

HinglishEval Presentation

Introduction

- Code-mixing is the phenomenon of mixing words and phrases from multiple languages in a single utterance of a text or speech.
- A significant challenge to research is that there are no formal sources like books or news articles in code-mixed languages, and studies have to rely on sources like Twitter or messaging platforms.
- With no baseline set, it is widely dependent on people fluent in both languages to produce and rank the data that is available/ produced.
- With this task, the main objective is to propose and develop new strategies that cater to the broad requirement of the quality evaluation of the generated code-mixed text.

Dataset

- English-Hindi sentence pairs and corresponding synthetic and human generated Hinglish sentences.
- 2766 Synthetic and 6694 Human sentences in training data
- 791 Synthetic sentences in test Data
- Each having corresponding Average Rating and Disagreement Scores

English	Hindi	Human-generated Hinglish	WAC	PAC
The reward of goodness shall be nothing but goodness.	अच्छाई का बदला अच्छाई के सिवा और क्या हो सकता है?	The reward of achai shall be nothing but achai.	reward ka badla reward ke nothing aur kya ho sakta hai Rating1: 7 Rating2: 4	reward of goodness goodness ke siva aur kya ho sakta hai Rating1: 9 Rating2: 7
		Goodness ka badla goodness ke siva aur kya ho sakta hai.		
		Achai ka badla shall be nothing but achai.		

Problem Statement

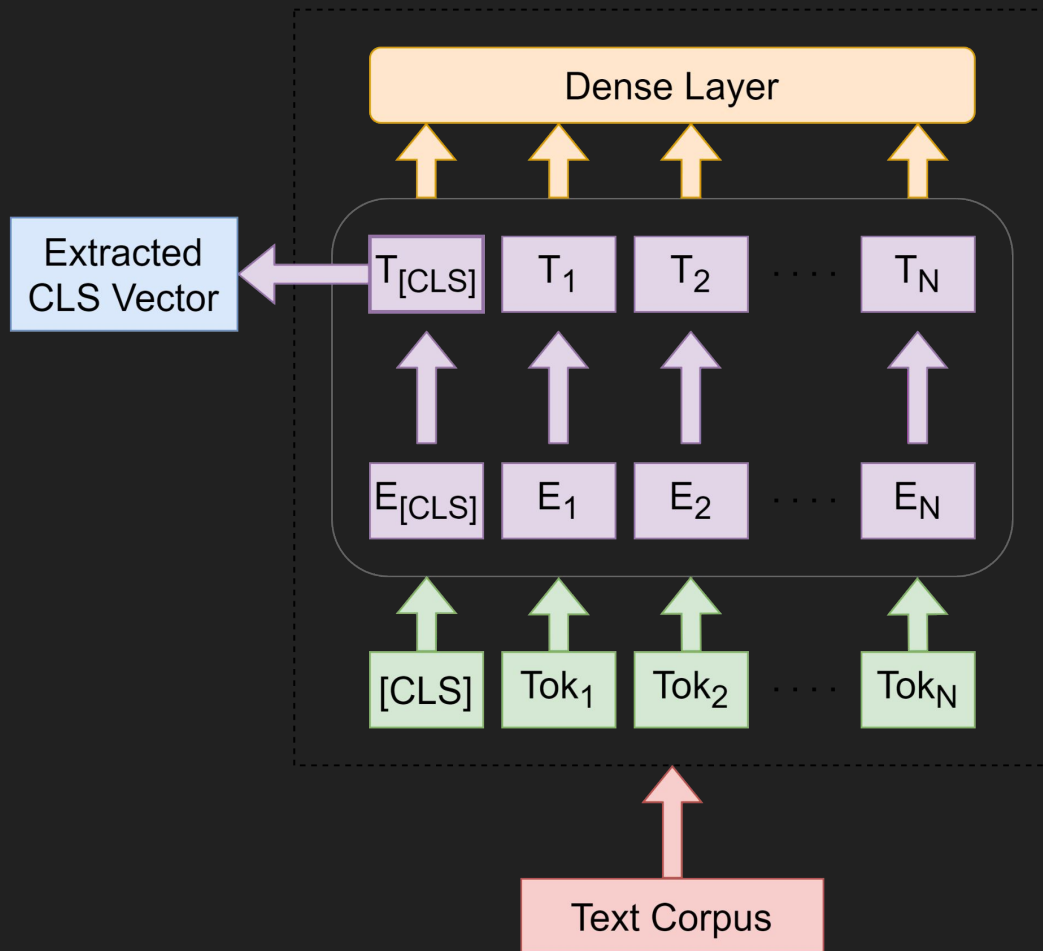
The goal of this project was to create a model that could accurately predict the following two things:

