

# CMPT470-816AssignmentMachineLenaring

This Repository contains the code for the machine learning assignment of CMPT 470/816.

## Getting Started

The assignemnt will give you a basic understanding of how deep neural network can be used for source code. Here we use the source code as a feature.

By this assignment we will try to solve the following Software engineering problem:

Method Call Recommendation using Reccurent Neural Network(RNN) and Long Short Term Memo

There are three parts of this assignment:

1. Run the existing code: The code in the repository collect the dataset from any number of java projects. Then you use the dataset and run the python written machine learning code. It will train a reccurent neural network based model, evaluate the top-1 reccomendation and return the accuracy value. Neural Network is developed using [Keras](#) library.
2. Train and test with LSTM model: In this task, you need to change the machine learning code in such a way that it will train a LSTM model and test using the same. To do that you need to first do understand the code -> concept location -> effect analysis -> update code. Hint: The change is very small
3. Write the machine learning code using DeepLearning4j: Now do the whole machine learning task using DeepLearning4j library. [DeepLearning4J](#) is a java based deep learning library maintained by Eclipse foundation. The architecture of the library is very similar to Keras. Therefore the tasks will be similar.

## Prerequisites

Before running the code, you must have following packages installed in your computer:

```
java8 or later
maven
python3
pip
```

## TASK 1

A step by step instruction is given in the following

Step 1 Import all dependencies of maven first.

```
Open src/main/java/config/Config.java
change the following global variables
public static final String LOG_PATH = ROOT_PATH+"cms500_codeparser.log";
```

Put your NSID in logfile name. Otherwise you will not get any marks.

Step 2: Delete all .log files.

Step 3:

```
Open src/main/java/extraction/MethodCallCollector.java
Run main function
```

If the main method is run properly, you will see 10 csv files is written in "./data/dataset/" folder. These will be used for the machine learning script. So dont change anything.

Step 4: Run the following comamnd:

```
pip install -r requirements.txt
```

This will install all the required library for the machine learning tasks.

Step 4:

```
open config.py
change the following global variables
log_file_path = root_folder+'cms500_keras_rnn.log'
```

next either run **main.py** file using IDE or run the following command:

```
python __main__.py
```

This will train the RNN based model on nine folds and will test with the remaining one fold of the dataset colelcted form step 3.

## TASK 2

Understand the code written in the **main.py** file. Find where you need to change to support LSTM training and testing (concept location). Try to analynis the effect of change of the code (effect analysis) and then update the code.

Hint: You dont need to install any new library. Just a very small change will do the work

## TASK 3

This is a but tough because you need to learn how DeepLearning4J work. However, the working mechanism of kepler and deeplearning4j is almost same. But you need do data collecttiona data data preprocessing in java too.

To learn how RNN please look the follwing code example:

[RNN Example](#)

## Another RNN Example

First try by yourself. If you fail then come to me. I will not help you if you dont anything.

### What to hand in

For each of the task provide follwoing three file: 1. Zip file of the full project 2\_codeparser.log 3. \_keras/deeplearning4j\_rnn/lstm.log

Lastly one file name as README.txt file that will have the followings: 1. Status of all three task 2. How to run each of your project(s): requirements, steps to run 3. Changes in python script for competing task 2. 4. Overall what you learn from the assignment and your assesment of the assignement out of 10 in three category: 1. Usability(10 as in very useful, 0 as in not required for future), 2. Complexity(10 as in very hard, 0 as in very easy), 3. New learning(10 as in everythin were new, 0 as in nothing learned) 5. Any additional comment. Your assesment will not be considered for marking and will be helpful for future.

### Marking

1. Dataset Creation: 5 (Considered as complete if and only if the your nsid\_codeparser.log file is provided that have all log statement)
2. Task 1: 10 (Considered as complete if and only if your nsid\_keras\_rnn.log file and the zip file of the project are provided)
3. Task 2: 20 (Considered as complete if and only if your nsid\_keras\_lstm.log file, the zip file of the project and the README.txt file are provided. For task 2, the main important line in README.txt is "4. Changes in python script for competing task 2.")
4. Task 3: 40 (Considered as complete if and only if your nsid\_deeplearning4j\_rnn/lstm.log file and the zip file of the project are provided.)
5. README.txt: 5 (Considered as complete if and only if descriptions of first two points are provided and TASK 1 is complete. Even if you write you cant finish Task 2 or 3, you will get full marks for README.txt file.)

Total Marks: 80

### Issues

If you have any problem with the code provided in the repositroy. Please create an issue. I will try to solve it as soon as possible

## Authors

---

- C M Khaled Saifullah - *Initial work* - [khaledkucse](#)

See also the list of [contributors](#) who participated in this project.

## License

---

This project is licensed under the GNU General Public License v3.0 - see the [LICENSE](#) file for details

