



IFT6285 (TALN) — Project 1
Classification

Contact :
Philippe Langlais +1 514 343 61 11 ext: 47494
RALI/DIRO felipe@iro.umontreal.ca
Université de Montréal <http://www.iro.umontreal.ca/~felipe/>

■ dernière compilation : 12 octobre 2021 (08:44)

Context

With the progress in AI, datasets such as [GLUE](#) or [SuperGLUE](#) have emerged, each of which contains about ten tasks with semantic flavor. These datasets are gathered in order to favor a transfer approach where a pre-trained model is adjusted on one or more tasks at the same time. Therefore, a model trained directly on a task has very little chance of beating a transfer approach.

However, you are under no obligation to consider a transfer approach in this project.

Data

Here are four tasks gathered in [GLUE](#) which are all binary classification tasks :

COLA [The Corpus of Linguistic Acceptability](#): is to identify whether a sentence is acceptable or not.

SST [The Stanford Sentiment Treebank](#) where the task is to identify whether the input sentence is positive or not

QQP Quora Question Pairs, is to identify whether two questions are paraphrases

MRPC [Microsoft Research Paraphrase Corpus](#), is to identify whether two sentences are paraphrases

You will consider (at least) two of these four tasks, making sure to consider COLA or SST (entry = one sentence) as well as QQP or MRPC (entry = two sentences). If you are turning in your work in a group of three, then at least three of these tasks should be considered.

For most tasks, the test corpus is reserved for official submissions (you do not have the answer). So you will evaluate your approaches on the developmental corpus for each task.

To do

1. Download the data from the four tasks. Analyze them.
2. Train one baseline system per task and report its performance on the development corpus associated with the task. You will take care to describe your baseline as well as the evaluation metric(s) used.
3. Describe your efforts to create a model that improves the performance achieved by your baseline. You will identify for the proposed model(s) the hyper-parameters that allow you to adjust them. You will show the impact of these hyper-parameters (or a subset).
4. Identify the configuration of your best system for each task studied (the same configuration can be used for all tasks) as well as its performance.
5. Analyze in detail the performance of your best model by creating new data to better understand it. You will describe the result of this analysis. For each task, you must create at least 50 new test data that you will save in a file of the same format as the development set accompanying each task and with the name `task-name.t/csv` where `task` $\in \{\text{qqp}, \text{mrpc}, \text{cola}, \text{sst}\}$ and `name` indicates the identity of the persons involved in the project. Your data should not all be of the same class.
6. If the analysis of your best model helps you identify a deficiency that you are able to fix, do so and indicate what it is.
7. Once your project is submitted (after November 12), a test set will be distributed for each task. You will run your best model for the tasks you have studied. The format of your output should be in the same format as the `dev` files for each task. More later.

The report (pdf format, in English or French) describing your experiences should be no more than five pages in length and should be named `report-name.pdf` where `name` indicates your identity.

Rating

The rating is not correlated to the performance of your approaches, but to the **curiosity** that you will develop and to your **analysis**. The clarity and informativeness of your reports are two important criteria.

Here are some of the questions we will be asking while grading your report :

- are the baseline models reasonable ?
- are the efforts put in place to exceed the performance of the baseline model justified, well described, consistent ?
- is the report clear, are the analyses informative ?
- are the annotated resources done with care ?

Because these datasets are so well studied, there is a lot of code you can use. Doing so is not a problem at all, but you should make it very clear which codes you are using.

Submitting

Upload your project on Studium under the label **projet1**. You have to submit your code, your report (pdf format, text in English or in French), and the annotation files (one for each task studied) in an archive (gzip, tar, tar.gz) whose name is prefixed with **project1-names**, where **names** is to be replaced by the identity of the people (**firstname**) involved in the project. So if I had to submit project1 alone, I would do it under the name **project1-philippe_langlais.tar.gz**.

Make sure that the names of the people involved in the project are indicated on the report. The project is due in groups of no more than three persons by Friday, November 12 at 11 :59pm.