

GPT-2 : Language Models are Unsupervised Multitask Learners

- fine tuning 과정이 더 이상 필요하지 않은 모델
- 문장만이 input 주어지고 주어진 단어들로 다음 단어 예측하는 방식으로 학습된 GPT1과 달리 추가적으로 task 정보(요약 task 등 task의 종류)를 함께 input으로 줌.
- fine tuning 과정 없이 output을 잘 생성해낼 수 있게 됨.

GPT-3 ---- few shot

- gpt2처럼 input에 task에 대한 정보를 특별한 토큰으로 명시
- 추가로 몇 가지 예제를 입력으로 넣어줌. (이를 few shot 이라고 함)
- 특징: attention 주는 부분에서 sparse하게 attention을 주고 있다.
- 학습방식 – meta-learning : 사람이 통제하던 기계학습 과정을 자동화 -> 기계 스스로 학습규칙(메타 지식)을 익힐 수 있도록 함.
- Language Modeling : penn tree bank, story completion(lambada)
- Knowledge Intensive Task: Reading Comprehension
- Structured problems that require multiple steps of reasoning: RTE, Arithmetic, Word problems, Analogy making
- 한계: Language is situated and GPT-3 is merely learning from text 'without being exposed to other modalities.'

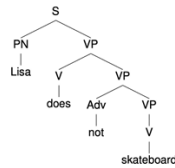
Compositional Representations and Systematic Generalization

* Systematicity(체계성): 사람이 이해하는 문장들 간에는 확실하고 예측 가능한 패턴이 있다.

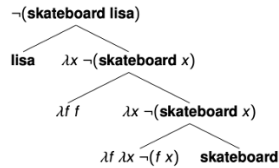
* Compositionality(구성성): 한 표현의 의미는 그 표현을 의미하는 구성 요소들의 의미와 구조로 이루어진다.

- According to Montague, Compositionality is about the existence of a homomorphism from syntax to semantics:

Lisa does not skateboard =
 $\langle \text{Lisa}, \langle \text{does}, \langle \text{not}, \text{skateboard} \rangle \rangle \rangle$



$m(\text{Lisa does not skateboard}) =$
 $\langle m(\text{Lisa}), \langle m(\text{does}), \langle m(\text{not}), m(\text{skateboard}) \rangle \rangle \rangle$



- Tree Reconstruction Error (TRE) [Andreas 2019]: Compositionality of **representations** is about how well the representation approximates an explicitly homomorphic function in a *learnt representation space*

compositional generalization: the capacity to understand and produce a potentially infinite number of novel combinations of known components

이 task에서 좋은 성능을 보이는 neural network가 있을까?

can we create a dataset split that explicitly tests for compositional generalization?

- similar atom distribution
- different compound distribution