

E0 334 - Deep Learning for NLP

Assignment 1

(due by 3rd Sept, 11:59 PM)

Note: Use the following link for submitting your results of Assignment 1.

<https://forms.office.com/r/AFQTteKZAR>

Problem:

The aim of this assignment is to study the use of different pre-trained word embeddings (word2vec/GloVe/fastText) for text representation and use them for text classification. You can download pre-trained word embeddings of your choice [1, 2, 3].

Design classifiers using any of the neural architectures discussed in the class till Aug 23 (including the one mentioned in [5]) for the three datasets provided (ClassificationDataset0,1,2). You could use different classifiers for each of the datasets.

Use the above-mentioned form to submit average validation set performance (F1-score) for each of the datasets.

Note: You can use libraries like Gensim (<https://radimrehurek.com/gensim/>) or spaCy (<https://spacy.io/>) for various text processing as well as other tasks. You can use t-SNE software[4] to visualize sentence/paragraph representations.

References

1. Word2vec (<https://code.google.com/archive/p/word2vec/>)
2. GloVe (<https://nlp.stanford.edu/projects/glove/>)
3. fastText (<https://fasttext.cc/>)
4. t-SNE Software (<https://lvdmaaten.github.io/tsne/>)
5. Iyyer *et al*, Deep Unordered Composition Rivals Syntactic Methods for Text Classification. https://people.cs.umass.edu/~miyyer/pubs/2015_acl_dan.pdf. Also see the code available at <https://github.com/miyyer/dan>