

Replication package explanatory file

Manuscript title:

Reliability of electric vehicle charging infrastructure: a cross-lingual deep learning approach

Manuscript ID: COMMTR-D-22-00115

Authors: Yifan Liu, Azell Francis, Catharina Hollauer, M. Cade Lawsons, Omar Shaikh, Ashley Cotsman, Khushi Bhardwaj, Aline Banboukian, Mimi Li, Anne Webb, and Omar Isaac Asensio.

Replication package access:

The replication code and data files have been deposited to: <https://doi.org/10.5281/zenodo.7130962>

Data description:

The paper comprises a three-stage methodology to analyze publicly available consumer reviews. The first stage involved machine translation, which utilized the Google API (<https://cloud.google.com/translate>). The input files for the next stage incorporated the English machine translated reviews. In the second stage, we applied a transformer-based deep learning model BERT to classify the machine translated unstructured reviews into predicted topics for analysis. For text pre-processing, we leverage the capabilities of BERT's open-source tokenizer to convert words into tokens that serve as an input for the model. The BERT-based-uncased tokenizer provided a reliable way to remove redundant terms such as punctuation and white spaces, which are not critical for classification. We only processed non-missing reviews. Use of the BERT tokenizer greatly simplified data pre-processing needed to accommodate a wide range of text inputs. Three separate .csv data files are required for classification: train_final.csv, valid_final.csv and test_final.csv. The predictions output .csv file named asia_test_predictions.csv contains the predictions for the test_final.csv input file. The output file from this stage was used as input for the third stage when statistical analysis was performed using R.

For the third stage, statistical modelling, the file asia_merged.csv served as the required input file for analysis in R. The output files contained the results for statistical adjustment of predicted station features based on observable characteristics. The asia_merged.csv is the main data file used to replicate the statistical analysis using the provided replication

code. In the code description below, we list the specific versions of Python and R packages needed for replication.

Code description:

We tested all replication code using Python version 3.7 and R version 4.1.2 environments, which are required to fully replicate the code. The necessary packages to have installed before code replication are provided below:

Python packages: torch==1.11.0 , os==3.1.0, numpy==1.14.3 , pandas==0.23.0, scikit-learn==0.19.1, tqdm==1.9.0, transformers==2.1.0.

R packages: frm, tidyverse, lubridate, fastDummies, stringr

Simulation

Not applicable.

software

description:

Experiment

Not applicable.

design

description:

Others:

Not applicable.