

# 特許ドメイン特化型 BERT による脱炭素関連特許技術の 「見える化」

“Visualization” of decarbonization-related patent technologies by patent domain-specific BERT

前原義明<sup>1</sup> 久々宇篤志<sup>1</sup> 長部喜幸<sup>1</sup>

Yoshiaki MAEHARA<sup>1</sup>, Atsushi KUKU<sup>1</sup>, and Yoshiyuki OSABE<sup>1</sup>

<sup>1</sup> 日本特許情報機構

<sup>1</sup>Japan Patent Information Organization

**Abstract:** We trained a classifier to determine whether a technology in the patent literature corresponds to a decarbonization-related technology or not, using BERT pre-trained on the patent documents and the corpus of decarbonization-related patents we collected. The trained BERT model was also used to conduct macro analysis of Japanese published patent gazettes applied after 2000 to visualize the decarbonization-related patent technologies in Japan.

## 1. はじめに

日本政府は 2050 年カーボンニュートラル実現を宣言し[1], 2030 年度の温暖化ガス排出削減目標として, 46%削減 (2013 年度比) することを目指し, さらに 50%の高みに向けて挑戦を続けるとの新たな方針を示した.

これを受け, 経済産業省は「2050 年カーボンニュートラルに伴うグリーン成長戦略」を公表し[2], この挑戦を「経済と環境の好循環」につなげるための産業政策として策定した. 経済産業省は, エネルギー政策及びエネルギー需給の観点から, 「エネルギー関連産業」「輸送・製造関連産業」「家庭・オフィス関連産業」において成長が期待される 14 分野を導き出し, 2050 年に向けた高い目標を設定している.

2050 年カーボンニュートラルの実現は, エネルギー・産業部門の構造転換, 大胆な投資によるイノベーションの創出といった取組を, 産学官が一体となり, 大きく加速させることが必要である. 産業界は, これまでのビジネスモデルや戦略を根本的に変えていく必要がある一方で, もともと省エネ技術に長けている多くの日本企業にとって, 新しい時代をリードしていくチャンスでもある. そして, 目標の実現には, イノベーションに密接に関連する特許情報の整理・俯瞰が欠かせない.

しかし, 脱炭素技術はカバーする技術範囲・産業分野が広く, また, 通常の特許検索で使用されてい

る特許分類には脱炭素技術に係る特別の分類コードが付与されておらず, 脱炭素に関連する技術を網羅的に捉えることは困難であった. さらに, 扱うデータ量が膨大なことから, 自然言語処理技術を用いて特許明細書から直接脱炭素技術を判定できる手法の開発が望まれていたところである.

そこで, 本研究では, 欧州特許庁 (EPO) で付与されている Y02 (気候変動緩和技術), Y04 (スマートグリッド関連技術) という共通特許分類 (CPC) に着目し, 当該分類が付与されている日本語特許公報を収集してコーパスを作成, このコーパスを利用して, 特許明細書の記載が Y02, Y04 の技術に包含されるか否かを判定する機械学習モデルを作成した. そして, この機械学習モデルを用いることによって, 2000 年以降に日本に出願された特許のマクロ解析を行い, 脱炭素関連特許技術の「見える化」を行った.

このマクロ解析の結果は, ESG 投資をはじめとする「脱炭素マネー」の行方を決める点においても, 非常に重要な情報となり得る.

## 2. 分析手法

### 2. 1. 解析モデル

機械学習モデルとしては BERT[3]を採用した. BERT は Bidirectional Encoder Representations from Transformers の略であり, 各コーパスにおいて, 少ないコーパスで高い精度を出すことができるものとし

て有名だが、特許分類タスクにおいて最も高い精度を出しているモデルでもある。[4][6]

また、BERT はドメイン特化によって、その性能が向上することが知られている[5]。そのため、本研究においては、2018 年から 2019 年の日本国公開特許公報を用いて、BERT の事前学習を行った（ここで事前学習を行った特許ドメイン特化型 BERT のことを、以下「PatentBERT\_japanese」と呼ぶ。）。PatentBERT\_japanese の学習には、NVIDIA が考案した LAMB Optimizer[7]を用いた。

表 1 に BERT の事前学習結果を示す。非常に高いマスク語予測精度から、マスク語予測に特許文献を用いることによって、特許文献のマスク語予測が適切にできるようになったことが理解できる。

表 1 特許ドメイン特化型 BERT の事前学習結果

	MLM acc.	NSP acc.
PatentBERT_japanese	93.5	98.7

また、以下の図 1 は、PatentBERT\_japanese によるマスク語予測のデモである。

**マスク語予測タスク**

ルテイン、ゼアキササンチン, [MASK], アスタキサンチン, ドコサヘキサエン酸, ドコサペンタエン酸, エイコサペンタエン酸, ビタミンE,  $\beta$ -カロテン, 及びコエンザイムQ-10から選ばれる1種以上を含有することを特徴とする請求項1~8のいずれか1項に記載の視力回復用組成物

**PatentBERT\_japanese によるマスク語予測トップ3**

1. リコピン (正解)
2. アスタキサンチン
3. カロテノイド

図 1 PatentBERT\_japanese によるマスク語予測デモ

これにより、PatentBERT\_japanese は、特許明細書中に出現する非常に専門的な語であっても、その意味概念を概ね正しく捉えることができることが分かる。

そして、この PatentBERT\_japanese の最下層に、入力された特許明細書の文章が Y02, Y04 の各 CPC メイングループ(CPC はセクション (例: Y)、クラス (例: Y02)、サブクラス (例: Y02B)、メイングループ (例: Y02B10) に階層化されており、メイン

