Semi-Supervised Event Prediction with Graph Networks

Peter Nemec
Event Registry
Slovenia
peter@eventregistry.org

Gregor Leban
Event Registry
Slovenia
gregor@eventregistry.org

Jože Rožanec
Jožef Stefan Institute
Slovenia
joze.rozanec@ijs.si

Marko Grobelnik
Jožef Stefan Institute
Slovenia
marko.grobelnik@ijs.si

ABSTRACT

This paper describes how we envision classifying events into the United Nations Sustainable Development Goals (SDGs) by utilizing machine learning techniques on global news data. We propose extracting data from a media intelligence platform using an ontology and a classifier to assign each event to its corresponding SDG. To minimize the labeling effort, a few-shot classification approach is employed. Additionally, a labeling tool is developed to facilitate event analysis and assign labels accurately. We envision this approach could be used for analyzing media events at large scale and track progress towards the SDGs.

CCS CONCEPTS

• **Applied computing → Computers in other domains**.

ACM Reference Format:
Peter Nemec, Gregor Leban, Jože Rožanec, and Marko Grobelnik. 2023. Semi-Supervised Event Prediction with Graph Networks. In *Companion of the 2023 ACM/SPEC International Conference on Performance Engineering (ICPE '23 Companion), April 15–19, 2023, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3578245.3585332

1 INTRODUCTION

This paper presents a novel approach to predicting future events by leveraging large-scale digital histories captured from ten years of news reported in the global news. We aim to develop and test methods for making real-time predictions about human and natural events of interest. In particular, we are interested in predicting how certain events relate to Sustainable Development Goals (SDGs).

How to best monitor and analyze ongoing media news events, to later predict their consequences, has been a long-standing challenge for the political science, intelligence, and finance communities [3, 9]. To address this challenge, tools have been developed for event analysis and forecasting primarily based on understanding natural language descriptions of events from text [5]. Such tools enable monitoring and analyzing news events. It can utilize generic knowledge extraction methods over event descriptions in publicly available text documents, enabling event analysis and forecasting Fig. 1. Such capabilities can be used in applications to address scientific discovery, finance, and risk management. We envision an additional step in news monitoring and analysis that can be included to determine how the events are related to SDGs.

ICPE '23 Companion, April 15–16, 2023, Coimbra, Portugal

ACM ISBN 979-8-4007-0072-9/23/04.
https://doi.org/10.1145/3578245.3585332

2 EVENT REGISTRY METHODOLOGY

Our current approach retrieves news from Event Registry [5]. Event Registry monitors global news and provides insights regarding the events and rich metadata. Event identification methods map news article body to an event type. They also capture the context of the event, e.g., locations, organizations, people, and products involved. The tool currently includes two implementations of this function, including a baseline method that relies on semantic parsing and concept linking [2], and a method using a supervised relation classification using deep learning library BERT [8]). New event matching can be introduced by labeling as few as a hundred data examples [4].

Using natural language processing to identify how a given text relates to SDGs has been explored by several authors. Amel et al. analyzed using a diverse set of classifiers consuming Word2Vec and Doc2Vec embeddings of corporate sustainability reports to predict SDG alignment [1]. Smith et al. analyzed reports presented by the United Nations Economic and Social Council about the SDGs and clustered them by computing the cosine distance between word2vec embeddings [7]. Finally, Matsui et al. reports on research on an Open Source SDG ontology and multi-label classification models [6].

3 EVENT FORESIGHT AND SDGS ALIGNMENT

Given an event of interest, such as a news event, the output of the event along with the identification and concept analysis can be utilized to analyze its potential causes and consequences and explore the relationship between two or more events. Additionally, an analysis of historical sequences of events involving the same type of entity (location, organization, person) in a given time frame can be used to understand the causes of an event and its potential consequences. This function will be implemented by analyzing pre-recorded event sequences learned using sequences of event types extracted from a historical knowledge base. Event types with recurring patterns on an entity level will have a high probability of recurrence. Our model will include capabilities for identifying sets of event types that are triggers for future entity-specified event

