

# Contributions

- Sarthak Agrawal: Ran all the experiments. Includes the baseline models, hyperparameter tuning with different tasks in validation, experiments with different sentence encoders, and the final results.
- Aryan Kharbanda: Reviewed Literature on Sentence Encoders and tried implementing a Neural Sentence Encoder.
- Sarthak Mahajan: Tried using roberta and T5 as sentence encoders. T5 is a generative encoder, and the paper says, "... generative coherence model can be leveraged by our model to benefit from the advantages of both generative and discriminative training." Did literature review to see the various coherence scoring methods.

## Literature Review

### Main

- [A Cross-Domain Transferable Neural Coherence Model](#)
- [Neural Net Models for Open-Domain Discourse Coherence](#)

### Others (mainly for scoring methods)

\*\* : Uses sigmoid for scoring

- <https://aclanthology.org/D14-1218.pdf>
- <https://aclanthology.org/P17-1121/>
- <https://arxiv.org/abs/1909.00349>
- <https://aclanthology.org/D18-1464/> \*\*
- <https://aclanthology.org/2020.coling-main.194/>
- <https://ojs.aaai.org/index.php/AAAI/article/view/12045> \*\*
- <https://aclanthology.org/W16-3407.pdf>
- [https://link.springer.com/chapter/10.1007/978-3-030-01716-3\\_32](https://link.springer.com/chapter/10.1007/978-3-030-01716-3_32) \*\*
- <https://aclanthology.org/P11-1100.pdf>
- <https://aclanthology.org/C14-1089.pdf>
- <https://arxiv.org/abs/2005.10389>
- <https://aclanthology.org/P08-2011.pdf>
- <https://arxiv.org/abs/1804.06898> \*\*
- <https://dl.acm.org/doi/abs/10.1145/3132847.3133047>

## Evaluation

The following methods are known of

- Discrimination
- Insertion
- Paragraph Reconstruction
- Readability Assessment

The first two are used.

## Experiments

### Comparison of baseline results

- Results with the recommended hyperparameters

Output function	Encoder	Discrimination	Insertion
None	Avg_Glove	0.92537	0.29847
Sigmoid	Avg_Glove	0.80393	0.21469
TanH	Avg_Glove	0.10488	0.72060
None	SBERT	0.93851	0.33034

```
"hparams": {
  "input_dropout": 0.6,
  "hidden_layers": 1,
  "hidden_dropout": 0.3,
  "margin": 5.0,
  "weight_decay": 0.0,
  "dpout_model": 0.0
}
```

# Analysis of Hyperparameter tuning

## Hyperparameters tuned:

- input\_dropout: [0.5, 0.6, 0.7]
- hidden\_layers: [1, 2]
- hidden\_dropout: [0.2, 0.3, 0.4]
- margin: [4.0, 5.0, 6.0]
- weight\_decay: [0.0, 0.1]
- dpout\_model: [0.0, 0.05, 0.1]

## Discrimination validation

Results for best 20 are taken

### Scores

- Discrimination

Description	Value
mean	0.92649
std	0.00107
min	0.925
max	0.9291
25%	0.92585
50%	0.92635
75%	0.9269

- Insertion

Description	Value
mean	0.30562
std	0.00156
min	0.3036
max	0.3091
25%	0.304275
50%	0.3054
75%	0.306825

### Trends

- input\_dropout: 0.5 > 0.6
- hidden\_layers: 2 > 1
- margin: 6.0 > 4.0 > 5.0

- weight\_decay: 0.0
- dpout\_model: 0.0 > 0.1 > 0.05
- hidden\_dropout
  - Discrimination: 0.3 > 0.2 > 0.4
  - Insertion: 0.3 > 0.4 > 0.2

### Best model

- Same model performs best in both tasks

```
"hparams": {
  "input_dropout": 0.5,
  "hidden_layers": 2,
  "hidden_dropout": 0.3,
  "margin": 6.0,
  "weight_decay": 0.0,
  "dpout_model": 0.0
}
```

## Insertion validation

Results for best 20 are taken

### Scores

- Discrimination

Description	Value
mean	0.926410
std	0.001753
min	0.923900
max	0.930300
25%	0.925100
50%	0.925900
75%	0.927425

- Insertion

Description	Value
mean	0.3041
std	0.0029
min	0.301
max	0.3124
25%	0.3024
50%	0.3033
75%	0.305

### Trends

- input\_dropout: 0.5
- hidden\_layers: 1 > 2
- hidden\_dropout: 0.4 > 0.2 > 0.3
- margin: 6.0 > 5.0 > 4.0
- weight\_decay: 0.0

- dpout\_model: 0.05 > 0.1 > 0.0

### Best model

- Same model performs best in both tasks

```
"hparams": {
  "input_dropout": 0.5,
  "hidden_layers": 1,
  "hidden_dropout": 0.4,
  "margin": 6.0,
  "weight_decay": 0.0,
  "dpout_model": 0.05
}
```

## Final Results

### After Hyperparameter tuning

Validation Task	Bidirectional	Output function	Encoder	Discr	Ins	Avg
discrimination	False	None	average_glove	0.9204	0.3028	0.6116
discrimination	False	None	sbert	0.9232	0.3139	0.61855
discrimination	False	tanh	average_glove	0.0952	0.8153	0.45525
discrimination	False	tanh	sbert	0.2989	0.2798	0.28935
discrimination	False	sigmoid	average_glove	0.6148	<b>0.4915</b>	0.55315
discrimination	False	sigmoid	sbert	0.2711	0.3196	0.29535
discrimination	True	None	average_glove	0.9259	0.3091	0.61749
discrimination	True	None	sbert	<b>0.9268</b>	0.313	<b>0.6199</b>
discrimination	True	tanh	average_glove	0.0001	1.0	0.50005
discrimination	True	tanh	sbert	0.807	0.3134	0.5602
discrimination	True	sigmoid	average_glove	0.713	0.3809	0.54695
discrimination	True	sigmoid	sbert	0.8113	0.2484	0.52985
insertion	False	None	average_glove	0.9185	0.2993	0.6089
insertion	False	None	sbert	0.9226	0.3164	0.61949
insertion	False	tanh	average_glove	0.0	1.0	0.5
insertion	False	tanh	sbert	0.5872	<b>0.4079</b>	0.49755
insertion	False	sigmoid	average_glove	0.0581	0.9511	0.50459
insertion	False	sigmoid	sbert	0.6704	0.1794	0.4249
insertion	True	None	average_glove	0.9191	0.2929	0.606
insertion	True	None	sbert	<b>0.9245</b>	0.3239	<b>0.6242</b>
insertion	True	tanh	average_glove	0.7525	0.3163	0.5344
insertion	True	tanh	sbert	0.7852	0.3211	0.55315

Validation Task	Bidirectional	Output function	Encoder	Discr	Ins	Avg
insertion	True	sigmoid	average_glove	0.6717	0.3519	0.5118
insertion	True	sigmoid	sbert	0.8125	0.3122	0.56235