グループは、8 デジットある分類情報のうちの 6 デジットまでの情報を示す)に該当するか否かを判定する分類層を追加することによって、本研究で用いる機械学習モデルを構築した。なお、対象とした Y02, Y04 のメイングループは 39 種類あるため、最下層に追加した分類層は 39 クラスの分類層である。

なお、特許明細書に対する Y02, Y04 の CPC メイングループの付与は重複的に行われるため、当該 CPC メイングループの推定はマルチラベルタスクである。そのため、損失関数には BCEWithLogitsLoss を用いることとした。

## 2. 2. コーパス

コーパスについては、Y02, Y04 が付与されている日本語特許公報を収集し、さらに、特許技術者が当該収集した特許公報の明細書中身を目視で確認することによって収集を行った。コーパスのサイズは、合計 8,724 件である。

## 2. 3. 計算条件

計算条件は以下の通りである。

表 2 訓練における各種パラメータ

	BCEWithLogitsLoss
損失関数	BCEWithLogitsLoss
学習率	$5.0 \times 10^{-5}$
バッチ数	256
エポック数	50

## 3. 結果

コーパスの 10% をバリデーション用データとして、ホールドアウト法によって精度の検証を行った。結果は以下の通りである。

表 3 学習結果

Precision(Micro)	0.89
Recall(Micro)	0.86
F1(Micro)	0.88
F1(Macro)	0.86

## 4. 2000 年以降に日本に出願された特許出願のマクロ分析

学習したモデルを用いて、2000 年以降に日本に出願された特許出願 6,700,799 件全てについて推論を行った。そのうち、2019 年に出願された特許 226,180

件について、その結果を示す。なお、対象を2019年に出願された特許に絞った理由は、通常、特許出願は出願日から1年6月経たないと公開されず、現時点で2020年以降に出願された特許すべてを分析できないためである。

当該モデルに特許の明細書を入力すると、Y02、Y04のCPCメイングループからなる39のCPCメイングループの推定値が算出される。ここで、本研究では、39のCPCメイングループと経済産業省が定める「エネルギー関連産業」「輸送・製造関連産業」「家庭・オフィス関連産業」の3産業とのコンコードダンスを作成し、当該コンコードダンスに基づいて、39のCPCメイングループの推定値を、特許明細書に記載されている技術が、当該3産業のうちどの産業の脱炭素に資するものであるのかについての推定値へ変換を行った。

そして、各出願人について3産業それぞれの推定値の合計をスコアとして算出することによって、2019年に出願された特許における各3産業の出願人スコアランキングを算出した。これらを以下に示す。

表4 企業ランキング  
エネルギー関連産業，2019年出願

1	トヨタ自動車株式会社
2	パナソニックIPマネジメント株式会社
3	株式会社豊田自動織機
4	エルジー・ケム・リミテッド
5	本田技研工業株式会社
6	株式会社GSユアサ
7	株式会社東芝
8	株式会社デンソー
9	積水化学工業株式会社
10	株式会社豊田中央研究所

表5 企業ランキング  
輸送・製造関連産業，2019年出願

1	トヨタ自動車株式会社
2	本田技研工業株式会社
3	株式会社デンソー
4	キヤノン株式会社
5	パナソニックIPマネジメント株式会社
6	株式会社ディスコ

7	株式会社豊田自動織機
8	株式会社SUBARU
9	マツダ株式会社
10	三菱電機株式会社

表6 企業ランキング  
家庭・オフィス関連産業，2019年出願

1	パナソニックIPマネジメント株式会社
2	三菱電機株式会社
3	株式会社日立製作所
4	キヤノン株式会社
5	ダイキン工業株式会社
6	株式会社東芝
7	株式会社半導体エネルギー研究所
8	シャープ株式会社
9	ファナック株式会社
10	株式会社デンソー

## 5. まとめと今後の課題

機械分類モデルでは、コーパスの品質及びサイズが非常に重要であるため、コーパスの整備及び整備後のコーパスによるモデルの再学習を今後も継続的に行っていく必要がある。

また、本研究では日本に出願された特許の公開特許公報を解析対象としたが、弊機構は高精度の機械翻訳エンジンを持っていることから、海外の特許明細書を日本語化して、本モデルの解析対象とすることも可能である。今後は引き続き、国際公開の特許明細書や米国特許公開公報等の解析を行っていく。

## 参考文献

- [1] 環境省, 2050年カーボンニュートラルの実現に向けて, [https://www.env.go.jp/earth/2050carbon\\_neutral.html](https://www.env.go.jp/earth/2050carbon_neutral.html), (参照 2021年9月2日)
- [2] 経済産業省, 2050年カーボンニュートラルに伴うグリーン成長戦略を策定しました, <https://www.meti.go.jp/press/2021/06/20210618005/20210618005.html>, (参照 2021年9月2日)
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. (2018)

- [ 4 ] Jieh-Sheng Lee, Jieh Hsiang.: PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model. arXiv:1906.02124. (2019)
- [ 5 ] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. arXiv:1901.08746. (2019)
- [ 6 ] Manzil Zaheer, Guru Guruganesh, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, Amr Ahmed.: Big Bird: Transformers for Longer Sequences. arXiv:2007.14062. (2020)
- [ 7 ] Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, Cho-Jui Hsieh.: Large Batch Optimization for Deep Learning: Training BERT in 76 minutes. arXiv:1904.00962. (2019)