

IPCCAT を用いた IPC 予測:

特許庁付与 IPC との比較

○藤田明¹⁾

JFEテクノリサーチ株式会社¹⁾

〒210-0855 神奈川県川崎市川崎区南渡田町 1 番1号

Tel: 044-322-6305 FAX: 044-322-6521

E-mail: ak-fujita@jfe-tec.co.jp

Predicted IPC by using IPCCAT: The Comparison of IPC Allotted by Patent Offices

FUJITA Akira¹⁾

JFE Techno-Research Corp.¹⁾

1-1, Minamiwatarida-cho, Kawasaki-ku, Kawasaki-shi, Kanagawa 210-0855 Japan

Phone: +81-44-322-6305 Fax: +81-44-322-6521

E-mail: ak-fujita@jfe-tec.co.jp

【報告概要】

世界知的所有権機関 (WIPO) の公開している IPCCAT (Categorization Assistant in the International Patent Classification) という分類ツールを用いると、要約等を入力して IPC を予測することができ、公報に記載されている技術がどの特許分類に属するかを予測することができる。最近は人工知能を用いて IPC の予測精度があがってきているとの報告がある。そこで米国特許の要約ならびに日本特許の PAJ 要約を入力として予測した IPC が特許庁付与の IPC とどのくらい一致するかを調べた。米国特許では Subgroup レベルで 93%まで一致したが、日本特許で PAJ を入力した際には Group レベルで 80%、Subgroup レベルでは 64%の一致率であり、中には Section レベルで一致しない特許公報もあった。英語翻訳手段を変えた場合には一部異なる結果が得られた。

【キーワード】

IPCCAT, IPC, 特許情報, 特許庁, 国際特許分類, PATENTSCOPE

1. はじめに

世界知的所有権機関(WIPO)が公開している IPCCAT (Categorization Assistant in the International Patent Classification)は、特許文献を国際特許分類(IPC)に分類するツールであり[1]、要約あるいは請求項を入力すると関連性の高い IPC を予測してくれる。2018年5月1日付の WIPO のニュースでは、以下のように報じている。ニューラルネットワーク技術を使用する IPC の自動分類は、72,137 のサブグループの中で最も関連性の高いシンボルを選択することができるようになり、その上位3つの予測の少なくとも1つが特許分類専門家による分類と一致する確率は英語で80%である[2]。IPCCAT のサイトから閲覧できる、IPC Publication Help (IPCPUB Version 7.7:2018年12月3日に更新)によると、上記の一致確率は、Maingroup レベルで 88.9%、Subgroup レベルで 81.9%となっている[3]。

INFOPRO2016 で、アジア新興国の各国特許庁が付与する IPC の正確度の実態を評価する基準として IPCCAT による分類との一致の度合いを調査した[4]。その時のデータも参照しつつ、今回は最近の IPCCAT による分類の一致度の進化について調査した。

2. 調査方法

WIPO の提供するこの IPCCAT は、テキストを所定の欄に入力することで、その内容の技術カテゴリーに近い IPC を提示するものである[1]。入力言語として英語、フランス語を選択することができる。分類レベルとして Class, Subclass, Maingroup, Subgroup の4種類を選択でき、また、3個あるいは5個の IPC の予測結果を得ることができる。

入力するテキストは発明主題を的確に

表している内容にすべきであるが、前に行った調査結果から特許文献の要約(Abstract)を入力することにした[4]。対象特許文献は、WIPO のサイトで検索した2015年8月6日公開の米国特許80件を対象に、Maingroup レベル3個を IPCCAT により分類し、それらを実際に付与された IPC を比較した。なお、技術分野が偏らないよう、Section の A から H まで各10件ずつ抽出した。また、直近では Subgroup レベルでの比較も行った。

一方、日本特許についても、IPCCAT の予測する分類と日本国特許庁の付与する分類との一致度を見るため、公開特許公報英文抄録(PAJ)を用いて、Section の A から H まで各10件ずつ計80件を対象として比較した。

双方の IPC が一致するレベルは表1で説明すると、Subgroup まで一致、すなわちすべて一致するレベルから、Section すら一致しないレベルまで6段階に分けられる。ここで IPCCAT が予測

