

# 特許ドメイン特化型 BERT による SDGs 関連特許技術の「見える化」

前原義明<sup>†1</sup> 久々宇篤志<sup>†1</sup> 長部喜幸<sup>†1</sup>  
一般財団法人 日本特許情報機構 知財 AI 研究センター<sup>†1</sup>

## 1. はじめに

SDGs とは「Sustainable Development Goals (持続可能な開発目標)」の略称であり、2015 年 9 月の国連サミットで採択された、国連加盟 193 カ国が 2016 年から 2030 年の 15 年間で達成するために掲げた目標である [1]。SDGs は以下の通り、17 の大きな目標と、それらを達成するための具体的な 169 のターゲットで構成されている。



図 1 持続可能な開発目標

機械学習を用いた SDGs 関連の文章分類についてはいくつかの先行研究があるものの[8][11]、特許文献を SDGs の観点から分類した先行研究は、今まで存在しなかった。すなわち、SDGs の実現には技術イノベーションが不可欠であるものの、現状では SDGs 関連技術がどこ（国や企業）にどの程度あるのか不明であった。

本研究では、自然言語処理技術である BERT、及び、日本特許情報機構（以下、「弊機構」という）が所有する日本国特許公報を用いて、我が国における SDGs 技術の「見える化」について取り組んだ。

## 2. 分析手法

### 2.1 解析モデル

解析に用いるモデルは BERT[4]を選択した。BERT は Bidirectional Encoder Representations from Transformers の略であり、少ないコーパスながら、最も成果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。[5][7]

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[6]。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。なお、学習には NVIDIA が考案した LAMB Optimizer[9]を用いた。

表 1 に示した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 accuracy	NSP accuracy
PatentBERT_japanese	93.5	98.7

以下の図 2 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、β-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 2 PatentBERT\_japanese によるマスク語予測デモ

### 2.2 マルチラベルタスク化

我々が以前に行った研究[10]では、損失関数として CrossEntropyLoss を選択し、本問題をシングルラベルタスクとして学習を行った。しかし、SDGs においては、この問題設定は最適であるとはいえない。これは、1つの特許文献に対して複数の SDGs が該当することがあるからである。そのため、本研究においては、損失関数を BCEWithLogitsLoss として、タスクをマルチラベル化した。

なお、推論を行う際は、Pooler 層から出力された Logits を sigmoid 関数で正規化した値について、0.5 を閾値として、ラベルの正負を決定している。

<sup>†1</sup> IP AI Research Center, Japan Patent Information Organization

## 2.3 訓練データ

SDGsのゴールについて、16（平和と公正をすべての人に）等、一部は特許ないし科学技術が達成すべき目標としてそぐわないものがある。そのため、本研究においては、世界知的所有権機関（WIPO）が特許ないし科学技術が寄与すべきゴールとして挙げていた[3]、2（飢餓をゼロに）、3（すべての人に健康と福祉を）、6（安全な水とトイレを世界中に）、7（エネルギーをみんなにそしてクリーンに）、8（働きがいも経済成長も）、9（産業と技術革新の基盤をつくろう）、11（住み続けられるまちづくりを）、13（気候変動に具体的な対策を）についてのみコーパスを収集した。

具体的には、以前に我々がシングルラベルタスクとして学習を行った際に、2020年公開の日本特許公報に対して付与したラベルを参考にしながら、目視で各ゴールについての代表的な文献をそれぞれ100件抽出した。したがって、抽出した全ての目標とラベルの対のデータは800件である。

## 2.4 計算条件

計算条件は以下の通りである。なお、早期終了はバリデーションロス値を監視して行った。

表 2 訓練における各種パラメータ

損失関数	BCEWithLogitsLoss
学習率	$5.0 \times 10^{-5}$
バッチ数	256
エポック数	34(早期終了)

## 3. 結果及び考察

### 3.1 検証データによる精度検証

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)	0.949
Recall(Micro)	0.938
F1(Micro)	0.943
F1(Macro)	0.946

## 4. 2020年に公開された日本国公開特許公報のマクロ分析

学習させたモデルを用いて、2020年に公開された日本国特許公開公報のデータ288,196件全てに対して推論を行った。各目標に該当すると判定された文献の件数を図3に示す。

