**Assignment 4: Movie Review Prediction**

**Due: December 3 11:59pm (AOE)**

1. Task: In this assignment you will be implementing a movie review prediction system which predicts a numerical rating score (range from [0, 1]) for a movie review text.

● Dataset: Movie Review Data ([website](https://www.cs.cornell.edu/people/pabo/movie-review-data/))

○ You may download the preprocessed version from [here](https://1drv.ms/u/s!AjiSpuwVTt09gipGGMCIf1QH2ecV?e=KVmvLJ), which contains 3,003/1,001/1,002 train, development and test examples.

● Format - index, rating, text

*3550 0.8 like water for chocolate is a wonderful , magical movie from mexico . it is a tragedy of shakespearean proportions . as i watched it , i keep thinking it reminded me of shakespeare's play romeo and juliet and of wager's opera gotterdammerung . within the confines of the basic plot setup is the heart of the movie .*

*...*

● Evaluation: Compute R^2 (coefficient of determination) scores on test set and report them in the final submission. You can call the metric function implemented by sklearn ([link](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html)).

2. [5pts] You should implement a GNN-based system which takes as input a text graph, and predicts a graph-level score as output. Some design choices to consider:

* Graph construction, e.g., how to convert the input text to a graph which captures its rich structural information
* Graph representation learning, e.g., how to design the GNN encoder which suits the text graph, how to compute the graph-level representation from the node-level representations
* Node embedding initialization, e.g., how to initialize the node embeddings

You should implement your system using the [Graph4NLP library](https://github.com/graph4ai/graph4nlp) which provides off-the-shelf implementations of many variants of the above design choices. You can find other [demos](https://github.com/graph4ai/graph4nlp/tree/stable_nov2021/examples/pytorch) built upon Graph4NLP and borrow their implementation ideas. More details covered in the lecture slides.

3. [2pts] Possible Enhancements to consider:

● Compare and report the performance of various options for graph construction, graph representation learning and node embedding initialization. You might need to tune other hyperparameters (e.g., number of GNN layers, hidden size) as well to get better performance.

* Any other enhancements to improve the model performance.

4. [3pts] Qualitative Analysis and Report Writing

● If you have done any successful enhancements to the baseline methods, report gains and concrete examples to explain why they worked.

● Look into your remaining errors, categorize remaining challenges and suggest possible solutions with concrete examples.

5. Submission:

● Please submit your **{netid}.zip** to CANVAS

● Your zip file should include your **source code folder** and your **report pdf**. Any supplementary materials can be uploaded as external downloadable links.

6. Graph4NLP library environment setup

1) Create virtual environment

```

conda create --name graph4nlp python=3.7

conda activate graph4nlp

```

2) Install [graph4nlp](https://github.com/graph4ai/graph4nlp) library

* Clone the github repo

```

git clone -b **stable\_nov2021** https://github.com/graph4ai/graph4nlp.git

cd graph4nlp

```

* Then run `./configure` (or `./configure.bat` if you are using Windows 10) to config your installation. The configuration program will ask you to specify your CUDA version. If you do not have a GPU, please choose 'cpu'.

```

./configure

```

* Finally, install the package

```

python setup.py install

```

3) Install other packages

```

pip install torchtext

```

4) Set up StanfordCoreNLP (for static graph construction)

* Download [StanfordCoreNLP](https://stanfordnlp.github.io/CoreNLP/)
* Go to the root folder and start the server

```

java -mx4g -cp "\*" edu.stanford.nlp.pipeline.StanfordCoreNLPServer -port 9000 -timeout 15000

```

If you have any questions regarding implementing the system using Graph4NLP library, please leave your questions in this FAQ [doc](https://docs.google.com/document/d/1rkUwOyWddcnn45XyeI4PikiZ9QR8qve8LHMd0-gZQcc/edit?usp=sharing). I will answer your questions there.

References:

* Graph4NLP library <https://github.com/graph4ai/graph4nlp>
* Graph4NLP documentation <https://graph4ai.github.io/graph4nlp/>
* DLG4NLP website <http://dlg4nlp.github.io>
* Graph Neural Networks for Natural Language Processing: A Survey <https://arxiv.org/pdf/2106.06090.pdf>