

BA5 自然言語処理と機械学習技術を用いた自治体総合計画の SDGs 対応化支援ツールの開発

Developing Assist Tool for Upgrading Master Plan of Local Government to Adapt to SDGs by Using Natural Language Processing and Machine Learning

地球循環共生工学領域 08E16005

安藤響太 (Kyota ANDO)

Abstract: SDGs localization is essential for comprehensive achievement of SDGs, but the master plans of local governments are still not adapted to the SDGs' context. From this background I developed a classifier to identify the SDGs associated to the master plan sentences automatically and visualizer for SDGs nexus of the master plans by using natural language processing and machine learning. As a result, top1 and top3 accuracies of the classifier were 52.9%, and 73.0%, respectively. The classifier could visualize a nexus structures and coverage of SDGs in the master plans. The future tasks are improving accuracy and social implementation.

Keywords: SDGs, local government, natural language processing, machine learning

1. 背景と目的

2030年のSDGsの達成に向け、全セクターでSDGsの社会実装が重要である¹⁾。また基礎自治体がローカルSDGs²⁾に取り組み成功事例を共有することで包括的なSDGsの達成に貢献する。しかし現在多くの基礎自治体の総合計画はSDGsに未対応である。これはSDGsを熟知していないと計画・戦略とSDGsの対応関係を評価できないためであり、既存の計画・戦略をSDGsに対応づけるロジックの開発が必要である³⁾。そこで、本研究ではローカルSDGsの達成を支援するため、①分類器を構築し既存の総合計画をSDGs目標に写像すること、②各種自治体の総合計画に含まれるSDGs目標の連環関係(nexus)を可視化すること、③分類器の有用性と今後の開発ニーズを評価することを目的とした。

2. 手法

2.1 学習データの準備

図1に本研究の全体像を示す。2018年4月から2019年4月に各省庁が公開したSDGsに関する文書と、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」の民間セクターの資料から1,241文章を収集し、各文章にSDGs目標をラベリングした。次に、各文章を形態素解析したのちに、単語集合をストップワードでフィルタリングし、処理対象外の単語を除外した。単語集合の50%をWikipedia2vecでランダムに類義語置換し、データ量を3倍の3,723に水増しして学習用データセットを構築した。

2.2 分類器の構築

学習用データを文章をベクトル空間に分散表現するDoc2vecアルゴリズムでベクトル化した。次にこの分散ベクトルとSDGs目標ラベルを、教師あり機械学習であるXGBoostで学習させた。最終的に文章を入力すると、その内容がどのSDGs目標に対応するかを確率値で出力するSDGs目標分類器を構築した。分類器の未知データに対する汎化性能は、目視による分類結果と比較することで評価した。

2.3 WordCloudの作成

各目標を象徴する単語を学習データから抽出し、SDGs目標別の単語集合を作成した。学習データ内の単語とこの単語集合をDoc2vecアルゴリズムでベクトル化しコサイン類似度を計算した。各目標と類似度が高い500単語を抽出しWordCloudを作成した。

2.4 分類器の適用とSDGsネクサスの可視化

SDGs目標は1つで完結するものではなく、互いに連環関係(nexus)を持っている。分類結果の上位3つのSDGs目標がそれぞれnexusを持つと定義し、郡山市の総合計画文書内で現在想定されているSDGs目標のネクサス構造を可視化した。各目標の組み合わせが3位以内に同時に出力される共起関係を集計したのち、各目標間リンク構造をネットワーク図として可視化した。

