CS291A Deep Learning NLP Assignment 2: Fake News Detection Competition

Due: Monday 2/26, 23:59pm PT via CodaLab
Instructor: William Wang

1 Policy on Collaboration among Students

We follow UCSB's academic integrity policy from UCSB Campus Regulations, Chapter VII: "Student Conduct and Discipline"):

"It is expected that students attending the University of California understand and subscribe to the ideal of academic integrity, and are willing to bear individual responsibility for their work. Any work (written or otherwise) submitted to fulfill an academic requirement must represent a student's original work. Any act of academic dishonesty, such as cheating or plagiarism, will subject a person to University disciplinary action. Using or attempting to use materials, information, study aids, or commercial "research" services not authorized by the instructor of the course constitutes cheating. Representing the words, ideas, or concepts of another person without appropriate attribution is plagiarism. Whenever another person's written work is utilized, whether it be a single phrase or longer, quotation marks must be used and sources cited. Paraphrasing another's work, i.e., borrowing the ideas or concepts and putting them into one's "own" words, must also be acknowledged. Although a person's state of mind and intention will be considered in determining the University response to an act of academic dishonesty, this in no way lessens the responsibility of the student."

More specifically, we follow Stefano Tessaro and William Cohen's policy in this class:

- You cannot copy the code or answers to homework questions or exams from your classmates or from other sources;

- You may discuss course materials and assignments with your classmate, but you cannot write anything down.

- You must write down the answers / code independently.

- The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved, on the first page of their assignment.

Specifically, each assignment solution must start by answering the following questions:

1. Did you receive any help whatsoever from anyone in solving this assignment? Yes / No.
   If you answered 'yes', give full details: _____
   (e.g. "Jane explained to me what is asked in Question 3.4")

2. Did you give any help whatsoever to anyone in solving this assignment? Yes / No.
   If you answered 'yes', give full details: _____
   (e.g. "I pointed Joe to section 2.3 to help him with Question 2".)

Academic dishonesty will be reported to the highest line of command at UCSB. When you are not sure, ask the teaching staff before you do so. Students who engage in plagiarism activities will receive an F grade automatically.

2 Programming Assignment (80%)

In this past election cycle for the 45th President of the United States, the world has witnessed a growing epidemic of fake news. The plague of fake news not only poses serious threats to the integrity of journalism, but has also created turmoils in the political world. The worst real-world impact is that fake news seems to create real-life fears: last year, a man carried an AR-15 rifle and walked in a Washington DC Pizzeria, because he recently read online that "this pizzeria was harboring young children as sex slaves as part of a child-abuse ring led by Hillary Clinton"[1]. The man was later arrested by police, and he was charged for firing an assault rifle in the restaurant.

Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

Table 1: The LIAR dataset statistics.

So, a question that we are asking ourselves: What can Computer Scientists do to help? In this homework assignment, you will be given a fake news detection dataset collected from POLITIFACT.COM[2], a Pulitzer Prize-winning website. We consider six fine-grained labels for the truthfulness ratings: *pants-fire, false, barely-true, half-true, mostly-true, and true*. The distribution of labels in the LIAR dataset is relatively well-balanced: except for 1,050 pants-fire cases, the instances for all other labels range from 2,063 to 2,638. The speakers in the LIAR dataset include a mix of democrats and republicans, as well as a significant amount of posts from online social media. We include a rich set of meta-data for each speaker—party affiliations, current job, home state etc. These statements are sampled from various of contexts/venues, and the top categories include *news releases, TV/radio interviews, campaign speeches, TV ads, tweets, debates, Facebook posts, etc.* To ensure a broad coverage of the topics, there is also a diverse set of subjects discussed by the speakers. The top-10 most discussed subjects in the dataset are *economy, health-care, taxes, federal-budget, education, jobs, state-budget, candidates-biography, elections, and immigration.* The statistics of the dataset is shown in Table 1.

Your task is to design a deep learning algorithm that predicts the labels, given the short statement and the meta data. There are no constraints on the type of machine learning model.

2.1 Data Set, Input, and Output

1. For this assignment, you should download the data from https://drive.google.com/file/d/0B76iNC1BIl5GMGVHQVplZFp6ZWM/view?usp=sharing.

   For the train and validation tsv files, here's what each column means:

   Column 1: the label.
   Column 2: the statement.
   Column 3: the subject(s).
   Column 4: the speaker.
   Column 5: the speaker's job title.
   Column 6: the state info.
   Column 7: the party affiliation.
   Column 8: the context (venue / location of the speech or statement).

   The goal is to use statements and other columns to predict the label. Note that the test tsv file does not include the column 1 (label), and your are suppose to generate the predictions based on statements and other meta data for this test set.

   predictions.txt: an example of the predictions file that you need to generate. It's just for your reference.

---

[1]http://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html
[2]http://www.politifact.com/

2. The input to your program should be the training, validation, and testing tsv files. Note that the testing tsv file does not include labels.

3. The output should be named as "predictions.txt", which **only includes the labels for statements that your system predicted**, in the following one label per example/line format:

   true
   false
   . . .

Note that you are not allowed to find additional training and testing datasets for this assignment.

2.2 CodaLab Environment

For this assignment, we publish a CodaLab competition (https://competitions.codalab.org/competitions/18185?secret_key=c89d9748-115c-433a-ba7d-05cc93b4dd6f), where you can submit your code, your generated predictions, and get evaluation scores that will be listed on a leader-board. If you are not familiar with CodaLab Competitions, check out here: participating in a competition.

Remember to register a CodaLab Competitions account using your umail account, so that the username will be your UCSBNetID. After that, log into CodaLab Compections and set up your team name (whatever nickname you like). To protect your privacy, only the team names will be shown on the leader-board and your usernames will be anonymous. After your submission finishes running, please choose to submit it to the leader-board. Note that here team name is equivalent to your nickname, and it is still an independent homework assignment.

If you are using your late days, please submit at a separate CodaLab late submission URL (https://competitions.codalab.org/competitions/18186?secret_key=77226ea3-21cf-49b1-8545-68a64f947cd7).

2.3 Implementation & Submission

You may use any existing machine learning and natural language processing tools, such as Gensim, TensorFlow, Theano, MXnet, Caffe, (Py)Torch, and scikit-learn. You cannot use additional training data that you have collected on the Internet. **You must avoid searching for and making use of the test set of the LIAR dataset for training purposes. If you use test set in any form during training, or use test labels in your prediction, it is a violation of honor code, and your assignment will receive 0 points automatically.** Your code must be written in Python.

Submission format   The final submission format should be:
hw2.zip

- run.py

- predictions.txt

- report.pdf

- other python scripts you wrote

Note that to create a valid submission, zip all the file with 'zip -r zipfilename *' starting from this directory.

- **DO NOT zip the directory itself, just its contents.**

- **Also, DO NOT submit the corpus: only submit your learned predictions.txt and your code.**

Here is a sample 'run.py' file: https://www.cs.ucsb.edu/~william/courses/s17_292f/run.py. You must submit the 'run.py' file with the exactly same format. For this challenge, you have a total of 30 possible submissions.

2.4 Evaluation Criteria

To evaluate the quality of your fake news detector, we will use accuracy as the metric. Note that even though your submitted predictions and your code are automatically evaluated on CodaLab Competitions, we will run a Plagiarism detection program (Stanford Moss) on them.

3 Short Report (20%)

You must write a short report (at least one page) to document your approach. You need to clearly describe the technical approach, and your experimental settings and implementation details.