Natural Language Inference

Machine Learning for Language Processing Project 2018/2019

1 Project Description

This project actually contains 2 tasks : sentence entailment (classification task) and sentence relatedness (regression task)

The task of sentence entailment (SICK-E) is to predict whether two sentences are **entailed**, **neutral** or **contradictory**. The task of sentence relatedness (SICK-R) is to predict the **relatedness score** between two sentences. This score ranges from 0.0 to 5.0.

The goal of this project is to implement a machine learning model for SICK-E and/or SICK-R.

2 Data Description

Details for this dataset are available at the following address : http ://clic.cimec.unitn.it/composes/sick.html
File format (tab separated) :

pair_ID	sentence_A	sentence_B	relatedness_score	entailment_judgment
93	A lone biker is jumping in the air	A man is jumping into a full pool	1.7	NEUTRAL

The provided files are described in Table 1.

Name	File	# Sent. pairs
Train	SICK_train.txt	4501
Dev	SICK_trial.txt	501
Test	SICK_test.txt	4928

TABLE 1 – Description of the data

3 Evaluation

For SICK-E, systems are evaluated on classification accuracy (the percent of labels that are predicted correctly) for every sentence pairs. We are also interested in the precision/recall scores for each class as well as a confusion matrix.

For SICK-R, systems are evaluated using the Pearson correlation coefficient : see scipy.stats.pearsonr.

4 Project Roadmap

1. Preprocess and prepare the training data
2. Train, optimize and evaluate a baseline deep recurrent neural network (RNN) using pytorch. **[one per group]**
3. Each student should propose **one** enhancement to the baseline model (additional data, regularization, network initialization, new architecture, etc.) **[one per student]**
4. Prepare the final defense : present your model and the obtained results

5 References

— SICK webpage : http ://clic.cimec.unitn.it/composes/sick.html
— Conneau and Kiela, 2018 **SentEval : An Evaluation Toolkit for Universal Sentence Representations**
  — https ://arxiv.org/abs/1803.05449