Using Natural Language Processing for Automating the Identification of Climate Action Interlinkages within the Sustainable Development Goals

Xin Zhou[1], **Kshitij Jain**[2], **Mustafa Moinuddin**[1], **Patrick McSharry**[3,4,5]
[1]Institute for Global Environmental Strategies, 2108-11 Kamiyamaguchi, Hayama, Kanagawa, 240-0115 Japan; [2]Google Inc.;
[3]Carnegie Mellon University Africa, Kigali, Rwanda; [4]African Centre of Excellence in Data Science, University of Rwanda, Kigali, Rwanda; [5]Oxford Man Institute of Quantitative Finance, Oxford University, Oxford, UK.
zhou@iges.or.jp, kshitijj@google.com, moinuddin@iges.or.jp, patrick@mcsharry.net

Abstract

Climate action, Goal 13 of the UN Sustainable Development Goals (SDG), cuts across almost all SDGs. Achieving climate goals can reinforce the achievements in many other goals, but at the same time climate mitigation and adaptation measures may generate trade-offs, such as levelling the cost of energy and transitioning away from fossil fuels. Leveraging the synergies and minimizing the trade-offs among the climate goals and other SDGs is an imperative task for ensuring policy coherence. Understanding the interlinkages between climate action and other SDGs can help inform about the synergies and trade-offs. This paper presents a novel methodology by using natural language processing (NLP) to automate the process of systematically identifying the key interlinkages between climate action and SDGs from a large amount of climate literature. A qualitative SDG interlinkages model for climate action was automatically generated and visualized in a network graph. This work contributes to the conference thematic topic on using AI for policy alignment for climate change goals, SDGs and associated environmental, social and governance (ESG) frameworks.

Introduction

The UN 2030 Agenda charts out 17 interlinked Sustainable Development Goals (SDGs) that require a holistic approach to their implementation. However, such a holistic approach is new and challenging due to the broad coverage of social, economic and environmental dimensions and complicated relations among 169 targets. This is particularly true for SDG 13 (climate action) which links with all major sectors including energy, industry, transport and agriculture, and poses severe challenges to human and animal life, the planet's water, marine and terrestrial ecosystems and biodiversity. Achieving climate goals can reinforce the achievements in many other goals, but at the same time climate actions may generate trade-offs, such as levelling the cost of energy and transitioning away from fossil fuels. Leveraging the synergies and minimizing the trade-offs among climate goals and other SDGs is an imperative task for ensuring policy coherence. Understanding the interlinkages of climate action within the SDGs can help inform about the synergies and trade-offs.

There is a gap in the scientific knowledge about how the SDGs are interlinked, making it difficult to know whether targets are mutually reinforcing or competing with one another. This knowledge gap inhibits the adoption of an integrated approach. There are emerging attempts to understand SDG interlinkages (e.g., Allen, Metternicht, and Wiedmann 2018; ICSU 2017; Le Blanc 2015; Miola, Borchardt, and Neher 2019; Nilsson, Griggs, and Visback 2016; Zhou and Moinuddin 2017; Zhou, Moinuddin, and Li 2021). These attempts took different approaches, such as expert opinion (e.g., ICSU 2017), literature review (e.g., Zhou et al. 2022) and statistical analysis (e.g., Dörgő, Sebestyén, and Abonyi 2018). Among these attempts, the Institute for Global Environmental Strategies (IGES) developed the SDG Interlinkages Analysis & Visualisation Tool (https://sdginterlinkages.iges.jp/) to identify, quantify and visualise the SDG interlinkages (Zhou, Moinuddin, and Li 2021). The IGES method uses a literature review to identify the causal relations based on scientific evidence. The tool covering 27 countries from Asia and Africa has been used to assist policy making on SDG priority setting, institutional arrangement and sectoral policy development in countries such as Indonesia, Bangladesh and Viet Nam.

However, the lack of well-defined evidence on the causations between SDG targets poses a major challenge for the application of the IGES tool. The tool's current literature review is conducted manually based on a limited amount of literature, making it difficult to systematically capture the full range of SDG interactions. For example, a keyword-based search on "climate change" in titles from Scopus

---

found more than 85,000 relevant references. Such a large amoung of literature data makes a compelling case for a systematic approach to literature review. A manual approach is also laborious, potentially subjective based on the judgement of different subject experts and difficult to scale to the full set of SDGs. A recent study extended the IGES model at the river basin scale by exploring the causal relations through a systematic review using software-based text analysis and human intervention (Zhou et al. 2022). However, the human intervention process is still heavily dependent on expert judgment and its improvement requires a systematic and automated approach.

