

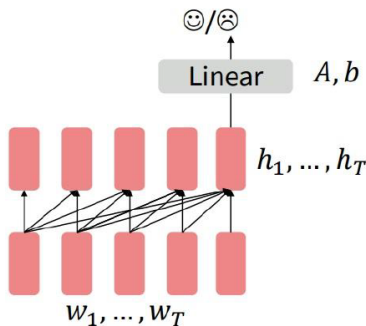
Plan of the Tutorial

- 1 Plan of the Tutorial
- 2 Introduction to NLP
- 3 Overview of Distributional Representation Learning for NLP
- 4 Overview of Transformer based Language Model
- 5 Overview of Large Language Models**
- 6 Concept of in-context learning and its application
- 7 Conclusion

Pretraining decoders

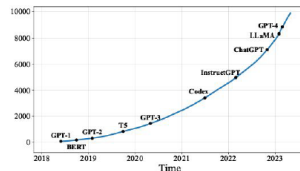
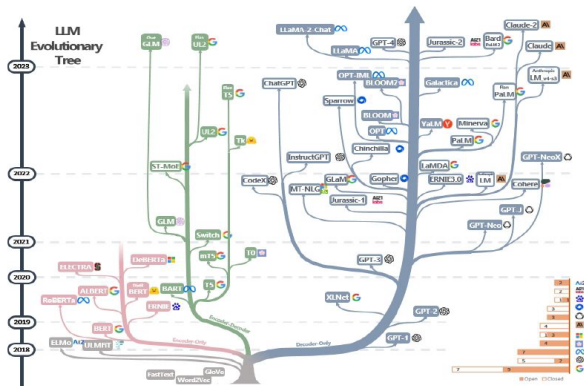
- We can finetune them by training a classifier on the last word's hidden state.

$$h_1, \dots, h_T = \text{Decoder}(w_1, \dots, w_T)$$
$$y \sim Ah_T + b$$

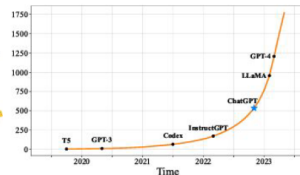


Where A and b are randomly initialized and specified by the downstream task.

LLM Evolution Tree



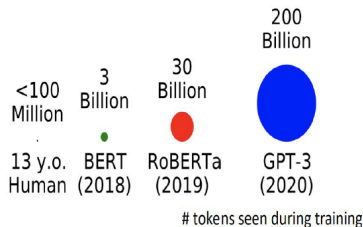
(a) Query="Language Model"



(b) Query="Large Language Model"

Bigger the Better

- GPT (117M parameters): Transformer decoder with 12 layers. Trained on BooksCorpus: over 7000 unique books (4.6GB text).
- Showed that language modeling at scale can be an effective pretraining technique for downstream tasks like natural language inference.
- GPT-2 (1.5B parameters): Same architecture as GPT, just bigger (117M \rightarrow 1.5B). Trained on much more data: 4GB \rightarrow 40GB of internet text data
- GPT-3 (175B parameters): Same architecture as GPT, just bigger (1.5B \rightarrow 175B). Trained on much more data: (40GB \rightarrow over 600GB) of internet text data.



Prompt → Predict

GPT3 shows an unique ability!

- Perform the task by generating natural language tokens, aka, prompting

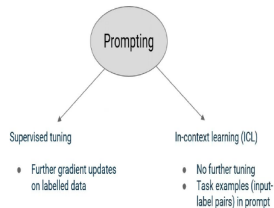
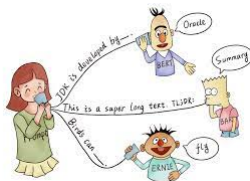
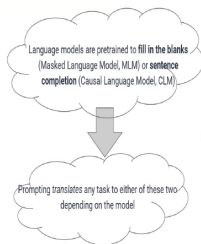
That movie was great. Sentiment: Positive

It was a horrible day! Sentiment: Negative

This is an absolute mess. Sentiment:



Prompting



Example

Language models may do rudimentary modeling of agents, beliefs, and actions:

SP

You

Pat watches demonstration of a bowling ball and a leaf being dropped at the same time in a vacuum chamber. If Pat is a Physicist, what he will predict?

< 1 / 2 >



ChatGPT

As a physicist, Pat would predict that in a vacuum chamber where air resistance is eliminated, both the bowling ball and the leaf would fall at the same rate, accelerating equally due to gravity. This concept is known as the equivalence principle, famously demonstrated by Galileo and later confirmed by Newton's laws of motion.

Example



You

Pat watches demonstration of a bowling ball and a leaf being dropped at the same time in a vacuum chamber. If Pat who has never seen this demonstration before, what he will predict?

< 2 / 2 >



ChatGPT

If Pat has never seen the demonstration before and is unaware of the concept of the equivalence principle (which states that in a vacuum, all objects fall at the same rate regardless of their mass), they might predict that the bowling ball would fall faster than the leaf due to its larger size or mass. This prediction is based on everyday observations where heavier objects often seem to fall faster than lighter ones. However, in a vacuum chamber where air resistance is eliminated, both the bowling ball and the leaf would fall at the same rate, defying Pat's initial prediction.



Example .. Code



You

Write a code in Python to arrange a sequence of numbers in ascending order.

< 2 / 2 >



ChatGPT

Certainly! You can use the `sorted()` function in Python to arrange a sequence of numbers in ascending order. Here's an example:

python

Copy code

```
# Define a sequence of numbers
numbers = [9, 4, 7, 2, 11, 5]

# Arrange the sequence in ascending order
ascending_order = sorted(numbers)

# Display the sequence in ascending order
print("Numbers in ascending order:", ascending_order)
```

Replace the `'numbers'` list with your desired sequence of numbers. When you run this code, it will output the sequence arranged in ascending order.

Example .. mathematical reasoning



You

What is the volume of a cube with edge $2a$?



ChatGPT

The volume of a cube is calculated by cubing the length of one of its edges.

So, if the edge of the cube is $2a$, the volume would be:

$$\text{Volume of cube} = (\text{Edge length})^3 = (2a)^3 = 8a^3$$

Therefore, the volume of the cube with an edge length of $2a$ is $8a^3$.



Emergent few-shot learning

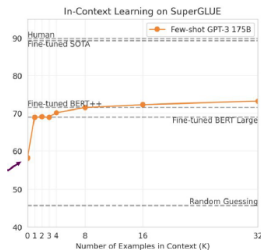
- Specify a task by simply prepending examples of the task before your example
- no gradient updates are performed

In-context learning

1	gaot => goat
2	sakne => snake
3	brid => bird
4	fsih => fish
5	douk => duck
6	cmihp => chimp

In-context learning

1	thanks => merci
2	hello => bonjour
3	mint => menthe
4	wall => mur
5	otter => loutre
6	bread => pain



Example

Zero-Shot Learning

Classify the given text using Sentiment Analysis.
Possible classes: POSITIVE, NEGATIVE, NEUTRAL.

Input: "I watched La La Land... yeah it was OK I guess"
Class:

Few-Shot Learning

Classify the given text using Sentiment Analysis.
Possible classes: POSITIVE, NEGATIVE, NEUTRAL.

Input: "Unbelievable, I wasted 2h on 'The Room'!"
Class: NEGATIVE

Input: "Dr Strangelove is my favorite movie <3 <3"
Class: POSITIVE

Input: "I watched La La Land... yeah it was OK I guess"
Class: