

1 环境搭建

面向GPU的环境搭建

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

有部分第三方功能,如代码能力基准测试 Humaneval 以及 Llama格式的模型评测,可能需要额外步骤才能正常运行,如需评测,详细步骤请参考[安装指南](#)。

2 数据准备和模型下载

```
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
```

查看支持的数据集和模型:

```
1 # 列出所有跟 internlm 及 ceval 相关的配置
2 python tools/list_configs.py internlm ceval
```

结果如下:

```
1 +-----+
2 | Model | Config Path |
3 |-----+-----|
4 | hf_internlm2_20b | configs/models/hf_internlm/hf_internlm2_20b.py |
5 | hf_internlm2_7b | configs/models/hf_internlm/hf_internlm2_7b.py |
6 | hf_internlm2_chat_20b | configs/models/hf_internlm/hf_internlm2_chat_20b.py |
7 | hf_internlm2_chat_7b | configs/models/hf_internlm/hf_internlm2_chat_7b.py |
8 | hf_internlm_20b | configs/models/hf_internlm/hf_internlm_20b.py |
9 | hf_internlm_7b | configs/models/hf_internlm/hf_internlm_7b.py |
10 | hf_internlm_chat_20b | configs/models/hf_internlm/hf_internlm_chat_20b.py |
```

```

11 | hf_internlm_chat_7b      |
    | configs/models/hf_internlm/hf_internlm_chat_7b.py      |
12 | hf_internlm_chat_7b_8k   |
    | configs/models/hf_internlm/hf_internlm_chat_7b_8k.py    |
13 | hf_internlm_chat_7b_v1_1 |
    | configs/models/hf_internlm/hf_internlm_chat_7b_v1_1.py |
14 | internlm_7b              | configs/models/internlm/internlm_7b.py
    |
15 | ms_internlm_chat_7b_8k   |
    | configs/models/ms_internlm/ms_internlm_chat_7b_8k.py    |
16 +-----+-----+
    |-----+
17 +-----+-----+
    |-----+
18 | Dataset                  | Config Path
    |
19 |-----+-----+
    |-----|
20 | ceval_clean_pp1         | configs/datasets/ceval/ceval_clean_pp1.py
    |
21 | ceval_gen               | configs/datasets/ceval/ceval_gen.py
    |
22 | ceval_gen_2daf24        | configs/datasets/ceval/ceval_gen_2daf24.py
    |
23 | ceval_gen_5f30c7        | configs/datasets/ceval/ceval_gen_5f30c7.py
    |
24 | ceval_pp1               | configs/datasets/ceval/ceval_pp1.py
    |
25 | ceval_pp1_578f8d        | configs/datasets/ceval/ceval_pp1_578f8d.py
    |
26 | ceval_pp1_93e5ce        | configs/datasets/ceval/ceval_pp1_93e5ce.py
    |
27 | ceval_zero_shot_gen_bd40ef |
    | configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py |
28 +-----+-----+
    |-----+

```

接着下载internlm2-chat-7b模型（记得下载huggingface_hub）：

```

1 | import os
2 | # 设置环境变量
3 | os.environ['HF_ENDPOINT'] = 'https://hf-mirror.com'
4 | # 下载模型
5 | os.system('huggingface-cli download --resume-download internlm/internlm2-
    chat-7b --local-dir /root/opencompass/internlm2_chat_7b')

```

3 启动评测

GPU配置一定要足够：本次测评采用2* 1/4 A100，否则部分指标会因为现存不够无法测出来。

```

1 python run.py --datasets ceval_gen --hf-path
  /root/opencompass/internlm2_chat_7b --tokenizer-path
  /root/opencompass/internlm2_chat_7b --tokenizer-kwarg padding_side='left'
  truncation='left' trust_remote_code=True --model-kwarg
  trust_remote_code=True device_map='auto' --max-seq-len 2048 --max-out-len 16
  --batch-size 4 --num-gpus 1 --debug

```

--datasets: 需要测试的数据集

--hf-path: hf模型的路径

--tokenizer-path: 分词器路径

--max-out-len: 设置比较小是为了方便客观题回答

--debug: 调试模式, 终端会在终端打印log的所有信息 (输出信息均在/opencompass/output下)

4 评测结果

gen评测结果:

1	dataset	version	metric
	mode		
	opencompass.models.huggingface.HuggingFace_opencompass_internlm2_chat_7b		
2	-----	-----	-----

