User Manual - Major II

Short Text Classification using Memory Networks

Submitted By:

Aman Parmar (16103221)

Palak Arora (16103046)

Anjali Sharma (16103015)

Submitted To:

Mr. Prashant Kaushik

User Manual

How to acquire the data set?

The data set used for the project can be acquired from various sources mentioned below

- Google database search by using the keyword short text instances
- GitHub repositories
- Kaggle databases by using the keyword short text classification

The datasets acquired from the above three sources are the primary datasets. Furthermore preprocessed versions of the same can be downloaded from the GitHub repository specified below :

https://github.com/amanparmar17/short-text-classification

The complete preprocessed dataset for each of the models used in the project specifically mentioned inside the folder title with the same name.

How to run the complete project?

- Create a Python version 3.6+ virtual environment in the machine
- Install the required packages :
    - **Numpy**
        - Scientific Python library for advanced mathematical and array related operations
    - **Pandas**
        - Advanced Python library to handle data frames and series
    - **Matplotlib**
        - Python package used for high dimensional plotting of data points
    - **Scikit learn**
        - Python package specifically designed for pre-processing and machine learning operations
    - **Gensim**

        - Scientific Python package for handling text classification tasks
    - **Keras**
        - Python package built to handle tasks related to artificial neural networks and other deep learning models
    - **NLTK**
        - Python package used for NLP related tasks like tokenization and removal of stop words
- Clone the GitHub repository titled "Short text classification", from the link mentioned below :

    https://github.com/amanparmar17/short-text-classification

Models present in the project:

- Bag of Words
- N-grams
- Random Forest
- Long Short Term Memory
- Topic Memory Network

How to run the models?

To use the pre-trained models implemented in the project, navigate to the "models" directory inside the cloned repository, and use the models as per need.

To handle the complete task of training and testing the machine learning and deep learning models from scratch the skeleton code for each of the models can be found inside a specific folder titled "Scripts".

NOTE: To run the models, use your own dedicated machine or can use any other cloud provider like google colab or Microsoft azure.