

基础作业：

使用 OpenCompass 评测 InternLM2-Chat-7B 模型在 C-Eval 数据集上的性能  
Github 连不上，用了 Gtee。

查看支持的数据集和类型：

```
(opencompass) root@intern-studio:~/opencompass# python tools/list_configs.py internlm ceval
```

Model	Config Path
hf_internlm_7b	configs/models/hf_internlm/hf_internlm_7b.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
internlm_7b	configs/models/internlm/internlm_7b.py

  

Dataset	Config Path
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_ppl	configs/datasets/ceval/ceval_ppl.py
ceval_ppl_578f8d	configs/datasets/ceval/ceval_ppl_578f8d.py
ceval_ppl_93e5ce	configs/datasets/ceval/ceval_ppl_93e5ce.py

过程中：\*注意把 batchsize 改小

```
Loading checkpoint shards: 100%| 8/8 [00:11<00:00, 1.41s/it]
01/22 20:40:47 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-college_economics]
100%| 55/55 [00:00<00:00, 1325785.75it/s]
[2024-01-22 20:40:47,830] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100%| 14/14 [00:35<00:00, 2.54s/it]
01/22 20:41:23 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-accountant]
100%| 49/49 [00:00<00:00, 1533738.03it/s]
[2024-01-22 20:41:23,538] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100%| 13/13 [00:38<00:00, 2.98s/it]
01/22 20:42:02 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-tax_accountant]
100%| 49/49 [00:00<00:00, 1388654.70it/s]
[2024-01-22 20:42:02,547] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100%| 13/13 [00:35<00:00, 2.70s/it]
01/22 20:42:37 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-physician]
100%| 49/49 [00:00<00:00, 1427228.44it/s]
[2024-01-22 20:42:37,788] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100%| 13/13 [00:23<00:00, 1.80s/it]
01/22 20:43:01 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-civil_servant]
100%| 47/47 [00:00<00:00, 1418217.90it/s]
[2024-01-22 20:43:01,445] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100%| 12/12 [00:42<00:00, 3.55s/it]
01/22 20:43:47 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-urban_and_rural_planner]
```

评测结果：

ceval-urban_and_rural_planner	95b885	accuracy	gen
ceval-accountant	45.65	002837	accuracy gen
ceval-fire_engineer	26.53	bc23f5	accuracy gen
ceval-environmental_impact_assessment_engineer	22.58	c64e2d	accuracy gen
ceval-tax_accountant	64.52	3a5e3c	accuracy gen
ceval-physician	34.69	6e277d	accuracy gen
ceval-stem	44.9	-	naive_average gen
ceval-social-science	35.87	-	naive_average gen
ceval-humanities	52.2	-	naive_average gen
ceval-other	52.1	-	naive_average gen
ceval-hard	44.73	-	naive_average gen
ceval	24.57	-	naive_average gen
	44.32		

```
01/22 23:51:40 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240122_224139/summary/summary
_20240122_224139.txt
01/22 23:51:40 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240122_224139/summary/summary_202
40122_224139.csv
(opencompass) root@intern-studio:~/opencompass#
```

### 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

参考了大佬的过程：[第 6 节课作业\(2 班\) · InternLM/tutorial · Discussion #411 \(github.com\)](#)

首先部署模型

然后 python run.py configs/eval\_internlm2-7b-deploy2.0.py --debug --num-gpus 1

```
# config for internlm-chat-7b
internlm2_chat_7b = dict(
    type=TurboMindModel,
    abbr='internlm2-chat-7b-turbomind',
    path='/root/model/internlm7b_lmdeploy2.0/',
    engine_config=dict(session_len=2048,
                        max_batch_size=32,
                        rope_scaling_factor=1.0),
    gen_config=dict(top_k=1,
                    top_p=0.8,
                    temperature=1.0,
                    max_new_tokens=100),
    max_out_len=100,
    max_seq_len=2048,
    batch_size=32,
    concurrency=32,
    meta_template=internlm_meta_template,
    run_cfg=dict(num_gpus=1, num_procs=1),
)

models = [internlm2_chat_7b]
```

调小了些 batchsize。

最终结果：

```
01/29 22:35:53 - OpenCompass - INFO - time elapsed: 27.31s
01/29 22:35:54 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:35:54 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
in opencompass.tasks.openicl_eval
01/29 22:36:35 - OpenCompass - INFO - Task [internlm2-chat-7b-turbomind/ceval-marxism]: {'accuracy': 84.21052631578947}
01/29 22:36:35 - OpenCompass - INFO - time elapsed: 25.04s
01/29 22:36:36 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:36:36 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
in opencompass.tasks.openicl_eval
01/29 22:37:17 - OpenCompass - INFO - Task [internlm2-chat-7b-turbomind/ceval-mao_zedong_thought]: {'accuracy': 70.83333333333334}
01/29 22:37:17 - OpenCompass - INFO - time elapsed: 25.74s
01/29 22:37:18 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:37:18 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
in opencompass.tasks.openicl_eval
01/29 22:37:56 - OpenCompass - INFO - Task [internlm2-chat-7b-turbomind/ceval-education_science]: {'accuracy': 79.3103448275862}
01/29 22:37:56 - OpenCompass - INFO - time elapsed: 23.03s
01/29 22:37:57 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:37:57 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
in opencompass.tasks.openicl_eval
01/29 22:38:35 - OpenCompass - INFO - Task [internlm2-chat-7b-turbomind/ceval-teacher_qualification]: {'accuracy': 77.27272727272727}
01/29 22:38:35 - OpenCompass - INFO - time elapsed: 24.22s
01/29 22:38:36 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:38:36 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
in opencompass.tasks.openicl_eval
01/29 22:39:16 - OpenCompass - INFO - Task [internlm2-chat-7b-turbomind/ceval-high_school_politics]: {'accuracy': 89.47368421052632}
01/29 22:39:16 - OpenCompass - INFO - time elapsed: 26.68s
01/29 22:39:16 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:39:16 - OpenCompass - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found
```