作业：OpenCompass

1、查看支持的数据集和模型

2、启动评测

输入如下shell指令，评测internlm2-chat-1_8b模型。

```
python run.py --datasets ceval_gen \
--hf-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b \
--tokenizer-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b \
--tokenizer-kwargs padding_side='left' truncation='left' trust_remote_code=True \
--model-kwargs device_map='auto' trust_remote_code=True \
--max-seq-len 2048 \
--max-out-len 16 \
--batch-size 4  \
--num-gpus 1 \
--debug
```

3、评测结果

dataset	version	metric	mode	opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1_8b
ceval-computer_network	db9ce2	accuracy	gen	47.37
ceval-operating_system	1c2571	accuracy	gen	47.37
ceval-computer_architecture	a74dad	accuracy	gen	23.81
ceval-college_programming	4ca32a	accuracy	gen	13.51
ceval-college_physics	963fa8	accuracy	gen	42.11
ceval-college_chemistry	e78857	accuracy	gen	33.33
ceval-advanced_mathematics	ce03e2	accuracy	gen	15.79
ceval-probability_and_statistics	65e812	accuracy	gen	22.22
ceval-discrete_mathematics	e894ae	accuracy	gen	25
ceval-electrical_engineer	ae42b9	accuracy	gen	29.73
ceval-metrology_engineer	ee34ea	accuracy	gen	54.17
ceval-high_school_mathematics	1dc5bf	accuracy	gen	16.67
ceval-high_school_physics	adf25f	accuracy	gen	42.11
ceval-high_school_chemistry	2ed27f	accuracy	gen	47.37
ceval-high_school_biology	8e2b9a	accuracy	gen	26.32
ceval-middle_school_mathematics	bee8d5	accuracy	gen	36.84
ceval-middle_school_biology	86817c	accuracy	gen	80.95
ceval-middle_school_physics	8accf6	accuracy	gen	47.37
ceval-middle_school_chemistry	167a15	accuracy	gen	80
ceval-veterinary_medicine	b4e08d	accuracy	gen	43.48
ceval-college_economics	f3f4e6	accuracy	gen	34.55
ceval-business_administration	c1614e	accuracy	gen	36.36
ceval-marxism	cf874c	accuracy	gen	68.42
ceval-mao_zedong_thought	51c7a4	accuracy	gen	70.83
ceval-education_science	591fee	accuracy	gen	55.17
ceval-teacher_qualification	4e4ced	accuracy	gen	59.09
ceval-high_school_politics	5c0de2	accuracy	gen	57.89
ceval-high_school_geography	865461	accuracy	gen	47.37
ceval-middle_school_politics	5be3e7	accuracy	gen	71.43
ceval-middle_school_geography	8a63be	accuracy	gen	75
ceval-modern_chinese_history	fc01af	accuracy	gen	52.17
ceval-ideological_and_moral_cultivation	a2aa4a	accuracy	gen	73.68
ceval-logic	f5b022	accuracy	gen	31.82
ceval-law	a110a1	accuracy	gen	29.17
ceval-chinese_language_and_literature	0f8b68	accuracy	gen	47.83
ceval-art_studies	2a1300	accuracy	gen	42.42
ceval-professional_tour_guide	4e673e	accuracy	gen	51.72
ceval-legal_professional	ce8787	accuracy	gen	34.78
ceval-high_school_chinese	315705	accuracy	gen	36.84
ceval-high_school_history	7eb30a	accuracy	gen	65
ceval-middle_school_history	48ab4a	accuracy	gen	86.36
ceval-civil_servant	87d061	accuracy	gen	42.55
ceval-sports_science	70f27b	accuracy	gen	52.63
ceval-plant_protection	8941f9	accuracy	gen	40.91
ceval-basic_medicine	c409d6	accuracy	gen	68.42
ceval-clinical_medicine	49e82d	accuracy	gen	31.82
ceval-urban_and_rural_planner	95b885	accuracy	gen	47.83
ceval-accountant	002837	accuracy	gen	36.73
ceval-fire_engineer	bc23f5	accuracy	gen	38.71
ceval-environmental_impact_assessment_engineer	c64 e2d	accuracy	gen	51.61
ceval-tax_accountant	3a5e3c	accuracy	gen	36.73
ceval-physician	6e277d	accuracy	gen	42.86
ceval-stem	-	naive_average	gen	38.78
ceval-social-science	-	naive_average	gen	57.61
ceval-humanities	-	naive_average	gen	50.16
ceval-other	-	naive_average	gen	44.62
ceval-hard	-	naive_average	gen	30.57
ceval	-	naive_average	gen	46.04

```
04/26 19:43:25 - OpenCompass -      - write summary to /root/opencompass/outputs/default/20240426_193517/summary/summary_20240426_193517.txt
04/26 19:43:25 - OpenCompass -      - write csv to /root/opencompass/outputs/default/20240426_193517/summary/summary_20240426_193517.csv
(opencompass) root@intern-studio-117673:~/opencompass#
```