表1 各 IPC の一致レベルに分類例

レベル	IPC	IPCCAT
SubGroup	<u>B65H37/06</u> <u>D21F7/00</u> <u>B42C19/02</u>	<u>B42C19/02</u> <u>B65H37/06</u> <u>D21F7/00</u>
Group	<u>F01N3/08</u> F02D45/00 B01D53/94	<u>F01N3/20</u> <u>F01N11/00</u> <u>F01N9/00</u>
SubClass	<u>E04D13/00</u> <u>E04B1/41</u>	F24J2/52 H01Q1/12 <u>E04D1/34</u>
Class	<u>H04B10/116</u> <u>H04B10/80</u> <u>H01S5/022</u>	<u>H04J14/02</u> A61B5/01 <u>H04N5/235</u>
Section	<u>B32B27/00</u> <u>B32B27/04</u> <u>B32B27/42</u>	<u>B41M5/52</u> <u>B41M5/50</u> C09D11/322
一致せず	H05B33/08 H05B37/02	G06F3/041 G06F3/046 A43B13/12

するIPCの個数は3個とし、このうち一致レベルの最も高いIPCで評価した。表1のGroupレベルの例では特許庁付与のIPCに対してIPCCATの3個のIPCのうちGroupレベルで一致しているのは1個、Sectionレベル一致しているのが1個、残りは全く一致していないが、ここでは「Groupレベルで一致」と評価した。

3. 米国特許IPCとの比較

米国特許庁(USPTO)はこれまで、米国独自の分類であるUSPCを用いており、直接IPCを付与しているわけではないが、英語を母国語としていることから、その要約をIPCCATに入力してIPCを予測することにした。

IPCCATはバージョンアップを幾度が繰り返しており、2014年12月3日更新の画面では「IPCCAT v2014.11」という版で示されていたのが、2017年10月5日更新では「IPCPUB v7.3」と変わっており、画面も大きく変わっている。最近のバージョンでは、「IPCPUB v7.7」と表示が出ており、バージョンの数字が変わらない期間でも更新日が変わっており、部分的に更新になっている可能性もある。今回は、「IPCCAT v2014.11」という時期である2017年9月12日、「IPCPUB v7.6」の2018年6月14日、さらに「IPCPUB v7.7」の2018年11月1日にそれぞれ調査した結果を示す。2017年9月12日時点では、各Sectionから10個ずつ抽出した特許についてIPCCATで予測しているが、平均でおよそ6割がMaingroupまで一致しているものの、それ以外は一致レベルが低く、またSectionによる差も大きい。ところが、2018年6月14日時点では、同じ特許80件を対象にした結果で9割以上がMaingroupレベルで一致している。さらに2018年11月1日では一致レベルが上がっている。

少なくとも2018年4月27日更新の「IPCPUB v7.6」から、Subgroupまで予測できる仕様となっている。図1と同じ特許を対象に2018年11月1日時点で調べた結果、Subgroupレベルまで一致率は高く、図2に示したように、10件すべて一致している分野も見られた。

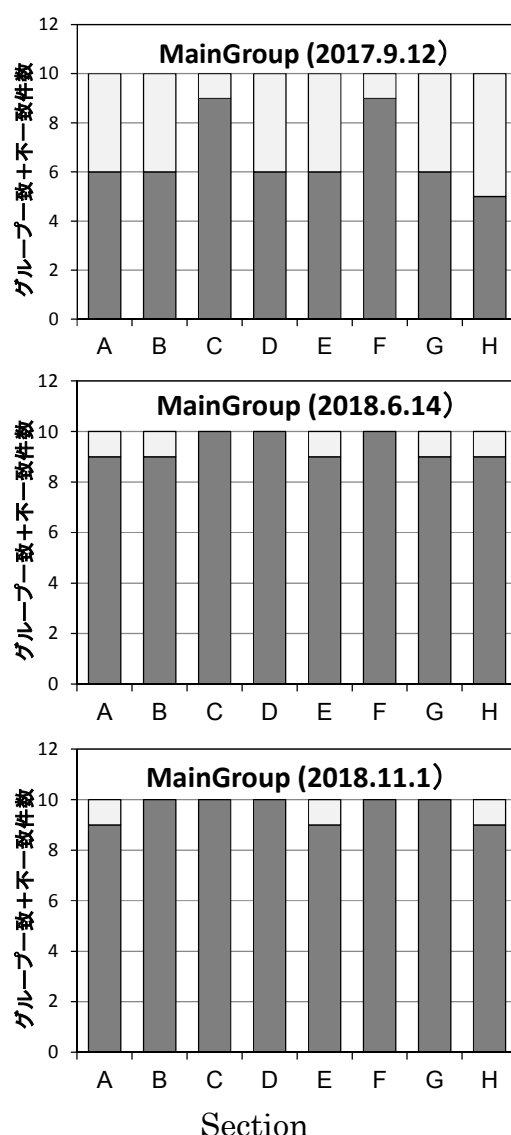


図1 米国特許を対象にしたIPCの一致レベル (ハッチングはGroupまで一致している件数)

