Final Report

Project: Textual Coherence
Team Number: 24
Team: Advaith Malladi (2021114005), Patanjali B (2021114014), Revanth Gundam
(2021114018)

What has been accomplished:

- Predict Coherence based on Topic of Paragraph (95% accuracy)
- Predict Coherence based on Topic of Paragraph and Temporal coherence of paragraph (whether events occur sequentially and temporally or not) (80% accuracy)
- Assign Coherence Score based on Topic of Paragraph and Temporal coherence of paragraph
- If a paragraph was shuffled, predict the most coherent re-ordering (76% accuracy)

What makes our final submission special:

- We accomplished predicting and evaluating topic and temporal coherence, which are arguably the most important metrics for overall textual coherence.
- Given shuffled paragraphs, our model predicts which permutation of sentences is the correct and original paragraph based on the metrics we considered earlier.

Dataset: We have used the CNN dataset for textual coherence

Model design:

LSTM design:
- input size: 300
- hidden layer size: 300
- number of stacks: 2

Followed by Hidden2Tag layer:
- input size: 300
- output size: 2

Followed by log_softmax

I have also tried to use GRU, the LSTM outperforms GRU by 10% in our case

Task 1:

- For, the first task we trained the model as we have explained in our interim report.
- My dataset consisted two kinds of samples, positive coherent samples, negative samples, that is paragraphs with random sentences inserted in the middle.
- By training on this dataset, our model learned to check if the paragraph was coherent or not, based on the topic.

- It achieved the following scores:
- Training:
  - 50 epochs of training
  - average accuracy: 0.959704061109581
  - average F1 score: 0.9592171332083778
  - average recall: 0.9598704778547373
  - average precision: 0.9598770763790266
- Testing:
  - average accuracy: 0.9530687785781591
  - average F1 score: 0.9526936217142479
  - average recall: 0.9530222501983382
  - average precision: 0.9531950044037375

Task 2:
- In this task, our dataset consisted of two kinds negative sampling:
  - paragraphs with random sentences in middle (topic coherence violated)
  - paragraphs shuffled into random order (temporal/sequential coherence violated)
- Results:
- Training:
  - 150 epochs of training
  - average accuracy: 0.909
  - average F1 score: 0.899
- Testing:
  •After training, i added more negative samples to my dataset to make my metrics more robust:
  - average accuracy: 0.80
  - average F1 score: 0.70

Task 3:
- In this task , our model had to give a coherence score to a paragraph, ranging from 0, 1.
- 1 is completely coherent, 0 is completely incoherent.
- If our model gives a score greater than 0.5, the paragraph could be considered coherent.
- In this task also, our model achieved an accuracy of 0.80

Task 4:
- In this task, I only had to testing on a different kind of task (shuffling) to evaluate earlier trained model.
- For each paragraph, I made 4 random shufflings of the paragraph.
- Our model was given 5 instances of a paragraph, 1 correct one and 4 randomly shuffled paragraph, our model had to predict the correct paragraph based on textual coherence.
- This was done by comparing the coherence score our model predicted.
- If it predicted correctly, that counted as one positive prediction.

- Our testset has around 35000 unseen samples.
- Out of this, our model gave an accuracy of 0.758322900431

Link to Presentation: Presentation18.pptx