

1 创建环境

```
1 bash
2 conda create --name opencompass --clone=/root/share/conda_envs/internlm-base
3 conda activate opencompass
4 git clone https://github.com/open-compass/opencompass
5 cd opencompass
6 pip install -e .
```

2 数据准备

```
1 # 解压评测数据集到data/处
2 cp /share/temp/datasets/OpenCompassData-core-20231110.zip /root/opencompass/
3 unzip OpenCompassData-core-20231110.zip
4
5 # 将会在opencompass下看到data文件夹
```

3 查看支持的数据集和模型

```
1 python tools/list_configs.py internlm ceval
```

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

Model	Config Path
hf_internlm2_20b	configs/models/hf_internlm/hf_internlm2_20b.py
hf_internlm2_7b	configs/models/hf_internlm/hf_internlm2_7b.py
hf_internlm2_chat_20b	configs/models/hf_internlm/hf_internlm2_chat_20b.py
hf_internlm2_chat_7b	configs/models/hf_internlm/hf_internlm2_chat_7b.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
ms_internlm_chat_7b_8k	configs/models/ms_internlm/ms_internlm_chat_7b_8k.py

Dataset	Config Path
ceval_clean_ppl	configs/datasets/ceval/ceval_clean_ppl.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_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
ceval_zero_shot_gen_bd40ef	configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py

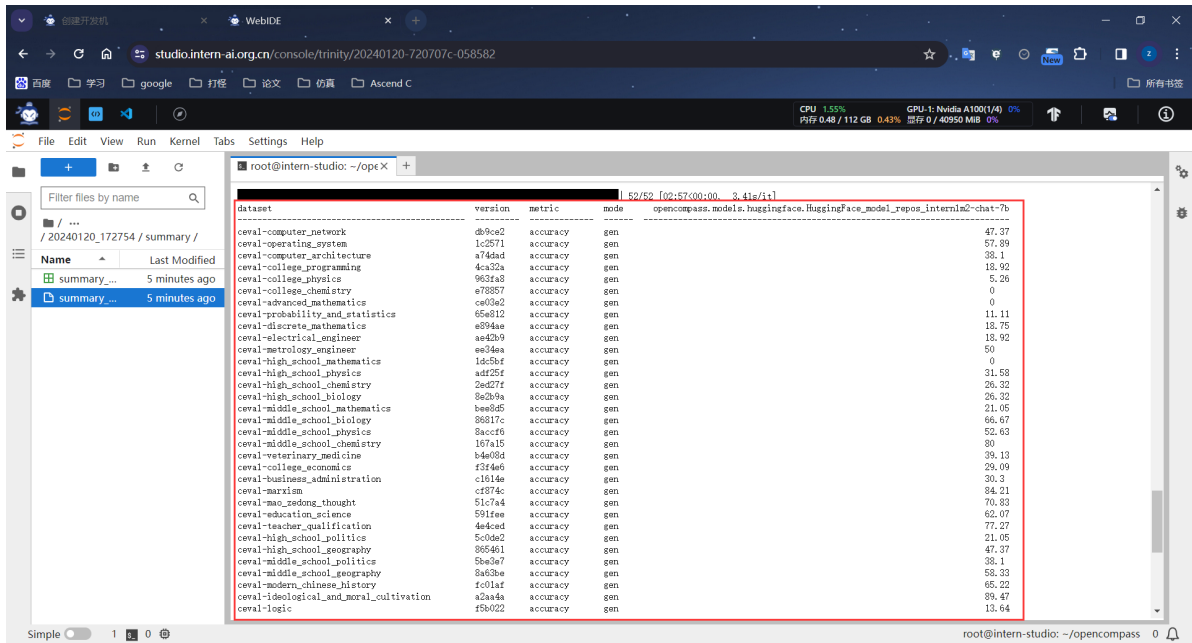
4 启动评测

```

1 python run.py \
2     --datasets ceval_gen \
3     --hf-path /root/share/model_repos/internlm2-chat-7b/ \
4     --tokenizer-path /root/share/model_repos/internlm2-chat-7b/ \
5     --tokenizer-kwarg padding_side='left' truncation='left'
trust_remote_code=True \
6     --model-kwarg trust_remote_code=True device_map='auto' \
7     --max-seq-len 2048 \
8     --max-out-len 16 \
9     --batch-size 4 \
10    --num-gpus 1 \

```

5 可视化结果展示



dataset	version	metric	mode	score
ceval-computer_network	d99ce2	accuracy	gen	47.37
ceval-operating_system	1c2571	accuracy	gen	57.89
ceval-computer_architecture	a74da4	accuracy	gen	38.1
ceval-college_programming	4ca32a	accuracy	gen	18.92
ceval-college_physics	9631a8	accuracy	gen	5.26
ceval-college_chemistry	e78857	accuracy	gen	0
ceval-advanced_mathematics	ce03e2	accuracy	gen	0
ceval-probability_and_statistics	65a812	accuracy	gen	11.11
ceval-discrete_mathematics	e894a0	accuracy	gen	18.75
ceval-electrical_engineer	ae42b9	accuracy	gen	18.92
ceval-metrology_engineer	ee34ea	accuracy	gen	50
ceval-high_school_mathematics	1dc5bf	accuracy	gen	0
ceval-high_school_physics	adr25f	accuracy	gen	31.58
ceval-high_school_chemistry	2e427f	accuracy	gen	26.32
ceval-high_school_biology	8e2b9a	accuracy	gen	26.32
ceval-middle_school_mathematics	bee8d5	accuracy	gen	21.05
ceval-middle_school_biology	96817c	accuracy	gen	66.67
ceval-middle_school_physics	8ac2cf	accuracy	gen	52.63
ceval-middle_school_chemistry	167a15	accuracy	gen	80
ceval-veterinary_medicine	b4e08d	accuracy	gen	39.13
ceval-college_economics	f3f4e0	accuracy	gen	29.09
ceval-business_administration	c1614e	accuracy	gen	30.3
ceval-marxism	cf874c	accuracy	gen	84.21
ceval-mao_tsedong_thought	51c7a4	accuracy	gen	70.83
ceval-education_science	591fee	accuracy	gen	62.07
ceval-teacher_qualification	4a4ced	accuracy	gen	77.27
ceval-high_school_politics	5c0de2	accuracy	gen	21.05
ceval-high_school_geography	8054b1	accuracy	gen	47.37
ceval-middle_school_politics	9e3e37	accuracy	gen	38.1
ceval-middle_school_geography	8a83be	accuracy	gen	58.33
ceval-modern_chinese_history	f001af	accuracy	gen	65.22
ceval-ideological_and_moral_cultivation	a2aa4a	accuracy	gen	89.47
ceval-logic	f5b022	accuracy	gen	13.64