评测后保存summary的csv和txt文件如下所示：

	dataset	version	metric	mode	ory_internlm2-chat-1_8b
1	eval-computer_network	db9ce2	accuracy	gen	47.37
2	ceval-operating_system	1c2571	accuracy	gen	47.37
3	l-computer_architecture	a74dad	accuracy	gen	23.81
4	il-college_programming	4ca32a	accuracy	gen	13.51
5	ceval-college_physics	963fa8	accuracy	gen	42.11
6	ceval-college_chemistry	e78857	accuracy	gen	33.33
7	advanced_mathematics	ce03e2	accuracy	gen	15.79
8	robability_and_statistics	65e812	accuracy	gen	22.22
9	al-discrete_mathematics	e894ae	accuracy	gen	25.00
10	eval-electrical_engineer	ae42b9	accuracy	gen	29.73
11	val-metrology_engineer	ee34ea	accuracy	gen	54.17
12	gh_school_mathematics	1dc5bf	accuracy	gen	16.67
13	val-high_school_physics	adf25f	accuracy	gen	42.11
14	-high_school_chemistry	2ed27f	accuracy	gen	47.37
15	val-high_school_biology	8e2b9a	accuracy	gen	26.32
16	dle_school_mathematics	bee8d5	accuracy	gen	36.84
17	-middle_school_biology	86817c	accuracy	gen	80.95
18	-middle_school_physics	8accf6	accuracy	gen	47.37
19	niddle_school_chemistry	167a15	accuracy	gen	80.00
20	val-veterinary_medicine	b4e08d	accuracy	gen	43.48
21	eval-college_economics	f3f4e6	accuracy	gen	34.55
22	business_administration	c1614e	accuracy	gen	36.36
23	ceval-marxism	cf874c	accuracy	gen	68.42
24	l-mao_zedong_thought	51c7a4	accuracy	gen	70.83
25	ceval-education_science	591fee	accuracy	gen	55.17
26	val-teacher_qualification	4e4ced	accuracy	gen	59.09
27	val-high_school_politics	5c0de2	accuracy	gen	57.89
28	high_school_geography	865461	accuracy	gen	47.37
29	l-middle_school_politics	5be3e7	accuracy	gen	71.43
30	ddle_school_geography	8a63be	accuracy	gen	75.00
31	modern_chinese_history	fc01af	accuracy	gen	52.17
32	l_and_moral_cultivation	a2aa4a	accuracy	gen	73.68
33	ceval-logic	f5b022	accuracy	gen	31.82
34	ceval-law	a110a1	accuracy	gen	29.17
35	language_and_literature	0f8b68	accuracy	gen	47.83
36	ceval-art_studies	2a1300	accuracy	gen	42.42
37	professional_tour_guide	4e673e	accuracy	gen	51.72
38	ceval-legal_professional	ce8787	accuracy	gen	34.78
39	val-high_school_chinese	315705	accuracy	gen	36.84
40	val-high_school_history	7eb30a	accuracy	gen	65.00
41	l-middle_school_history	48ab4a	accuracy	gen	86.36
42	ceval-civil_servant	87d061	accuracy	gen	42.55
43	ceval-sports_science	70f27b	accuracy	gen	52.63
44	ceval-plant_protection	8941f9	accuracy	gen	40.91
45	ceval-basic_medicine	c409d6	accuracy	gen	68.42
46	ceval-clinical_medicine	49e82d	accuracy	gen	31.82
47	urban_and_rural_planner	95b885	accuracy	gen	47.83

root@intern-studio-11767 × | summary_20240426_1935· × | summary_20240426_1935· × | requirements.txt × | +

