

MissionForest における活動目標データと SDGs 目標の対応関係推定手法の検討

ZHANG Xiangyu, 白松 俊, JIN Yuxi, 神谷 晃

Xiangyu ZHANG, Shun SHIRAMATSU, Yuxi JIN, Akira KAMIYA

名古屋工業大学 大学院工学研究科 情報工学専攻

Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology

Abstract: 本研究では、市民の活動や学生の研究など、世の中を良くする活動についての目標（ミッション）を構造化してオープンデータ化する Web アプリケーション MissionForest を開発中である。様々な活動目標を収集できたら、類似する問題意識を持つ活動同士で連携を検討できる機能を推薦したいと考えている。本稿ではそのために、SDGs の 17 つの目標を活用する。具体的には、MissionForest における活動目標やサブタスクと SDGs 目標の対応関係を ALBERT で推定する手法を検討し、実験によりその精度を検証した。

1. はじめに: 背景と目的

近年、社会の持続可能性を脅かす様々な課題が顕在化している。そのような課題には多様な利害関係者が存在するので、解決には組織横断的な協働が不可欠である。特に、お互いに似たような課題意識を持つ相手とは、協働できる可能性がある。そこで本稿では、似た目標を持つ主体同士のペアを発見するため、SDGs における 17 の目標 (図 1) との対応関係を自動推定する手法を検討する。

これまで本研究では、市民の活動や学生の研究など、世の中を良くする活動についての目標（ミッション）を構造化してオープンデータ化する Web アプリケーション MissionForest を開発してきた[1,2]。図 2 にそのスクリーンショットを示す。MissionForest は、まだ完全に手動でデータを入力するようになっており、外部の目標と自動的に関連付けたり、似た目標を推薦する機能は備えていない。そこで、MissionForest に入力された活動目標およびそのサブタスクを、SDGs の 17 の目標に自動的

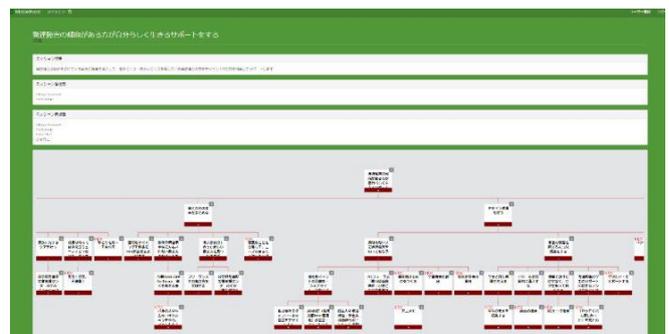


図 2 MissionForest

に対応付ける手法を開発する。これにより、協働できる可能性のある活動同士を推薦する機能を開発したい。具体的には、図 3 のように共通の SDGs 目標に対応付けられた活動同士は、互いに連携できる場合があると考え、将来的には連携の可能性を検討できるような機能へと発展させたいと考えている。

2. 対象とするデータ

MissionForest には、NPO やシビックテック団体、学生などのミッション（活動目標）が 221 個入力されている。ミ



図 1 SDGs の 17 の目標

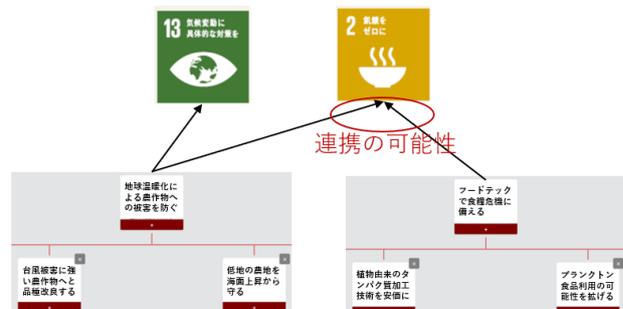


図 3 SDGs 目標との対応付けによる連携可能性の例

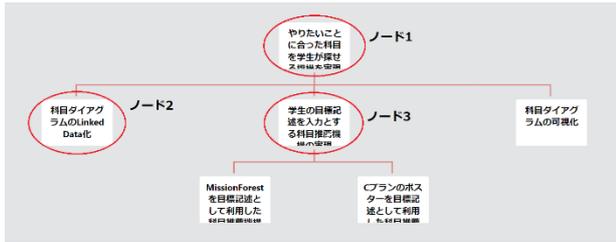


図 4 ミッションの階層構造

ミッションは、階層的にサブタスクに細分化され、構造化されている。MissionForest に入力されているサブタスクの数は 2217 個である。図 4 を例にとると、ノード 2 とノード 3 はノード 1 のサブタスク(子ノード)であり、ノード 1 はノード 2, 3 の上位タスク(親ノード)である。データを処理する場合、このようなノードのペアを「親子ノード」と呼ぶ。さらに、各ノードにはタイトルと説明文が付与されている。

本稿では、既に MissionForest に入力されている親子ノードのペアを機械学習の訓練データとして用いることで、対応する SDGs 目標を自動的に判別可能であるかどうかを検証する。

3. 提案手法と実験設定

3.1 提案手法: ALBERT を用いた親子ノード対応関係の判別

本稿で用いる機械学習のアルゴリズムは、ALBERT (A Lite BERT) [3] である。ALBERT は、BERT (Bidirectional Encoder Representations from Transformers) [4] の軽量版であり、学習すべきパラメータの数を削減しつつ高速・高精度化を図ったものである。