With the motivation of extending the IGES tool to systematically include all relevant interlinkages from a large amount of literature, the current paper proposes an artificial intelligence (AI) based natural language processing (NLP) methodology to identify the causal links among the SDGs. The methodology was applied to Goal 13 on climate action due to its cross-cutting nature in the SDG framework. The methodology, replicable to other SDGs, and its application, including limitations, is explained here in detail. The outcome of this research has the potential to greatly contribute to the understanding of SDG interlinkages, which is of utmost importance for the sustainable development agenda in particular and to ensure the wellbeing of global citizens and the planet in general. This is consistent with the conference thematic topic on using AI for policy alignment of climate change goals, the SDGs, and ESG frameworks.

Methodology

The methodology of using NLP for automating the process of systematically extracting the key SDG interlinkages from climate change literature follows six steps (Figure 1).

Step 1 is selection of literature. We use five assessment reports of the Intergovernmental Panel on Climate Change (IPCC 1990, 1995, 2001, 2007, 2014) which provide comprehensive knowledge on climate change, its causes, impacts and response options based on the assessment of scientific, technical and socio-economic literature.

Step 2 is processing the corpus data using NLP. NLP is used to extract top frequent terms and map their relations. We first clean the data by removing new lines, blank lines, and digit terms. Further, we tokenize the text and remove punctuation tokens and stop words. We use Porter Stemmer to stem the tokens to ensure that different forms of a given word map together.

Step 3 is identifying top frequent terms. We run a hyper-parameter search on the set of n-grams from the cleaned tokens and find that 2, and 3 n-grams are best suited for our purpose. We calculate the frequency of all bigrams and trigrams and identify top 500 terms (Figure 2).

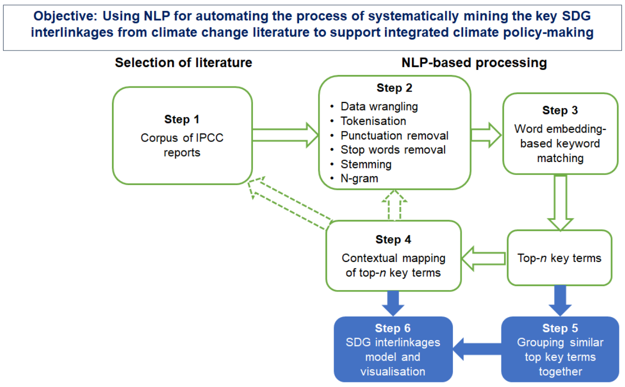


Figure 1: An NLP-based automated process for systematically extracting key SDG interlinkages

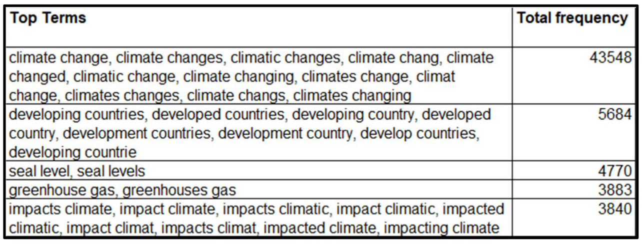
Step 4 is mapping the linkages among the top terms. We set each top term as a centre term and search if other top terms occur in 10 tokens either before or after the centre term. The identified terms are defined as the contextual terms of the centre term. Each contextual term and its centre term form a pair and the associated 10 tokens is defined as a contextual sentence describing their linkage. For each centre term, we identify top-*k* contextual terms based on the frequency. Additionally, to better understand each linkage, we identify the most common pre- and post-verb between the centre term and the contextual term. A pre-verb is defined where the contextual term occurs followed by the verb and then the centre term, while a post-verb is where the centre term is followed by the verb and then the contextual term. To do so, we rely on NLP and use part-of-speech tagging to identify different grammatical components of the sentence.

Step 5 is grouping similar top terms together. The 500 top terms have many similarities, such as "adaptation climate change, adapt climate change, …" and "climate change adapting, climate change adapt, …". To remove the noise caused by different wording, similar terms were grouped together. The frequency of each group is the sum of the frequency of individual terms in this group. Currently, we are exploring the use of word embeddings to automate the grouping exercise. As part of an experimental process, we used GloVe (Pennington, Socher, and Christopher 2014) embeddings followed by k-means clustering to group similar top terms. For this study, we used Human-in-the-Loop (using domain experts to refine our NLP system) process to group similar top terms.

