

Inspecting the Accuracy of the SDG Text Mining Approach

Co-developed by Yingjie Li, Veronica F. Frans, Yuqian Zhang, and Meng Cai

Inspection Protocol

To examine and improve the accuracy of the SDG Text Mining approach, we took random text samples from the corporate reports and conducted three-round expert inspections. For each round, we asked four experts to identify if the machine-coded results are correct or not. If not, the experts should advise which SDG targets were missing or should not be coded. The experts are required to add detailed explanations and comments, followed by a 2-hour group meeting to discuss ways to improve the search terms and approach. We selected these four experts because they have rich research experience in SDGs studies, and have multi-disciplinary backgrounds. The detailed inspection procedures are as follows.

Random sample selection

We first randomly selected 0.11% ($n = 4200$) of the sentences from the data pool. Of the selected 4200 sentences, 50% are with SDG labels while the other 50% are without any SDG labels (indicating the text mining approach did not detect any SDG commitments in the sentences). In addition, among the 4200 samples, 200 are selected as common samples that all experts need to inspect. This subset can be used to evaluate inter-expert reliability. The rest of the 4000 is evenly and randomly allocated to each expert (**Figure S1**).

Inspection

Sentence samples for inspection were sorted by SDG labels so that experts can better distinguish different targets under each SDG goal. Although all the experts have sufficient knowledge and substantial research experience on SDGs, we requested that experts should refer to the [UN official SDG indicator framework](#) for inspecting the sentence samples.

Experts should read each sentence carefully and check if the SDG targets are correctly labeled or not (if yes, put 1, otherwise put 0 into the column `correct_or_not`). In most cases, one single sentence only matches one SDG target; but in some cases, one single sentence can match multiple SDG targets. Experts need to read the text carefully, then double-check and decide if ALL of the machine-coded results are correct (again, if yes, put 1, otherwise put 0 into the column `correct_or_not`).

For sentences with incorrect SDG labels, or without any SDG labels, experts should refer to the [UN official SDG indicator framework](#) and identify which SDG targets can be labeled based on their knowledge and best judgment (`which_target_ifnot`).

Experts were asked to add detailed explanations and comments to the `coder_notes` column for further group discussion. These notes can be further used to refine the database of SDG search terms.

Notes for a few special cases

- For **SDG13_2**, any statements related to climate mitigation, including adopting any type of renewable energy, can be counted. This is partly because of the broad implication of this target “13.2 Integrate climate change measures into national policies, strategies, and planning”, as well as the broader inclusion standards by UNGSII (the NGO).
- For a certain statement that is not specific enough regarding commitment to SDGs, experts can count it as an SDG commitment **as long as the action/plan/prospect is related and can potentially lead to a better result**. Alternatively, experts can put "Y maybe" or "No maybe" in the `correct_or_not` column.

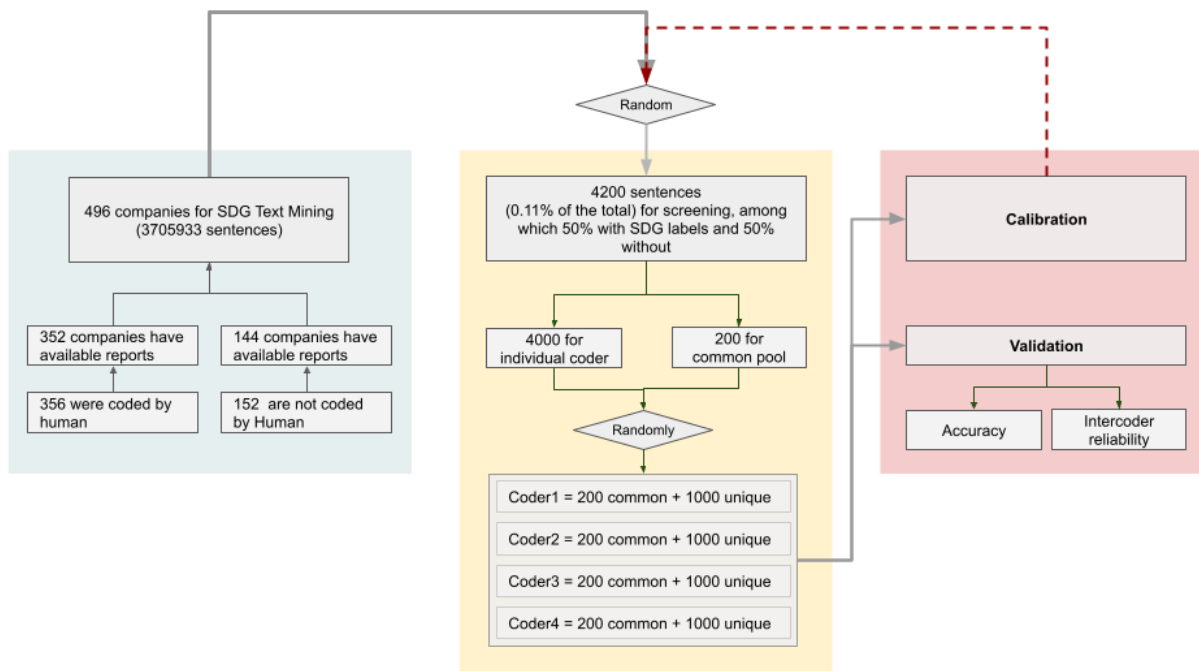


Figure S1. Flowchart for selecting samples for inspection

Inspection Results

We calculated the accuracy of the SDG Text Mining approach as well as inter-expert reliability to evaluate our approaches. Overall, our approach has achieved high accuracy in detecting SDG-related statements (> 75%, measured by the alignment between the Text Mining approach and four experts' manually-coded results. See Figure S2 and S3)

