

CMPT470-816AssignmentMachineLenaring

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

Getting Started

The assignment will give you a basic understanding of how deep neural networks can be used for source code. Here I use the source code as a feature and feed them into the neural network.

By this assignment, you will 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 collects 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 recurrent neural network based model, evaluate the top-1 recommendation and return the accuracy value. Neural Network is developed using [Keras](#) library.
2. Train and test with the LSTM model: In this task, you need to change the machine learning code in such a way that it will train an LSTM model and test using the same. To do that, you need to first understand the code, locate where the change is required (concept location), determine the lines of code that are going to be affected due to the change (effect analysis), and update the code. Hint: The change is very small
3. Write the machine learning code using DeepLearning4j: Now you need to write the whole machine learning code using the DeepLearning4j library in Java. [DeepLearning4J](#) is a java based library maintained by Eclipse foundation. The architecture of the library is very similar to Keras. Therefore the coding style will be similar.

Prerequisites

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

```
java 8
maven
python 3.x
pip
```

TASK 1

A step by step instruction is given in the following

Step 1 Import all dependencies of maven first.

Next,

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

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

Step 2: Delete all .log files(if any).

Step 3:

```
Open src/main/java/extraction/MethodCallCollector.java
Run the 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 don't change anything.

Step 4: Run the following command:

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

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

Step 4:

```
open config.py
change the following global variable
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 collected from step 3.

TASK 2

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

Hint: You don't need to install any new library. Just a small change will do the work

TASK 3

This is a bit tricky because you need to learn how DeepLearning4J works. However, the working mechanism of Keras and deeplearning4j are almost the same. However, you need to do data collection and data

preprocessing in java before configuring neural network, training, and testing.

To learn how RNN works, please look at the following code examples:

[RNN Example](#)

[Another RNN Example](#)

First, try for yourself. If you fail then come to me. I will not help you if you don't anything.

What to hand in

1. your_nsid_codeparser.log
2. For each of the task provide the following three files:
 - Zip file of the full project
 - your_nsid_keras/deeplearning4j_rnn/lstm.log
3. Lastly one file name as README.txt file that will have the followings:
 - Status of all three tasks.
 - How to run each of your project(s): requirements, steps to run
 - Changes in python script for completing task 2.
 - Overall what you learn from the assignment out of 10 in three categories: 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 everything were new, 0 as in nothing learned)
 - Any additional comment. Your assessment will not be considered for marking and will be helpful for the future.

Marking

1. Dataset Creation: 5 (will be considered as complete if your_nsid_codeparser.log file is provided with all log statements)
2. Task 1: 10 (will be considered as complete if your_nsid_keras_rnn.log file and the zip file of the project are provided)
3. Task 2: 20 (will be considered as complete 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 completing task 2.")
4. Task 3: 40 (will be considered as complete if your_nsid_deeplearning4j_rnn/lstm.log file and the zip file of the project are provided)
5. README.txt: 5 (will be considered as complete if the 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 repository. 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