BERT:
**B**idirectional
**E**ncoder
**R**epresentations
from
**T**ransformers

Baseado em transformers

Possui apenas a parte do Encode, pois o objetivo é gerar uma representação da linguagem

Bidirecional: capaz de ler o texto em ambas as direções

# BERT

- A representação do embedding para cada palavra é contextualizada
- O contexto é obtido buscando a relação entre as palavras utilizando multi-head attention

# Variações: 24

| | Camadas de Encoders | Unidades na Camada Oculta |
|---|---|---|
| Tiny | 2 | 128 |
| Mini | 4 | 256 |
| Small | 4 | 512 |
| Medium | 8 | 512 |
| Base | 12 | 768 |
| Large | 24 | 1024 |

# Pré treinamento

Treinamento não supervisionado

Bert em inglês, utiliza Toronto BookCorpus e Wikipidia

4 dias utilizando 16 TPUs

# Pré treinamento

- Modelos pre-treinados estão disponíveis (não é necessário anotação)
- Como o treino é realizado?
  - Masked language modeling
  - Next sentence prediction
- Ambos são utilizados

- Para uma tarefa, usamos o modelo pré-treinado
- Pode-se fazer ajuste fine (ajuste de pesos – custo baixo)

# Masked language modeling

- 15% dos tokens são mascarados, recebendo [MASK]
- O modelo tenta prever o token mascarados através de um processo de treinamento (ajuste de pesos) baseado nas palavras não mascaradas
- O modelo retorna uma lista de previsões com probabilidades

# Next sentence prediction

- São utilizadas pares de sentenças
- O objetivo é prever se a segunda sentença é continuação da primeira
- Problema de classificação binário: IsNext, NotNext

# BERT

- Texto deve ser convertido em 3 embedding layers
  - Token
  - Segment
  - Position

# Tokenizer

- Se a palavra não estiver presente no vocabulário, é dividida n vezes até que seja encontrada no vocabulário
- Palavras divididas são sinalizadas com ##

# Token Embedding

- Tokenização
- Token [CLS] no inicio da primeira sentença
- Token [SEP] no final de todas as sentenças

| Token | [CLS] | Data | Science | Is | Cool | [CLS] | But | I | Am | afraid | [SEP] |
|---|---|---|---|---|---|---|---|---|---|---|---|

# Segment Embedding

- Separa as sentenças em diferentes segmentos

| | [CLS] | Data | Science | Is | Cool | [CLS] | But | I | Am | afraid | [SEP] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Segment | A | A | A | A | A | A | B | B | B | B | B |

# Position Embedding

- Usado para informar posição, uma vez que o processamento é em paralelo

| | [CLS] | Data | Science | Is | Cool | [CLS] | But | I | Am | afraid | [SEP] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Segment | A | A | A | A | A | A | B | B | B | B | B |
| Position | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |

# Bert em Português

- Bert padrão é um modelo em inglês
- Alternativas?
  - Multilingual Bert
    - Treinando utilizando a Wikipedia em mais de 100 idiomas
    - Open Source
    - Cased e Uncased
  - BERTimbau
    - Base e Large