

自然言語処理技術と機械学習を用いた SDGs 達成のためのマッチング支援ツールの開発

A Development of Machine Learning and Natural Language Processing Model for Supporting SDG

地球循環共生工学領域 28H21035 鈴木かの子 (Kanoko SUZUKI)

Abstract: Achieving the SDGs requires the participation and cooperation of all stakeholders. In this study, using the natural language processing model RoBERTa, a tool was developed, which consists of (1) a classifier that interprets the meaning of sentences and labels SDG goals, (2) a method for visualizing linkages based on co-occurrence relationships among goals, and (3) a tool for matching local issues with solutions. To develop the classifier, a dataset was created with documents issued by official organizations or companies etc. and corresponding SDG labels. A pretrained Japanese RoBERTa model was fine-tuned on a multi-label text classification task, while nested cross-validation was conducted to optimize the hyperparameters and estimate cross-validation accuracy.

Keywords: SDGs, nexus and interlinkages, matchmaking stakeholders, text classification, RoBERTa

1. 背景と目的

2030年のSDGs達成に向けて、産官学民のあらゆる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。現在、国内ではステークホルダーたちが取り組みや成功事例を共有するためのオンラインプラットフォーム；Platform Cloverが開発されている。開発上の課題はステークホルダーが達成に貢献できる目標を提案すること、目標間の相互作用を読み解くこと、地域課題とその解決策になりうる技術やノウハウを持つステークホルダーとのマッチングを行うことである。本研究では、深層学習型の自然言語処理モデル RoBERTa¹⁾をSDGsコーパスを用いて学習し、(1)ステークホルダーが抱える課題や技術をSDGsに紐づけるテキスト分類器、(2)目標間の共起関係から連環を可視化する手法、(3)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 SDGs コーパスの収集と学習用データフレームの構築

産官学民のステークホルダーたちが発行した、SDGsの定義、SDGsの各目標への取り組みを評価する指標、大学や大阪万博協会のSDGs関連プロジェクト、企業や大学が公開するSDGsに関する取り組み、さらにNPOの活動や高校生の提案するSDGsへの取組、などを表す文章を12,400個収集した。各文章とそれに対応する複数の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。

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で日本語 Wikipedia で事前学習済みの RoBERTa モデルをファインチューニングし、入力された文章に関連するSDGsの目標をマルチラベルで推論すると同時に、文章の意味を768次元ベクトルに埋め込む自然言語処理モデルを構築した。RoBERTaモデルの入力トークン長は512に設定されているためコーパスを512トークン長に分割し、学習用データベースの文章は15450個となった。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て交差検証精度を評価した。また、テキスト分類器が文章にラベルを付与するときに注目したトークンを可視化するため、モデルの12のtransformers層のAttention Weightの平均値をトークンのハイライトの強さとして可視化した。

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

地方創生SDGs官民連携プラットフォーム²⁾に掲載された地域課題のニーズと産官学民の様々なステークホルダーが有する技術のシーズについての文書(N=791)をテキスト分類器で意味ベクトルに変換した。そして意味ベクトルのコサイン類似度を求め、類似度の大きいものとマッチングを行った。

表 1 SDG 別の分類精度

SDG	precision	recall	f1-score	support	SDG	precision	recall	f1-score	support
GOAL 01: 貧困をなくそう	0.91	0.88	0.89	980	GOAL 10: 人や国の不平等をなくそう	0.94	0.93	0.94	1527
GOAL 02: 飢餓をゼロに	0.93	0.92	0.93	1019	GOAL 11: 住み続けられるまちづくりを	0.95	0.94	0.94	4268
GOAL 03: すべての人に健康と福祉を	0.97	0.95	0.96	3587	GOAL 12: つくる責任、つかう責任	0.96	0.96	0.96	6174
GOAL 04: 質の高い教育をみんなに	0.96	0.94	0.95	2636	GOAL 13: 気候変動に具体的な対策を	0.96	0.95	0.96	4623
GOAL 05: ジェンダー平等を実現しよう	0.95	0.94	0.94	1841	GOAL 14: 海の豊かさを守ろう	0.96	0.95	0.96	4571
GOAL 06: 安全な水とトイレを世界中に	0.96	0.95	0.95	2454	GOAL 15: 陸の豊かさを守ろう	0.93	0.92	0.93	3101
GOAL 07: エネルギーをみんなに、そしてクリーンに	0.96	0.94	0.95	3195	GOAL 16: 平和と公正をすべての人に	0.95	0.94	0.95	1490
GOAL 08: 働きがいも経済成長も	0.95	0.93	0.94	3808	GOAL 17: パートナースhipで目標を達成しよう	0.95	0.93	0.94	3609
GOAL 09: 産業と技術革新の基盤をつくろう	0.95	0.94	0.95	3970	Overall	0.95	0.94	0.94	52853

