Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Transformers
BERT and GPT examples

Learning Objectives

In this lesson, you will learn about two well known transformers-based models => BERT and GPT

- Encoder-based model
- Decoder-based model

BERT

Bidirectional Encoder Representation from Transformers

- Introduced to improve the previous uni-directional language models (standard LSTM or RNN); models that are trained on sentences using only one direction (left-to-right or right-to-left). Uni-directional models are highly limited in processing the input to make predictions
- Bi-directional models use information from both the past and future for better prediction, while uni-directional models only use past information
- Good for tasks that need an understanding of Language. E.g., Natural Machine Translation, Question answering, sentiment analysis, text summarization, etc.
- Created by stacking up encoders
- Trained on BooksCorpus (800M words) and English Wikipedia (2,500M words)
- Using two unsupervised tasks for training:
  - Masked Language Modeling (MLM)
  - Next Sentence Prediction (NSP)

BERT

Bidirectional Encoder Representation from Transformers

- Introduced to improve the previous uni-directional language models (standard LSTM or RNN); models that are trained on sentences using only one direction (left-to-right or right-to-left). Uni-directional models are highly limited in processing the input to make predictions
- Bi-directional models use information from both the past and future for better prediction, while uni-directional models only use past information
- Good for tasks that need an understanding of Language. E.g., Natural Machine Translation, Question answering, sentiment analysis, text summarization, etc.
- Created by stacking up encoders
- Trained on BooksCorpus (800M words) and English Wikipedia (2,500M words)
- Using two unsupervised tasks for training:
  - Masked Language Modeling (MLM)
  - Next Sentence Prediction (NSP)

Masked Language Modeling (MLM)

- Masking out words in the input and training the model to predict the masked words
  - **Input**: The man went to the [MASK_1]. He bought a [MASK_2] of milk
  - **Labels**: [MASK_1] = store, [MASK_2] = gallon

- BERT is a bidirectional model, i.e., to predict a masked word, the wo rds after and before the masked word are considered

Next Sentence Prediction (NSP)

- Giving the model two sentences and training the model to learn the order of the sentences
- Important because in many NLP tasks such as question answering and Nature Language Inference (NLI), understanding the order of the sentences is crucial
- Given two sentences, A and B, does B come after A?
  - E.g.
    - Sentence A: The man went to the store.
    - Sentence B: He bought a gallon of milk.
    - Label: IsNextSentence

    - Sentence A: The man went to the store.
    - Sentence B: Penguins are flightless.
    - Label: NotNextSentence

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

Example

- BERT model (using Hugging Face Transformer library):
  - fill-mask: filling the blank ([MASK]) with an appropriate word
  - Bert-base-uncased: BERT model, pretrained on English language using a masked language modeling (MLM) objective
  - top_k: number of outputs
  - token: The predicted token id
  - Token_str: The predicted token

```
from transformers import pipeline


unmasker = pipeline("fill-mask", model="bert-
base-uncased")

unmasker("This course will teach you all
about [MASK] models.", top_k=2)
```

```
>> [
{'score': 0.196198508143425,
 'sequence': 'This course will teach you all
about mathematical models.',
 'token': 30412,
 'token_str': ' mathematical'},

{'score': 0.040527332574129105,
 'sequence': 'This course will teach you all
about computational models.',
 'token': 38163,
 'token_str': ' computational'}
]
```

GPT

Generative Pre-Training

- It is a decoder-based model
- It has different versions GPT-1, GPT-2, GPT-3 as of 2023
- It is used to generate human-like text
- It is an auto-regressive model

Example

```
from transformers import pipeline, set_seed

generator = pipeline('text-generation', model='gpt2')
set_seed(42)

generator("Hello, I'm a language model,", max_length=30,
num_return_sequences=5)
```

```
[{'generated_text': "Hello, I'm a language model, a language for thinking, a
language for expressing thoughts."},
 {'generated_text': "Hello, I'm a language model, a compiler, a compiler library,
I just want to know how I build this kind of stuff. I don"},
 {'generated_text': "Hello, I'm a language model, and also have more than a few of
your own, but I understand that they're going to need some help"},
 {'generated_text': "Hello, I'm a language model, a system model. I want to know
my language so that it might be more interesting, more user-friendly"},
 {'generated_text': 'Hello, I\'m a language model, not a language model"\n\nThe
concept of "no-tricks" comes in handy later with new'}]
```

https://huggingface.co/gpt2

Summary

We learned about two models:

- BERT
- GPT