関する特許出願の明細書を読んで、最適な記号を付与している。例えば、スマートウォッチのアンテナ配置に関する特許出願については、時計の技術の専門家とアンテナの技術の専門家が、それぞれ明細書の内容を理解し、それぞれが担当する分類の範囲で、最適な

分類記号を付与している。

ここで、分類付与の業務を効率的に行うためには、どの分野の担当者に案件を割り振るかということが重要となる。つまり、せっかく明細書を読んだのに、自身の担当する分類の範囲には付与すべき分類がないということになると、全くの無駄読み (空振り) になってしまう。IPCC では、最初に案件を読んだ人が、自身の担当する分類を付与するとともに、自身の担当する以外の適当な分類があるかどうかを判断し、それがあつ場合には、その担当者に案件を回すというようなやり方を採用しており、このように、技術の専門家の判断により回し先を決定するようにすれば、案件の二人目以降の回し先では、「空振り」が発生しにくくなるが、案件の最初の回し先をどのように決定するかが、残る課題となる。この最初の回し先の決定を「大分け」と呼び、IPCC ではこの「大分け」を計算機による自動分類で行なっている。

以下では、IPCC における大分けシステムの開発の歴史と今後の展望について紹介する。

2 大分けシステムの開発の歴史

IPCC が初めて大分けシステムの開発に着手したのは、一元付与業務の本格実施を2年後に控えた1998年のことである。その後、一元付与業務の試行運用を行っていた1999年4月に、対象を絞って稼働を開始し、2001年度には対象を全特許出願に拡大している。当時、IPCC が一元付与業務を本格的に受注する以前は、特許庁において、分野を代表するベテラン審査官が週ご

とに集まって、全出願を38の審査室に分ける作業を行っており、これが「大分け作業」と呼ばれていたが、この作業はベテラン審査官がいる特許庁では可能であっても、IPCCでは難しく、効率性を欠くことが予想されたため、この「大分け作業」を機械的に行うシステムの開発が必要となったのである。なお、このような、あらゆる技術分野の発明を機械的に分類するシステムは当時他に無く、WIPOや日米欧の特許庁会合などでも紹介され、欧州では「OWAKE(独語読みで「オバケ」)」と呼ばれて注目を浴びることとなった^[1]。

稼働当初、大分けシステムは以下のような処理により、その出願に最も相応しいテーマ(筆頭テーマ)を特定するものであった(図1)。

- ① 予め、過去に発行された公開公報の内容及び付与されている「テーマ」の記号を解析して大分けデータベースを作成する。

- ② 大分け対象の出願の明細書から単語抽出を行い、大分けデータベース内の単語と比較して、IPCクラス別に類似度を計算して粗ぶるい処理(最も近いIPCクラスの特定)を行う。

- ③ 特定されたIPCクラスに属する各公報との類似度計算を行い、公報に付与されているテーマコード毎に類似度を積算してヒストグラムを作成し、それをもとにテーマを特定する。

そして、稼働開始後も、以下のように継続的に精度向上を目的とした開発を行っている。

- (I) 2002~2008年度(第1世代の改良)

