Name : Souleymane Sankara #300100940

Part 2 Report: **Evaluation of pre-trained sentence embedding models**

1. Sentence Embedding models
   1. All-MiniLM-L6-v2

* Maps a sentence or short paragraph to 384-dimensional dense vector
* Built with 6 transformers layers
* Based on Bert architecture
  1. All-MPNet-Base-V2
* Maps sentences and paragraphs into 768-deimensional dense vector
* Fined-tuned from pre-trained Microsoft MPNet-Base model
* Built with 12 layers
  1. T5-Large
* A variant of the T5 model family, which include T5-small, T5-medium and T5-large
* Used 24 layers
* Maps sentences and paragraphs into 1024-dimensional vector
  1. MSMARCO-RoBERTa-Base-v2
* Produce embedding in a 768-dimensional dense vector space
  1. All-DistilRoBERTa-v1
* Maps text into 768-dimensional vector space

1. Performance Evaluation

To compute similarity, we used cosine similarity as the metric. The results were then scaled using the formula:  
This approach allowed us to derive a performance score ranging from 0 to 5

1. Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Datasets | SE1 | SE2 | SE3 | SE4 | SE5 | Best score |
| STS2016.gs.answer-answer.txt | 0.725 | 0.748 | 0.840 | 0.719 | 0.732 | 0.840 |
| STS2016.gs.headlines.txt | 0.796 | 0.843 | 0.837 | 0.782 | 0.828 | 0.843 |
| STS2016.gs.plagiarism.txt | 0.829 | 0.823 | 0.864 | 0.804 | 0.849 | 0.864 |
| STS2016.gs.postediting.txt | 0.857 | 0.880 | 0.880 | 0.834 | 0.885 | 0.885 |
| STS2016.gs.question-question.txt | 0.805 | 0.821 | 0.805 | 0.784 | 0.794 | 0.821 |

1. Discussion

T5-Large (SE3), a larger model, gives strong performance achieving then the best score in 2 of the 5 trials. For the answer-answer dataset it outperforms all other models by significant margin. Although, we can notice that All-MPNet-Base-v2 (SE2) which is a smaller model gives competitive performance with also the best score in 2 cases out of the 5. This means that SE2 can be an alternative, especially in scenarios requiring smaller computational characteristics.

SE1 ((All-MiniLM-L6-v2) and SE5 (All-DistilRoBERTa-v1) shows promising results despite their smaller architecture. They offer a good trade-off between accuracy and efficiency