

CS4248 Natural Language Processing

Assignment 2: Solving fact checking

Distributed on 6 Feb 2021

Due in LumiNUS Files by 25 Feb 2021 11:59 PM SGT

This assignment contributes 15 marks towards your final mark for the class, and is graded out of a rubric of 100 points.

Integrity Note. Since this assignment is a common Natural Language Processing task and uses public datasets, there are (undoubtedly) solutions posted somewhere. Under the NUS Code of Conduct, you must follow class policy in working on this individual assignment. When in doubt of whether an action would constitute a violation of policy, please ask us on Slack on the general channel or by private Direct Message to Min, **before** attempting the action.

1 Instruction

The task for Assignment 2 is to attempt to solve a fact-checking task. In this fact-checking task, you need to classify whether a text is an important fact (labelled 1), an unimportant statement (labelled 0), or a non-factual statement (labelled -1).

1. Sign-in or Sign-up to Kaggle using your **NUSNET email** *If you have an existing Kaggle account under a different name, please create a new account using your NUSNET email.*
2. Go to <https://www.kaggle.com/t/6554cc27afb443e9853ebff98d6c1a50> and join the competition.
3. Change your team name to your student number.
4. Read the description, understand the data, and the evaluation metric.
5. Start coding and have fun :)

Please be aware that due to Kaggle's restriction, you can only submit up to 20 submissions per day.

You can use the optional skeleton code for this assignment: <https://drive.google.com/drive/folders/19KDKbRjGMrMi0JByBd0W7a1TFLW8QG0A?usp=sharing>

2 Submission

You need to submit a **single .zip files**, named with your student number (A01234567.zip), which contains your assignment write-up and the code that can reproduce your final (highest scoring) Kaggle submission. All files should be stored within a directory named for your student number. Your report should be named A01234567X/writeup.pdf and your code should be in Python and the main file should be named A01234567X/assignment2.py. Please use Python 3.6 and explain your code dependencies (including the library version) and the instruction to run it in your report. Both files are to be uploaded to LumiNUS Files before 25 Feb 2021 11:59 PM SGT.

3 Grading criteria

The grading consists of 70 points from the report and 30 points from the competitive level of the Kaggle final submission. The write-up grading criteria consists of:

1. Model [20%]
You should explain the reason behind the choice of the machine learning algorithm and hyper-parameters that are used to train your model. You do not need to explain how the algorithm works, but should explain any deviations from any standard model that you used or re-implemented.
2. Pre-processing [10%]
Explain all pre-processings that you have tried (including those that do not work), and how they affect the model.
3. Feature Engineering [30%]
Explain all feature engineering that you have tried (including those that do not work), and how they affect the model.
4. Analysis [30%]
Make an ablation study (analysis of how each part of your model contributes to the final score), or make a baseline model for empirical comparison.
5. Clarity [10%]
The report should be written in a clear manner and contains all necessary information of your work.

At the bottom of your report, you must include the text of the two statements below in your submitted work and digitally sign your homework using your Student Number (starting with A. . .; N.B., not your NUSNET email identifier). Make sure you have attached this statement to your submission either in written or typed form.

1A. Declaration of Original Work. *By entering my Student ID below, I certify that I completed my assignment independently of all others (except where sanctioned during in-class sessions), obeying the class policy outlined in the introductory lecture. In particular, I am allowed to discuss the problems and solutions in this assignment, but have waited at least 30 minutes by doing other activities unrelated to class before attempting to complete or modify my answers as per the Pokémon Go rule.*

1B. Exception to the Class Policy. *I did not follow the CS4248 Class Policy in doing this assignment. This text explains why and how I believe I should be assessed for this assignment given the circumstances explained.*

Signed, [Enter your A... Student ID here]

2. References *I give credit where credit is due. I acknowledge that I used the following websites or contacts to complete this assignment (but please note that many uses of Web search and detailed discussion are not allowed:*

- *Sample. Website 1, for following mathematical proofs.*
- *Sample. My friend, A000000X, whom helped me figure out the course deadlines*