Compare Results

Total Changes

67

Content

4   Replacements

5   Insertions

32   Deletions

Styling and Annotations

12   Styling

14   Annotations

Go to First Change (page 2)

MTA4DPR: Multi-Teaching-Assistants Based Iterative Knowledge Distillation for Dense Passage Retrieval

Qixi Lu[1,2], Endong Xun[1], and Gongbo Tang[2]

[1]Beijing Advanced Innovation Center for Language Resources, Beijing Language and Culture University, China
[2]School of Information Science, Beijing Language and Culture University, China
lqxaixxh@gmail.com, edxun, gongbo.tang@blcu.edu.cn

Abstract

Although Dense Passage Retrieval (DPR) models have achieved significantly enhanced performance, their widespread application is still hindered by the demanding inference efficiency and high deployment costs. Knowledge distillation is an efficient method to compress models, which transfers knowledge from strong teacher models to weak student models. Previous studies have proved the effectiveness of knowledge distillation in DPR. However, there often remains a significant performance gap between the teacher and the distilled student. To narrow this performance gap, we propose MTA4DPR, a Multi-Teaching-Assistants based iterative knowledge distillation method for Dense Passage Retrieval, which transfers knowledge from the teacher to the student with the help of multiple assistants in an iterative manner; with each iteration, the student learns from more performant assistants and more difficult data. The experimental results show that our 66M student model achieves the state-of-the-art performance among models with same parame[MISSING]

1 Introduction

[MISSING CONTENT]

2 Methodology

[MISSING CONTENT]

3 Experiments

[MISSING CONTENT]

4 Conclusion

[MISSING CONTENT]

A Algorithm

Algorithm 1 MTA4DPR Training Process

Require: : the teacher model; : the assistant models; : the student model; : the query set; : the passage set; : maximum number of training iterations; : maximum number of training steps; : Learning rate;

Ensure:

1:
2: **while do**
3:    {[ILLEGIBLE]}
4:    [MISSING]
5: **end while**

---

*Corresponding author