BioBERT

(Bidirectional Encoder Representations from Transformers for Biomedical Text Mining)

BERT

- Fully trained language model that Google released few months ago
- BERT uses two training strategies:

  - **Masked LM (MLM) :** Before feeding word sequences into BERT, 15% of the words in each sequence are replaced with a [MASK] token. The model then attempts to predict the original value of the masked words, based on the context provided by the other, non-masked, words in the sequence
  - **Next Sentence Prediction (NSP) :** In the BERT training process, the model receives pairs of sentences as input and learns to predict if the second sentence in the pair is the subsequent sentence in the original document.

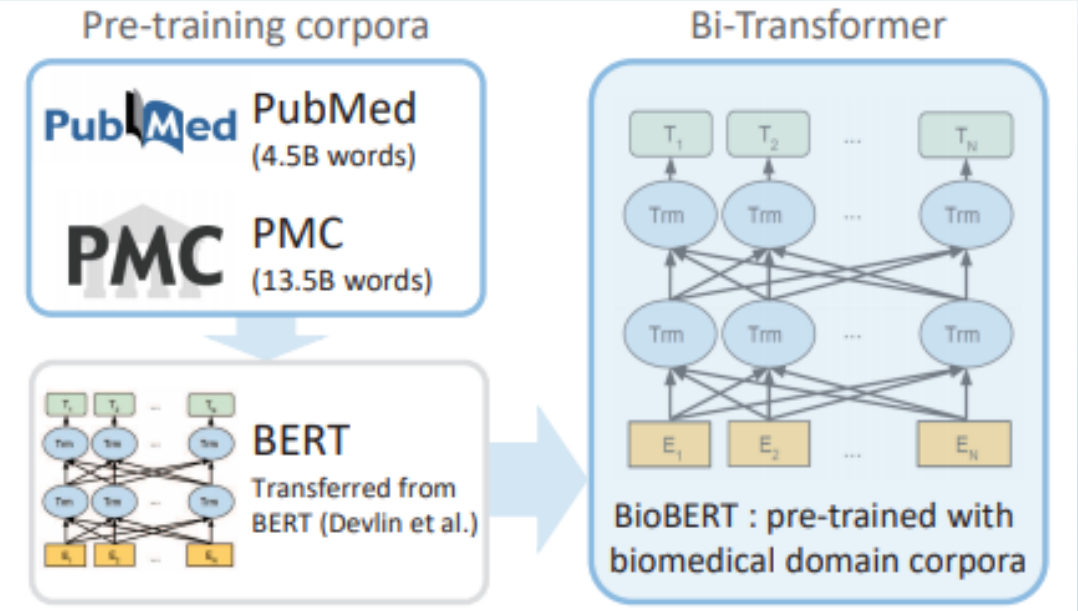
BioBert :

- A pre-trained biomedical language representation model for biomedical text mining
- A domain specific language representation model pre-trained on large-scale biomedical corpora
- BioBERT significantly outperforms BERT on the following three representative biomedical text mining tasks:

1. Biomedical question answering (9.61% absolute improvement)
2. Biomedical relation extraction (3.49% absolute improvement)
3. Biomedical named entity recognition (0.51% absolute improvement)

Data

- BERT's original training data which includes English Wikipedia and BooksCorpus
- Domain specific data which are PubMed abstracts and PMC full text articles



BioBERT Architecture


Input Data
1. Symptoms
2. Medications
3. Side Effects
BioBERT
Output Data
1.Diagnosis
2.Treatment Plan

Health Assistant

My head hurts after listening to music loudly for 3 hours

Any other symptoms or medications you take

No, just a slight pain in my forehead

You have migraine, rest, take aspirin & drink water

THANK YOU