School of Computer Sciences

CPC353/CPT346 Natural Language Processing

ASSIGNMENT 2 : Natural Language Processing Semester I 2021/2022

Objectives

- Build natural language processing and deep learning approaches in problems involving prediction, classification and sequence modeling in text and speech.
- Design customized solutions for different problems in text and speech analytics.

Title

Sentiment analysis on news headlines

Specification

Annotate the given news headlines. Create a bag-of-words feedforward classifier and a bidirectional LSTM classifier to model sentiment given the news headlines.

Specification

The assignment will be carried out by a group of 1 student.

Part 1

Annotate 100 headlines with positive, neutral or negative sentiment.

Part 2

Create a Bidirectional LSTM classifier to predict the sentiment. Convert the words to word embedding using Glove (https://nlp.stanford.edu/projects/glove/). Do not use word embedding layer. Suggest and add improvement on the model.

Mark distribution

i. Part 1:20%
ii. Part 2: 80%

Deadlines:

Part 1: Data submission deadline: Sunday, 9/1/2022 11.59 pm.

Download combined dataset from elearning: Monday, 10/1/2022

Part 2: Program, report and video submission deadline: Sunday, 30/1/2022 11.59 pm.

Reference:

Aurélien Géron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly Media, 2017.

Note:

This assignment is to be carried out individually. Any references taken from any source must be quoted and declared.

.