

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：

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

```
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#
```