DOCUMENTATION

B21PV02

Team members:

Swathi Kedarasetty S20180010172
Akula Hari Chandana S20180010009
Usarthi Harshini S20180010180

Our project is a news recommendation system with Attentive Multi-view learning.

Steps to run the code

Basic Setup

Go into NewsRecommendation directory and install the requirements

```
cd NewsRecommendation
pip3 install -r requirements.txt
```

Download and preprocess the data

```
cd data
# Download GloVe pre-trained word embedding
wget https://nlp.stanford.edu/data/glove.840B.300d.zip
sudo apt install unzip
unzip glove.840B.300d.zip -d glove
rm glove.840B.300d.zip


# Use the MIND Small dataset (Note MIND Small doesn't have a test set, so
we just copy the validation set as test set )
wget
https://mind201910small.blob.core.windows.net/release/MINDsmall_train.zip
https://mind201910small.blob.core.windows.net/release/MINDsmall_dev.zip
unzip MINDsmall_train.zip -d train
```

```
unzip MINDsmall_dev.zip -d val
cp -r val test # MIND Small has no test set :)
rm MINDsmall_*.zip

# Preprocess data into appropriate format
cd ..
python3 src/data_preprocess.py
```

Run

Load the latest checkpoint

I've already trained the model and hence, we can load the latest check-point and directly evaluate on the test set using the following command
Download the latest check-point that I've saved from the following link
https://drive.google.com/drive/folders/1-JBHYgaeAZx7iqRtaskO41O6z8OsDDUy?usp=sharing

```
# Load latest checkpoint and evaluate on the test set
python3 src/evaluate.py
```

To do a fresh run

```
# Train and save checkpoint into `checkpoint/{model_name}/` directory
python3 src/train.py
# Load latest checkpoint and evaluate on the test set
python3 src/evaluate.py
```

Previous run

The news_run.ipynb in the src folder has the previously run session along with the results obtained on the MIND-small dataset.