11/11/24

GPT 1 (Paper)

Abstract : generative pre training → discriminative fine tuning

While fine tuning → Use task aware input transformations to achieve transfer while aiming at minimal changes in the model architecture

models make use of unlabelled data - evidence = trained word embeddings

Challenges

① Unclear of optimization objectives

② No consensus on way to transfer these learned representations to target task.

GPT 1 : Semi Supervised approach for language understanding task using combination of unsupervised pre training and supervised fine tuning

Training → 2 stages

① LM objective on unlabeled data
to learn initial parameters of NN model

② adapt those parameters to target task
using corresponding supervised object
(fine-tuning) on labelled data

Transformers :

(why) - Perform strongly on
- Machine Translation
- Document Generation
- Syntactic Parsing

| Captures longer range linguistic structure

- Provides structured memory for handling long term dependencies in text.
- Robust transfer performance across tasks
- Processes structured text as single contagious sequence of tokens

Unsupervised pre training :-

goal : to find good initialization point instead of modifying supervised learning objective