

# 作业

---

## 基础作业

- 使用 OpenCompass 评测 InternLM2-Chat-7B 模型在 C-Eval 数据集上的性能

## 进阶作业

- 使用 OpenCompass 评测 InternLM2-Chat-7B 模型使用 LMDeploy 0.2.0 部署后在 C-Eval 数据集上的性能

备注：由于进阶作业较难，完成基础作业之后就可以先提交作业了，在后续的大作业项目中使用这些技术将作为重要的加分点！

## 整体实训营项目：

时间周期：即日起致课程结束

即日起开始可以在班级群中随机组队完成一个大作业项目，一些可提供的选题如下：

- 人情世故大模型：一个帮助用户撰写新年祝福文案的人情事故大模型
- 中小学数学大模型：一个拥有一定数学解题能力的大模型
- 心理大模型：一个治愈的心理大模型
- 工具调用类项目：结合 Lagent 构建数据集训练 InternLM 模型，支持对 MMYOLO 等工具的调用

其他基于书生·浦语工具链的小项目都在范围内，欢迎大家充分发挥想象力。

```
inflating: data/mmlu/dev/astronomy_dev.csv
inflating: data/mmlu/dev/college_computer_science_dev.csv
inflating: data/mmlu/dev/high_school_biology_dev.csv
inflating: data/mmlu/dev/high_school_physics_dev.csv
inflating: data/mmlu/dev/high_school_psychology_dev.csv
inflating: data/mmlu/dev/high_school_statistics_dev.csv
inflating: data/mmlu/dev/professional_law_dev.csv
inflating: data/mmlu/dev/sociology_dev.csv
  creating: data/siqa/
inflating: data/siqa/dev.jsonl
inflating: data/siqa/train-labels.lst
inflating: data/siqa/dev-labels.lst
inflating: data/siqa/train.jsonl
  creating: data/TheoremQA/
inflating: data/TheoremQA/test.csv
  creating: data/triviaqa/
inflating: data/triviaqa/trivia-dev.qa.csv
inflating: data/triviaqa/trivia-test.qa.csv
  creating: data/lambada/
inflating: data/lambada/test.jsonl
  creating: data/strategyqa/
inflating: data/strategyqa/strategyQA_train.json
(opencompass) root@intern-studio-50023492: ~/opencompass# python tools/list_configs.py internlm ceval
```

Model	Config Path
hf_internlm2_1_8b	configs/models/hf_internlm/hf_internlm2_1_8b.py
hf_internlm2_20b	configs/models/hf_internlm/hf_internlm2_20b.py
hf_internlm2_7b	configs/models/hf_internlm/hf_internlm2_7b.py
hf_internlm2_base_20b	configs/models/hf_internlm/hf_internlm2_base_20b.py
hf_internlm2_base_7b	configs/models/hf_internlm/hf_internlm2_base_7b.py
hf_internlm2_chat_1_8b	configs/models/hf_internlm/hf_internlm2_chat_1_8b.py
hf_internlm2_chat_1_8b_sft	configs/models/hf_internlm/hf_internlm2_chat_1_8b_sft.py
hf_internlm2_chat_20b	configs/models/hf_internlm/hf_internlm2_chat_20b.py
hf_internlm2_chat_20b_sft	configs/models/hf_internlm/hf_internlm2_chat_20b_sft.py
hf_internlm2_chat_20b_with_system	configs/models/hf_internlm/hf_internlm2_chat_20b_with_system.py
hf_internlm2_chat_7b	configs/models/hf_internlm/hf_internlm2_chat_7b.py
hf_internlm2_chat_7b_sft	configs/models/hf_internlm/hf_internlm2_chat_7b_sft.py
hf_internlm2_chat_7b_with_system	configs/models/hf_internlm/hf_internlm2_chat_7b_with_system.py
hf_internlm2_chat_math_20b	configs/models/hf_internlm/hf_internlm2_chat_math_20b.py
hf_internlm2_chat_math_20b_with_system	configs/models/hf_internlm/hf_internlm2_chat_math_20b_with_system.py
hf_internlm2_chat_math_7b	configs/models/hf_internlm/hf_internlm2_chat_math_7b.py
hf_internlm2_chat_math_7b_with_system	configs/models/hf_internlm/hf_internlm2_chat_math_7b_with_system.py
hf_internlm_20b	configs/models/hf_internlm/hf_internlm_20b.py
hf_internlm_7b	configs/models/hf_internlm/hf_internlm_7b.py
hf_internlm_chat_20b	configs/models/hf_internlm/hf_internlm_chat_20b.py
hf_internlm_chat_7b	configs/models/hf_internlm/hf_internlm_chat_7b.py
hf_internlm_chat_7b_8k	configs/models/hf_internlm/hf_internlm_chat_7b_8k.py
hf_internlm_chat_7b_v1_1	configs/models/hf_internlm/hf_internlm_chat_7b_v1_1.py
internlm_7b	configs/models/internlm/internlm_7b.py
lmdeploy_internlm2_chat_20b	configs/models/hf_internlm/lmdeploy_internlm2_chat_20b.py
lmdeploy_internlm2_chat_7b	configs/models/hf_internlm/lmdeploy_internlm2_chat_7b.py
ms_internlm_chat_7b_8k	configs/models/ms_internlm/ms_internlm_chat_7b_8k.py

