Language Models (GPT, GPT-2 and GPT-3)

Advanced Techniques in Artificial Intelligence

Jon Ander Almandoz, Julen Etxaniz and Jokin Rodriguez

2020-09-28

Contents

1. Introduction

1. Introduction

- GPT, GPT-2 and GPT-3 are NLP's created by OpenAI.
- Generative Pre-trained Transformer.
- The release of the GPT-2 made a lot of echo.
- The ability to "understand".
- GPT-2 and GPT-3 controversy.
- Each GPT model has much more parameters than the previous one (150M, 1.5B, 175B).
- Performance upgrade.

2. GPT

2.1. Improving Language Understanding by Generative Pre-Training

2.1. Improving Language Understanding by Generative Pre-Training

- NLP challenges: question answering, document classification...
- Large improvements by generative pre-training, followed by supervised fine-tuning on each task.
- All datasets use a single language model, requiring very little tuning.

2.1.1. Why Unsupervised Learning?

- Although unlabeled text is abundant, labeled data for specific tasks is scarce.
- Therefore, training models with supervised learning is hard.
- Unsupervised learning can address these drawbacks.
- The language model begins to perform tasks without training on them.
- The performance is still quite low compared to supervised models.
- But this is robust across a broad set of tasks.

2.1.2. Drawbacks

- Expensive computer requirements
- The limits and bias of learning about the world through text
- Still fragile generalization

2.1.3. Future

- Scaling the approach
- Improved fine-tuning
- Better understanding of why generative pre-training helps

3. GPT-2

3.1. Language Models are Unsupervised Multitask Learners

3.2. Release Strategies and the Social Impacts of Language Models

3.1. Language Models are Unsupervised Multitask Learners

3.1. Language Models are Unsupervised Multitask Learners

- GPT-2 Objective: Predict the next word.
- GPT Vs GPT-2: 10X more parameters, trained on 10X more data.
- The model creates realistic samples based on the given input.

3.1.1. Samples

- 10 tries: scientist discovered a herd of unicorns
- 1 try: a train carriage containing controlled nuclear materials was stolen
- 2 tries: Miley Cyrus was caught shoplifting
- 1 try: a large language model called GPT-2 generates realistic paragraphs of text
- 1 try: Legolas and Gimli advanced on the orcs
- 5 tries: please describe the reasons for the US Civil War
- 25 tries: John F. Kennedy was just elected President of the United States
- 25 tries: recycling is bad for the world

3.1.2. Zero-shot examples

- **Reading Comprehension**: answer questions about given passages
- **Common Sense Reasoning**: resolution of an ambiguous pronoun
- **Question Answering**: answer questions without context
- **Language Modeling of Broad Contexts**: predict the last word of a passage
- **Summarization**: summarize news articles
- **Machine Translation**: translate French sentences to English

3.1.3. Our examples

- InferKit
- ***My parents told me I should stay at home because*** there are too many negative things going around in the city. They said I should stop being so impulsive and start doing things slowly. They're right. I should have listened to them. But, here I am. Once again, I'm in trouble. I've done something I shouldn't have, and it's only getting worse.

- ***What is 2001: A Space Odyssey about?*** The opening sequence of 2001 sees an Earth that is still thriving after the crash of 1967. The long-dead nuclear war has been forgotten, and mankind is now engaged in interstellar exploration. For their last home world, humanity has developed computer systems that allow the probes to travel through

- ***OpenAI GPT is*** a scalable, unsupervised, reinforcement learning environment for the training and testing of supervised models on unstructured data. It can also be used for feature engineering. Assembling ideas from earlier open source projects, OpenAI GPT is inspired by agent-based and reinforcement learning.

3.2. Release Strategies and the Social Impacts of Language Models

3.2.1. Policy Implications

Beneficial purposes:

- AI writing assistants
- More capable dialogue agents
- Unsupervised translation between languages
- Better speech recognition systems

Malicious purposes:

- Generate misleading news articles
- Impersonate others online
- Automate the production of abusive or faked content to post on social media
- Automate the production of spam/phishing content

3.2.2. Release Strategy

Due to concerns about malicious uses, they initially released a smaller version of GPT-2.