- The average quality rating given to a synthetically generated Hinglish sentence by two different annotators
- The disagreement between the scores given to each sentence by the two annotators

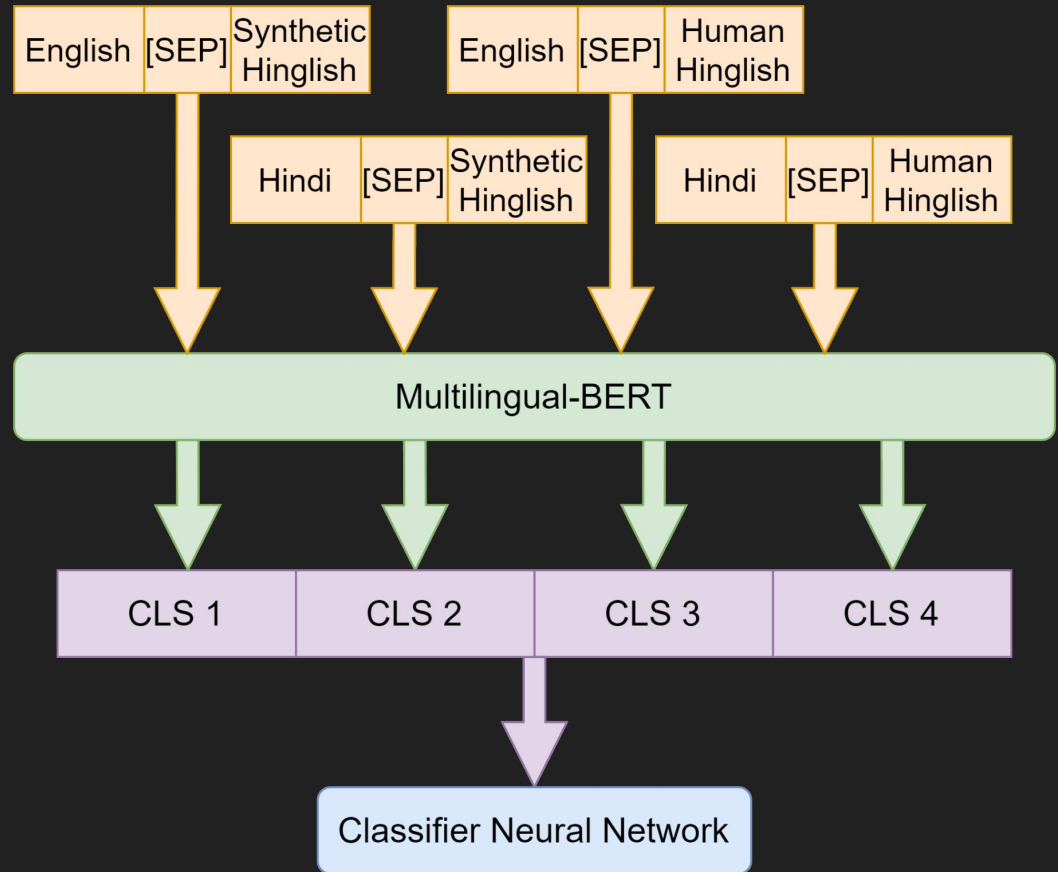
These are sub-task 1 and 2 respectively

Proposed Model

- **Multilingual-BERT:** We use a modified pretrained BERT encoder called bert-base-multilingual-cased. It was pretrained in a self-supervised fashion on the top 104 languages having the largest Wikipedias using the objective of masked language modelling. We extracted the CLS vectors for different sentence pairs from BERT.



- **Classifier Neural Network:** We trained two fully connected neural networks, one for each sub-task, using these concatenated vector sets as their respective inputs



Evaluation

Sub-Task 1 (Average Rating)			Sub-Task 2 (Disagreement)	
F1 Score	Cohen's Kappa	MSE	F1 Score	MSE
0.21796	0.07337	3.00000	0.24252	4.00000

Conclusion

- In this paper, we utilised a twofold technique to solve a text classification problem. This method involved using Multilingual BERT over pairs of different sentence types to extract deep semantic text features in the form of CLS vectors for each of them. The next step was to appropriately combine these vectors and run them through a fully connected classifier neural network.
- Upon training over 2766 synthetic sentences and testing over 791 synthetic sentences we achieved F1 scores of 0.218 on Rating classification and 0.242 on Disagreement classification.
- With improved fine tuning and a larger dataset these scores could significantly increase.

Thank You