

Natural Language Processing – Project Description

Tal Levi – 207084427, Omri Maoz - 205488661

1. Topic:

- Test hypothesis for classify text phrases based on their corresponding tagging phrases using 2 Deep learning contextualize models (DAN & LSTM).

2. Data:

- 2 datasets will be used to test the hypothesis and compare results:
 - The IMDB positive and negative reviews – binary label space. [Link](#).
 - News categories headlines – multi-class label space. [Link](#)
- Create a corresponding tagging text phrase for each phrase from the dataset using Python NLTK package. [Link](#).

3. Methods:

- We would like to test whether, and with which accuracy, can a text phrase be classify based on its tagging.
- We will train 2 models, on each dataset, with and without tagging:
 - Trained IMDB dataset with DAN\LSTM compared to trained 'IMDB' tagging-dataset with DAN\LSTM.
 - Trained 'News' dataset with DAN\LSTM compared to trained 'News' tagging-dataset with DAN\LSTM.

* Training IMDB dataset with DAN we did on homework 2.

4. Evaluation:

- Evaluate models by their accuracy on the test (validation) data.

5. Expectation:

- We believe that positive and negative semantics has some common tagging structures in natural language and that is an advantage we can harness to achieve better classification results.
- We also want to test the hypothesis with a multi-class classification which we believe will be a more “difficult” task for the models.