

# 10월 29일

## CSQA-200(bs=32,mbs=4,ebs=8,epoch 40,top50) max\_len = 512-GSC

### 1. only triple context score

- seed0(20231029\_015042) - 512(max\_len) → 이 자식 마무리하면 좀 실험 setting 정리를 하자
  - dev : 71.09, test : 71.07
- seed1
- seed2(20231028\_153644) → NAN

### 2. triple score + graph score

- a. seed0(20231030\_020451)
- b. seed1
- c. seed2(20231028\_152934)
  - i. dev : 75.76, test : 73.09

## CSQA-200(bs=64,mbs=4epoch 30,top50) max\_len = 512- qagnn

### 1. only triple context score

### 2. triple score + graph score

- a. seed0(20231028\_153350)

## CSQA-200(bs=128,mbs=4,ebs=8,epoch 40,top50) max\_len = 512-GSC

### 1. only triple context score

### 2. triple score + graph score

a. seed0(20231029\_015156) → 성능 bad

지금 실험 결과를 살펴보면 OpenBookQA 데이터셋은 graph score를 사용하면 성능이 많이 내려간다.

하지만 CommonsenseQA 데이터셋은 graph score를 사용하면 성능이 올라간다. 이것에 대해서 구체적인 실험을 정하고 진행해야할것 같다.

지금 나에게는 OpenBookQA 데이터셋은 Top-20 이 있고, CommonsenseQA는 Top-100, Top-50이 있다.

## **OBQA-200(bs=16, mbs=2,ebs=4,epoch 75) max\_len = 224**

### 1. only triple context score

- seed0(20231025\_043344)
  - dev : 69.4, test : 71.6
- seed1(20231025\_044204)
  - dev : 69.6, test : 69.6
- seed2(20231025\_044307)
  - dev : 67.6, test : 71

## **CSQA-200(bs=64,mbs=4,ebs=8,epoch 30) max\_len = 512(top-100)**

### 1. only triple context score

- seed0(20231025\_134848)
  - dev : 71.66, test : 69.54
- seed1(20231025\_135020)
  - dev :
- seed2(20231025\_135817)

### 2. triple context score + graph score

- seed0(20231026\_022502)
  - dev : 74.12, test : 74.21 → epoch를 늘리면 더 좋아지지 않을까?
- seed1(20231026\_022540)
  - dev : 73.05, test : 73.57
- seed2(20231026\_153619)
  - dev : 73.87, test : 73.33

### **CSQA-200(bs=32,mbs=4,ebs=8,epoch 30) max\_len = 512(top-100)**

#### 1. triple context + graph score

- seed0(2023102
- 9\_081747) → 성능 bad