

프리콘 보딩 AI 심화과정 week 1-1

1. 본인이 본 강의를 수강하는 목적에 대해서 자유롭게 적어보세요.

이번 강의에 바라는 가장 큰 목적은 경험이다. 내가 데이터 직군에 관련된 경험은 거의 대부분이 얼마 전 까지 들었던 부트 캠프에서 비롯된 경험이 거의 다 이다. 물론 부트 캠프에서도 생각 보다 많은 것을 배우고 여러 가지 특이한 경험을 할 수 있었다.

하지만 아직까지 어떤것을 해야할지 잘 모르겠다는게 문제였다. 또 포트폴리오를 정리하다 보니 내가 했던 프로젝트 들이 좀 많이 부족하다는 게 눈에 띄기 시작했다. 그것들을 보완할 프로젝트를 하거나 아니면 더 나은 프로젝트로 채워봄이 더 좋을 것 같았고 고민 중에 이런 기회가 있다는 것을 알고 수강하게 되었다.

또 취업이 될 확률을 높여줄 채용 의사가 있는 기업들과의 협약도 이 강의를 수강을 결정하는데 도움이 되었다.

2 . Paperswithcode(<https://paperswithcode.com/area/natural-language-processing>

)에서 NLP sub task 중에 2개를 선택하여 본인 블로그에 정리해보세요. task 별로 아래 3가지 항목에 대해서 정리하세요. (각 항목 고려 사항 참고)

1)본인 아이디 및 닉네임

노지훈

2) url

<https://blog.naver.com/nnnobaram/222653775008>

<https://blog.naver.com/nnnobaram/222653714151>

3) 게시물 캡처



1. 문제 정의

Machine Translation이란 원래의 글자를 다른 타겟의 글자로 번역하는 것이 해결할 문제이다. 기계번역에 대한 접근 방식은 규칙 기반에서 통계기반, 신경 기반에 이르기 까지 다양합니다.

2. 데이터 소개(대표적인 데이터 1개)

1) 생성 요약 task를 해결하기 위해 사용할 수 있는 데이터

언어 모델링을 하기위한 가장 좋은 데이터 셋은 WMT 데이터 셋입니다. WMT는 Ninth Workshop on Statistical Machine Translation에서 공유작업에 사용되는 데이터 세트 모음이다. 워크샵에서는 뉴스 번역, 품질 평가, 메트릭, 의료 텍스트 번역 4가지 작업이 포함되어있다.

2) 데이터 구조

데이터는 여러 언어 번역이 있다. 영어, 독일어, 프랑스, 힌디, 스페인어, 드등 여러개의 언어가 있고 매년 데이터가 새로나와서 추가가 되어가고 있다.

3.SOTA 모델 소개(대표적인 모델 1개)

1) task의 SOTA 모델은 무엇인가

[Transformer+BT](#)가 가장 성능이 좋게 나오는 모델이다.

2) 해당 모델 논문의 요약에서 주요 키워드는?

해당 모델의 논문은 Very Deep Transformers for Neural Machine Translation 이라는 논문이다.

논문은 훈련은 안정화 하는 효과적인 초기화 기술을 사용해서 최대 60개의 인코더 레이어와 12개의 디코더 레이어가 있는 표준 변환기 기반 모델을 구축했다는 것을 보여준다.

과제 - NLP subtask(1)-Question Answering



노지훈 · 2시간 전

URL 복사

👤 통계



1. 문제 정의

Question Answering은 묻는 문제에 답하는 게 만드는 것이다. 그러나 준비되지 않은 질문이 주어진다면 함복을 한다. Question Answering은 커뮤니티 질문 답변 및 지식 기반 질문 답변과 같은 도메인별 작업으로 나눌 수 있습니다.

2. 데이터 소개(대표적인 데이터 1개)

1) 생성 요약 task를 해결하기 위해 사용할 수 있는 데이터

언어 모델링을 하기위한 가장 좋은 데이터 셋은 SQuAD 데이터 셋입니다. SQuAD는 Stanford Question Answering Dataset은 위키피디아 글에서 질문과 답 쌍의 문치이다. SQuA에서 질문의 정확한 답은 주어진 글에서 토큰의 순서일수 있다. 문제와 답은 crowdsourcing을 통해서 사람에 의해 생산이 되어진다. 이것이 다른 데이터 셋보다 더 다양한게 해준다. 536 글에서 107,785개의 쌍의 질문과 답을 가지고 | o s s 다.

2) 데이터 구조

```
{
  "question": "When did Beyonc\u00e9 start recording albums?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "In the late 2000s",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
},
{
  "question": "What year did Beyonc\u00e9 compete in when she was growing up?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "2003",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
},
{
  "question": "When did Beyonc\u00e9 leave Destiny's Child and become a solo singer?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "2003",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
},
{
  "question": "In what city and state did Beyonc\u00e9 grow up?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "Houston, Texas",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
},
{
  "question": "In what decade did Beyonc\u00e9 become famous?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "2000s",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
},
{
  "question": "In what year did Beyonc\u00e9 become famous?",
  "id": "5b92c0a0a000000000000000",
  "answers": [
    {
      "text": "2003",
      "answer_start": 180,
      "is_impossible": false
    }
  ],
  "context": "When asked why Beyonc\u00e9 started recording albums, she said she was growing up.",
  "is_impossible": false
}
```

데이터의 구조는 json으로 되어있고 질문과 그에따른 아이디 그리고 그 답과 그에따른 아이디로 되어있다.

3.SOTA 모델 소개(대표적인 모델 1개)

1) task의 SOTA 모델은 무엇인가

T5-11B이 가장 성능이 좋게 나오는 모델이다.


2) 해당 모델 논문의 요약에서 주요 키워드는?

해당 모델의 논문은 Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer 이라는 논문으로 주요 키워드는 transfer learning일 것이다.

transfer learning이란 전이 학습으로 이전에 학습시켰던 모델의 가중치를 가지고와서 재보정해서 사용하는것입니다.

3.댓글

1) <https://blog.naver.com/codenavy94/222653364998>

**노지훈**


문제 정의에서 개체명 인식에 대한 설명이 예시 까지 들어주어서 이해하는데 도움이 많이 되었습니다.

2022.2.21. 21:20

답글

♡ 0

2)<https://blog.naver.com/codenavy94/222653405430>


**노지훈**

생성 요약과 추출 요약과의 차이점을 설명해주어서 좋았습니다. 잘 읽었습니다.

2022.2.21. 19:17

답글

3)<https://blog.naver.com/cardchan/222653567329>


**노지훈**

초록요약을 통해서 논문에 대한 정리를 해주셔서 좋았습니다. 특히 사전 학습으로 인해서 모델끼리 비교하기가 어려워 졌다는점이 흥미롭네요.

2022.2.21. 21:01

답글

4)<https://blog.naver.com/cardchan/222653484009>


**노지훈**

관련 논문이 전이학습의 한계를 돌파하려는 시도라는 것을 처음 알게되었네요. 논문에 대한 정리를 잘해 주셔서 좋았습니다.

2022.2.21. 20:53

답글

5)<https://blog.naver.com/kackpooh/222653750269>


**노지훈**

1. bert 보다 뛰어난 XLnet 알고리즘이 있군요
2. hasa라는 방법이 있다는게 신기합니다.

2022.2.21. 21:10

답글

6)<https://blog.naver.com/ctk456/222653835941>

 **노지훈**

모델의 성능을 설명해주셔서 모델의 효능을 알려주어서 좋았습니다.

2022.2.21. 21:16

답글