

BF6 機械学習を用いたSDGs達成に向けたマッチング支援ツールの開発

A Development of Multi-label Text Classification and Matching System for Achieving SDGs with BERT

地球循環共生工学領域 08E17031 鈴木かの子 (Kanoko SUZUKI)

Abstract: Working on SDGs and sharing successful practices with wider stakeholders are important to achieve SDGs. In this study, with a deep-learning natural language processing model, BERT, we aimed to (1) build a classifier that enables to map the meanings of practices and issues to the SDGs context, (2) build a matching system between local issues and initiatives which can be solutions. A data frame with each document which was published by official organizations and multi-labels that are corresponded to SDGs was constructed. Pretrained Japanese BERT model was fine-tuned by a multi-label text classification task, and nested cross-validation was conducted to optimize the hyperparameters and estimate the cross-validation accuracy. Finally, a matching system was developed by obtaining cosine similarities between embedded vectors of local issues and initiatives. This system helps stakeholders take action for the Sustainable Development Goals.

Keywords: Sustainable Development Goals (SDGs), natural language processing, multi-label classification, BERT

1. 背景と目的

2030年のSDGs達成に向けて、自治体や企業・団体・個人など幅広いステークホルダーがSDGsに取り組み、成功事例を共有することが重要である。そこで本研究では、深層学習型の自然言語処理モデルで、(1)ステークホルダーが抱える社会課題やノウハウをSDGsに写像する分類器を構築し、(2)SDGsの視点で地域のニーズとシーズをマッチングするシステムを構築することを目的とした。

2. 手法

2.1 学習用データフレームの構築

(a) 2018年4月から2019年4月に各省庁が公開したSDGsに関する文書、(b)環境省の「すべての企業が持続的に発展するために 資料編」¹⁾の民間セクターの資料、(c)内閣府が収集するSDGsの課題解決等に関する提案文書²⁾、(d)SDGsに官民連携で取り組む事例³⁾から1,604個の文章を収集した。各文章と対応する複数のSDGsが対になったマルチラベル形式の学習用データフレームを構築した。文章中のランダムな単語を削除、WordNetで同義語に置換の2種類のデータオーグメンテーションを施した。

2.2 テキスト分類器の学習と評価

学習用データフレームを用いて、訓練済み日本語BERTモデル⁴⁾をマルチラベルテキスト分類タスクでファインチューニングし、入力された文章にSDGsの17個の目標のラベルを付与すると同時に、入力した文章を768次元の意味ベクトルに埋め込むことができるテキスト分類器を構築した。入れ子式交差検証でハイパーパラメータをベイズ最適化し、交差検証精度を評価した。テキスト分類器が対応するSDGsに分類するときに注目したトークンを可視化するために、BERTモデルの12の中間層のうち、ファインチューニングした最終層の12のAttention Weightの平均値を可視化した。

2.3 地域課題ニーズとステークホルダーが持つシーズのマッチング方法

テキスト分類器でデータ(c)に含まれる地域課題のニーズと様々な産官学界が有する技術やノウハウのシーズが記述された文書を意味ベクトルに変換し、(1)のコサイン類似度でマッチングした。

$$\text{Similarity} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

ここで、Similarityはコサイン類似度、 A_i, B_i はベクトルA, Bのi番目の要素、nは次元(768)を表す。

表 1 SDGs 目標別の分類精度

SDGs Goal	precision	recall	f1-score	support	SDGs Goal	precision	recall	f1-score	support
GOAL 01: No Poverty	0.94	0.91	0.93	311	GOAL 10: Reduced Inequality	0.93	0.91	0.92	352
GOAL 02: Zero Hunger	0.94	0.88	0.91	270	GOAL 11: Sustainable Cities and Communities	0.94	0.94	0.94	655
GOAL 03: Good Health and Well-being	0.94	0.92	0.93	530	GOAL 12: Responsible Consumption and Production	0.94	0.92	0.93	446
GOAL 04: Quality Education	0.96	0.94	0.95	497	GOAL 13: Climate Action	0.91	0.92	0.91	417
GOAL 05: Gender Equality	0.95	0.93	0.94	320	GOAL 14: Life Below Water	0.96	0.92	0.94	351
GOAL 06: Clean Water and Sanitation	0.94	0.90	0.92	244	GOAL 15: Life on Land	0.94	0.91	0.92	402
GOAL 07: Affordable and Clean Energy	0.95	0.95	0.95	378	GOAL 16: Peace and Justice Strong Institutions	0.92	0.92	0.92	309
GOAL 08: Decent Work and Economic Growth	0.97	0.95	0.96	641	GOAL 17: Partnerships to achieve the Goal	0.96	0.94	0.95	648
GOAL 09: Industry Innovation and Infrastructure	0.95	0.94	0.95	574	Overall	0.94	0.92	0.93	7345