Step 6 is building a qualitative SDG interlinkages model by mapping grouped terms with SDG targets and visualising the results in a network graph. The model is generic at the global level. Further validation and customization to local contexts can be conducted through stakeholders' engagement, but this is not included in the current study.

Results

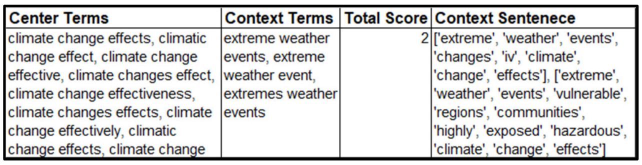
An excerpt of the identified top 500 terms and their frequency is shown in Figure 2. Since we have stemmed the tokens, different forms of a term are grouped together.



Top Terms	Total frequency
climate change, climate changes, climatic changes, climate chang, climate changed, climatic change, climate changing, climates change, climat change, climates changes, climate changs, climates changing	43548
developing countries, developed countries, developing country, developed country, development countries, development country, develop countries, developing countrie	5684
seal level, seal levels	4770
greenhouse gas, greenhouses gas	3883
impacts climate, impact climate, impacts climatic, impact climatic, impacted climatic, impact climat, impacts climat, impacted climate, impacting climate	3840

Figure 2: An excerpt of top 500 terms identified through an NLP-based semantic analysis

From Step 4, 4,600 paired terms among 500 top terms were extracted (see Figure 3 as an example). The total score is calculated by term frequency-inverse document frequency (TF-IDF) for each pair (See Step 4). Each contextual sentence or phrase (separated by a comma) describes the linkage of the pair. For example, "'extreme', 'weather', 'events', 'vulnerable', 'regions', 'communities', 'highly', 'exposed', 'hazardous', 'climate', 'change', 'effects'" implies that extreme weather events affected the vulnerable regions and communities who are exposed to hazardous climate effects.



Center Terms	Context Terms	Total Score	Context Sentenece
climate change effects, climatic change effect, climate change effective, climate changes effect, climate change effectiveness, climate changes effects, climate change effectively, climatic change effects, climate change	extreme weather events, extreme weather event, extremes weather events	2	['extreme', 'weather', 'events', 'changes', 'iv', 'climate', 'change', 'effects'], ['extreme', 'weather', 'events', 'vulnerable', 'regions', 'communities', 'highly', 'exposed', 'hazardous', 'climate', 'change', 'effects']

Figure 3: An excerpt of the identified contextual terms of a centre term

After grouping similar terms together (Step 5), the number of top terms was reduced from 500 to 104 and the number of pairs was reduced from 4,600 to 962. After mapping grouped terms with relevant SDG targets, a qualitative SDG interlinkages model, including 104 nodes and 962 links, was established for Goal 13 (climate action). The model was visualised in a network graph arranged by the SDGs (see Figure 4). For example, there is a link between "vulnerability" and "climate change" (indicated in a blue edge). To understand how they link, the contextual sentence, "'vulnerability', 'climate', 'change', 'determined', 'exposure', 'impacts', 'climate', 'change'", can inform us that vulnerability to climate change is affected by the exposure to the impacts of climate change, among other factors.

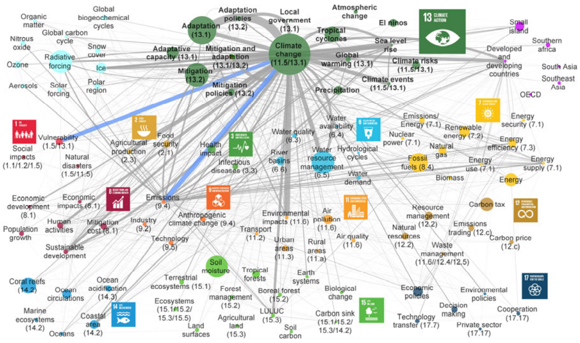


Figure 4: Visualization of the SDG interlinkages model for SDG 13 (climate action)

Note: The visualization was generated using Cytoscape. Each node indicates a grouped term which size indicates the frequency. The code in parentheses indicates relevant SDG targets. Nodes under the same SDG are placed close to each other. The edge between a pair represents a linkage, which width is measured by TF-IDF.

