Pre-Training Tasks for BERT

Pre-Training Tasks

- **Pre-training** refers to training a model in one task to help it learn a set of parameters that can be used in other tasks

Pre-Training Tasks

- **Pre-training** refers to training a model in one task to help it learn a set of parameters that can be used in other tasks

- Pre-training tasks for BERT are based on self-supervised learning

Pre-Training Tasks

- **Pre-training** refers to training a model in one task to help it learn a set of parameters that can be used in other tasks
- Pre-training tasks for BERT are based on self-supervised learning
- BERT Pre-training tasks:-
  - Masked Language Modeling (MLM)
  - Next Sentence Prediction (NSP)

Pre-Training Tasks

- **Pre-training** refers to training a model in one task to help it learn a set of parameters that can be used in other tasks
- Pre-training tasks for BERT are based on self-supervised learning
- BERT Pre-training tasks:-
  - Masked Language Modeling (MLM)
  - Next Sentence Prediction (NSP)
- BERT is trained on both the tasks simultaneously

Masked Language Modeling (MLM)

- Input Sequence = “how are [MASK] doing?”

  Target Sequence = “how are you doing?”

Analytics Vidhya

Masked Language Modeling (MLM)

- Input Sequence = “how are [MASK] doing?”

  Target Sequence = “how are you doing?”

- 12% of the tokens are replaced randomly with the mask ([MASK]) token

Next Sentence Prediction (NSP)

- A sentence pair dataset is formed

Next Sentence Prediction (NSP)

- A sentence pair dataset is formed
- For 50% of these pairs, the second sentence would actually be the next sentence to the first sentence

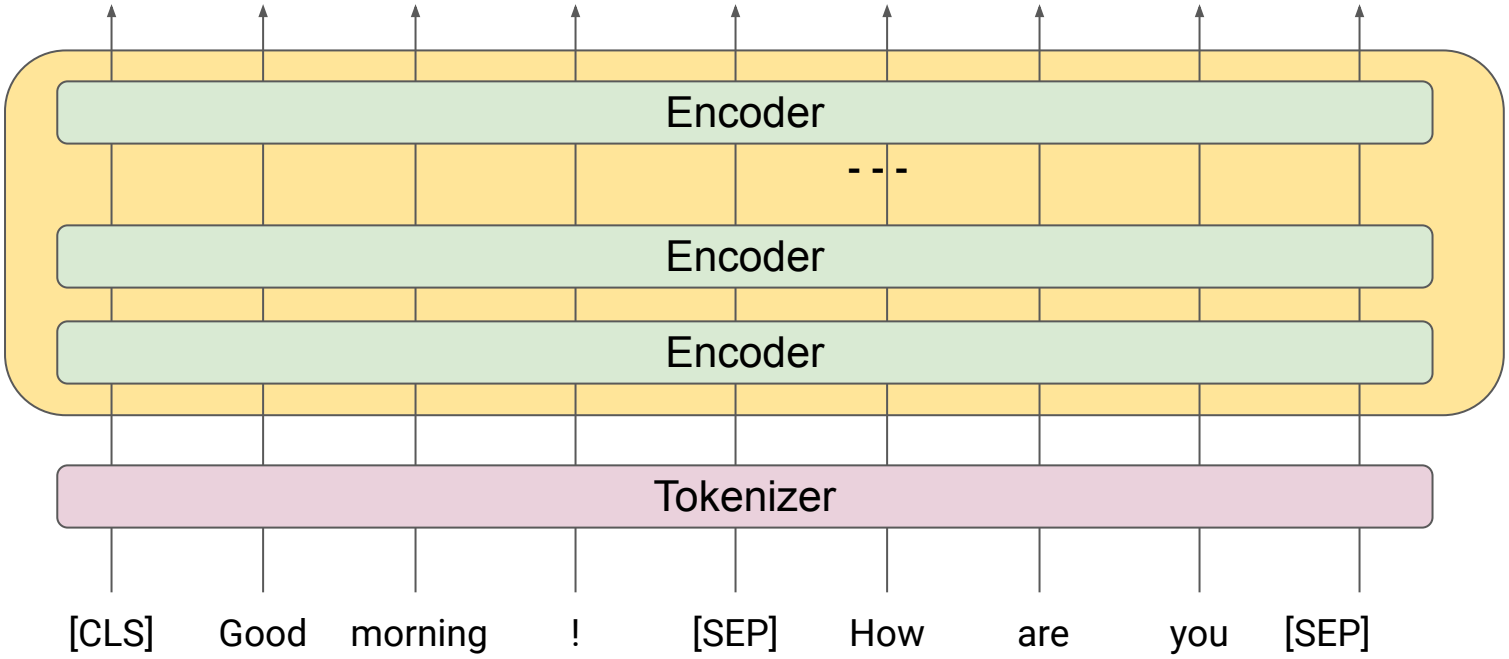
Next Sentence Prediction (NSP)