本稿では、これまでに MissionForest に入力された SDGs 目標を含まない親子ノードだけで ALBERT を再学習し、SDGs 目標との対応関係の判別に用いることを提案する。すなわち、「訓練データに SDGs 目標との対応関係を含まなくても、これまでの MissionForest のデータだけで SDGs 目標との対応関係判別が可能である」という仮説を検証する。

3.2 データセットの作成

データセットを作成するため、MissionForest からすべての親子ノードを抽出する。その中で、記述が曖昧であるようなノードを含む親子ノードのペアは訓練データとして相応しくないため、削除する。親子関係ではないデータペアをランダムに組み合わせて、正事例とほぼ同じ数の負事例を作成する。

MissionForest ノードと SDGs 目標の対応関係を 25 件用意し、テストデータとして用いる。表 1 にその一部を示す。

3.3 実験設定

BERT や ALBERT は事前学習フェーズと再学習フェーズがあり、事前学習には訓練データとして巨大なコーパスが必要である。本稿では、日本語 Wikipedia 記事を

表 1 SDGs 目標との対応関係のテストデータ例

MissionForest 中の目標	SDGs 目標
世代間格差問題を解決する	不平等：国内及び各国家間の不平等を是正する。
熱中症問題を解決する	保健：あらゆる年齢のすべての人々の健康的な生活を確保し、福祉を促進する。
大学・大学院の授業料免除	教育：すべての人々への包摂的かつ公正な質の高い教育を提供し、生涯学習の機会を促進する。
分散型電源でエネルギー問題を解決する	エネルギー：すべての人々の安価かつ信頼できる持続可能な近代的なエネルギーへのアクセスを確保する。

訓練データとして事前学習したモデル[5] を用いる。また、事前学習モデルでは単語分割に SentencePiece [6] が用いられていたため、そのまま踏襲し、SentencePiece を単語分割に用いる。

学習回数は 10 回と 20 回、Batch Size は、16 と 32 で 4 つのモデルを訓練する。データ数が 4005 で、トレーニングデータ・バリデーションデータ・テストデータがそれぞれ 6 割・2 割・2 割(2403 件・801 件・801 件)と設定した。

4. 実験結果と考察

4.1 実験結果

表 2 に、実験結果を示す。3 列目は MissionForest 中の (SDGs 目標を含まない) 801 件でテストした精度であり、4 列目は SDGs 目標との対応関係 25 件でテストした精度である。SDGs 目標との対応関係の精度は、学習回数 20 回、Batch Size 16 のとき、最大値 0.68 となった。

表 2 学習パラメータと精度

学習回数	Batch Size	MissionForest 中のデータ (801 件) でテストした精度	SDGs 目標との対応関係 (25 件) でテストした精度
10	16	0.753	0.64
20	16	0.703	0.68
10	32	0.765	0.64
20	32	0.759	0.56

4.2 考察

今回組み合わせを変えたパラメータだけでは、精度に大きな変化がなかった。今後、精度を上げるため以下の 4 つの方面から精度改善を試みる予定である。

1. 記述が極めて短い・長いノードを訓練データから削除する。

2. 日本語 ALBERT モデルを使うので、URL や英語文があるデータは上手く処理できない可能性がある、そのようなデータを訓練データから削除する。
3. 前文の通りノードごとに説明文があるので、説明文も訓練データに加えてみる。
4. クラウドソーシングによりSDGs 目標との対応関係を含む訓練データを整備し、訓練データとして用いる。

SentencePiece for Japanese text, <https://github.com/alinear-corp/albert-japanese> (2020年11月12日アクセス)

- [6] Kudo, T., Richardson, J. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. arXiv preprint arXiv:1808.06226, 2018.

5. まとめと今後の展望

本稿では、MissionForest の課題と SDGs 目標の対応関係を、ALBERT を用いて自動推定する手法を試した。ただし、やはり SDGs を含まない訓練データだけでは、0.68 という不十分な精度しか達成できないという結果になった。今後は、考察で述べた 4 つの方策を試し、精度の改善を図る予定である。

特に、SDGs 目標との対応関係を含む訓練データについては、既に着手済みである。MissionForest に入力された 221 件の活動目標の中から、176 件を選び、クラウドソーシングサイトで SDGs 目標との対応関係のデータを整備した。1 つのサンプルを 2 人のクラウドワーカーに依頼したため、352 サンプルの訓練データを得た。今後、これらを整理し、訓練データとして用いる予定である。

謝辞

本研究の一部は、JST CREST (JPMJCR15E1) および NEDO (JPNP20006) の助成を受けて実施されました。

参考文献

- [1] 後藤誉昌, 渡辺賢, 白松俊. “MissionForest: 組織内外における協働支援のためのタスク構造化システムの試作”, 情報処理学会 第 79 回全国大会講演論文集, 6ZA-02, 2017.
- [2] Masaru Watanabe, Shun Shiramatsu, Yasuaki Goto. "Tag-based Approaches to Sharing Background Information regarding Social Problems towards Facilitating Public Collaboration," in eGose '17 Proceedings of the International Conference on Electronic Governance and Open Society: Challenges in Eurasia, pp. 113-118, 2017.
- [3] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, Radu Soricut. “ALBERT: A Lite BERT for Self-supervised Learning of Language Representations”, arXiv preprint arXiv:1909.11942, 2019
- [4] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [5] ALINEAR. alinear-corp/albert-japanese: BERT with