Model Info Sheet

**Section 1: Information about paper or report**

1) Author(s): Names of the authors of the paper or report

Liu, Yang and Wu, Yi Fang (2018)

2) Title of the paper or report which introduces the model

Early Detection of Fake News on Social Media Through Propagation Path Classification with Recurrent and Convolutional Networks

3) DOI or permanent link to the paper or report (for example, link to arxiv.org webpage)

<https://doi.org/10.1609/aaai.v32i1.11268>

4) License: Under which license(s) are the data and/or model shared?

n/a

5) Email address of the corresponding author

**[MX to check datasets referenced]**

**Section 2: Scientific claim(s) of interest**

6) Does your paper make a generalizable claim based on the ML model? If yes, what is the scientific claim? For example, “Our ML model can be used to diagnose Covid-19 using chest radiographs of adult patients”.

*Authors claim that their model, which employs common user characteristics, is more robust and generalizable for the early detection of fake news than state-of-the-art misinfo detection approaches that rely on linguistic and structural features.*

7) Is the scientific claim made about a distribution or population from which you can sample? If yes: (a) what is the population or distribution about which the scientific claim is being made? (b) What is the sample used for the study? For example, “(a) Population: adult patients with symptoms of Covid-19. (b) Sample: We use a random sample of adult patients who present at a U.S. based hospital between April 2020 and June 2020”.

*Yes – all three datasets consist of sample posts pulled from popular social media platforms (Weibo, Twitter). Features were extracted via Twitter API and were readily available on Weibo, as well.*

8) Does the scientific claim only apply to certain subsets of the distribution mentioned in Q6? For example, “Our model works on chest radiographs of U.S.-based adult patients and might not generalize to radiographs taken in other places or using different machines.”

*Greater method effectiveness is observed for those tweets and Weibo posts with more than a certain number of retweets / reposts on both platforms – more than 40 retweets and more than 30 reposts, respectively.*

**Section 3: Train-test split is maintained across all steps in creating the model**

9) Train-test split type: How was the dataset split into train and test sets? (For example, cross-validation; separate train and test sets).

*The authors followed the splits described in the original papers introducing all three datasets. In greater detail: they randomly divide each dataset into three folds. They hold out 10% of each dataset for validation, and split the rest 3:1 for training and testing.*

10) Are there duplicates in the dataset? If yes, explain how duplicates are handled to ensure the train-test split.

*Unclear – it seems implied that the posts in each dataset are unique.*

11) In case the dataset has dependencies (e.g., multiple rows of data from the same patient), describe how the dependencies were addressed (for example, using block-cross validation).

*Unclear if each post is from a unique user, or if multiple posts from a single user are permitted in all three original datasets.*

*Authors take pains to make sure no single user appears in both test and train sets (if a user appears in test set and is labeled misinfo spreader, would unfairly bias RNN on test set).*

12) List all the pre-processing steps used in creating your model. For example, imputing missing data, normalizing feature values, selecting a subset of rows from the dataset for building the model.

*The authors used the Twitter API (and the Weibo API? unspecified) to identify user features for those accounts associated with the posts in each dataset. ‘Propagation paths’ were constructed by considering each post as a ‘source’ and building a vector (propagation path P) of user characteristics for those accounts that interacted with the post. Following, the authors transform P into a fixed-length multivariate sequence through 1) truncating any sequences longer than some predetermined length n, or 2) oversampling any sequences shorter than n. Also: API requests made in 2017, or at time of writing, but datasets are several years old. ~~Are they differentiating between dupes?: there are several varieties: 1) a user retweets the same story multiple times; 2) a user tweets multiple different stories; 3) a user appears on multiple propagation paths.~~*

13) How was the train-test split observed during each pre-processing step? If applicable, use a separate line for each step mentioned in Q12.

*Unclear if test and train sets were treated differently during these preprocessing steps – if anything, it would seem that these steps were performed before splits were determined.*

14) List all the modeling steps used in creating your model. For example, feature selection, parameter tuning, model selection.

*Model selection, feature selection, propagation path construction, parameter tuning with validation set.*

15) How was the train-test split observed during each modeling step? If applicable, use a separate line for each step mentioned in Q14.

*First three steps were broadly applicable to each dataset in its entirety; validation sets for all three datasets were a set of randomly selected posts consisting of 10% of the dataset.*

16) List all the evaluation steps used in evaluating model performance. For example, cross-validation, out-of-sample testing.

*The authors compare their model performance to three state-of-the-art methods for the same task (abbreviated DTR, GRU, PTK), and graph no. of retweets vs detection deadline and accuracy vs no. of tweets for each dataset.*

*The authors perform precision-recall tests for all methods under consideration.*

*The authors graph detection deadline vs accuracy for all methods under consideration.*

17) How was the train-test split observed during each evaluation step? If applicable, use a separate line for each step mentioned in Q16.

*Performance on each test set was the average of the method’s performance over each stratum evaluated during three-fold cross-validation.*

**Section 4: Test set is drawn from the distribution of scientific interest.**

18) Why is your test set representative of the population or distribution about which you are making your scientific claims?

*Test sets were sampled from the most popular social media site in the U.S. and China. It seems possible that there might be platform-specific biases – maybe certain types of posts travel more quickly on Twitter vs Facebook vs something else – but the authors don’t make mention of that here.*

19) Explain the process for selecting the test set and why this does not introduce selection bias in the learning process.

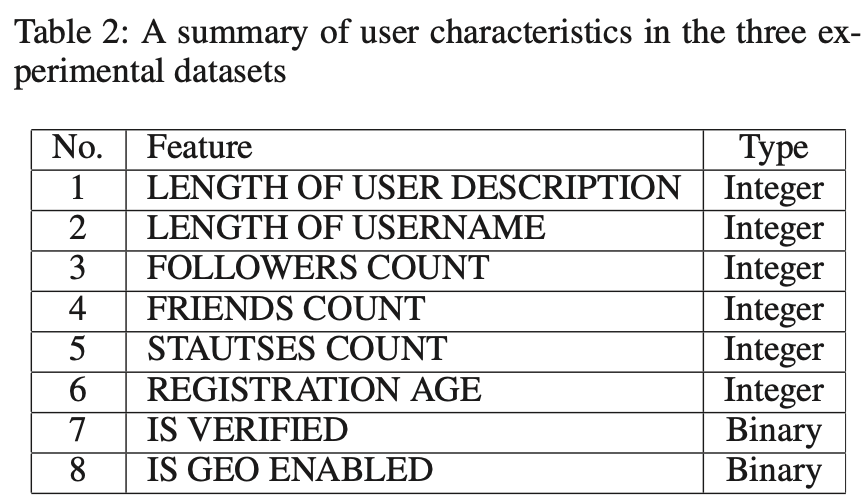
*The authors chose test sets that had been previously published. It is unclear if the authors were seeking out new / updated misinformative narratives, or if they were testing for the same misinformative news that the authors of the original datasets were flagging (the latter seems most likely). For reference, the datasets used were published in 2016 and 2017, while the paper was published in 2018.*

20) In case your model is used to predict a future outcome of interest using past data, detail how data in the training set is always from a date earlier than the data in the test set.

*Unclear what the date range for each dataset is, but they do seem to be time-boxed into single years (2016, 2017).*

**Section 5:** **Each feature used in the model is legitimate for the task**

21) List the features used in the model, alongside an argument for their legitimacy. A legitimate feature is one that would be available when the model is used in the real world and is not a proxy of the outcome being predicted. You can also include this list in an appendix and reference the relevant section of your Appendix here.



*These seem pretty standard – wondering what config of these features in propagation paths constitutes a signal for misinfo ??*