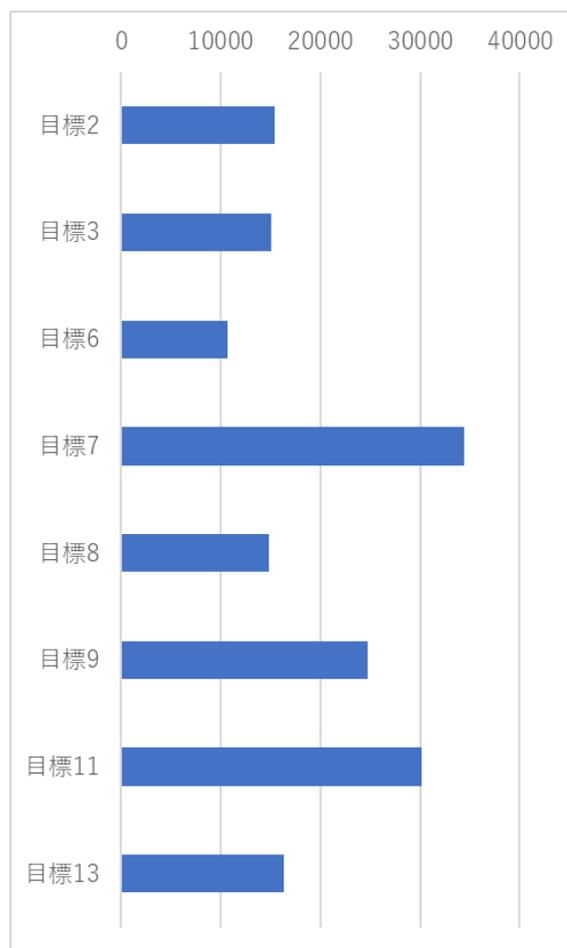


図 3 2010年に公開された日本国公開特許公報において、SDGsの各目標に該当すると判断された文献の件数

図3より、日本に出願された特許は、目標7（エネルギーをみんなにそしてクリーンに）、目標11（住み続けられるまちづくりを）等に注力していることがわかった。

また、各ラベルにおいて、最も推定値が高かった公開特許公報は、以下の通りである。

表 4 各ゴールについて最も推定値が高かった公報

SDGs	公報番号及び内容
目標 2	<u>JPWO2019066067S (積水化学工業株式会社)</u> 塩化ナトリウムを含む栽培用溶液を用いたナス科植物栽培において、前記載培用溶液中の塩化ナト...
目標 3	<u>JP2020509392T (アナバイオスコーポレーション)</u> 選択された少なくとも1つの心臓に関連する少なくとも1つの収縮性パラメータに基づき、少なくとも...
目標 6	<u>JP2020194666A (ダイニチ工業株式会社)</u> 【課題】水タンクへの給水量が低下している状況下においても、水張りや補水等の水タンクへの

	給水操...
目標 7	<u>JP2020034571A (アクアラミーターズリミテッド ライアビリティカンパニー)</u> 【課題】双方向通信プロトコルを含む、より先 進的なエネルギー監視を可能にするエネルギー メータ向...
目標 8	<u>JP2020095296A (パナソニック I P マネジメント 株式会社)</u> 【課題】対応可能な作業者の不在に起因する生 産設備の停止時間を推定することができる生産 管理装置...
目標 9	<u>JP2020123841A (株式会社リコー)</u> 【課題】有線通信と無線通信が併用可能であ って、ユーザの利便性を向上させることができ る通信装置...
目標 11	<u>JP2020095160A (旭化成ホームズ株式会社)</u> 【課題】振動に伴う建物の倒壊をユーザに疑似 的に体験させることが可能な振動体験システム を提供す...
目標 13	<u>JP2020517456T (ジョンソン、マッセイ、パブリ ック、リミテッド、カンパニー)</u> ディーゼルエンジンからの排気ガスを処理す るためのNO <sub>x</sub> 吸収体触媒。NO <sub>x</sub> 吸収体触媒 は、モ...

	向上させるための技術を提供する。【解決手段】 燃料電...
目標 8	<u>JP2020060698A (学校法人早稲田大学)</u> 【課題】誤りを含むプログラムコードを間違い 修正問題として学習者に提示することである。 【解決手...
目標 9	<u>JP2020129711A (国立大学法人徳島大学)</u> 【課題】データを迅速に転送するための経路を 選択するための経路選択方法を実現する。【解決 手段】 ...
目標 11	<u>JP2020003400A (国立大学法人金沢大学)</u> 【課題】移動体の横方向位置を精度良く推定で きる横方向位置推定装置等を提供する。【解決手 段】 横...
目標 13	<u>JP2020200287A (国立大学法人長崎大学)</u> 【課題】 第一または第二アルコール化合物を 酸化してカルボニル化合物を得る方法を、次亜塩 素酸...

