

BART

Abstract

- BART : denoising autoencoder for pertaining sequence to sequence model이다.
- BART는 noise data로 학습됨
- 성능 좋음
- replicate other pretraining schemes within the BART -> better performance

Introduction

- 기존 masked token 방식의 한계점 (task에 대한 부분- 특정 task에 집중된 방식)
- BART는 bidirectional auto regressive transformer의 약자이다.
- Denoting autoencoder built with a sequence to sequence model -> wide range of end task
- BART는 bidirectional encoder 부분과 autoregressive decoder로 이루어져있다.
- 주요 장점 : noise에 flexible하다.
- ablation analysis를 통해서 -> 'This study allows us to carefully control for a number of factors, including data and optimization parameters, which have been shown to be as important for overall performance as the selection of training objectives' 라고 알게 됨

앞으로 해결해야 되는 문제

- task에 대한 맞춤형 corrupting documents for pre-training에 대한 새로운 방법을 탐구하는 것

나의 생각

- 무엇을 해결하고 싶은 것인가?
 - 기존의 mask 방식은 특정 task에 대해서만 집중이 되었는데 이 부분을 해결하고자 -> 더 범용적인 task를 잘 수행할 수 있는 모델 만들고 싶어서
- 이전 모델과 달라진 점은?
 - GPT나 BERT와 다른 Encoder + Decoder 형식이고 이때 encoder는 bidirectional이고 decoder는 autoregressive 형태이다.
- 핵심 키워드 - 그렇게 생각한 이유
 - Denoting autoencoder - bart라는 모델이 pretraining Seq2Seq에 대한 denoting autoencoder라고 언급되었기 때문

- Noising - BART학습시 여러 noise data를 이용하였기 때문
- Bidirectional - BART의 B 부분
- Auto regressive transformer - BART의 ART 부분
- 논문 진행 방향(흐름)
 - 모델 구조를 자세히 설명하면서 bidirectional encoder + autoregressive decoder 형태로 만든 이유가 나올 것 같다.
 - input에서 기존과 다른 mask token방식에 대해서 설명(noise를 추가한?)
 - 위의 두 방식을 통해서 여러 task에서 좋은 성능을 냈음을 보여줌
 - 다른 pertained model과 bart를 함께 사용했을때도 좋은 성능이 나옴을 증명

모르는 개념

- autoencoder
- Self supervised method
- Auto regressive transformer
- Target-side language model.
- ablation analysis : 모델이나 알고리즘의 'feature'들을 제거해 나가면서 그 행위가 성능에 얼마나 영향을 미치는지 확인해보는 분석.

Question!