Dataset	Config Path
ceval_clean_pp1	configs/datasets/ceval/ceval_clean_pp1.py
ceval_contamination_pp1_810ec6	configs/datasets/contamination/ceval_contamination_pp1_810ec6.py
ceval_gen	configs/datasets/ceval/ceval_gen.py
ceval_gen_2daf24	configs/datasets/ceval/ceval_gen_2daf24.py
ceval_gen_5f30c7	configs/datasets/ceval/ceval_gen_5f30c7.py
ceval_internal_pp1_lcd8bf	configs/datasets/ceval/ceval_internal_pp1_lcd8bf.py
ceval_pp1	configs/datasets/ceval/ceval_pp1.py
ceval_pp1_lcd8bf	configs/datasets/ceval/ceval_pp1_lcd8bf.py
ceval_pp1_578f8d	configs/datasets/ceval/ceval_pp1_578f8d.py
ceval_pp1_93e5ce	configs/datasets/ceval/ceval_pp1_93e5ce.py
ceval_zero_shot_gen_bd40ef	configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py

## 启动评测

【评测 InternLM-Chat-7B 模型在 C-Eval 数据集上的性能。由于 OpenCompass 默认并行启动评估过程，我们可以在第一次运行时以 --debug 模式启动评估，并检查是否存在问题。在 --debug 模式下，任务将按顺序执行，并实时打印输出】

```
--datasets ceval_gen \  
--hf-path /share/temp/model_repos/internlm-chat-7b/ \ # HuggingFace 模型路径  
--tokenizer-path /share/temp/model_repos/internlm-chat-7b/ \ # HuggingFace tokenizer 路径（如果与模型路径相同，可以省略）  
--tokenizer-kwarg padding_side='left' truncation='left' trust_remote_code=True \ # 构建 tokenizer 的参数  
--model-kwarg device_map='auto' trust_remote_code=True \ # 构建模型的参数  
--max-seq-len 2048 \ # 模型可以接受的最大序列长度  
--max-out-len 16 \ # 生成的最大 token 数  
--batch-size 2 \ # 批量大小  
--num-gpus 1 # 运行模型所需的 GPU 数量  
--debug
```



```
CPU 34.03% GPU-1: 30% Nvidia A100 0%
内存 0.94 / 72 GB 1.31% 显存 0 / 24566 MiB 0%

04/25 18:27:26 - OpenCompass - DEBUG - An `OpenICLEvalTask` instance is built from registry, and its implementation can be found in opencompass.tasks.openicl_eval
04/25 18:27:28 - OpenCompass - DEBUG - Get class `CEvalDataset` from "load_dataset" registry in "opencompass"
04/25 18:27:28 - OpenCompass - DEBUG - An `CEvalDataset` instance is built from registry, and its implementation can be found in opencompass.datasets.ceval
04/25 18:27:28 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 239 - Task [opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-tax_accountant]: No predictions found.
04/25 18:27:28 - OpenCompass - DEBUG - Get class `OpenICLEvalTask` from "task" registry in "opencompass"
04/25 18:27:28 - OpenCompass - DEBUG - An `OpenICLEvalTask` instance is built from registry, and its implementation can be found in opencompass.tasks.openicl_eval
04/25 18:27:30 - OpenCompass - DEBUG - Get class `CEvalDataset` from "load_dataset" registry in "opencompass"
04/25 18:27:30 - OpenCompass - DEBUG - An `CEvalDataset` instance is built from registry, and its implementation can be found in opencompass.datasets.ceval
04/25 18:27:30 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 239 - Task [opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-physician]: No predictions found.
04/25 18:27:30 - OpenCompass - DEBUG - An `DefaultSummarizer` instance is built from registry, and its implementation can be found in opencompass.summarizers.default

dataset version metric mode opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b
-----
ceval-computer_network - - - -
ceval-operating_system - - - -
ceval-computer_architecture - - - -
ceval-college_programming - - - -
ceval-college_physics - - - -
ceval-college_chemistry - - - -
ceval-advanced_mathematics - - - -
ceval-probability_and_statistics - - - -
ceval-discrete_mathematics - - - -
ceval-electrical_engineer - - - -
ceval-metrology_engineer - - - -
ceval-high_school_mathematics - - - -
ceval-high_school_physics - - - -
ceval-high_school_chemistry - - - -
ceval-high_school_biology - - - -
ceval-middle_school_mathematics - - - -
ceval-middle_school_biology - - - -
ceval-middle_school_physics - - - -
ceval-middle_school_chemistry - - - -
ceval-veterinary_medicine - - - -
ceval-college_economics - - - -
ceval-business_administration - - - -
ceval-marxism - - - -
ceval-mao_zedong_thought - - - -
ceval-education_science - - - -
ceval-teacher_qualification - - - -
ceval-high_school_politics - - - -
ceval-high_school_geography - - - -
ceval-middle_school_politics - - - -
ceval-middle_school_geography - - - -
ceval-modern_chinese_history - - - -
ceval-ideological_and_moral_cultivation - - - -
ceval-logic - - - -
ceval-law - - - -
ceval-chinese_language_and_literature - - - -
ceval-art_studies - - - -
ceval-professional_tour_guide - - - -
ceval-legal_professional - - - -
ceval-high_school_chinese - - - -
ceval-high_school_history - - - -
ceval-middle_school_history - - - -
ceval-civil_servant - - - -
ceval-sports_science - - - -
ceval-plant_protection - - - -
ceval-basic_medicine - - - -
ceval-clinical_medicine - - - -
ceval-urban_and_rural_planner - - - -
ceval-accountant - - - -
ceval-fire_engineer - - - -
ceval-environmental_impact_assessment_engineer - - - -
ceval-tax_accountant - - - -
ceval-physician - - - -
04/25 18:27:30 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240425_182534/summary/summary_20240425_182534.txt
04/25 18:27:30 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240425_182534/summary/summary_20240425_182534.csv
(opencompass) root@intern-studio-50023492: ~/opencompass#
(opencompass) root@intern-studio-50023492: ~/opencompass#
```



