4S-08

A Method for Estimating Correspondence between Activity Goals in MissionForest and SDGs Goals using ALBERT

Xiangyu ZHANG† Shun SHIRAMATSU† Yuxi JIN† Akira KAMIYA†

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

1. Introduction

In recent years, various issues that threaten social sustainability have become apparent. Since there are various stakeholders in such issues, cross-organizational collaboration is indispensable for solving them. Particularly, people who have issues in similar field have possibility of cooperation. Therefore, in order to discover pairs of subjects who have similar goals, we study a method for automatically estimating the correspondence with 17 SDGs goals (Fig. 1).

SUSTAINABLE DEVELOPMENT GOALS
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS

Fig. 1. 17 Goals of SDGs

Up to now, a web application MissionForest has been developed in this research field. It structures the goals (missions) of activities that make the world better, such as citizens' activities and students' research, and converts them into open data [1]. Fig. 2 shows an image of MissionForest.

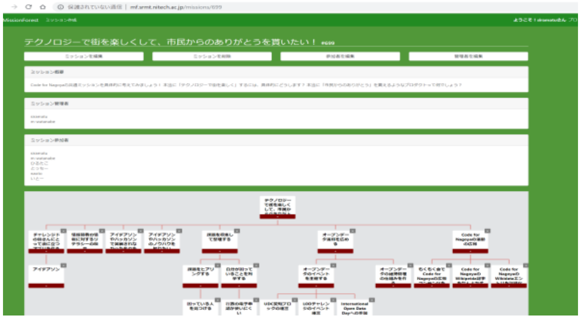

テクノロジーで街を楽しくして、市民からのありがとうを貰いたい！

Fig. 2. MissionForest

So far, users still need to completely manually input data into MissionForest. It has no function that automatically associate with external goals or recommend similar goals. Therefore, we want to study a method to automatically associate the activity goals and their subtasks input in the MissionForest with the 17 SDGs goals. In this way, we would like to develop a function that recommends activities to each other that may have possibility of cooperation. Specifically, as shown in Fig. 3, we think that activities associated with common SDGs goals may be able to cooperate with each other and develop the function so that the possibility of cooperation can be studied in the future.

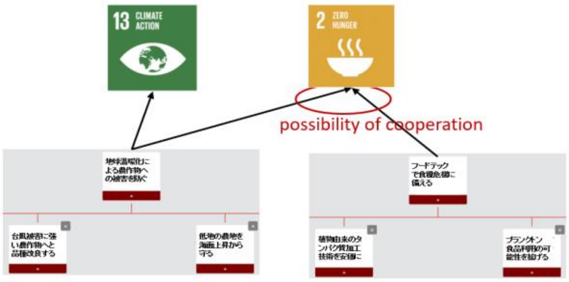

13 CLIMATE ACTION
2 ZERO HUNGER
possibility of cooperation
地球温暖化による農作物への被害を防ぐ
台風被害に強い農作物へと品種改良する
低地の農地を海面上昇から守る
フードテックで食糧危機に備える
植物由来のタンパク質加工技術を安価に
プランクトン食品利用の可能性を拡げる

Fig. 3. Two missions that have possibility of cooperation

2. Target Data

There are 221 missions (activity goals) in MissionForest. Missions are hierarchically subdivided and structured into subtasks. There are 2217 subtasks in MissionForest. There is an example in Fig. 4. Nodes 2 and 3 are subtasks (children nodes), meanwhile node 1 is a higher-level task (parent node) of nodes 2 and 3. When processing data, such a pair of nodes is called parent-child nodes. In addition, every node has a title and description.

In this time, pairs of parent-child nodes that have been input into MissionForest are used as first part of training data.

We also collect data by cloud sourcing. An order including activity goals in MissionForest is sent to cloud sourcing site and some data is collected as second part of training data. (Fig. 5)

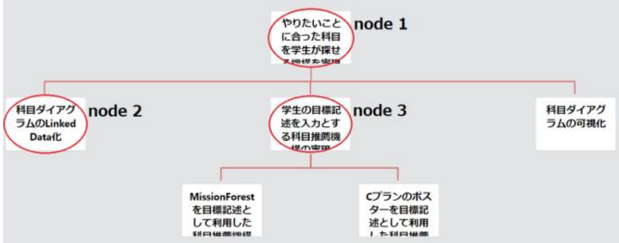
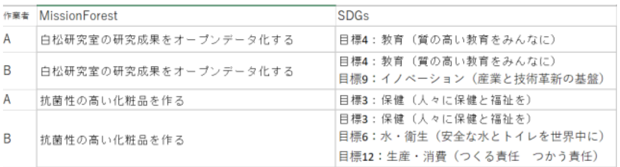

やりたいことに合った科目を学生が探せる機構を実現
node 1
科目ダイアグラムのLinked Data化
node 2
学生の目標記述を入力とする科目推薦機構の実現
node 3
科目ダイアグラムの可視化
MissionForestを目標記述として利用した科目推薦機構
Cプランのポスターを目標記述として利用した科目推薦