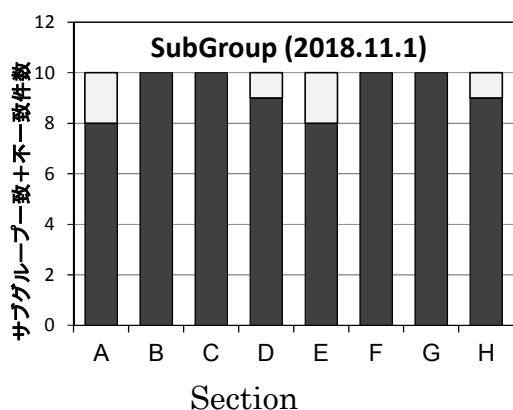


図2 米国特許を対象にした IPC の一致レベル (濃いハッチングは Subgroup まで一致している件数)

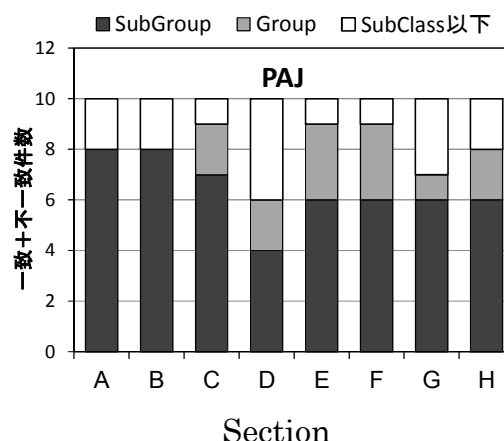


図4 PAJ の英文抄録を入力して IPCCAT で予測した IPC と、特許庁付与 IPC との一致件数 (SubGroup レベルで一致した件数を濃いグレーで、Group レベルで一致した件数を薄いグレーで、それ以外の件数を白で示した)

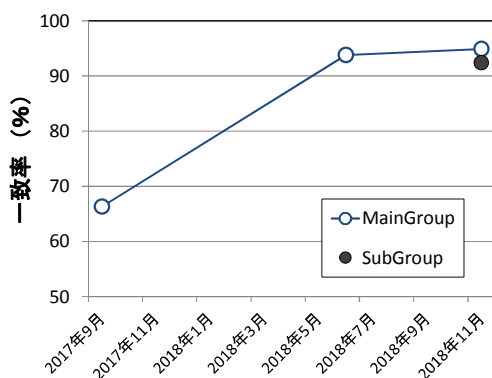


図3 米国特許を対象にした特許庁付与 IPC と IPCCAT 予測 IPC との一致率の年代変化 (○と●は、それぞれ MainGroup レベル、SubGroup レベルの指定で予測した結果)

特許庁付与 IPC と IPCCAT 予測の IPC が一致する割合を一致率と定義し、調査時期でプロットしたものが図3である。この1年で一致率は格段に上がっており、SubGroup のレベルでも92%を上回っている。WIPO が公表している数字に比べるとかなり高い数値となっている。

4. PAJ を用いた比較

米国特許で高い一致率を示したが、日本特許で IPCCAT を活用するにあたり、日本国特許庁の付与する IPC と比較した。もし、日本語で書かれた要約等を

入力して適確な IPC が予測できるのであれば先行技術調査や技術分類作業にも活用できる。ただし、入力は英語もしくはフランス語に限られているため、翻訳を介さなければならない。そこで、機械翻訳ではない公開特許公報英文抄録 (PAJ) を入力して IPCCAT の予測する IPC と特許庁付与の IPC と比較した。結果を図4に示す。Group レベルの一致率は80%であるが、Subgroup レベルでは64%まで低下する。特に D Section (分野: 繊維; 紙) では一致率が低い。Section すら一致しない件数が80件中5件あり、それらを表2に示した。

No.1 の「画像記録装置および画像記録方法」では、公報記載の IPC は、Group の B41J2 が共通でインクジェットプリンタに関する分類であるが、IPCCAT が予測した IPC は、G06K15/10 (マトリックス型プリンタ)、あるいは H04N1 (画像通信の伝送・再生) であり、まったく異なる技術範囲というわけではない。一方、No.2 の IPC は放射性汚染物質の処理、汚染除去装置に対

