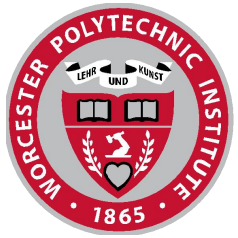
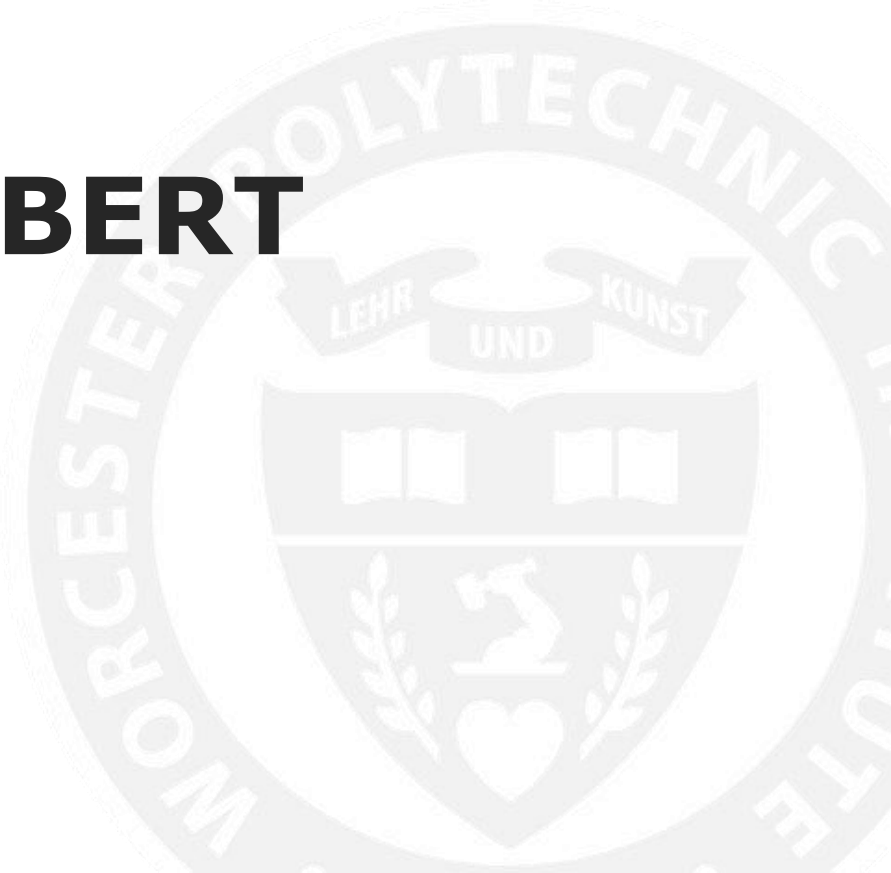


WPI

Question Matching using BERT

Ronak Sankaranarayanan, Siddhi Patil, Sitanshu Rupani



Question Matching

- What is Question Matching?
  - Identification of similar questions is known as Question Matching

- Need for Question Matching
  - Match answers to different sounding questions
  - Beneficial to users with high quality human-generated answers for solving their problems

The Dataset

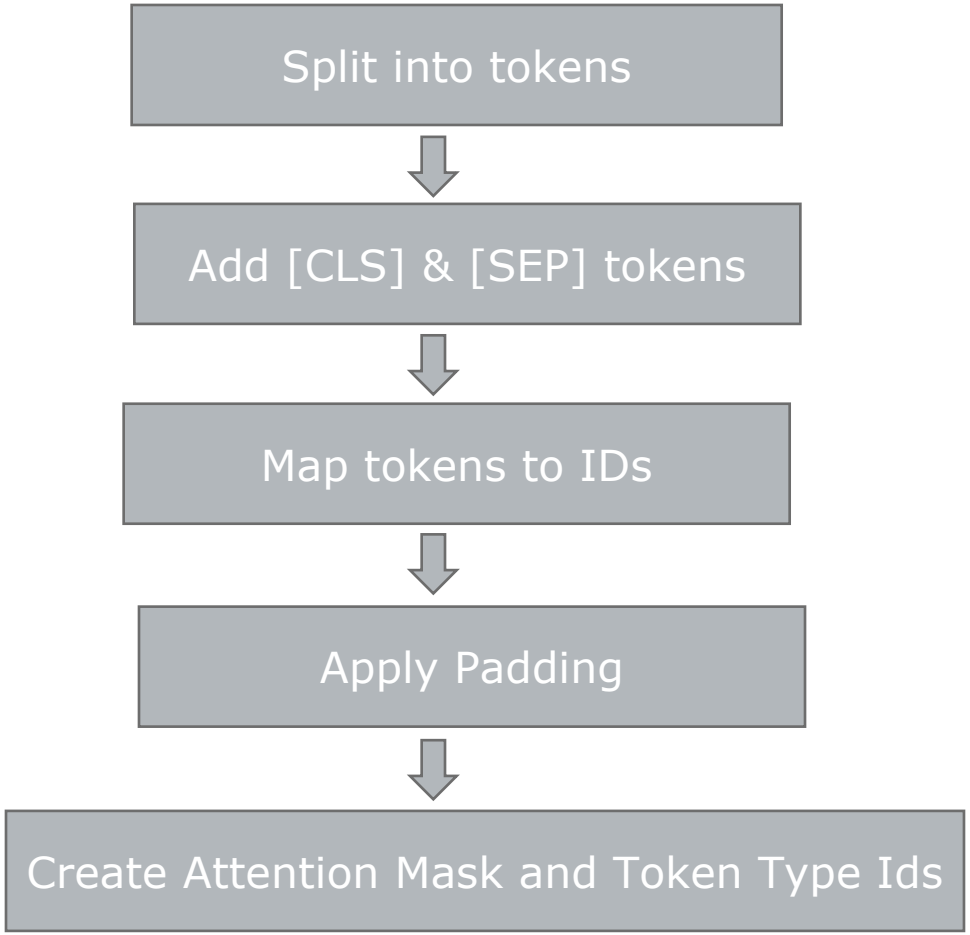
- Dataset taken from Kaggle
  - https://www.kaggle.com/c/quora-question-pairs/data
- 404,290 question pairs in the dataset