The frequency of the top terms (indicated by the size of the node), shows the scientific importance of relevant factors and actors. Figure 4 shows that climate change-related factors, such as global warming, extreme climate events, sea level rise and precipitation, are the most frequent terms. Other factors including GHG emissions, climate mitigation and adaptation, agricultural production, land use and land cover change (LULUC), environmental impacts and vulnerability, among others, are also on the top list.

The results can be shown in sub-models focusing on one or a few factors. Figure 5 is an example of climate adaptation and its linkages with the SDGs. Adaptation to climate change links with both adaptation policy and mitigation policy. Adaptation influenced by the adaptive capacity, climate risks and vulnerabilities, particularly of the poor. Key areas of adaptation include agricultural production, water resource management, natural resource management, rural development and coastal areas.

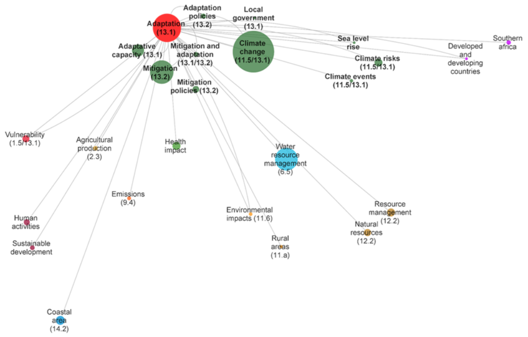


Figure 5: Linkages between climate adaptation and SDGs

Discussions and Future Research Agenda

This paper presents a novel methodology of using natural language processing to automating the identification of SDG 13 (climate action) interlinkages within the SDGs from a large corpus based on the IPCC assessment reports. A qualitative SDG interlinkages model for SDG 13 was built systematically and visualized in a network graph. The proposed methodology can be extended by systematically analysing the SDG interlinkages of all other SDGs. Previously, this process was highly manual and required an overwhelming effort from the experts. In this work, we have successfully reduced the time to build an interlinkage model by about 30 times. This could only have been achieved by using an NLP-based automating process. Additionally, this has enabled the use of a large corpus to find such linkages.

The SDG interlinkages model and its visualization can be used as a practical tool to communicate among the stakeholders in national and local climate policy development. This tool can be used but not limited to the following ways:

- To support systematic policy scoping in the early stage of policy development

Together with stakeholders' engagement, the list of top frequent terms can be used for systematic scoping and screening of key factors, sectors and actors to be considered in national and local climate policies based on the local context. For example, in vulnerable and low-income countries, adaptation policy development should take full account of the poor, particularly poor women and children.

- To support integrated policy making and cross-sector collaborations

In climate policy development, the key linkages including the synergies, trade-offs and feedback among major factors and sectors should be taken account to ensure policy integrity and inclusiveness. Furthermore, relevant departments, sectors and local governments related to the key synergies and trade-offs in the systems of climate change and climate action should be included in the policy-making process to ensure integrated governance and budget allocation by collectively addressing the trade-offs and leverage the synergies. For example, Figure 5 implies that effective governance for climate adaptation planning and implementation should include relevant competent departments in charge of agriculture, natural resource management, water resource management, and public health, etc. The results from this research provide scientific evidence on the synergies and trade-offs to support integrated institutional arrangement and budget allocation.

- To support climate policy making tailored to the local context

To support the development of a long-term climate mitigation strategy (LTS) in West Java Province, Indonesia, an engagement of more than 50 stakeholders from national and local governments, state-owned enterprises, the private sector, NGOs, academia, and the media, was conducted from June to July in 2022 to validate and contextualise the SDG interlinkages model for Goal 13 (Moinuddin and Zhou 2022). The activity was conducted through an online survey (55 participants) and a focus group discussion (40 participants) to receive feedback on four specific topics in West Java: renewable energy development, gender dimension of low carbon development, economic and employment impacts, and climate change policy measures and their impacts. This exercise not only raises the awareness of taking an integrated approach but also helps validate the model with the feedback on the linkages of climate change and climate action with SDGs from West Java's perspective. The tailored SDG interlinkages model for West Java can be used to inform the policy makers about key factors and key linkages in the development of the LTS.

