On the problem of automatically aligning indicators to SDGs

Mario Soriano[1], Rafael Berlanga[1], and Indira Lanza-Cruz[1]

Dep. de Llenguatges i Sistemes Informàtics, Universitat Jaume I, Spain
berlanga@uji.es, magregor@uji.es, lanza@uji.es

Abstract. In this paper we present a first approach to the application of transformer-based language models to the automatic alignment to sustainable development goals (SDGs). This task is quite relevant for the development of new tools that aim at measuring the engagement degree of the organization's indicators to the SGDs. Our first experiments show that this task is hard, and that even powerful large language models do not achieve a high accuracy as in other NLP tasks.

Keywords: Transformers · Indicators · Knowledge Graphs.

1 Introduction

In 2015, the UN defined the Agenda 2030 to establish a series of goals for the sustainable development of the World [6]. The sustainable development goals (SDGs) consist of 17 main objectives covering different perspectives related to sustainability. These goals are described in detail by defining their context and the main indicators that must be tracked to measure their achievement. As a consequence, all kind of organizations must align their indicators and goals towards the SDGs. Indeed, most public funding is currently conditioned to define appropriate indicators related to SGDs.

Aligning specific indicators to the global SDGs is a challenging task. Firstly, this is a very open scenario which involves many perspectives and knowledge areas. Secondly, we have to find a causal chain between indicators and SDGs, which requires a deep understanding of both the organization and the SDGs.

In this paper, we present a first approach to this problem by applying deep learning language models. The main idea is to measure the quality of automatic alignments produced by these models. We evaluate language models trained for English and Spanish over the indicators proposed in the Agenda 2030. We show in this paper that the connection between indicators and goals are not so evident and, in some cases, it requires a reasoning process to find out them.

2 Related Work

2.1 Goals, actions and indicators

Strategic maps usually represent the knowledge in terms of goals, actions and indicators. The goals express the future state we aim at in terms of some measurable quantities [5]. For example, the SDG-2 "Zero hunger" implies that some global measure for "hunger" must be reduced to 0. This global measure relies on many indicators, which express the different perspectives involved in such an abstract goal. For example, the indicator "Prevalence of anaemia in women aged 15 to 49 years, by pregnancy status" is used in the SDG-2 to account for the nutrition status of woman in reproductive age. Reducing this indicator directly implies approaching to the goal according to the following causal chain: if anemia has been reduced is because the nutrition status was improved, so hunger was indeed reduced.

2.2 Dynamic SLOD-BI

The activities and indicators of any organization must also be aligned to the SGDs, so that they can impact positively towards the sustainability goals. For this purpose, the context of this paper is the semantic infrastructure for business intelligence named Dynamic SLOD-BI [4]. This infrastructure deals with streams of open and corporate data which feed a live knowledge graph (KG) for analytical purposes. Goals and indicators are also represented in this graph, along with their connections to the streamed data. A first prototype of this infrastructure is being developed in the scenario of sustainable tourism, where the UNWTO defined its own indicators before the SDGs were established.

2.3 SDGs-based classification

There are some work about using text classifiers to assign SDGs to arbitrary documents. For example, in [3], authors apply NLP techniques to articles from peer-reviewed journals in order to classify them according to the 17 SDGs. They compare the performance of different multi-label text classification models with datasets of different characteristics. On the other hand, authors in [1] tried to map the Environmental Higher Education Ranking Systems Indicators (ESHERS) to the SGD indicators using NLP and document similarity techniques. Finally, the work in [2] fine-tune a BERT multi-class model to classify documents into the SDGs. All these approaches uses a labelled set of documents (e.g., scientific articles) to predict labels for arbitrary documents. Instead, in our work, we focus on classifying indicators and sub-goals, which are very short descriptions with clear semantics.

It is worth mentioning that there are some online tools like Escaner2030[1] provided by the Political Watch and the Spanish Ministry of Foreign Affairs, which classifies any paragraph into their most likely SDGs.

3 Methods and results

For the experiments, we directly use the inventory of sub-goals and indicators proposed in 2015 by the UN to define the 17 SDGs. Thus, the dataset contains

[1] https://escaner2030.es/

400 sentences describing different sub-goals and indicators labelled with their SDG code.

