



CS 329P : Practical Machine Learning (2021 Fall)

11.4 Prompt-based Learning

Qingqing Huang, Mu Li, Alex Smola

<https://c.d2l.ai/stanford-cs329p>

Prompt-based Learning



- BERT has different pre-training tasks and downstream tasks
 - Usually requires thousands of examples to fine tune for a task
- Prompting tries to convert downstream tasks to the same format of pre-training, i.e. language model
 - GPT made prompt-based learning popular
 - Sentiment analysis: I like this movie. **It was great**
 - Machine translation: Hello world! => Bonjour le monde!

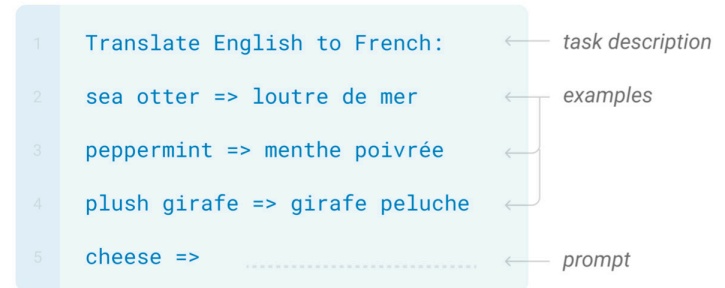
Prompt

Pos: great
Neg: terrible

Prompt with GPT-3



- A giant transformer decoder with 173B parameters
 - Trained on >500B tokens from CommonCrawl, WebText, books..
 - OpenAI spent ~\$12M to train the model
 - OpenAI provides paid API access (~\$0.06 per 1K tokens)...with a waitlist as of now
- General purpose LM with impressive text generation capabilities and zero-shot / few-shot learning
 - The LM also understands task specification
 - Prompting: Provide a task description with a few (~10) examples, and a prompt





GPT-3 Applications

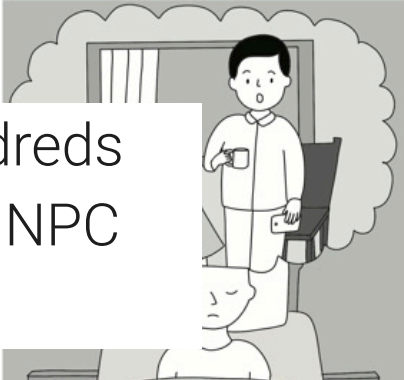
- Write code: a table of the richest countries in the worlds with column name and gdp

```
<table style={{width: '100%', border: '1px solid #ccc', borderRadius: '3px'}}>
<thead> <tr> <th>Name</th> <th>GDP</th> </tr> </thead> <tbody> <tr> <td>United
States</td> <td>18.6 trillion</td> </tr> <tr> <td>China</td> <td>11.2 trillion</td>
</tr> <tr> <td>Japan</td> <td>4.9 trillion</td> </tr> <tr> <td>Germany</td> <td>3.6
trillion</td> </tr> <tr> <td>United Kingdom</td> <td>2.6 trillion</td> </tr> <tr>
<td>France</td> <td>2.4 trillion</td> </tr> <tr> <td>Italy</td> <td>2.3
trillion</td> </tr> <tr> <td>Canada</td> <td>1.5 trillion</td> </tr> <tr>
<td>India</td> <td>1.4 trillion</td> </tr> <tr> <td>Australia</td> <td>1.3
trillion</td> </tr> <tr> <td>Russia</td> <td>1.1 trillion</td> </tr> <tr>
<td>Brazil</td> <td>1.0 trillion</td> </tr> <tr> <td>Mexico</td> <td>0.9
trillion</td> </tr> <tr> <td>South Korea</td> <td>0.8 trillion</td> </tr> <tr>
<td>Saudi Arabia</td> <td>0.7 trillion</td> </tr> <tr> <td>Italy</td> <td>0.7
trillion</td> </tr> <tr> <td>Switzerland</td> <td>0.6 trillion</td> </tr> <tr>
<td>Canada</td> <td>0.6 trillion</td> </tr> <tr> <td>Taiwan</td> <td>0.6
trillion</td> </tr></tbody> </table>
```

- Generate thought experiment with classic examples as input

The Dreaming Argument

Imagine you are a normal person living a normal life one day. However, everything you experience is actually a dream. You have not yet awoken.



- Check gpt3demo.com for hundreds demos such as search engine, NPC dialogues, and writing poems



Prompt-based Fine-tuning

- Goal: Fine-tune the weights of medium sized LM (e.g. <1B)
- Design task-specific prompts VS train a new output layer
 - Prompt-based FT is more example efficient than standard FT (100x)
- Automatic prompt search
 - Label words and template selection

Task	Auto template	Auto label words
SST-2	(positive/negative)	
	<S ₁ > A [MASK] one .	irresistible/pathetic
	<S ₁ > A [MASK] piece .	wonderful/bad
	<S ₁ > All in all [MASK] .	delicious/bad

SST-2 (positive/negative)		mean (std)
<S ₁ > It was [MASK] .	great/terrible	92.7 (0.9)
<S ₁ > It was [MASK] .	good/bad	92.5 (1.0)
<S ₁ > It was [MASK] .	cat/dog	91.5 (1.4)
<S ₁ > It was [MASK] .	dog/cat	86.2 (5.4)
<S ₁ > It was [MASK] .	terrible/great	83.2 (6.9)
Fine-tuning	-	81.4 (3.8)

The impact of templates and label words

Gao et.al. 2021

Summary



- Prompt-based learning uses prompt to present downstream tasks in a language model format
 - GPT-3 directly uses the pre-trained models for downstream tasks without updating parameters
 - Using it in fine-tuning leads to better example efficiency