Automatic Text Scoring

Akhilesh Soni (201530001)
Sagar Thakur (20172103)
Sai Teja Reddy (201564086)
Suma Reddy Duggenpudi (201525145)

Mentor : Rama Rohit Reddy

Introduction

Problem Statement : Implement an end-to-end Automatic Text Scoring System(ATS) using novel deep neural architectures.

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

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.

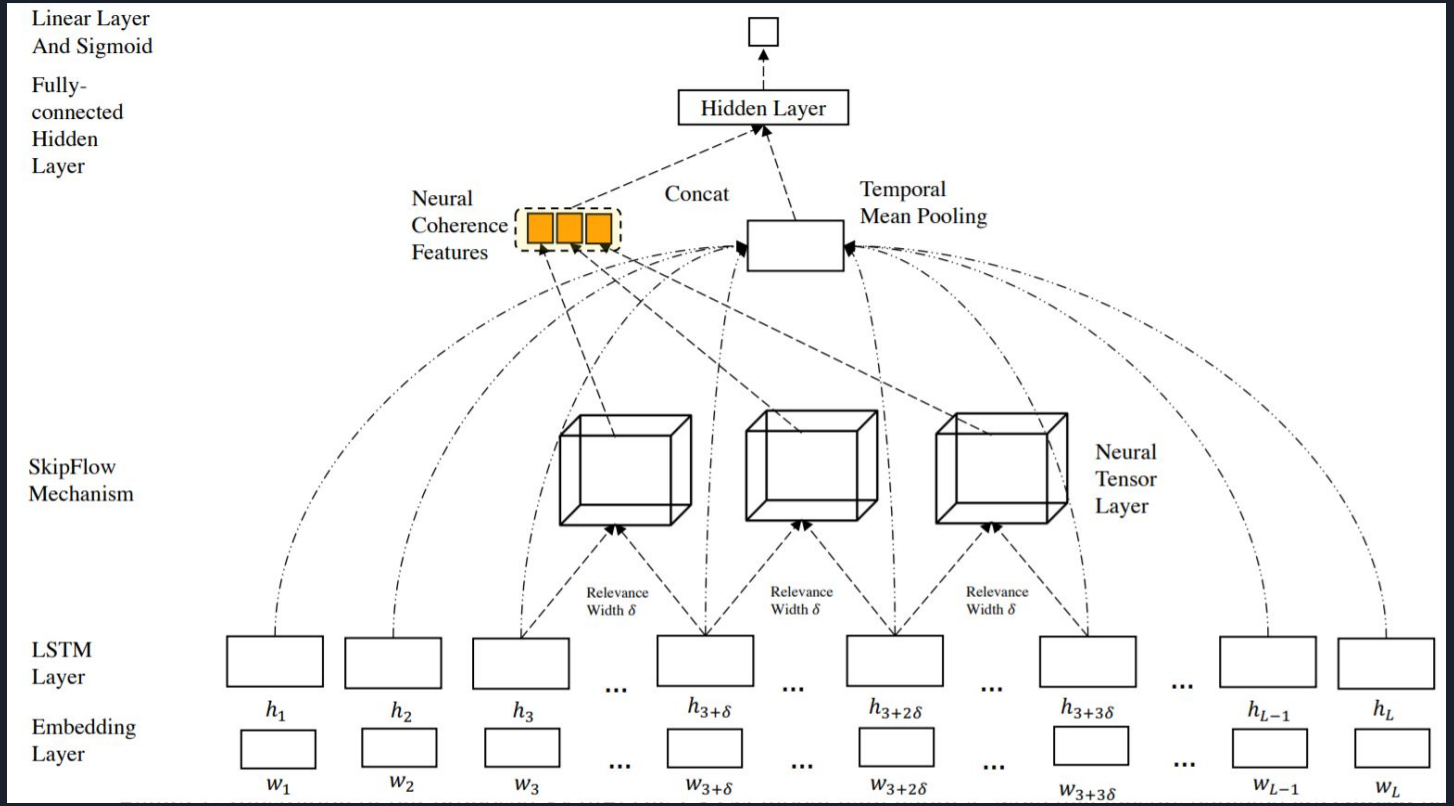
Dataset Used

We used ASAP(Automated Student Assessment Prize) dataset where we are provided access to hand scored essays , so that we can build , train and test scoring systems .

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

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

Architecture : Skipflow LSTM Model



Skipflow LSTM Model

Embedding Layer: The parameters of the embedding layer are defined as $W_e \in R^{|V| \times N}$ where |V | is the size of the vocabulary and N is the dimensionality of the word embeddings.

Long Short-Term Memory (LSTM): The sequence of word embeddings obtained from the embedding layer is then passed into a long short-term memory network.

Neural Tensor Layer: A tensor layer is used to model the relationship between two LSTM outputs. It is a parameterized composition which is defined using :

$$s_i(a, b) = \sigma(u^T f(v_a^T M^{[1:k]} v_b + V[v_a, v_b] + b))$$

The vector outputs of LSTM at two time steps of δ-width apart are passed through neural tensor layer and it returns a similarity score that determines the coherence feature between the two vectors.

Skipflow LSTM Model(Cont.)

Fully-Connected Hidden Layer

Linear Layer with Sigmoid

Tools Used:

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

Evaluation Metric

The experimental setup is a 5 fold cross validation to evaluate all systems with a 60-20-20 split for train , validation and test sets. We trained all models for 1000 epochs or lesser depending on whether the model keeps improving its performance . For word processing , we used tokenizer of NLTK library. The evaluation metric used was the **Quadratic Weighted Kappa (QWK)** which measures agreement b/w raters and it is a commonly used metric for ATS systems.

Observations

The Kappa scores obtained for each essay prompt are as follows :

Type of Essay	SKIP FLOW LSTM Architecture QWK Score
1	80.10
2	66.67
3	70.84
4	85.71
5	80.88
6	78.83
7	76.47
8	68.81
Average	75.91

Conclusion

This system serves the purpose of score prediction for a given essay . This has been mainly done by incorporating the intuition of textual coherence in neural ATS system. SKIPFLOW architecture adopts parameterized tensor compositions to model the relationships b/w different points within an essay , generating neural coherence features that can support predictions . These neural coherence features when combined with LSTM sentence representations can produce significantly better results.

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.

Goals Achieved

- Finished the implementation of the complete end-to-end system.
- Compared the results of the paper with the metrics obtained for the built system.
- Built an interactive demo (web-based) for users to input essays and get the predicted ATS score.

Demo

References

- SKIPFLOW: Incorporating Neural Coherence Features for End-to-End Automatic Text Scoring Paper .
- Automatic Student Assessment Prize (ASAP) Dataset from Kaggle
- Automatic Text Scoring Using Neural Networks

Thank You