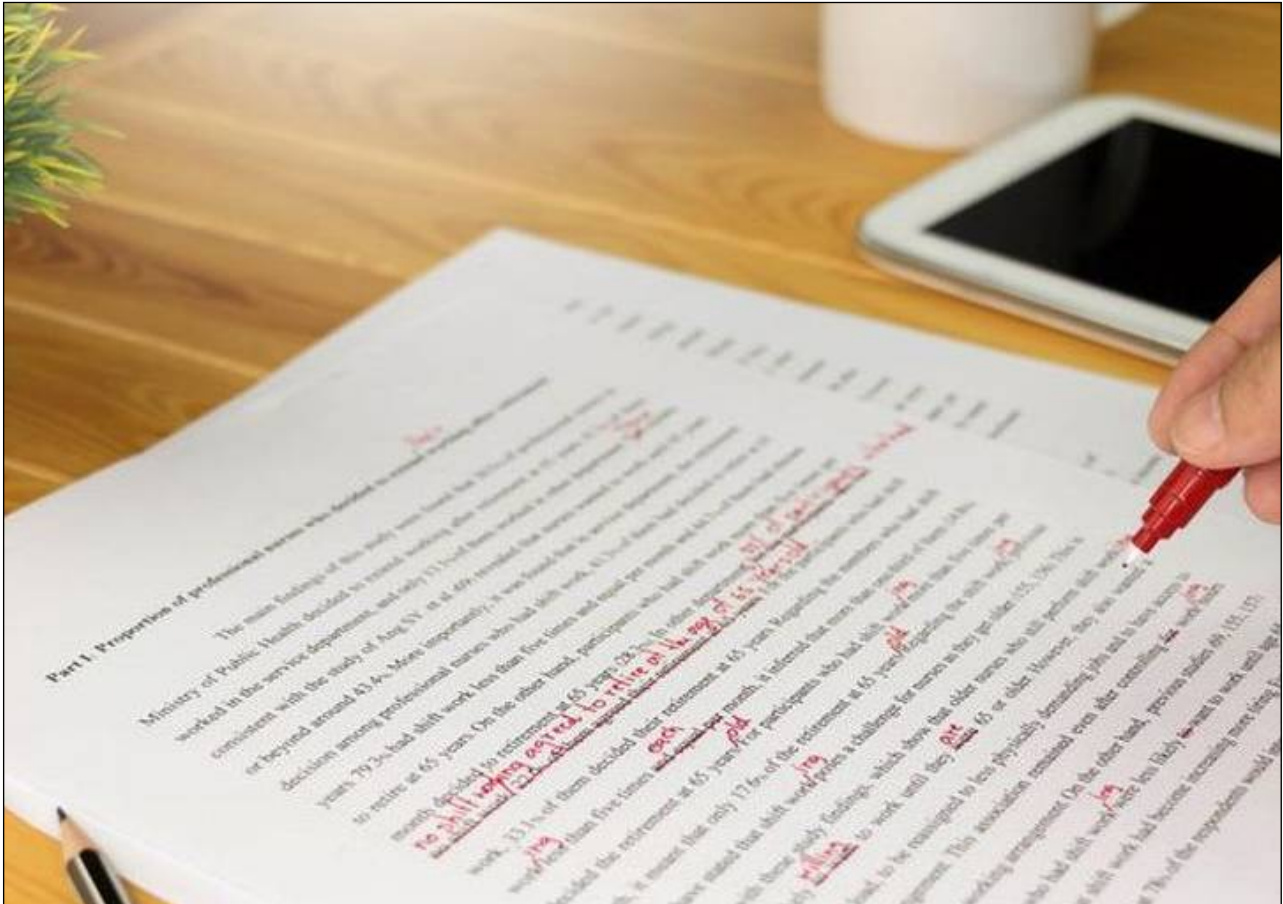

Automatic Text Scoring

IRE Scope Document

Group 23



Team Members:

Akhilesh Soni (201530001)

Sagar Thakur (20172103)

Sai Teja Reddy Moolamalla (201564086)

Suma Reddy Duggenpudi (201525145)

TA Mentor:

Rama Rohit Reddy

Project Statement

We have to implement an end-to-end Automatic Text Scoring (ATS) System, using novel deep neural architectures.

The major component would be the implementation of the paper, "[*SKIPFLOW: Incorporating Neural Coherence Features for End-to-End Automatic Text Scoring*](#)", whose approach has the state of the art performance on ATS tasks on the benchmark ASAP Dataset.

If time permits, we will be experimenting with various modifications to the above neural architecture and report the results.

As the name suggests, an Automatic Text Scoring (ATS) system looks at a piece of text / an essay and gives a score depending on the various features it has learned. Few notable features would be essay length, sentence length, grammar correctness and other lexical, semantic or syntactic features depending on the data. Handcrafting these features is not easy and requires a lot of human-involvement. The aim for the model is to learn the features by itself.

The suggested *SKIPFLOW* model generates neural coherence features and observes semantic relatedness over time, thus generating better scores than it's closely related models.

Dataset

We'll be using the ASAP (Automated Student Assessment Prize) Dataset where we are provided access to hand scored essays, so that we can build, train and test scoring engines.

Prompt	#Essays	Avg Length	Scores
1	1783	350	2-12
2	1800	350	1-6
3	1726	150	0-3
4	1772	150	0-3
5	1805	150	0-4
6	1800	150	0-4
7	1569	250	0-30
8	723	650	0-60

The dataset consists of various types/prompts of essays each with an individual score, with the statistics of data as shown above. The above dataset would be used for training and validation of the model.

Applications

- Due to the neural coherence features used to model the relationships between different parts of the essay, the *SKIPFLOW* model is able to generate more accurate ATS scores when compared to baseline LSTMs, multi-layered and attentional LSTMs.
- The model is able to identify key features for scoring the essays, without any human intervention for identification of features. Such a system is also able to largely outperform those systems with handcrafted features.
- Such ATS systems can be utilized in improving the feedback cycle in educational systems, scoring in high stake assessments like GRE and TOEFL.
- A successful ATS system brings about widespread benefits to society and the education industry.

Challenges

- Implementing the Neural Tensor Layer in the architecture, and tuning the hyperparameters of the model.
- Realizing a modified model which has better QWK (Quadratic Weighted Kappa) metrics than the model suggested in the paper, if time permits.

Second Deliverable / End-to-End System

- Obtain the ASAP dataset and preprocess it, to work on it.
- Finish the implementation of the complete end-to-end system.
- Show the Kappa scoring values for the trained model.
- Compare the results of the paper with the metrics obtained for the built system.

Third Deliverable / Demo

- Try to enhance the accuracy, by training on various slightly modified / tweaked models.
- Build an interactive demo (mostly web-based) for users to input essays and get the predicted ATS score.
- Make a presentation video briefly explaining the work done.
- Make a formal report of the project.

Tools

1. **Keras + Tensorflow** - for implementing the model and training the model using the ASAP dataset.
2. **NLTK Toolkit** - for preprocessing the data.

References

- *SKIPFLOW: Incorporating Neural Coherence Features for End-to-End Automatic Text Scoring - The Reference Paper*
- *Automatic Student Assessment Prize (ASAP) Dataset from Kaggle*