3	ceval-computer_network	db9ce2	accuracy
	gen		
	47.37		
4	ceval-operating_system	1c2571	accuracy
	gen		
	57.89		
5	ceval-computer_architecture	a74dad	accuracy
	gen		
	42.86		
6	ceval-college_programming	4ca32a	accuracy
	gen		
	51.35		
7	ceval-college_physics	963fa8	accuracy
	gen		
	36.84		
8	ceval-college_chemistry	e78857	accuracy
	gen		
	33.33		
9	ceval-advanced_mathematics	ce03e2	accuracy
	gen		
	15.79		
10	ceval-probability_and_statistics	65e812	accuracy
	gen		
	27.78		
11	ceval-discrete_mathematics	e894ae	accuracy
	gen		
	18.75		

12	ceval-electrical_engineer gen 40.54	ae42b9	accuracy
13	ceval-metrology_engineer gen 58.33	ee34ea	accuracy
14	ceval-high_school_mathematics gen 44.44	1dc5bf	accuracy
15	ceval-high_school_physics gen 47.37	adf25f	accuracy
16	ceval-high_school_chemistry gen 52.63	2ed27f	accuracy
17	ceval-high_school_biology gen 26.32	8e2b9a	accuracy
18	ceval-middle_school_mathematics gen 26.32	bee8d5	accuracy
19	ceval-middle_school_biology gen 66.67	86817c	accuracy
20	ceval-middle_school_physics gen 57.89	8accf6	accuracy
21	ceval-middle_school_chemistry gen 95	167a15	accuracy
22	ceval-veterinary_medicine gen 39.13	b4e08d	accuracy
23	ceval-college_economics gen 47.27	f3f4e6	accuracy
24	ceval-business_administration gen 51.52	c1614e	accuracy
25	ceval-marxism gen 84.21	cf874c	accuracy
26	ceval-mao_zedong_thought gen 70.83	51c7a4	accuracy
27	ceval-education_science gen 72.41	591fee	accuracy
28	ceval-teacher_qualification gen 79.55	4e4ced	accuracy
29	ceval-high_school_politics gen 21.05	5c0de2	accuracy
30	ceval-high_school_geography gen 47.37	865461	accuracy

31	ceval-middle_school_politics gen 42.86	5be3e7	accuracy
32	ceval-middle_school_geography gen 58.33	8a63be	accuracy
33	ceval-modern_chinese_history gen 65.22	fc01af	accuracy
34	ceval-ideological_and_moral_cultivation gen 89.47	a2aa4a	accuracy
35	ceval-logic gen 54.55	f5b022	accuracy
36	ceval-law gen 41.67	a110a1	accuracy
37	ceval-chinese_language_and_literature gen 56.52	0f8b68	accuracy
38	ceval-art_studies gen 69.7	2a1300	accuracy
39	ceval-professional_tour_guide gen 86.21	4e673e	accuracy
40	ceval-legal_professional gen 43.48	ce8787	accuracy
41	ceval-high_school_chinese gen 68.42	315705	accuracy
42	ceval-high_school_history gen 75	7eb30a	accuracy
43	ceval-middle_school_history gen 68.18	48ab4a	accuracy
44	ceval-civil_servant gen 55.32	87d061	accuracy
45	ceval-sports_science gen 73.68	70f27b	accuracy
46	ceval-plant_protection gen 77.27	8941f9	accuracy
47	ceval-basic_medicine gen 63.16	c409d6	accuracy
48	ceval-clinical_medicine gen 45.45	49e82d	accuracy
49	ceval-urban_and_rural_planner gen 58.7	95b885	accuracy

50	ceval-accountant gen 44.9	002837	accuracy
51	ceval-fire_engineer gen 38.71	bc23f5	accuracy
52	ceval-environmental_impact_assessment_engineer gen 45.16	c64e2d	accuracy
53	ceval-tax_accountant gen 51.02	3a5e3c	accuracy
54	ceval-physician gen 51.02	6e277d	accuracy
55	ceval-stem gen 44.33	-	naive_average
56	ceval-social-science gen 57.54	-	naive_average
57	ceval-humanities gen 65.31	-	naive_average
58	ceval-other gen 54.94	-	naive_average
59	ceval-hard gen 34.62	-	naive_average
60	ceval gen 53.55	-	naive_average
61	01/28 10:36:59 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240128_100530/summary/summary_20240128_100530.txt		
62	01/28 10:36:59 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240128_100530/summary/summary_20240128_100530.csv		

ppl评测结果:

