Project: Towards building foolproof Automated Essay Scoring system robust to Adversarial attacks

Author: Dheeraj Kumar Kondaparthi(dkonda2@uic.edu)

------------------------------------------------------------------------------------------------------

Extract the .zip and upload /StatNLP/... folder and all its files and sub-folders into Google drive

--------------------------------------------------------------------------------------------------------

Pre-requisites:

All libraries are required as listed below.

Python 3.7.3

Google Colab

Keras 1.14.0

Tensorflow 2.0

Anaconda 3

Jupyter Notebook

Matplotlib

Numpy

Pandas

GloVe Embeddings must be downloaded into “Data Sets/GloVe” folder using below URL

http://nlp.stanford.edu/data/glove.6B.zip

--------------------------------------------------------------------------------------------------

Part 1 - Creating a Dual Scoring CNN system, as a Black Box to test.

Open

- Blacbox.ipynb file

using Google Colab and set the notebox mode into "Blackbox Training" and execute all cells

Part 2 - Generating Adversarial Attacks.

Load

- ADVERSARIAL-ATTACK-GENERATOR.ipynb

- AnchorModel.ipynb and

using Google Colab and execute all cells.

(Please note Anchor model takes in terms of hrs to run, so be patient to get results)

Part 3 - Applying the attacks, and evaluating the results.\*\*

Open

- Blackbox.ipynb file using Google Colab and set the notebox mode into "Adversarial Testing" and execute all cells.

-------------------------------------------------------------------------------------------------------

Model Files : This folder has all the saved models and architecture diagram of dual CNN model.

Data Sets: It has dataset from ASAP challenge, GLove Embeddings,