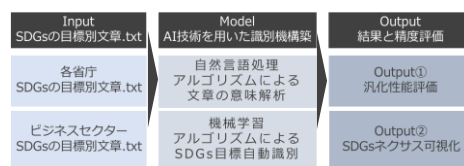


図1 研究の全体像

	GOAL_01	GOAL_02	GOAL_03	GOAL_04	GOAL_05	GOAL_06	GOAL_07	GOAL_08	GOAL_09	GOAL_10	GOAL_11	GOAL_12	GOAL_13	GOAL_14	GOAL_15	GOAL_16	GOAL_17	recall
GOAL_01	99	4	12	2	4	5	4	11	4	13	6	5	12	3	4	4	6	0.50
GOAL_02	5	99	10	1	1	2	2	2	5	2	5	9	7	6	8	2	2	0.59
GOAL_03	9	4	180	9	3	8	3	7	10	3	15	3	8	3	11	6	6	0.63
GOAL_04	4	4	14	106	10	2	3	13	8	11	5	6	4	4	0	0	4	0.54
GOAL_05	4	2	4	9	111	3	0	14	2	8	4	5	4	2	1	6	7	0.60
GOAL_06	0	5	8	1	2	103	6	4	7	5	5	10	6	9	3	3	3	0.57
GOAL_07	2	3	3	3	1	6	82	8	7	4	13	14	15	5	4	5	5	0.46
GOAL_08	15	4	7	11	13	3	10	124	11	6	11	7	6	0	6	14	16	0.47
GOAL_09	2	4	8	7	2	6	10	6	104	4	25	8	14	6	3	4	9	0.47
GOAL_10	13	2	4	7	4	5	4	9	4	81	3	3	1	1	0	10	11	0.50
GOAL_11	5	0	22	12	5	8	17	14	19	4	132	8	27	2	13	2	13	0.44
GOAL_12	4	6	12	2	4	10	11	13	10	3	6	122	16	7	15	4	7	0.48
GOAL_13	7	6	5	3	1	6	17	2	16	2	33	10	100	12	10	2	5	0.42
GOAL_14	1	3	3	5	1	10	4	4	2	4	3	9	8	85	18	2	6	0.51
GOAL_15	16	6	8	4	5	5	2	1	1	13	18	11	15	118	2	5	0.51	
GOAL_16	3	4	3	2	9	4	4	10	1	7	2	2	5	3	1	141	6	0.68
GOAL_17	2	3	8	6	3	3	3	13	6	7	13	5	9	4	5	5	181	0.66
precision	0.52	0.62	0.58	0.56	0.62	0.54	0.45	0.49	0.48	0.49	0.45	0.50	0.40	0.51	0.54	0.67	0.62	0.53

図 2 モデルの精度検証 confusion matrix

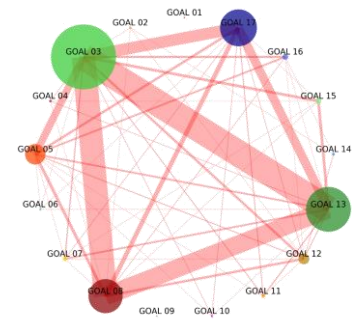


図 4 郡山市 SDGs ネクサネットワーク図



図 3 目標 03 (左) と目標 08 (中央) と目標 09 (右) の WordCloud

3. 結果と考察

3.1 分類器の精度と汎化性能評価

分類器の精度を 10 分割入れ子式交差検証で検証した結果を図 2 に示す。1 位で正解ラベルが出力された Top1_accuracy と 3 位以内に正解ラベルが出力された Top3_accuracy は、それぞれ 52.9%、73.0% の精度を得た。ランダムに 17 個のラベルを予測した場合の理論的な精度は Top1_accuracy が 5.9%、Top3_accuracy が 17.6% となるため、本分類器は高精度に予測が可能だといえる。

汎化性能は分類器と目視による分類結果を比較することで評価した。さらに各目標と単語のベクトルのコサイン類似度を基に WordCloud を作成した (図 3)。目標 03 健康、目標 08 経済成長、目標 13 気候変動の分類で高い汎化性能が得られた。これらの目標とコサイン類似度が高いベクトルを持つ”健康”、”産業”、”環境”が計画文章全体に見られたためである。逆に目標 07 エネルギー、目標 09 産業基盤、目標 11 都市計画では汎化性能が低かった。これらの SDGs 目標の分散ベクトルとコサイン類似度が高い単語ベクトルは一般的な単語があり、目標を表現できていないためである。これは一般的な単語や複数の目標と高い類似度を持つ単語をストップワードで除去することで解決できると期待される。

3.2 SDGs ネクサ可視化

郡山市の SDGs 未来都市計画内の SDGs ネクサスを可視化した (図 4)。ノードは各 SDGs 目標であり、その大きさは中心性を表す。エッジの幅は SDGs 目標間の連関の強さを表す。目標 03、目標 08、目標 13 が中心である一方で目標 01、目標 06、目標 09 は他の目標との nexus が弱く計画が不足しているといえるなど、計画内における SDGs 目標のネクサス構造と網羅率を可視化できた。

4 今後の課題

今後は学習データの増加と分析対象単語の精査によるモデルの分類精度向上を目指す。さらに新規の総合計画策定を支援するために各 SDGs 目標の特徴を表す単語を示す。また、自治体やシンクタンクへのヒヤリングにより有用性評価と開発課題の対話を行うことで実用化を目指す。

参考文献

- 1) SDGs 推進本部：SDGs アクションプラン 2019, <<https://www.kantei.go.jp/jp/singi/sdgs/pdf/actionplan2019.pdf>>, (2020.01.31 参照)。
- 2) 川久保俊：ローカル SDGs プラットフォーム, <<https://local-sdgs.jp/>>, (2020.01.21 参照)。
- 3) 松井孝典ほか：ネクサス・アプローチに基づいた SDGs 目標・ターゲット・指標間の構造解析, 土木学会論文集 G (環境), Vol. 75 (6), pp. II_39-47, 2019。