Variable Name	Description
qid1	Question ID 1
qid2	Question ID 2
question1	Full text of Question 1
question2	Full text of Question 2
is_duplicate	Ground truth, indicator if both questions are similar

Dataset & Pre-Processing

- Around 37% duplicate questions in dataset
- Random Sample of 40000 question pairs
- Remove rows with empty question texts
- Replaced contractions such as ‘what’s’ with ‘what is’, ‘can’t’ with ‘can not’, etc.
- Removed special characters
- Split Dataset into Train, Validation & Test Sets (80-20-20)

Pre-Processing Steps



question1: How can I recover old gmail account?

question2: How can I access my old gmail account?

tokens: [CLS] how can i recover old gmail account ?
[SEP] how can i access my old gmail account ? [SEP]

input_ids: 101 2129 2064 1045 8980 2214 20917 4014 4070 1029
102 2129 2064 1045 3229 2026 2214 20917 4014 4070 1029 102

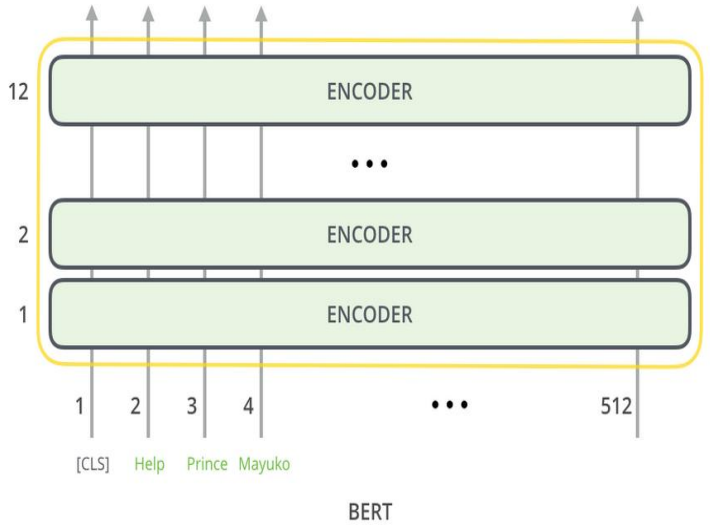
attention_mask: 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

token type_ids: 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1

is_duplicate: 1

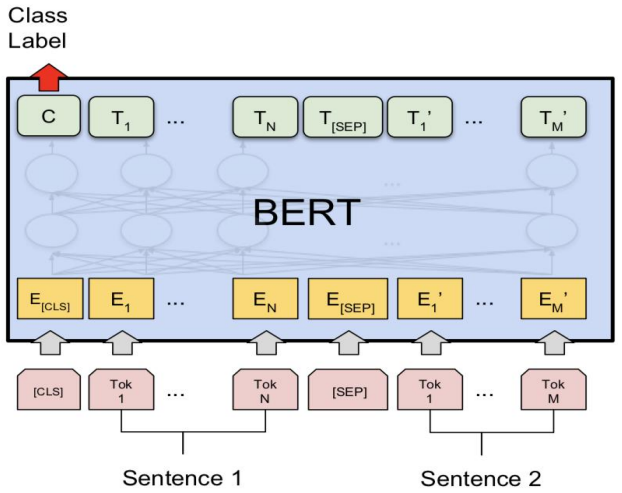
BERT

- Bidirectional Encoder Representations from Transformers (BERT) released and pre-trained by Google for NLP tasks in late 2018
- 12 Transformer Encoding layers or 24 for large BERT
- Pre-trained using plain text corpus
- Takes a sequence of words as input
- Each layer applies self-attention and passes results through a feed-forward network to the next encoder



BERT Model

- Used Bert-Base-Uncased model :
  - 12 layers, 12 attention heads, 110 million parameters
- Fine-tuned for the classification task :
  - Added softmax last layer on top of the BERT model to classify duplicate questions
- With input data, the pre-trained BERT model and the additional untrained classification layer is trained using gra



Hyper parameter Tuning

- Number of layers
- Activation functions
- Number of units
- Drop out rate
- Number of attention heads
- Learning rate
- Epochs

Results

Model	Validation accuracy	Testing accuracy	Precision	Recall	F1 score
Relu activation	72	71	0.75	0.76	0.71
Gelu activation	73.60	73.90	0.76	0.77	0.74
7 Layers	79%	77%	0.77	0.78	0.77
12 Layers	80%	80.8%	0.8	0.82	0.8
24 Attention heads	78.9%	78.25%	0.77	0.78	0.77

Demo

Thank You