3. 結果と考察

3.1 テキスト分類器の精度

テキスト分類器の目標別の精度を表 1 に示す。Overall では Precision (正予測の正答率) は 0.94, Recall (正に対する正答率) は 0.92 となり高い交差検証精度を示した。また目標別の分類精度をみると、f1-score は 0.91~0.96 となり、どの目標でも高い分類性能を示した。

3.2 目標の予測と Attention の可視化

大阪大学が性的マイノリティの取組指標「PRIDE 指標 2020」で最高評価を受賞したときのニュース記事の文章³⁾をテキスト分類器に入力すると、最も意味が近似している目標である「Goal 05: Gender Equality」が予測された。また Goal 05 に分類した判断根拠となるトークンへの Attention Weight を可視化すると、[性的, 男女, 多様]などのジェンダー平等や女性のエンパワーメントに関連が強いと考えられるトークンが注目されていることが示された (図 1)。

3.3 地域のニーズとシーズのマッチングの結果

廃校などの遊休施設を有効活用し、移住者、2拠点居住者の増加や企業のサテライトオフィスの開設や企業版ふるさと納税を通じて地域を活性化させたいというニーズを持つ岡山県真庭市に対して、地方創生 SDGs 官民連携プラットフォームの 761 のシーズとマッチングを試みた結果を表 1 に示す。平均、最大、最小のコサイン類似度は(0.189, 0.719, -0.224)となった。表 2 にコサイン類似度の上位 5 つのシーズを示した。ICT 技術を利用した観光客の誘致や中小企業と自治体の連携を述べたシーズがマッチした。

4. 今後の課題

今後は学習用データの増強を行う。また、全球規模で SDGs 達成に向けた社会課題解決のためのマッチング支援を行うために、本システムの学習用データや分類器を多言語対応させることを目指す。

参考文献

- 1) 環境省: すべての企業が持続的に発展するためにー持続可能な開発目標 (SDGs) 活用ガイドー, <<https://www.env.go.jp/policy/SDGsguide-siryu.rev.pdf>>, (2020.01.23 参照).
- 2) 内閣府: 「マッチング支援 | 地方創生 SDGs 官民連携プラットフォーム (内閣府)」, <<https://future-city.go.jp/platform/matching/>>, (2020.12.01 参照).
- 3) 内閣府: 「官民連携の事例 | 地方創生 SDGs 官民連携プラットフォーム (内閣府)」, <<https://future-city.go.jp/platform/case/>>, (2020.12.03 参照).
- 4) 鈴木正敏: 「Pretrained Japanese BERT Models / 日本語 BERT 訓練済みモデル」, <<https://github.com/cl-tohoku/bert-japanese>>, (2021.02.05 参照).
- 5) 大阪大学: 「性的マイノリティの取組指標「PRIDE 指標 2020」において、大学で唯一、2年連続で最高評価の「ゴールド」を受賞!」, <<https://www.osaka-u.ac.jp/ja/news/topics/2020/11/1201>>, (2021.02.02 参照).



性的マイノリティの取組指標「PRIDE指標2020」において、大学で唯一、2年連続で最高評価の「ゴールド」を受賞した2020年11月11日、本学は、LGBTQなどの性的マイノリティに関する取組を評価する指標「PRIDE指標2020」(1)において、大学で唯一、2年連続で、最高評価の「ゴールド」を獲得しました。「PRIDE指標」は、任意団体w#orkwithpride(2)により、「企業団体等の枠組みを超えてLGBTQが働きやすい職場づくりを日本で実現する」ことを目指して、2016年に策定されました。本学は、2017年7月11日、『大阪大学「性的指標」(SexualOrientation)」と「性自認(GenderIdentity)の多様性に関する基本方針」を公表するとともに、「ALLGEN#DE#R」トイレを設置するなど、積極的にSO#G#I(3)に関する取組、啓発活動を進めてまいりました。2020年9月15日には、学内外に対して男女協働推進SO#G#Iの理解をさらに拡げるため、総長をはじめとする役員と幹部職員(本部長級職員)22名が率先し、「イクク#ボス#言SO#G#Iアラ#イ宣言」を行いました。本学は、今後一層、総長はじめ学内構成#員が#丸#となって、多様な個性が輝くキャンパスの実現に向け、取組を推進します。そして、このダイバー#シティ&インクルー#ジョンの姿勢が社会にも広がっていくよう、発信を続けていきます。

図 1 Attention の可視化の例

表 2 マッチングの上位 5 つのシーズの例

Rank	キーワード	Cosine similarity
I	スマートフォンを活用した観光型MaaS	0.719
II	自治体にかかわる民間企業による市民の生活トラブルへの対応	0.653
III	地域観光業・産業IOT化推進プラットフォームの提供	0.624
IV	中小企業を主役とした雇用・経済・安全の維持	0.599
V	宿泊施設を利用した地域経済における雇用創出と観光客の誘致	0.558