また、大学のみ絞ったデータを図4及び表5に示す。  
これにより、大学は主に目標3（すべての人に健康と福祉  
を）や目標13（気候変動に具体的な対策を）に注力して  
いることがわかった。

表 5 各ゴールについて最も推定値が高かった公報 (大学  
による出願)

SDGs	公報番号及び内容
目標 2	<u>JP2020143047A (学校法人君が淵学園)</u> 【課題】管理が容易且つ低コストな植物成長促 進剤、及び該植物成長促進剤の製造方法を提供す る。【...
目標 3	<u>JP2020196704A (学校法人帝京大学)</u> 【課題】インフルエンザ感染症またはコロナウ イルス感染症を予防および／または治療するた めの新規...
目標 6	<u>JP2020026933A (公立大学法人大阪)</u> 【課題】上部帯水層及び下部帯水層の利用に際 し、井戸の閉塞が抑制される地中熱利用システム 及び地...
目標 7	<u>JP2020205212A (国立大学法人東京大学)</u> 【課題】燃料電池システムのエネルギー効率を

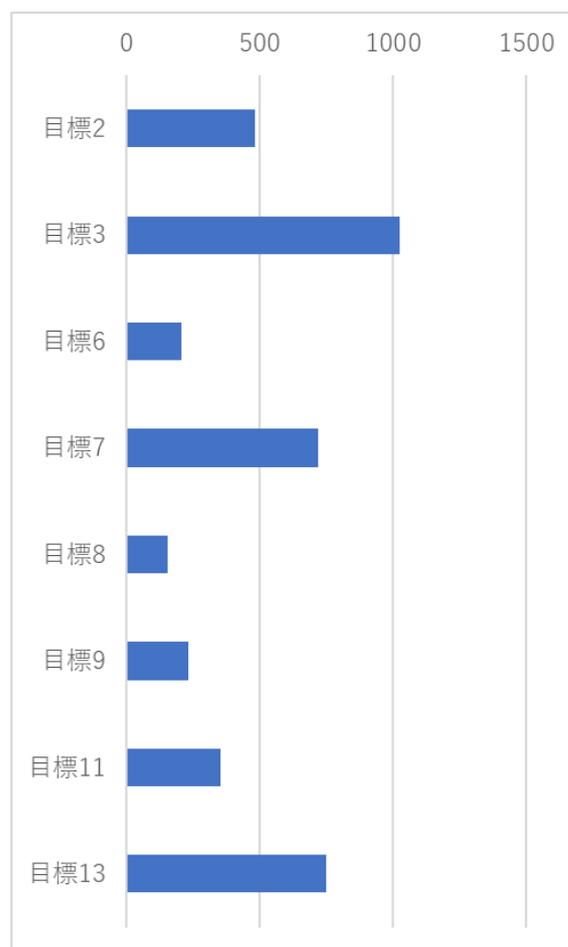


図 4 2010年に公開された日本国公開特許公報 (大学による出願) において、SDGsの各目標に該当すると判断された文献の件数

## 5. まとめと今後の課題

特許ドメインに特化した BERT を用いることによって、SDGs 関連特許技術がどこ（国・出願人）にどの程度あるのかを可視化することに成功した。今後は、コーパスの増強を図ることによって、本研究において対象とした 8 種類のゴール以外のゴールの解析も行えるようにしていきたい。

### 参考文献

- [1] “THE 17 GOALS | Sustainable Development” .  
<https://sdgs.un.org/goals>, (参照 2021-06-07).
- [2] “国連の開発目標と知財 関連技術、特許で可視化を” .  
<https://www.nikkei.com/article/DGXXKZO30075030S8A500C1KE8000/>, (参照 2021-06-07).
- [3] “The Impact of Innovation: WIPO and the Sustainable Development Goals” . <https://www.wipo.int/sdgs/en/story.html>, (参照 2021-06-07).
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805. 2018.
- [5] Jieh-Sheng Lee, Jieh Hsiang. PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model. arXiv:1906.02124. 2019.
- [6] 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.
- [7] 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.
- [8] MedinaRodríguezSamuel. Multi-Label Text Classification with Transfer Learning for Policy Documents. UPPSALA UNIVERSITET. 2019.
- [9] 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.
- [10] 前原 義明, 久々宇 篤志, 長部 喜幸. 自然言語処理技術による SDGs 関連特許技術の「見える化」. 言語処理学会第 27 回年次大会 発表論文集. 2021.
- [11] 張鑫. 日本語データセット及び分類タスク ベースラインの作成. 一般社団法人人工知能学会. 2020.