

# TEXTUAL COHERENCE

Eshika Khandelwal - 2020114018

Yash Agrawal - 2020114005

Adith John Rajeev - 2020114010

# Today's Agenda

◆ Data-sets used

◆ Models

◆ Approaches used

◆ Result and Conclusion

# Data-sets used

## Grammarly Corpus of Discourse Coherence

The dataset is annotated into classes 1,2 and 3 where 3 denotes the most coherent paragraph. It consists of four training datasets (1000 paragraphs each) and four testing datasets (200 paragraphs each). We have merged the four training datasets along with three testing datasets to train our model (4600 paragraphs) and used the remaining test data to test the accuracy of our model. Some datasets being more open-domained than the others, we decided to vary the test dataset chosen in order to compare the results.

# Data-sets used

## Wikipedia-CNN Dataset

The dataset consists of coherent text from Wikipedia and the CNN/Daily news sets. The text dataset also contains respective replacements to make every paragraph incoherent. These replaced sets are used as negative samples in our model.

# Models

- 01 Simple RNN - A fully-connected RNN where the output from previous timestep is to be fed to next timestep**
- 02 GRU - The GRU is like a long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate**
- 03 LSTM - LSTM units include a 'memory cell' that can maintain information in memory for long periods of time. This memory cell lets them learn longerterm dependencies.**

# Approaches used

01

**Varying Test Data** - Some test data in the GCDC corpus was more closed-domained than the others leading to varying accuracies.

02

**Three way classifier to binary classifier** - We trained our model on the GCDC corpus using the default annotation to make a three-way multi-classifier. Later, we switched this approach, remodelled our data to binary labels- coherent and incoherent- and used this data on a binary classifier to observe better results

03

**Similarity** - Similarity between sentences is useful in scoring coherence as a higher degree of similarity between neighbouring sentences indicates a more coherent text

# RESULTS AND CONCLUSION



# Accuracy obtained

Varying test data in GCDC Corpus: Results obtained after running the LSTM model with Binary classifier and minimum similarity as a parameter were as follows:

TABLE I  
ACCURACY RESULTS WITH DIFFERENT TEST DATA

	Corpus used for Test Data			
	<i>Clinton</i>	<i>Enron</i>	<i>Yahoo</i>	<i>Yelp</i>
<i>Accuracy %</i>	61.0%	66.0%	54.5%	56.5%





## LSTM



## GCDC Corpus

- Without using similarity as a parameter
- 3000 training data, 600 testing data, three-way classifier: **approx 30%**
- 4600 training data, Yahoo test data (200), three-way classifier: **approx 36.5%**
- 4600 training data, Yahoo test data (200), binary classifier: **approx 55%**
- 4600 training data, Clinton test data (200), binary classifier: **approx 64.99%**



## LSTM



## GCDC Corpus

- Using average similarity as a parameter
  - 4600 training data, Clinton test data (200), three-way classifier: **approx 34%**
- Using minimum similarity as a parameter
  - 4600 training data, Clinton test data (200), three way classifier: **approx 39.5%**
  - 4600 training data, Clinton test data (200) , binary classifier: **approx 67%**



## LSTM



## Wikipedia-CNN Corpus

- Without using any similarity parameter: **approx 71.66%**
- Using minimum similarity parameter: **approx 74.55%**



## GRU

.....

### GCDC Corpus

- Without using similarity as a parameter
  - 4600 training data, Clinton test data (200), three-way classifier: **approx 41.99%**
  - 4600 training data, Clinton test data (200), binary classifier: **approx 57.99%**
- Using minimum similarity as a parameter
  - 4600 training data, Clinton test data (200), binary classifier: **approx 63.99%**



## GRU

.....

## Wikipedia-CNN Corpus

- Binary classifier was used for analysis.
  - Without using any similarity parameter: **approx 69.89%**
  - Using minimum similarity parameter: **approx 77.76%**



## Simple RNN

.....

### GCDC Corpus

- Without using similarity as a parameter
  - 4600 training data, Clinton test data (200), three-way classifier: **approx 32.49%**
  - 4600 training data, Clinton test data (200), binary classifier: **approx 55.5%**
- Using minimum similarity as a parameter
  - 4600 training data, Clinton test data (200), binary classifier: **approx 53.5%**

# Observations

Using Minimum Similarity as a parameter, we observe a significant rise of around 2-5% in the accuracy. This is in line with the definition of coherence which demands a level of similarity in the flow of the text.

While varying the test data in the GCDC corpus on the LSTM Binary classifier with minimum similarity as a parameter we observe that the Enron dataset has the most accuracy, which implies that it probably has the most closed domain, while Yahoo has the most open domain.

Expanding our dataset by including some test data into training increased the accuracy of our model around by 6- 7%. Using a much larger dataset, such as the Wikipedia CNN corpus (100x larger than GCDC) showed an even larger increase in the accuracy.

# REFERENCES





# Resource Page

- Jiwei Li and Dan Jurafsky. 2017. Neural Net Models of Open-domain Discourse Coherence. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 198–209, Copenhagen, Denmark. Association for Computational Linguistics
- Peng Xu, Hamidreza Saghir, Jin Sung Kang, Teng Long, Avishek Joey Bose, Yanshuai Cao, and Jackie Chi Kit Cheung. 2019. A Cross-Domain Transferable Neural Coherence Model. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 678–687, Florence, Italy. Association for Computational Linguistics
- Farag, Y. (2020). Neural approaches to discourse coherence: modeling, evaluation, and application (Doctoral thesis).<https://doi.org/10.17863/CAM.70024>
- Abhishek, Tushar Rawat, Daksh Gupta, Manish Varma, Vasudeva. (2021). Transformer Models for Text Coherence Assessment
- Aili Shen, Meladel Mistica, Bahar Salehi, Hang Li, Timothy Baldwin, Jianzhong Qi. Evaluating Document Coherence Modeling. Transactions of the Association for Computational Linguistics 2021; 9 621–640. doi: <https://doi.org/10.1162/tacla00388>