Winnow法の採用による粗ぶるい処理の精度向上、各種情報を用いた公報への重み付けによるヒストグラムの調整、分割出願の場合の特別ルールの採用等。

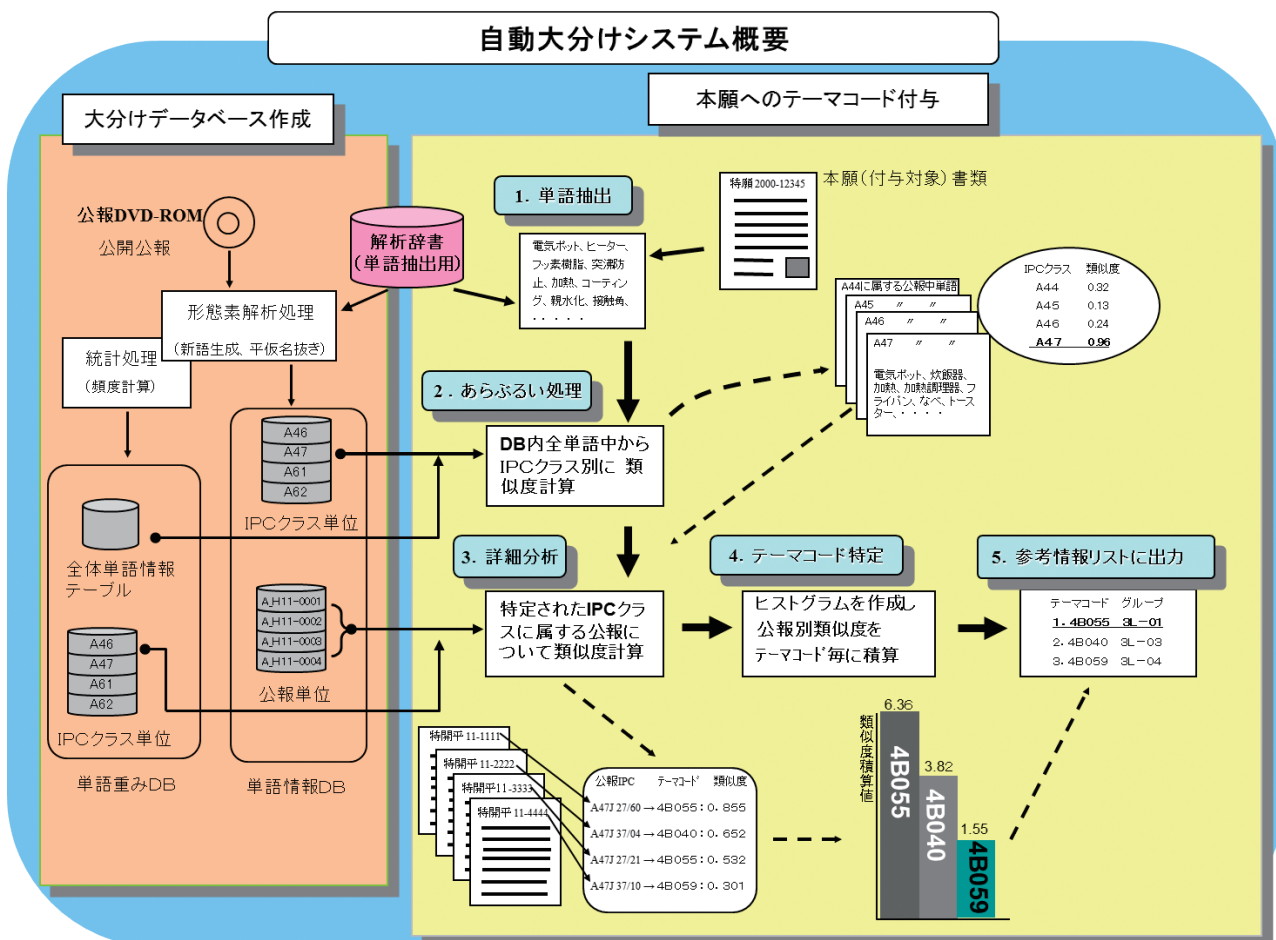


図1 大分けシステム(稼働当初)の処理

(II) 2009～2011年度(第2～4世代の開発)

中核語に基づく文献集合の作成、IPCクラスとは異なる小グループへの絞り込み、エントロピーに基づく類似度算出などの新方式の採用等。

(III) 2012～2014年度(第5世代の開発)

小グループへの絞り込みを行わない複数の小分類器による多数決方式の採用等。

このような継続的な開発により、稼働当初は筆頭テーマの正解率が40%程度であったのが、2012年度には70%に達し、業務の効率化に大きく寄与することとなった。しかし、2012年度の新方式の採用時点付近をピークに、正解率は徐々に低下傾向にあり、また、システムの世代ごとの得意不得意に応じて、複数の世代の特定結果を選択的に採用する方式をとっていたことで、システム全体が複雑化してシステム改修やメンテナンスに多くのコストがかかるという問題も顕在化してくることとなった。そこで、2019年度より、深層学習方式を採用し、残存する旧世代を廃止していくための開発を進めることとなった。