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。全体と目標別の f1-score はそれぞれ 0.94, 0.89~0.96 となった。大阪大学が LGBTQ が働きやすい職場づくりの評価指標である「PRIDE 指標 2022」で 4 年連続ゴールドを受賞したニュース記事³⁾を分類器に入力すると、100.0%の確率で Goal 05, 99.7%で Goal 10, 98.4%で Goal 04 が予測された。Attention Weight を可視化すると図 1 のように [Pride, 学] 等のトークンがハイライトされた。

3.2 目標間の連環の可視化

図 2 には、SDGs の 169 のターゲットの定義文を分類器でベクトル化し、コサイン類似度と共起関係から連環を可視化した。

3.3 地域のニーズシーズのマッチングの結果

コロナ禍をうけて行政サービスをオンライン化するなどの新しい持続可能な行政サービスを提供したいというニーズを持つ関東の基礎自治体に対し、地方創生 SDGs 官民連携プラットフォームの 761 の民間団体のシーズとマッチングを試みた。最大、最小のコサイン類似度はそれぞれ (0.518, -0.356) となった。高いランキングでマッチしたシーズは、地域デジタル化推進のサポート、ICT で地域回覧板の電子化等を行う事業を展開する企業であり、意味的にも近似するマッチングだと考えられる。

4. 今後の課題

今後の課題には (1) 学習用データを増強する、(2) ローカル SDGs 推進のためのオンラインプラットフォームである Platform Clover⁴⁾に本分類器とマッチングアルゴリズムの実装を行う、(3) 利用者からのレビューをもとに分類精度の評価、マッチングの有用性など実用性の評価を行う。

参考文献

- 1) Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, (2019), RoBERTa: A Robustly Optimized BERT Pretraining Approach, <https://doi.org/10.48550/arXiv.1907.11692>
- 2) 内閣府, 地方創生 SDGs 官民連携マッチングシステム, <https://future-city.go.jp/platform/>, 2023/2/2 アクセス
- 3) 大阪大学, 「セクシュアル・マイノリティの取組指標「PRIDE 指標 2022」において、4 年連続「ゴールド」を受賞!」, <https://www.osaka-u.ac.jp/ja/news/topics/2022/11/17001>, 2023/1/28 アクセス
- 4) サステナブルトランジション, 「Platform Clover」, <https://platform-clover.net/>, 2023/1/28 アクセス



セクシュアルマイノリティの取組指標「pride指標2022」において4年連続「ゴールド」を受賞! 本学はlgbtq+などのセクシュアルマイノリティに関する取組を評価する指標「pride指標2022」において最高評価の「ゴールド」を受賞しました応募を始めた2019年から4年連続のゴールド受賞を達成し全国の大学で唯一の連続受賞となります本学ではダイバーシティ&インクルージョン推進宣言に基づき一人ひとりの個性と潜在能力を十分に引き出しその強みを大学全体として活かすことを目指し様々な取組を行っています「pride指標(1)」ではそれぞれ取組が評価されゴールド受賞につながりました今後も多様な人材がいきいきと輝き尊重されその力を発揮できる環境整備の促進と徹底に尽力します「pride指標2022」結果発表の詳細はこちら(1)「pride指標」は任意団体work_withpride(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旨として、2016年に策定されました各企業団体等の取組状況を以下5つの評価指標に基づき評価しその達成度に応じてゴールドシルバーブロンズが与えられます評価指標policy_(行動宣言)representation_(当事者コミュニティ)inspiration_(啓発活動)development_(人事制度プログラム)engagement/empowerment(社会貢献渉外活動)(2)企業などの団体においてlgbtq+すなわちレスビアンゲイバイセクシュアルトランスジェンダーなどの性的マイノリティに関するダイバーシティマネジメントの促進と定着を支援する任意団体日本の企業内で「lgbtq+の人々が自分らしく働く職場づくりを進めるための情報を提供し各企業が積極的に取り組むきっかけを提供することを目的としています2016年に日本初の職場におけるlgbtq+などのセクシュアルマイノリティへの取組の評価指標「pride指標」を策定

図 1 Attention の可視化の例

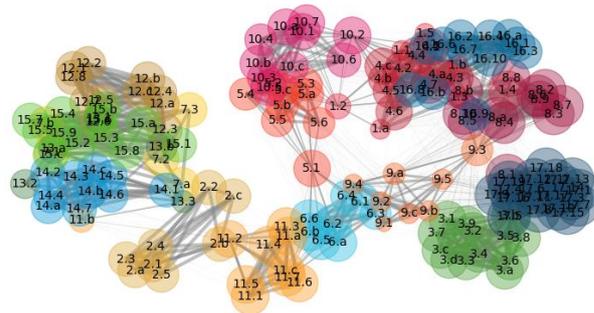


図 2 目標間の連環の可視化の例