

自然言語処理技術による SDGs 関連特許技術の「見える化」

前原 義明 久々宇 篤志 長部 喜幸
 一般財団法人 日本特許情報機構
 {yoshiaki_maehara, atsushi_kuku, yoshiyuki_osabe} @ japio.or.jp

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 ([2], [3]), 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 [4]。

本研究では、自然言語処理技術である BERT [5]、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

本研究においては、機械学習モデルとして BERT を採用した。BERT とはマスク語予測 (Masked Language Model) と次文予測 (Next Sentence Prediction) という 2 つのタスクでトレーニングされた、双方向性を持つ Transformer であり、感情分析や文章分類等の特定タスクに応用した場合に、少ないコーパスで非常に高い精度が達成できるのが特徴である。

本研究においては、まず、東北大学が開発・配布している日本語 BERT [6] の最下層に 17 クラスの分類層を追加することによってモデルを作成した。なお、タスクはシングルラベルタスクとし、最終層の出力は Softmax 関数で正規化されている。

2.2 訓練データ

本研究では、BERT を用いるため、訓練データとなるコーパスのサイズはそれほど大きくなくてもよい。そのため、目視で収集した各 SDGs に関連する約 150 個の文章を訓練データとして採用した。

2.3 分析データ

弊機構が所有する日本国特許公報に対して分析を行った。なお、東北大学が配布する日本語 BERT モデルの入力層は入力トークン列が 512 と固定されているため、特許の特徴が現出しやすい「発明の課題」についての記載がある箇所を、部分選択的に BERT に読み込ませた。

2.4 訓練及び分析対象

交差エントロピーを損失関数とし、100 エポック訓練を行ったモデルで分析を行った。Softmax で正規化された各ラベル(ゴール)の推定確率のうち、最も確率が高かったラベルをその特許文献のラベルとして採用した。訓練後におけるモデルの精度検証は、交差検証(クロスバリデーション)ではなく、訓練データ自身の精度をみたがラベル正解率 100% であった。

3 結果及び考察

2010 年以降に出願された特許出願の特許公報の全てにラベルを付与した結果、SDGs のラベルは図 2 に示す分布で付与された。

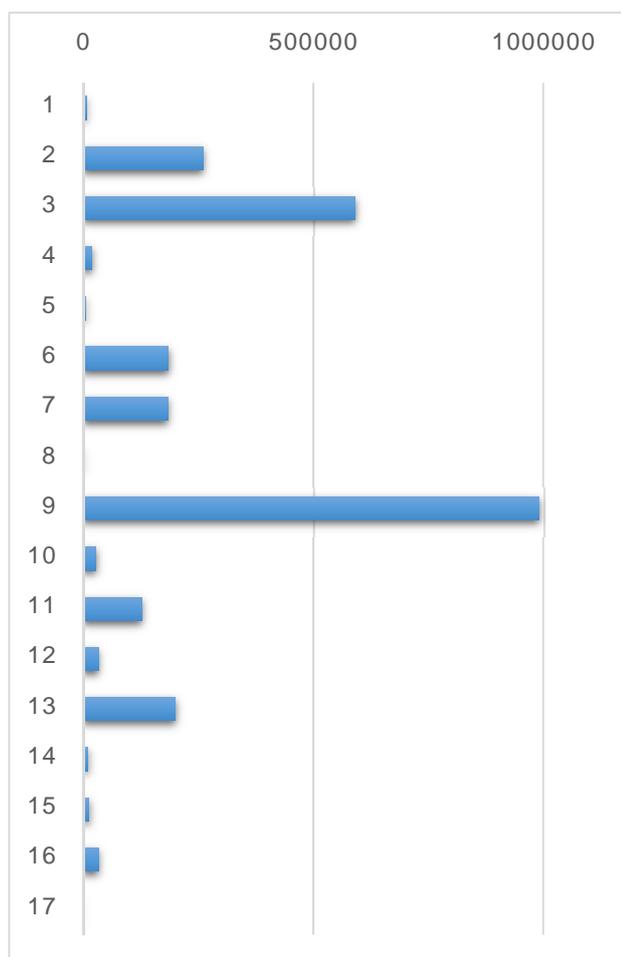


図 2 2010 年以降に出願された日本国特許公報に付与された SDGs ラベル (縦軸) の分布 (件数)