(IV) 2019～2020年度(第6世代の開発)

深層学習方式の採用、残存する旧世代の廃止等。

3 今後の展望

深層学習方式を採用し、定期的に最新の公報を含むデータにより学習を行うようにすることで、将来にわたって持続的な精度向上を行うことができると考えている。また、深層学習による自然言語処理は、近年著しい発展を遂げており、新たなモデルを採用することで、更なる精度向上や、F1、Fターム等のより細かい分類の特定も可能となることが期待される。

例えば、2017年6月にGoogle社の研究チームが“Attention Is All You Need”というタイトルで発表したTransformerモデル^[2]は、アテンションと呼ばれる機構を用いた機械翻訳のモデルであるが、学習に要する時間が大幅に短縮され、機械翻訳の精度も格段に向上している。このモデルのエンコーダは、アテンション機構によって入力系列の単語の関係性をうまく捉えること

ができることから文書分類にも利用されている。さらに、このTransformerモデルをベースとして2018年10月に発表されたBERT^[3]は、様々なタスクで著しい精度向上を達成している上に、ある領域の知識を別の領域の学習に適用させる転移学習も可能となるなど、自然言語処理における深層学習活用にブレイクスルーをもたらすこととなった。

このように、自然言語処理に有効な深層学習モデルが、近年次々とツールとして使えるようになってきていることに加え、深層学習への利用を想定し、学習に適したGPU等の計算資源を備えた構成をデフォルトで提供するクラウドサービスが増えてきていることも追い風といえる。なかには、Google Colaboratoryのように、開発環境も含め無料で提供されるものもあり、高価なシステムと開発環境を自前で準備しなくても深層学習を行うための高度な環境が簡単に手に入るようになった。

このような状況の中で、IPCCの研究所においても、クラウドサービス等を活用しつつ、深層学習モデルを利用した分類技術の研究に取り組んでいる。上述のようなアテンション機構を利用したモデルでは、アテンションの重みから、どの項目(単語、文章等)が出力に寄与しているかが可視化され(図2)、寄与した入力項目を考慮した分析が可能となるが、このような分析は、分類付与や検索を主要業務とし、膨大なノウハウを保有しているIPCCの知見が大いに活かせるところであり、その強みを活かして、モデル前処理やパラメータ等の検討が進展していくことを期待している。

左右一対のプリズムを本体固定部に対し一体的に配置しているため、プリズムの保持が大掛かりになりコンパクト化には不向きである。また、図2の一部を左右で共通に使うことで偏心駆動させているので、照度調整用に専用のプリズムを必要とする。光軸を曲げて焦点像をさせるための光学部品を必要とし、コスト及びサイズの面で課題があった。左右対物レンズ1a、1bを一体的に保持する保持手段4と、前記左右一対の対物レンズを光軸と垂直な単一平面内でヨー方向及びピッチ方向に回転成分無く移動させるヨー方向案内手段6a、6bとピッチ方向案内手段8a、8bと、観察時に発生するブレを挿出して、前記保持手段を該ブレを打ち消す方向へ前記ヨー方向案内手段及びピッチ方向案内手段に沿って移動させる駆動する手段9～12とを備えたものである。左右一対の対物光学系が形成する像を左右一対の対物光学系により観察する光学機器において、前記左右

図2 Fタームを付与した文献のアテンションの重みの可視化

参考文献

- [1] 保倉行雄, 特許庁の施策とIPCCのあゆみ, IPCC創立20周年記念誌—IPCC20年のあゆみ—, 2006, pp.275-277
- [2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosikhin, “Attention Is All You Need,” NIPS 2017

[3] J. Devlin, M-W. Chang, K. Lee, K. Toutanova,
“BERT : Pre-training of Deep Bidirectional
Transformers for Language Understanding.”
2018