表2 特許庁付与のIPCとPAJを入力したIPCCAT推測IPCとが一致しなかった特許

No.	公開番号	発明の名称	公報記載 IPC	IPCCAT予測 PAJ 入力	* IPCCAT予測 WIPO Translate
1	特開2017-35793	画像記録装置および画像記録方法	B41J2/525 B41J2/205 B41J2/21	G06K15/10 H04N1/54 H04N1/405	G06K15/10 H04N1/60 H04N1/54
2	特開2017-36972	放射性ストロンチウムを含有する被処理水の処理方法及び処理装置	G21F9/12 G21F9/06 G21F9/10	C02F1/44 C02F5/02 B01D65/08	C02F1/44 G21F9/12 B01D65/08
3	特開2017-37105	前面板および前面板付き表示装置	G09F9/00 G06F3/041 G02F1/1333	B41M3/14 H01Q1/38 F16J15/06	G02F1/13363 H01L51/52 G09F9/30
4	特開2017-29631	繊維製品の減圧沸騰攪拌式洗浄システム	D06F35/00	A47L15/00 H01L21/67 A61B1/12	A47L15/00 D06F39/02 B08B3/08
5	特開2016-112012	細胞吸引システム及びこれを用いた細胞内物質の吸引作業を行う方法	C12M1/00 C12M1/26 C12Q1/02	G01N35/10 G01N33/558 G01B11/24	G01N35/10 A61J1/03 B25J9/00

して、予測はC02F1/44(廃水, 下水または汚泥の処理)は近い印象を与えるが、他のIPCは、C02F5/02(水の軟化)、B01D65/08(半透膜の汚れ防止)である。No.4では洗浄というキーワードでは類似しているが対象が異なるように思われる。

これらの不一致は入力した抄録の言葉に基づいて判断されているため、別の翻訳文を入力して比較することを試みた。WIPOのPatentscopeは、翻訳を従来の統計的機械翻訳からニューラル機械翻訳に切り替えてその品質が改善されていると報告があった[5]。そこで、表2に挙げた公報の和文要約をこの翻訳機能[6]を用いて英文に翻訳して、IPCCATに入力して予測したところ、表2の右端(*IPCCAT予測WIPO Translateの列)に示した結果が得られ、No2~No.4の3文献で一部SubclassからMaingroupレベルの範囲で一致したIPCがあった。このことから、日本語で書かれた公報からIPCCATを用いたIPC予測は英文翻訳の品質が影響を及ぼすことは確かである。

5. まとめ

英文要約を入力するIPCCATによりその特許のIPC候補を予測することができ、米国と日本特許の公報記載のIPCと比較した結果、以下のことがわかった。

- (1) 年度を経るごとにIPCCATが予測するIPCと米国特許の公報記載のIPCとの一致度は向上しており、任意抽出した80件を対象に2018年11月時点で予測した結果、Subgroupレベルで92%に達している。
- (2) 日本特許を対象にして、現時点での同様の調査を、PAJ要約を入力することにより行ったところ、Groupレベルで80%、Subgroupレベルで64%であった。一致率は翻訳にも依存する。

6. おわりに

本報告を完成するにあたり、アジア特許情報研究会の皆様には情報の提供および指導を頂きました。ここにあらためて御礼申し上げます。

7. 参考文献

- [1] WIPO の IPCCAT のサイト
<https://www3.wipo.int/ipccat/faces/pages/SessionTimeout.jsp>
- [2] WIPO News
https://www.wipo.int/classifications/en/news/ipc/2018/news_0004.html
- [3] IPC Publication Help (IPCPUB Version 7.7)
<https://www.wipo.int/classifications/ipc/ipcpub/?notion=help&version=20180101&symbol=none&menulang=en&lang=en&viewmode=f&fipcpc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=advance>
- [4] 第 13 回情報プロフェッショナルシンポジウム (INFOPRO2016) C21
「IPCCAT を用いた新興国特許庁 IPC 付与の実態調査」
- [5] Magdalena Zelenkovska 「パテントスコープと特許情報検索のためのツール」2018.Nov7 特許・情報フェア&コンフェレンス
あるいは、WIPO Press Releases:
https://www.wipo.int/pressroom/en/articles/2016/article_0014.html
- [6] WIPO Patentscope Translate
<https://patentscope.wipo.int/translate/translate.jsf?interface>