この結果から、17 個ある SDGs ゴールのうち、特許が貢献できるものは、目標 2 (飢餓をゼロに)、目標 3 (すべての人に健康と福祉を)、目標 6 (安全な水とトイレを世界中に)、目標 7 (エネルギーをみんなにそしてクリーンに)、目標 9 (産業と技術革新の基盤をつくろう)、目標 11 (住み続けられるまちづくりを)、目標 13 (気候変動に具体的な対策を)であることがわかった。これは、WIPO (世界知的所有権機関) によって、特許が寄与できる SDGs のゴールとして挙げられていたもの [7] とおおむね整合するも

のである。

また、各目標に対応する特許として抽出されたものをサンプル的に抽出すると、表 1 の通りになる。

表 1 各 SDGs に対応する特許文献の例

SDGs	対応する特許文献の記載の一部
1	顧客の与信判断を支援するシステムを提供する。...本システムは、顧客の顧客情報...
2	ロボット又は IT を利用して、農業の生産性をさらに向上させることが期待されている。...
3	本発明は、血管炎症の新たな機能的バイオマーカー、および全死因死亡率または心臓死亡率を予測する...
4	知識ユニットに基づいて新しい形の教育サービスを提供すること。...本発明の一実...
5	複数の居住者が同時に効率的に家事労働を行うことができ、また、衣類などの生活用品が...
6	今後発生するインフラ設備老朽化による使用料金アップ及び地球温暖化による水資源の枯渇な...
7	ハイブリッドエネルギー供給回路、ハイブリッドエネルギー供給システムの使用及びハイブリッド...
8	近年、AI の発達により雇用者の仕事の多くが自動化されると言われている。また、成果主義...
9	複数のサービスを連携させて提供される連携サービスの品質の低下を抑制しながら、連携サー...
10	迅速かつ効率的に仮想通貨に相当する貨幣を入手可能とすること。...仮想通貨取引...
11	建造予定地の想定災害に対して建造想定及び、建造物建造単価倍率を対応させるなる建造想定...
12	炭素繊維を含有する廃棄物を円滑に処理することを可能にする廃棄物の分別装置を提供するこ...
13	順化に要する期間を定量的かつ事前に評価することができ、熱中症の予防対策と

	して事前に有...
14	小型で簡単な構成で、その進路を制御することが可能な漂流型海洋観測ブイを提供する。...
15	害獣の捕獲に関する情報の管理を効率的に行うこと。...管理システムは、ユーザの...
16	趣向性の高い対戦ゲームを提供する。...対戦ゲームである。ゲームフィー...
17	各グループに共通のテーマを与える手法に比して、グループ全体の成果を向上させる。...

そして、モデルの推定精度を検証するために、各ゴールにつき、Softmax で正規化された推定値が高い順に 100 件、目視で推定が正解かどうか確認を行った。

目標 2, 3, 6, 7, 8, 9, 11 については、推定値が高いものについては問題なく推定できていた。しかし、目標 1 (貧困をなくそう)、等の科学技術が設定する課題としてそぐわないものについては、推定度が高いものであってもほとんどがノイズであった。また、目標 4 (質の高い教育をみんなに) については、「学習」という単語に引きずられて、機械学習に関する特許文献がノイズとして混入していた。また、目標 13 についても、「温度」という単語に引きずられたためか、空気調和機に関する特許文献が多数混入していた。

今後はコーパスを増強するか、モデルに特許ドメインに特化した BERT を用いるなどして、分類精度の向上を図っていきたい。

参考文献

1. **United Nations.** THE 17 GOALS. (オンライン) (引用日: 2020 年 12 月 24 日.) <https://sdgs.un.org/goals>.
2. **張鑫.** SDGs 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. 出版地不明: 一般社団法人人工知能学会, 2020.
3. **MedinaRodríguezSamuel.** Multi-Label Text Classification with Transfer Learning for Policy Documents. 出版地不明: UPPSALA UNIVERSITET, 2019.

4. **日本経済新聞社.** 国連の開発目標と知財 関連技術、特許で可視化を. 2018 年.

5. **DevlinJacob.** Pre-training of Deep Bidirectional Transformers for Language Understanding. 出版地不明: arXiv, 2018.

6. **東北大学 乾・鈴木研究室.** Pretrained Japanese BERT models released / 日本語 BERT モデル公開. (オンライン) 2019 年 12 月 13 日. (引用日: 2020 年 12 月 24 日.) <https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>.

7. **WIPO.** The Impact of Innovation: WIPO and the Sustainable Development Goals. (オンライン) (引用日: 2020 年 12 月 24 日.) <https://www.wipo.int/sdgs/en/story.html>.