

Programming Assignment 4

(Submit via Blackboard on April 30th, 2020)

In this assignment, you will build a semantic role labeler. Many, in fact all, aspects of the system you will develop are left up to you. You can solve this task with neural or traditional ML approaches. To set up your system, first download datafile data.wsj.zip from blackboard under assignment 4, and then go to <https://www.cs.upc.edu/~srlconll/soft.html> where you will find the documentation of each data file in the data.wsj.zip file and scripts you will need to use such as `srl-eval.pl`, `make-trainset.sh`, `make-devset.sh`. You will need `make-trainset.sh` and `make-devset.sh` to create an input file. Do not use the data provided on the website, you must use the data.wsj.zip from the blackboard. You can also find a perl script(`srl-eval.pl`) that you can use for evaluation. Your task is to tackle semantic role labeling task on your own, with design and implementation decisions that you justify based on what you learned in class.

Expected output format of your system is as follows:

```
compare      *      (V*)
-            *      (A2*
-            *      *
-            *      *
-            *      *
-            *      *
-            *      *
-            *)     *)
-            *      *
-            *      *
-            (A0*)  (A0*)  (A0*)  *      *
-            *      (AM-ADV* *      *
take          (V*)   *      *      *
-            (A1*)  *      *      *
-            (A2*)  *      *      *
```

Documentation: Use the same documentation format from previous assignments. However, this time provide more extensive documentation on your machine learning algorithm of choice, features, and any other design decisions along with their justifications. Then, describe your code. As always, write a short description of each function on top of it.

Evaluation: Evaluate your system output on test data. This is the only place you should use test data. You must not touch it or look at it before your system is final. Any parameter tuning or architecture adjustment must be carried out on training data. Save your system output in a log file.

Deliverables: Submit a zip file named with student1[firstname initial][lastname]_student2[firstname initial][lastname]_[hw#].zip (i.e. student 1 jamie lee, student 2 kahyun lee: jlee_klee_hw1.zip). Zip file should include: Your code(s), models, model output, evaluation output with test data and loss values generated during training in graph (.png or bmp or any format). Give descriptive names to your models (if any). E.g., rnn_258_softmax... Indicate in model name your best model that produces the best output in the log file.

****** We will check if your program is running properly on **your** machine. One of your team members will need to meet with the TA at one of the times listed in the following link in order to show your program to the TA. Please reserve a time slot to meet with the TA using the google doc below. If none of you are available at the time slot, please contact TA.

<https://docs.google.com/spreadsheets/d/14jvlkce80dnu2W54SU4PNTNVpwQuV4rBAyoFlrYHO7g/edit?usp=sharing>