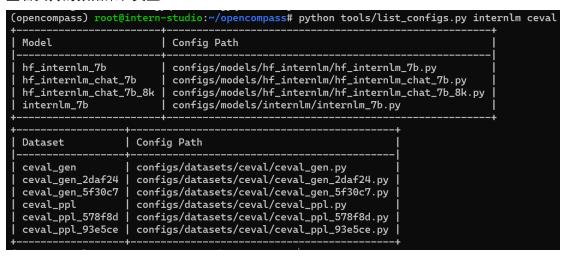
基础作业:

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:



过程中: *注意把 batchsize 改小

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_as		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 - <u>INFO</u> - write csv to /root/opencompass/outputs/default/20240122_224139/summary/summary_202				
40122_224139.csv (opencompass) root@intern-studio:~/opencompass#				
Ilcompass 0:bash*				"intern-studio" 10:51 23-Jan-24

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: <u>第 6 节课作业(2 班) · InternLM/tutorial · Discussion #411 (github.com)</u> 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 [internIm2-chat-7b-turbomind/ceval-marxism]: {'accuracy': 84.21052631578947}
01/29 22:36:36 - 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 - time elapsed: 25.74s
01/29 22:37:17 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' from "task" registry in "opencompass"
01/29 22:37:18 - OpenCompass - DEBUG - Get class 'OpenICLEvalTask' instance is built from registry, and its implementation can be found in opencompass.tasks.openicl_eval
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 - 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 - DEBUG - Task [internIm2-chat-7b-turbomind/ceval-education_science]: {'accuracy': 79.3103448275862}
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 - 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 - DEBUG - An 'OpenICLEvalTask' instance is built from registry, and its implementation can be found in opencompass: tasks.ope
```