The experiments in this paper make use of the pre-trained BERT-like models from HuggingFace. We conducted two experiments. The first one consists of directly encoding the sentences of the whole dataset to find out if the database is consistently distributed in a metric semantic space. That is, sentences should be somehow clustered around their goals. The second experiment consists of fine-tuning pre-trained models as multi-class models for SDGs. The first experiment is intended to link sentences to goals in an unsupervised way, whereas the latter needs a train/test partitions of the dataset (we used a 90/10 partition).

Table 1 shows the precision when linking sentences by using a similarity text (ST) encoder. Precision is calculated with the K-nearest sentences to each sentence, considering success if they belong to the same goal. Notice that the best score always occur at K=1, which indicates that the semantic space is quite heterogeneous. Notice that ST performs similarly in the two languages.

Table 2 shows the results of the trained multi-class models. For these experiments we use 9:1 train/test partitions and 6 epochs. We show best models and best scores. Surprisingly, these results are not much better than those in Table 1. Moreover, quality results of the trained models are affected by the language.

Text Similarity Models			
hiiamsid/sentence_similarity_spanish_es		distiluse-base-multilingual-cased-v1	
K	Score	K	Score
1	**0.66268**	1	**0.66427**
2	0.62081	2	0.60312
3	0.58214	3	0.57234

Table 1. Precision for text similarity in both languages.

We performed further experiments over the zero-shot pipelines of Huggingface resulting in very poor results. Also, the tool Escaner2030 showed very poor results when classifying indicators from this dataset (below 0.4). Finally, we report the results of the large language model chatGPT. We prepared 5-folds for each language, containing 30 random samples each. We used the new API with the following prompt:

Prompt: "Please, assign sustainability goals (SDGs) to a series of texts. When I write a text you must assign it to one SDG code followed by the words from the text that are relevant to your decision (relevant words). You have to answer simply with this format: <SDG Code> | <relevant words>"

With this prompt we also extract the explanations for the chosen SDGs. Usually, chatGPT returned the right keywords involved in the connection. In Table 2, we report the accuracy with confidence intervals for the 5-folders. Each fold takes approximately one minute, making this method no scalable. Like supervised methods, chatGPT performs much better in English than Spanish.

Language Model	**Language**	**Accuracy**
Multi-class Fine-tuned Models		
XLM-Roberta	Spanish	0.6415
Roberta	English	0.6981
Large Language Model		
chatGPT	Spanish	0.76 ± 0.07
chatGPT	English	$\mathbf{0.84 \pm 0.04}$

Table 2. Accuracy of supervised and chatGPT.

4 Conclusions

This paper presents a first approach to the SGD alignment problem for indicators. Results show that scores are still far from the ideal ones. Even the powerful chatGPT does not achieve high enough scores, being them also dependent on the target language. Future work will focus on enhancing text similarity methods and few-shot classifiers by means of explainable methods that rely on knowledge graphs. These techniques will allow us to align texts and indicators at a larger scale than chatGPT can currently do.

Acknowledgments

This project has been funded by the Ministry of Science and Innovation with project contract PDC2021-121097 and the Generalitat Valencia project INVESTIGO with contract number INVEST/2022/308.

References

1. A. Buzaboon, H. Alboflasa, W. Alnaser, S. Shatnawi and K. Albinali, "Automated Mapping of Environmental Higher Education Ranking Systems Indicators to SDGs Indicators using Natural Language Processing and Document Similarity".
2. J. Guisiano, R. Chiky. Automatic classification of multilabel texts related to Sustainable Development Goals (SDGs). TECHENV EGC2021, Jan 2021, France.
3. R. C. Morales-Hernández, J. G. Jagüey and D. Becerra-Alonso, "A Comparison of Multi-Label Text Classification Models in Research Articles Labeled With Sustainable Development Goals"
4. I. Lanza-Cruz, R. Berlanga, M. J. Aramburu: Modeling Analytical Streams for Social Business Intelligence. Informatics 5(3): 33 (2018)
5. Parmenter, D. (2015). Key performance indicators: developing, implementing, and using winning KPIs. John Wiley & Sons.
6. UN General Assembly, Transforming our world : the 2030 Agenda for Sustainable Development, 21 October 2015, A/RES/70/1 available at: https://www.refworld.org/docid/57b6e3e44.html