手机屏幕的截图

中度可信度描述已自动生成

文本数据挖掘大作业

成员1：李一鸣 202228013229030

成员2：张 兆 202228013229029

成员3：陈国鑫 2022E8013282125

代码开源地址：https://github.com/zhangzhao219/Multimodal-Sentiment-Analysis

2023年05月

目录

[1 引言 1](#_Toc120436406)

[1.1 任务描述 1](#_Toc120436407)

[1.2 赛题分析 1](#_Toc120436408)

[1.3 评价指标 1](#_Toc120436409)

[2 实现方案 3](#_Toc120436410)

[2.1 技术原理 3](#_Toc120436411)

[2.1.1 Re-ranker 3](#_Toc120436412)

[2.1.2 BERT 3](#_Toc120436413)

[2.1.3 ColBERT 4](#_Toc120436414)

[2.1.4 Warmup 5](#_Toc120436415)

[2.2 预训练模型 6](#_Toc120436416)

[2.2.1 ELECTRA 7](#_Toc120436417)

[2.2.2 SimLM 7](#_Toc120436418)

[2.2.3 CoT-MAE 8](#_Toc120436419)

[2.3 实验效果对比 9](#_Toc120436420)

[3 实验步骤 11](#_Toc120436421)

[3.1 提交说明 11](#_Toc120436422)

[3.2 文件说明 11](#_Toc120436423)

[3.3 训练流程 13](#_Toc120436424)

[3.4 预测流程 15](#_Toc120436425)

[4 参考文献 17](#_Toc120436426)

1. 引言
   1. 任务描述

其中的时间戳会去寻找刚刚对应时间戳下训练好的模型，测试的数据均为2020年的数据，其余的参数与训练大致相同。测试过程中也会同时在文件中和控制台输出对应的预测信息，并且在最后一行打印测试指标。

图10 测试日志示意图

1. 参考文献

[1] Yu Y, Xiong C, Sun S, et al. COCO-DR: Combating Distribution Shifts in Zero-Shot Dense Retrieval with Contrastive and Distributionally Robust Learning[J]. arXiv preprint arXiv:2210.15212, 2022.

[2] Khattab O, Zaharia M. Colbert: Efficient and effective passage search via contextualized late interaction over bert[C]//Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 2020: 39-48.

[3] Santhanam K, Khattab O, Saad-Falcon J, et al. Colbertv2: Effective and efficient retrieval via lightweight late interaction[J]. arXiv preprint arXiv:2112.01488, 2021.

[4] Wu X, Ma G, Lin M, et al. Contextual mask auto-encoder for dense passage retrieval[J]. arXiv preprint arXiv:2208.07670, 2022.

[5] Wang L, Yang N, Huang X, et al. Simlm: Pre-training with representation bottleneck for dense passage retrieval[J]. arXiv preprint arXiv:2207.02578, 2022.

[6] Craswell N, Mitra B, Yilmaz E, et al. Overview of the TREC 2019 deep learning track[J]. arXiv preprint arXiv:2003.07820, 2020.

[7] Craswell, N., Mitra, B., Yilmaz, E., and Campos, et al. Overview of the TREC 2020 deep learning track[J]. arXiv preprint arXiv: 2102.07662, 2021.

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[9] Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555.