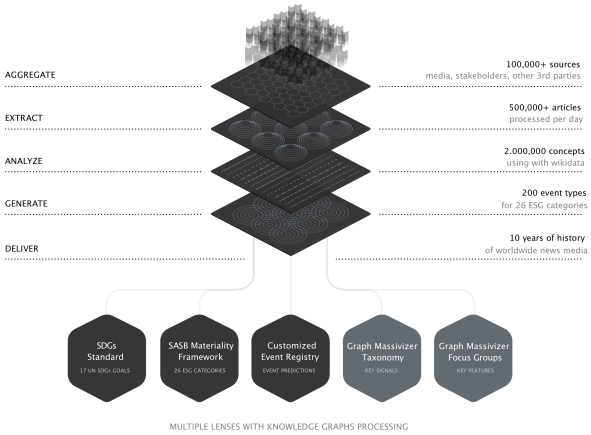

Figure 1: The diagram shows multilevel event registry data processing with knowledge graphs.

types and forecasting probabilities of the subsequent events. An ontology will be developed to align event types with the SDGs and monitoring entities' performance concerning the SDGs (see Fig. 2).

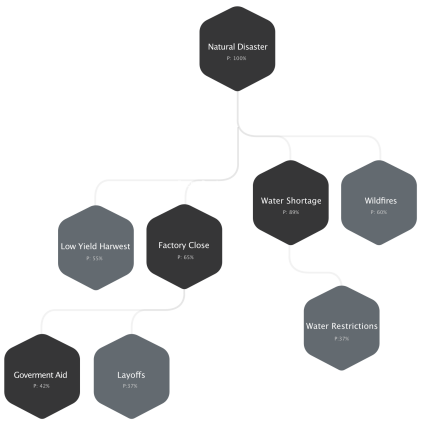

Figure 2: The chart represents events timeline for natural disaster future event-type prediction.

4 SUMMARY AND OUTLOOK

We intend to formally showcase our tool's diverse capabilities by employing a knowledge graph environment that allows us to access each node level using customized parameters and navigate different sections of the knowledge graph through an interactive visualization interface. To facilitate this, we will select recent or ongoing events to demonstrate the analysis of their potential outcomes. We may use examples such as the Ebola outbreak in 2022 or the natural disaster events such as the Indian floods caused by monsoon rain in 2016. Additionally, we will emphasize complex examples and noisy extractions and discuss various avenues for future development that can enhance this initial prototype into a reliable and robust AI assistant for analysts.

ACKNOWLEDGMENT

The Graph-Massivizer project receives funding from the Horizon Europe research and innovation program of the European Union under grant agreement 101093202. The project started on January 1st 2023 and lasts three years.

REFERENCES

[1] Amir Amel-Zadeh, Mike Chen, George Mussalli, and Michael Weinberg. 2021. NLP for SDGs: Measuring Corporate Alignment with the Sustainable Development Goals. *Columbia Business School Research Paper* (2021).
[2] Janez Brank, Gregor Leban, and Marko Grobelnik. 2017. Annotating documents with relevant wikipedia concepts. *Proceedings of SiKDD* 472 (2017).
[3] Graham Elliott and Allan Timmermann. 2016. Forecasting in economics and finance. *Annual Review of Economics* 8 (2016), 81–110.
[4] Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2019. FewRel 2.0: Towards more challenging few-shot relation classification. *arXiv preprint arXiv:1910.07124* (2019).
[5] Gregor Leban, Blaz Fortuna, Janez Brank, and Marko Grobelnik. 2014. Event registry: learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*. 107–110.
[6] Takanori Matsui, Kanoko Suzuki, Kyota Ando, Yuya Kitai, Chihiro Haga, Naoki Masuhara, and Shun Kawakubo. 2022. A natural language processing model for supporting sustainable development goals: translating semantics, visualizing nexus, and connecting stakeholders. *Sustainability Science* 17, 3 (2022), 969–985.
[7] Thomas Bryan Smith, Raffaele Vacca, Luca Mantegazza, and Ilaria Capua. 2021. Natural language processing and network analysis provide novel insights on policy and scientific discourse around Sustainable Development Goals. *Scientific reports* 11, 1 (2021), 22427.
[8] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158* (2019).
[9] Shirin Sohrabi, Michael Katz, Oktie Hassanzadeh, Octavian Udrea, Mark D Feblowitz, and Anton Riabov. 2019. IBM scenario planning advisor: Plan recognition as AI planning in practice. *Ai Communications* 32, 1 (2019), 1–13.