

Deep Learning and Applications

Homework 2

Due data: Mar. 31, 2020

1 Task Description

In the first homework, we have learned how to implement a basic recurrent model (e.g., RNN, LSTM) for text classification. In this homework, we will further explore this direction, and try to develop more advanced models for text classification.

Towards this goal, we create a [Kaggle in-class competition](#), where the goal is to predict the category of scientific research papers. We collect those papers from the DBLP database, where we select five different domains, including machine learning, natural language processing, data mining, database, and programming language. For each domain, we select several representative conferences, and collect a number of papers from the conferences.

We provide several data files for the competition:

- **train.csv**: This file provides a set of paper ids and the corresponding labels for training. More specifically, the header of the file is “**id,label**”. Each line provides a paper id together with the label of this paper.
- **test.csv**: This file provides a set of paper ids for testing.
- **text.csv**: This file provides the titles for all the papers. The header of the file is “**id,title**”. Then each line provides a paper id together with the title of this paper.
- **reference.csv**: This file provides the references for each paper. The header of the file is “**id,id**”. Each following line provides two paper ids, e.g., “1111,2222”, meaning that paper 1111 cites paper 2222.
- **sample.csv**: This file provides an example of the file you are required to submit. More specifically, the file should have a header “**id,label**”. Then each following line should give the label of a test paper, e.g., “1111,0”.

2 Registration

Below is some instructions for registering the competition:

- **Register an account:** Before joining the competition, please register an account for Kaggle.
- **Register an account:** Next, you may register the competition by clicking “Joint Competition” at the [homepage of the competition](#).
- **Fill in a form:** Afterwards, please fill in the [Google form](#) to inform us your Kaggle account, so that we can identify every account.

3 Rules

Below is some rules for this competition:

- For each account, only two submissions are allowed everyday, meaning that you could only evaluate your results twice within every 24 hours.
- Every student is only allowed to have one account. We will filter out those additional accounts.

4 Evaluation

For this homework, we will decide your score based on your results and also your codes.

To submit your codes, please fill in the [Google form](#), and provide the link in the section “The link of your Colab code assignment”.