1	dataset	version	metric
	mode		
	opencompass.models.huggingface.HuggingFace_opencompass_internlm2_chat_7b		
2	-----	-----	-----

3	ceval-computer_network ppl 63.16	9b9417	accuracy
4	ceval-operating_system ppl 63.16	b2b8cf	accuracy
5	ceval-computer_architecture ppl 52.38	1bd275	accuracy

6	ceval-college_programming pp1 56.76	2d0833	accuracy
7	ceval-college_physics pp1 52.63	fb7e04	accuracy
8	ceval-college_chemistry pp1 37.50	916b7d	accuracy
9	ceval-advanced_mathematics pp1 42.11	5cad2a	accuracy
10	ceval-probability_and_statistics pp1 38.89	a6b30e	accuracy
11	ceval-discrete_mathematics pp1 31.25	68be68	accuracy
12	ceval-electrical_engineer pp1 40.54	056c2e	accuracy
13	ceval-metrology_engineer pp1 62.50	4a757a	accuracy
14	ceval-high_school_mathematics pp1 27.78	a8ed21	accuracy
15	ceval-high_school_physics pp1 42.11	e1fc86	accuracy
16	ceval-high_school_chemistry pp1 52.63	9021c6	accuracy
17	ceval-high_school_biology pp1 36.84	c7f5a1	accuracy
18	ceval-middle_school_mathematics pp1 57.89	213989	accuracy
19	ceval-middle_school_biology pp1 80.95	ce0420	accuracy
20	ceval-middle_school_physics pp1 63.16	78f3af	accuracy
21	ceval-middle_school_chemistry pp1 95.00	d071d2	accuracy
22	ceval-veterinary_medicine pp1 47.83	cd3a07	accuracy
23	ceval-college_economics pp1 47.27	a35346	accuracy
24	ceval-business_administration pp1 51.52	69dd6a	accuracy
25	ceval-marxism pp1 84.21	283ce0	accuracy
26	ceval-mao_zedong_thought pp1 70.83	f38cd1	accuracy
27	ceval-education_science pp1 79.31	fbf65c	accuracy
28	ceval-teacher_qualification pp1 77.27	c77f1f	accuracy
29	ceval-high_school_politics pp1 94.74	bbac37	accuracy
30	ceval-high_school_geography pp1 68.42	730a30	accuracy
31	ceval-middle_school_politics pp1 76.19	15b2d7	accuracy
32	ceval-middle_school_geography pp1 66.67	b00167	accuracy
33	ceval-modern_chinese_history pp1 78.26	5a04cd	accuracy
34	ceval-ideological_and_moral_cultivation pp1 94.74	0829ff	accuracy

35	ceval-logic pp1 50.00	c9c394	accuracy	
36	ceval-law pp1 45.83	cbd3c5	accuracy	
37	ceval-chinese_language_and_literature pp1 60.87	716ab3	accuracy	
38	ceval-art_studies pp1 72.73	476114	accuracy	
39	ceval-professional_tour_guide pp1 86.21	70f30f	accuracy	
40	ceval-legal_professional pp1 39.13	f19cf5	accuracy	
41	ceval-high_school_chinese pp1 68.42	931614	accuracy	
42	ceval-high_school_history pp1 75.00	4d6364	accuracy	
43	ceval-middle_school_history pp1 77.27	7f6356	accuracy	
44	ceval-civil_servant -	-	-	-
45	ceval-sports_science pp1 73.68	192553	accuracy	
46	ceval-plant_protection pp1 77.27	f7ff86	accuracy	
47	ceval-basic_medicine pp1 63.16	a95a09	accuracy	
48	ceval-clinical_medicine pp1 54.55	664b54	accuracy	
49	ceval-urban_and_rural_planner -	-	-	-
50	ceval-accountant -	-	-	-
51	ceval-fire_engineer pp1 41.94	bb924d	accuracy	
52	ceval-environmental_impact_assessment_engineer pp1 48.39	d59200	accuracy	
53	ceval-tax_accountant -	-	-	-
54	ceval-physician -	-	-	-
55	ceval-stem pp1 52.25	-	naive_average	
56	ceval-social-science pp1 71.64	-	naive_average	
57	ceval-humanities pp1 68.04	-	naive_average	
58	ceval-hard pp1 40.61	-	naive_average	
59	01/28 10:22:47 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240128_093748/summary/summary_20240128_093748.txt			
60	01/28 10:22:47 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240128_093748/summary/summary_20240128_093748.csv			