A couple of limitations of the existing methodology were identified which can be set as future research agenda. These include: i) building a methodology for selecting relevant and high quality scientific literature from large bibliographic databases (e.g. Google Scholar and Scopus); ii) improving the pre-processing of the corpus data (Step 2) including identifying effective stop words, using part-of-speech tagging to help select effective terms and remove irrelevant terms, and avoiding truncation in the full phrase, etc.; and iii) automating the process of effectively grouping similar terms by leveraging word embeddings and large language models such as BERT (Devlin et al. 2018) (Step 5) and mapping with the SDGs (Step 6). All the above-mentioned challenges can be effectively solved by using state-of-the-art NLP methods, and we are confident that this work will open several possible avenues for further research in the area of using AI to respond to climate challenges.

Acknowledgments

This research was conducted under the project of 'Using AI to Impute Missing Data in the SDG Indicators and Test Causality', funded by Google's AI for Social Good program. This research was also supported by FY2021 Strategic Research Fund of the Institute for Global Environmental Strategies (IGES).

References

Allen, C.; Metternicht, G.; and Wiedmann, T. 2018. Prioritising SDG Targets: Assessing Baselines, Gaps and Interlinkages. *Sustainability Science* 14(2): 421–438.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and

Short Papers): 4171–4186. Minnesota: Association for Computational Linguistics. 10.18653/v1/N19-1423.

Dörgő, G.; Sebestyén, V.; and Abonyi, J. 2018. Evaluating the Interconnectedness of the Sustainable Development Goals Based on the Causality Analysis of Sustainability Indicators. *Sustainability* 10(10): 3766. https://doi.org/10.3390/su10103766.

ICSU. 2017. A Guide to SDG Interactions: From Science to Implementation. Paris: International Council for Science.

IPCC. 1990. First Assessment Report of the Intergovernmental Panel on Climate Change. Switzerland: Intergovernmental Panel on Climate Change (IPCC).

IPCC. 1995. Second Assessment Report of the Intergovernmental Panel on Climate Change. Switzerland: Intergovernmental Panel on Climate Change (IPCC).

IPCC. 2001. Third Assessment Report of the Intergovernmental Panel on Climate Change. Switzerland: Intergovernmental Panel on Climate Change (IPCC).

IPCC. 2007. Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Switzerland: Intergovernmental Panel on Climate Change (IPCC).

IPCC. 2014. Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Switzerland: Intergovernmental Panel on Climate Change (IPCC).

Le Blanc, D. 2015. Towards Integration at Last? The Sustainable Development Goals as a Network of Targets. *Sustainable Development* 23(3): 176–187. https://doi.org/10.1002/sd.1582.

Miola, A.; Borchardt, S.; Neher, F.; and Buscaglia, D. 2019. Interlinkages and Policy Coherence for the Sustainable Development Goals implementation: An Operational Method to Identify Trade-offs and Co-benefits in a Systemic Way. Brussels: European Commission, Joint Research Centre. https://data.europa.eu/doi/10.2760/472928.

Moinuddin, M; and Zhou, X. 2022. Interlinkages of the Transition to a Low-Carbon Society and Achieving Sustainable Development Goals in West Java, Indonesia. Presentation at the IGES-SDG Center UNPAD Focus Group Discussion Workshop. Online, July 14th.

Nilsson, M.; Griggs, D.; and Visback, M. 2016. Policy: Map the Interactions between Sustainable Development Goals. *Nature* 534: 320–322. https://doi.org/10.1038/534320a.

Pennington, J.; Socher, R.; and Manning, C. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP): 1532–1543. Doha, Qatar: Association for Computational Linguistics. 10.3115/v1/D14-1162.

Zhou, X.; and Moinuddin, M. 2017. Sustainable Development Goals Interlinkages and Network Analysis: A practical tool for SDG integration and policy coherence. Research Report No. RR1602. Hayama: Institute for Global Environmental Strategies (IGES). https://pub.iges.or.jp/pub/sustainable-development-goals-interlinkages. Accessed: 2021-12-03.

Zhou, X.; Moinuddin, M.; and Li, Y. 2021. SDG Interlinkages Analysis & Visualisation Tool (V4.0). Hayama: Institute for Global Environmental Strategies (IGES). https://sdginterlinkages.iges.jp/visualisationtool.html. Accessed: 2022-01-22.

Zhou, X.; Moinuddin, M.; Renaud, F.; Barrett, B.; Xu, J.; Liang, Q.; Zhao, J.; Xia, X.; Bosher, L.; Huang, S.; and Hoey, T. 2022. Development of an SDG Interlinkages Analysis Model at the River Basin Scale: A Case Study in the Luanhe River Basin, China. *Sustainability Science* 17: 1405-1433. https://doi.org/10.1007/s11625-021-01065-z.