The image shows a VS Code editor interface. On the left is the Explorer sidebar showing a file tree. The selected file is `summary_20240425_182534.txt` inside the `summary` folder. The main editor area displays the content of this file, which is a table with 5 columns: `dataset`, `version`, `metric`, `mode`, and `opencompass.models.huggingface.HuggingFace_model_repos_internl`. The table contains 57 rows of data, listing various datasets from the `ceval` (Chinese Evaluation) suite, such as `ceval-computer_network`, `ceval-operating_system`, `ceval-college_programming`, etc. The status bar at the bottom indicates the current cursor position is at line 1, column 1, and the file encoding is UTF-8 with LF line endings.

dataset	version	metric	mode	opencompass.models.huggingface.HuggingFace_model_repos_internl
20240425_182534				
tabulate format				
dataset				
ceval-computer_network	-	-	-	-
ceval-operating_system	-	-	-	-
ceval-computer_architecture	-	-	-	-
ceval-college_programming	-	-	-	-
ceval-college_physics	-	-	-	-
ceval-college_chemistry	-	-	-	-
ceval-advanced_mathematics	-	-	-	-
ceval-probability_and_statistics	-	-	-	-
ceval-discrete_mathematics	-	-	-	-
ceval-electrical_engineer	-	-	-	-
ceval-metrology_engineer	-	-	-	-
ceval-high_school_mathematics	-	-	-	-
ceval-high_school_physics	-	-	-	-
ceval-high_school_chemistry	-	-	-	-
ceval-high_school_biology	-	-	-	-
ceval-middle_school_mathematics	-	-	-	-
ceval-middle_school_biology	-	-	-	-
ceval-middle_school_physics	-	-	-	-
ceval-middle_school_chemistry	-	-	-	-
ceval-veterinary_medicine	-	-	-	-
ceval-college_economics	-	-	-	-
ceval-business_administration	-	-	-	-
ceval-marxism	-	-	-	-
ceval-mao_zedong_thought	-	-	-	-
ceval-education_science	-	-	-	-
ceval-teacher_qualification	-	-	-	-
ceval-high_school_politics	-	-	-	-
ceval-high_school_geography	-	-	-	-
ceval-middle_school_politics	-	-	-	-
ceval-middle_school_geography	-	-	-	-
ceval-modern_chinese_history	-	-	-	-
ceval-ideological_and_moral_cultivation	-	-	-	-
ceval-logic	-	-	-	-
ceval-law	-	-	-	-
ceval-chinese_language_and_literature	-	-	-	-
ceval-art_studies	-	-	-	-
ceval-professional_tour_guide	-	-	-	-
ceval-legal_professional	-	-	-	-
ceval-high_school_chinese	-	-	-	-
ceval-high_school_history	-	-	-	-
ceval-middle_school_history	-	-	-	-
ceval-civil_servant	-	-	-	-
ceval-sports_science	-	-	-	-
ceval-plant_protection	-	-	-	-
ceval-basic_medicine	-	-	-	-
ceval-clinical_medicine	-	-	-	-
ceval-urban_and_rural_planner	-	-	-	-
ceval-accountant	-	-	-	-
ceval-fire_engineer	-	-	-	-
ceval-environmental_impact_assessment_engineer	-	-	-	-
ceval-tax_accountant	-	-	-	-
ceval-physician	-	-	-	-