Filter files by name

/ ··· / 20240426_193517 / summary /

Name	Last Modified
summary_20240426_193517.csv	8 minutes ago
summary_20240426_193517.txt	8 minutes ago

```
1  20240426_193517
2  tabulate format
3  ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
4  dataset                                        version   metric         mode    opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1_8b
5  ---------------------------------------------  --------  -------------  ------  ----------------------------------------------------------------------------------------------
6  ceval-computer_network                         db9ce2    accuracy       gen                                                                                          47.37
7  ceval-operating_system                         1c2571    accuracy       gen                                                                                          47.37
8  ceval-computer_architecture                    a74dad    accuracy       gen                                                                                          23.81
9  ceval-college_programming                      4ca32a    accuracy       gen                                                                                          13.51
10 ceval-college_physics                          963fa8    accuracy       gen                                                                                          42.11
11 ceval-college_chemistry                        e78857    accuracy       gen                                                                                          33.33
12 ceval-advanced_mathematics                     ce03e2    accuracy       gen                                                                                          15.79
13 ceval-probability_and_statistics               65e812    accuracy       gen                                                                                          22.22
14 ceval-discrete_mathematics                     e894ae    accuracy       gen                                                                                          25
15 ceval-electrical_engineer                      ae42b9    accuracy       gen                                                                                          29.73
16 ceval-metrology_engineer                       ee34ea    accuracy       gen                                                                                          54.17
17 ceval-high_school_mathematics                  1dc5bf    accuracy       gen                                                                                          16.67
18 ceval-high_school_physics                      adf25f    accuracy       gen                                                                                          42.11
19 ceval-high_school_chemistry                    2ed27f    accuracy       gen                                                                                          47.37
20 ceval-high_school_biology                      8e2b9a    accuracy       gen                                                                                          26.32
21 ceval-middle_school_mathematics                bee8d5    accuracy       gen                                                                                          36.84
22 ceval-middle_school_biology                    86817c    accuracy       gen                                                                                          80.95
23 ceval-middle_school_physics                    8accf6    accuracy       gen                                                                                          47.37
24 ceval-middle_school_chemistry                  167a15    accuracy       gen                                                                                          80
25 ceval-veterinary_medicine                      b4e08d    accuracy       gen                                                                                          43.48
26 ceval-college_economics                        f3f4e6    accuracy       gen                                                                                          34.55
27 ceval-business_administration                  c1614e    accuracy       gen                                                                                          36.36
28 ceval-marxism                                  cf874c    accuracy       gen                                                                                          68.42
29 ceval-mao_zedong_thought                       51c7a4    accuracy       gen                                                                                          70.83
30 ceval-education_science                        591fee    accuracy       gen                                                                                          55.17
31 ceval-teacher_qualification                    4e4ced    accuracy       gen                                                                                          59.09
32 ceval-high_school_politics                     5c0de2    accuracy       gen                                                                                          57.89
33 ceval-high_school_geography                    865461    accuracy       gen                                                                                          47.37
34 ceval-middle_school_politics                   5be3e7    accuracy       gen                                                                                          71.43
35 ceval-middle_school_geography                  8a63be    accuracy       gen                                                                                          75
36 ceval-modern_chinese_history                   fc01af    accuracy       gen                                                                                          52.17
37 ceval-ideological_and_moral_cultivation        a2aa4a    accuracy       gen                                                                                          73.68
38 ceval-logic                                    f5b022    accuracy       gen                                                                                          31.82
39 ceval-law                                      a110a1    accuracy       gen                                                                                          29.17
40 ceval-chinese_language_and_literature          0f8b68    accuracy       gen                                                                                          47.83
41 ceval-art_studies                              2a1300    accuracy       gen                                                                                          42.42
42 ceval-professional_tour_guide                  4e673e    accuracy       gen                                                                                          51.72
43 ceval-legal_professional                       ce8787    accuracy       gen                                                                                          34.78
44 ceval-high_school_chinese                      315705    accuracy       gen                                                                                          36.84
45 ceval-high_school_history                      7eb30a    accuracy       gen                                                                                          65
46 ceval-middle_school_history                    48ab4a    accuracy       gen                                                                                          86.36
47 ceval-civil_servant                            87d061    accuracy       gen                                                                                          42.55
48 ceval-sports_science                           70f27b    accuracy       gen                                                                                          52.63
49 ceval-plant_protection                         8941f9    accuracy       gen                                                                                          40.91
50 ceval-basic_medicine                           c409d6    accuracy       gen                                                                                          68.42
51 ceval-clinical_medicine                        49e82d    accuracy       gen                                                                                          31.82
52 ceval-urban_and_rural_planner                  95b885    accuracy       gen                                                                                          47.83
53 ceval-accountant                               002837    accuracy       gen                                                                                          36.73
54 ceval-fire_engineer                            bc23f5    accuracy       gen                                                                                          38.71
55 ceval-environmental_impact_assessment_engineer c64e2d    accuracy       gen                                                                                          51.61
56 ceval-tax_accountant                           3a5e3c    accuracy       gen                                                                                          36.73
57 ceval-physician                                6e277d    accuracy       gen                                                                                          42.86
58 ceval-stem                                     -         naive_average  gen                                                                                          38.78
59 ceval-social-science                           -         naive_average  gen                                                                                          57.61
60 ceval-humanities                               -         naive_average  gen                                                                                          50.16
61 ceval-other                                    -         naive_average  gen                                                                                          44.62
62 ceval-hard                                     -         naive_average  gen                                                                                          30.57
63 ceval                                          -         naive_average  gen                                                                                          46.04
64 $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
```