Fig. 4. Hierarchical structure of missions



作業者	MissionForest	SDGs
A	白松研究室の研究成果をオープンデータ化する	目標4：教育（質の高い教育をみんなに）
B	白松研究室の研究成果をオープンデータ化する	目標4：教育（質の高い教育をみんなに） 目標9：イノベーション（産業と技術革新の基盤）
A	抗菌性の高い化粧品を作る	目標3：保健（人々に保健と福祉を）
B	抗菌性の高い化粧品を作る	目標3：保健（人々に保健と福祉を） 目標6：水・衛生（安全な水とトイレを世界中に） 目標12：生産・消費（つくる責任　つかう責任）

Fig. 5. A part of data by cloud sourcing

3. Proposed Method and Experimental Settings

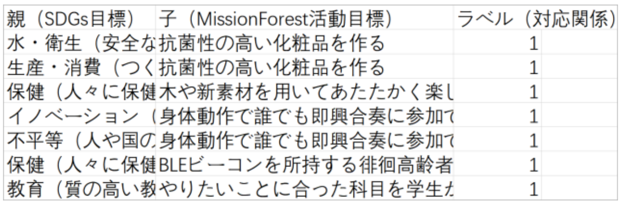
3.1. Proposed Method: Estimate Correspondence between parent and child using ALBERT

In this time, the algorithm of machine learning is ALBERT (A Lite BERT) [2]. ALBERT is a lite version of BERT (Bidirectional Encoder Representations from Transformers) [3]. It reduced parameters to make learning process faster and keep high accuracy meanwhile.

We relearn ALBERT with parent-child nodes that have no SDGs, and a small amount of data from cloud sourcing. The purpose is that using data that is not completely related to SDGs to study a method and use it to estimate the correspondence with SDGs goals. It means that, the hypothesis is: The ALBERT model trained by MissionForest data and a small amount of data that have correspondence between MissionForest nodes and SDGs, can estimate the correspondence between MissionForest nodes and SDGs goals.

3.2. Dataset

The dataset is composed of two parts (Table 1). Firstly, all parent-child nodes are extracted from MissionForest to create a dataset. Nodes with ambiguous description, URLs, English sentences, and extremely short or long content, cannot be processed well, so those data will be deleted. And nodes' title and description will be used meanwhile. Secondly, an order including 176 activity goals (samples) in MissionForest is sent to the cloud sourcing site. One sample asks two cloud sourcing workers to choose one or more SDGs to it. So, 352 samples are made up. These sample are be used as positive examples (Fig.6). Negative examples are manually complemented. This is the second part of dataset.



親（SDGs目標）	子（MissionForest活動目標）	ラベル（対応関係）
水・衛生（安全な	抗菌性の高い化粧品を作る	1
生産・消費（つく	抗菌性の高い化粧品を作る	1
保健（人々に保健	木や新素材を用いてあたたかく楽し	1
イノベーション（	身体動作で誰でも即興合奏に参加て	1
不平等（人や国の	身体動作で誰でも即興合奏に参加て	1
保健（人々に保健	BLEビーコンを所持する徘徊高齢者	1
教育（質の高い教	やりたいことに合った科目を学生た	1

Fig. 6. an image of positive data from cloud sourcing

One more test dataset including 151 positive examples and 149 negative examples is also prepared. These 300 data are exported from data of cloud sourcing.

Table 1. Dataset

Dataset composition		Number of data
Positive example	Data of MissionForest	1802
	Data of cloud sourcing	151
Negative example	Data of MissionForest	1800
	Data of cloud sourcing	150

3.3. Experimental Settings

We use a pre-learned ALBERT model using Japanese Wikipedia articles as training data. In addition, since SentencePiece was used for word division in the pre-learned model, we will not change it and still use SentencePiece for word division.

The epoch is 10 and 20. The batch size is 16 and 32, and learning rate is 2e-5, then 4 models are trained. The dataset is randomly divided to 60% training data, 20% validation data and 20% test data (2341, 780 and 782).

4. Result and Discussion

The result is shown in Table 2.

There no significant change in accuracy only with parameters used this time. Compared with 20 epochs of learning, 10 epochs also can achieve similar accuracy.

Table 2. Result

Epoch	Batch size	accuracy	Test dataset (300 data including correspondence between MissionForest and SDGs)
10	16	0.7910	0.8500
20	16	0.7769	0.8667
10	32	0.7718	0.8500
20	32	0.7731	0.8667

There are not only 150, but also over 2000 negative examples data of cloud sourcing. Because there is not enough positive example including correspondence between MissionForest nodes and SDGs, so we don't input more negative examples into dataset. It will be cause data balance problem.

5. Summary and Future Work

The experiment result showed that the accuracy of our method to estimate the correspondence between MissionForest goals and SDGs goals was over 0.85.

Due to the lack of positive examples, it is difficult to add a larger proportion of the data including correspondence between MissionForest and SDGs when making dataset. We are looking forward to finding a way to mix a larger proportion of data including correspondence to dataset.

Acknowledgements. This work was supported by NEDO (JPNP20006), JST CREST (JPMJCR15E1, JPMJCR20D1), and JSPS KAKENHI (17K00461).

Reference

[1] Masaru Watanabe, Shun Shiramatsu, Yasuaki Goto. "Tag-based Approaches to Sharing Background Information regarding Social Problems towards Facilitating Public Collaboration," Proc. of eGose '17, pp. 113-118, 2017.

[2] Zhenzhong Lan, et al. "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations", arXiv preprint arXiv:1909.11942, 2019

[3] Devlin, J., et al. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.