United Nations Educational, Scientific and Cultural Organization

International Research Centre on Artificial Intelligence under the auspices of UNESCO

IRCAI AI AWARD 2021 and Round table on Artificial Intelligence (AI) against Modern Slavery – a case study for SDG 8

17 March 2022, 9.00-11.00 (CET), register here: https://ircai.org/ircai-ai-award-2021/

IRCAI Award 2021

Explanation of the machine learning side of the project and the methods used, including research publications

There are an estimated 16,000 - 17,000 modern slavery statements published each year by UK businesses alone, with thousands more expected from other countries. Manually assessing statements is laborious and time-consuming. Currently, it takes approximately 1 hour for volunteers to evaluate manually one statement and for the researchers to validate their results. To respond to the need to accelerate this analysis and create accountability at scale, The Future Society launched a partnership with the Walk Free Initiative to automate the analysis of modern slavery statements produced by businesses to boost compliance and help combat and eradicate modern slavery. This work builds on the initial prototype developed by Adriana Bora.

The analysis of the statements takes into account a list of sixteen metrics, adapted from requirements set by the UK Modern Slavery Act. Each metric is a question about the quality of a company's modern slavery statement. For example, is the statement signed? Is it approved by the Board of Directors? Does the company have a mechanism to facilitate whistle-blowing or the reporting of suspected incidents of slavery?

The statements are mainly stored in PDFs and HTML formats. After extracting the text from PDFs and HTML, and understanding the corpus, the project moved to its second phase, focusing on creating the understanding needed to prepare for the training of the machine learning solutions. This project phase analyses the sixteen metrics used to benchmark the statements and identifies and extract their associated supporting text (quotes, context, meaning, claims, or facts) from the statements.

Different qualitative and quantitative methods were used in the supporting text extraction procedure. Yet, before starting to use any computational techniques to explore the statements, the process started with a semantic workshop to collect and imbed the domain knowledge of the specialists in modern slavery reporting.

This workshop looked at creating a set of guiding methodologies and keywords that will facilitate the identification and exploration of the metrics in the corpus. It brought together modern slavery research experts who have designed and worked with the metrics for the last few years. Based on their advice, a list of keywords was formed for each of the metrics to search for it in the document. Those keywords informed the computational methods to identify the metrics' associated supporting text in the statements.

Some exploratory tests were executed using diverse computational methods to understand the complexity of identifying the metrics and their associated supporting text in the statements. On top of this, those tests allowed preliminary classifications of the statements.

As a first step, Project AIMS considered a set of binary classification tasks whose goal is to distinguish statements addressing a given metric (no matter its actual labels) from those which do not. This step aims at understanding and assessing the complexity of the problem at a high level, and at potentially facilitating the labelling task to increase the amount of labeled examples that can be used to train machine learning models.

Those methods were:

- Labelling functions using Snorkel framework
- Rule-based and Random Forest approach
- Hierarchical Attention Network
- Transformer-based approach (using the 'Comments' from WikiRate dataset)

Then, a multi-class classification experiment was performed using the *(BHRRC + MSA) Statement approval metric* with the goal of predicting the corresponding labels for each statement.

For more details, please read *Chapter 3.3 Statement analysis using machine learning*, in the Technical Project Evaluation conducted by Mila.

Publication and Report

- Technical Project Evaluation conducted by Mila 2022, available on request. Please email Adriana at adriana.bora@thefuturesociety.org.
- AI against Modern Slavery: Digital Insights into Modern Slavery Reporting - Challenges and Opportunities - AAAI Publication 2019: The full paper can be found here; the proceedings can be found here.