

READ ME FILE

Title: Central Bank Communication: One Size Does Not Fit All

Authors: Joan Huang and John Simon

Description

This 'read me' file contains general instructions on how to replicate the results presented in RDP 2021-05.

If you make use of any of these files, you should clearly attribute the authors in any derivative work.

Folder structure

The zip file 'rdp-2021-05-supplementary-information' contains this read me file ('rdp-2021-05-read-me.pdf'), and the spreadsheet 'rdp-2021-05-graph-data.xlsx' that provides the data used to plot figures in the main paper in an excel format. All data is publically available.

It also contains the following folders:

Survey Data

This folder records the primary text sources used to extract the 1,000 sample paragraphs for building the 5 online surveys as well as the raw survey results:

- 1_survey_text_raw_paragraphs – contains 11 spreadsheets that record all paragraphs extracted from 11 sources; they are candidate paragraphs for us to draw a random selection of 1,000 to form final online surveys
- 2_random_selection_result – contains 11 spreadsheets that record a random selection of 1,000 paragraphs from the original sources as discussed in Table 1 of the main paper
- 3_survey_group – includes 5 online survey spreadsheets with a random selection of 200 paragraphs for each
- 4_survey_results.xlsx – includes the survey results that are directly extracted from online surveys.

Code

This folder contains R scripts to reproduce our analysis. To help users replicate our work we have included the RStudio project file 'rdp-2021-05.Rproj'. You can load this project file into RStudio by double clicking it.

There are 4 main programs that are all written as R Markdown files. You can run the R script '0_Main.R' to replicate the 4 programs and generate output files in HTML format that will be automatically saved under this folder. Otherwise, you can run the following 4 R Markdown reports individually by clicking 'knit to HTML':

- 'P1_Survey_Preparation.rmd' – this code shows the process of how we select a random sample of 1,000 paragraphs to form the 5 online surveys as discussed in Section 3 of the main paper
- 'P2_Survey_Result_Analysis.Rmd' – this code provides survey results analysis as discussed in Section 4 of the main paper
- 'P3_Building_Models.Rmd' – this code includes the process of building 4 RF models that are presented in Section 6 of the paper
- 'P4_Model_Implementation.Rmd' – this code produces the prediction results by applying 4 models to out-of-sample text as discussed in Section 7 of the paper.

Two programs are also included as supplements:

- 'a1_nlp_extract_text_feature.Rmd' – this code shows the process of how we apply natural language process to extract text-related features for this paper. As this code requires lots of computing power to run, we limit the input data to be 10 paragraphs to ensure the code runs smoothly.
- 'a2_extract sample paragraphs.Rmd' – this code extracts sample paragraphs that are shown in Table 3 of the paper.

The subfolder 'r_function' includes 4 R programs containing some functions used in the main programs:

- 'function_sentence_feature.R' – functions to extract sentence-related features
- 'POS_tag_function.R' – functions to extract Part of Speech features for sample paragraphs
- 'text_stats.R' – functions to extract text related features for sample paragraphs
- 'tree_parse_feature_extract.R' – functions to extract parse tree features for sample paragraphs.

The subfolder 'data_input' contains data that is used by the R scripts in the 'Code' folder, and a 'model' folder within records the final 4 models used to score out-of-sample texts.

The subfolder 'data_output' includes 4 spreadsheets that record the prediction results of out-of-sample paragraphs as generated from the code 'P4_Model_Implementation.Rmd' as discussed in Section 7 of the paper. It also includes 6 HTML files that are generated using the 6 R Markdown programs listed in the 'Code' folder.

Software Versions:

- R Studio Version 1.2.5001

Contact information

Any questions relating to the details of code and data can be directed to Joan Huang.

24 May 2021