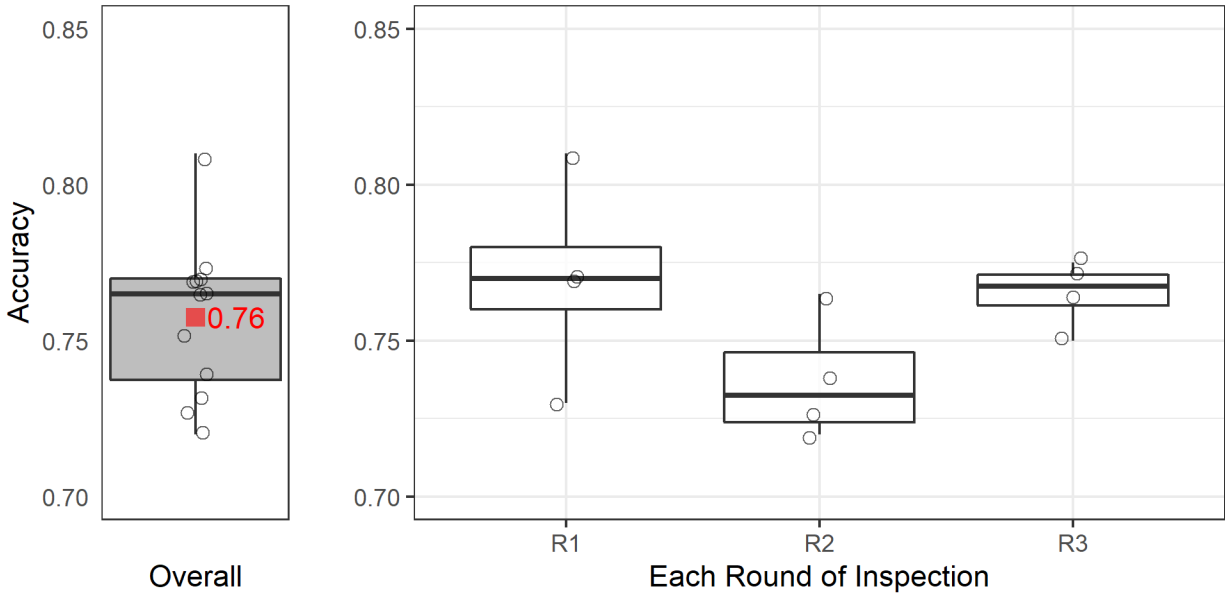


Figure S2. The overall accuracy (left) of the SDG Text Mining approach and the accuracy for each round of inspection (right). The red squared dot in the left plot indicates the mean value and the hollow round dots represent the accuracy values reported by each expert.

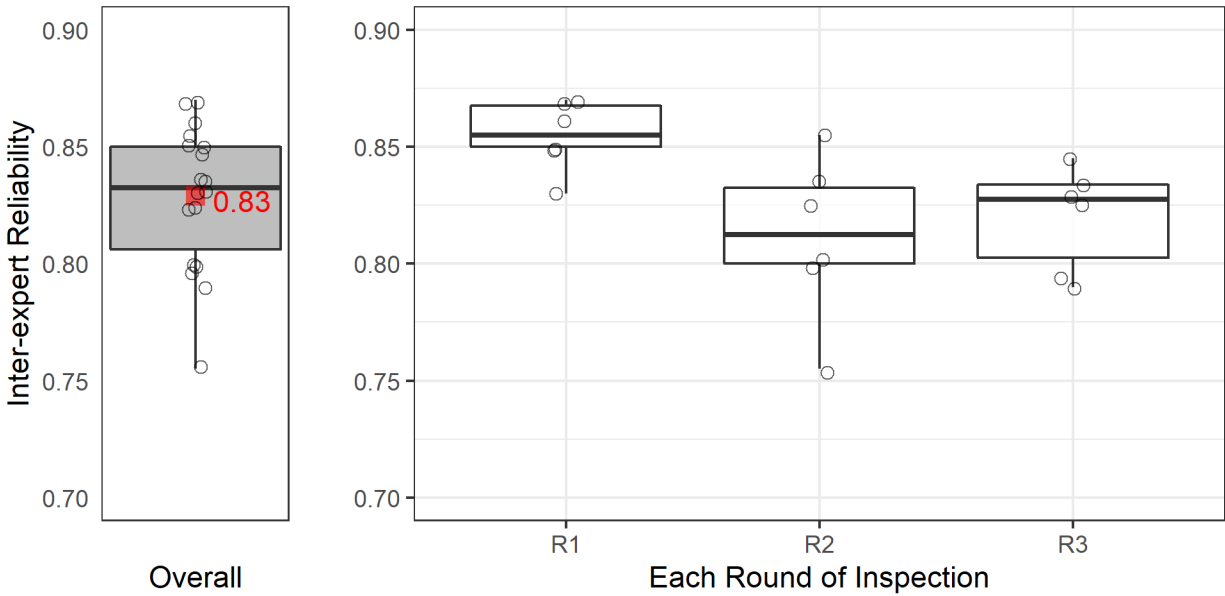


Figure S3. The overall inter-expert reliability (left) of the SDG Text Mining approach and the inter-expert reliability for each round of inspection (right). The red squared dot in the left plot indicates the mean value and the hollow round dots represent the accuracy values reported by each expert.