- A sentence pair dataset is formed
- For 50% of these pairs, the second sentence would actually be the next sentence to the first sentence
- For the remaining 50% of the pairs, the second sentence would be a random sentence from the corpus

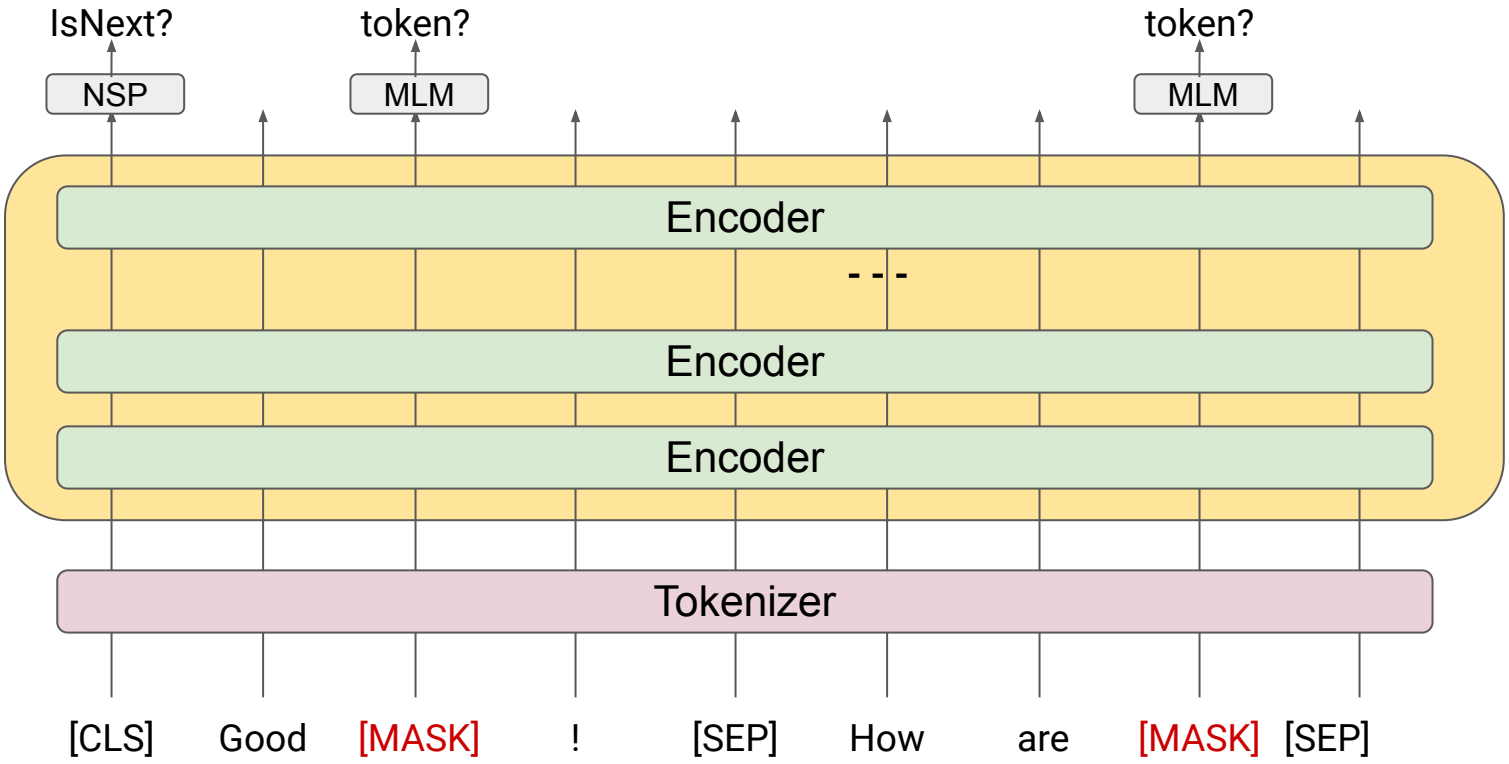
Next Sentence Prediction (NSP)

- A sentence pair dataset is formed
- For 50% of these pairs, the second sentence would actually be the next sentence to the first sentence
- For the remaining 50% of the pairs, the second sentence would be a random sentence from the corpus
- It is a binary classification problem. The labels for the first case would be ‘IsNext’ and ‘NotNext’ for the second case

BERT Pre-Training Tasks



BERT Pre-Training Tasks



Analytics Vidhya

Thank You

Analytics
Vidhya