This created a big controversy about OpenAI not being open.

Two mechanisms to responsibly publish next versions of GPT-2:

- Staged Release: Gradual release of a family of models over time.
- Partnership-based Sharing: Shared the 762M and 1.5B parameter versions to some Universities.

3.2.3. Staged Release

It involves the gradual release of a family of models over time.

To give people time to assess the properties of these models and evaluate the impacts after each release.

1. **February 2019:** Released small 124M parameter model and a technical paper.
2. **May 2019:** Released medium 355M parameter model.
3. **August 2019:** Released larger 774M parameter model and a technical report.
4. **November 2019:** Released largest 1.5B parameter model.

3.2.4. Partnerships

Partners in AI community who work to improve social preparedness for language models.

1. Cornell University studied **human susceptibility** to digital disinformation.
2. The Middlebury Institute of International Studies explored **misuse** by terrorists and extremists online.
3. The University of Oregon analyzed **bias** within GPT-2.
4. The University of Texas at Austin studied the **statistical detectability** of GPT-2.

3.2.5. Learnings

1. Coordination is difficult, but possible.
2. Humans find GPT-2 outputs convincing.
3. GPT-2 can be fine-tuned for misuse.
4. Detection is challenging.
5. There is no strong evidence of misuse so far.
6. Standards for studying bias are needed.

4. GPT-3

4.1. Language Models are Few-Shot Learners

- Humans perform well on new language tasks, just being given few instructions.
- NLP systems usually need thousands of example to perform well.
- GPT-3 has 175B parameters.
- GPT-3 few-shot is good at translations, question answering, unscrambling words…
- Few-shot learning does not succeed in all tasks.

4.2. OpenAI API

4.2. OpenAI API

- General-purpose interface, allowing users to try it on any English language task.
- It is the only way to use the GPT-3 model, as it hasn't been released.
- The API is in private beta and you have to request access and pay in order to use it.
- This continued the debate that started when OpenAI decided not to publish GPT-2.

4.2.1. Tasks

- **Semantic Search**: searching over documents based on meaning.
- **Chat**: fast, complex and consistent natural language discussions.
- **Customer Service**: dialogue to quickly give customers relevant information.
- **Generation**: generate complex and consistent natural language.
- **Productivity Tools**: parsing text into tables, summarizing email...
- **Content Comprehension**: help individuals consume content more efficiently.
- **Polyglot**: for translation or to chat with users in their preferred language.

4.2.2. Demo

- **Text generation**: generate a continuation to the given text
- **Q&A**: answer questions without context
- **Parse unstructured data**: extract structured data from text
- **Improve English**: improve badly written English sentences
- **Translate French**: translate English sentences to French
- **Speech to Bash**: translate speech to bash code

5. Conclusion

5. Conclusion

- No idea how will next generation NLP models perform.
- Really good on many tasks.
- Not capable of telling if a question makes sense or not.
- GPT-3 represents a paradigm shift.

6. References

6. References

1. Introduction:
   - Wikipedia: OpenAI
2. GPT:
   - Paper: Improving Language Understanding by Generative Pre-Training
   - Blog: Improving Language Understanding with Unsupervised Learning
3. GPT-2:
   - Paper: Language Models are Unsupervised Multitask Learners
   - Blog: Better Language Models and Their Implications
   - Report: Release Strategies and the Social Impacts of Language Models
   - Blog: GPT-2: 6-Month Follow-Up
   - Blog: GPT-2: 1.5B Release
   - Blog: OpenAI's GPT-2: the model, the hype, and the controversy | by Ryan Lowe
4. GPT-3:
   - Paper: Language Models are Few-Shot Learners
   - API: OpenAI API
   - Blog: OpenAI API
   - Blog: OpenAI reveals the pricing plans for its API — and it ain't cheap
5. Conclusion:
   - Blog: Por qué GPT-3, el nuevo modelo de lenguaje de OpenAI, es tan impresionante como poco útil