

Lesson 6 作业

OpenCompass 大模型评测

基础作业

- 使用 OpenCompass 评测 InternLM2-Chat-7B 模型在 C-Eval 数据集上的性能

```
(base) root@intern-studio-40069509:/opt/jupyterlab# conda create
--name opencompass --clone=/root/share/conda_envs/internlm-base
Retrieving notices: ...working... done
Source: /root/share/conda_envs/internlm-base
Destination: /root/.conda/envs/opencompass
Packages: 96
Files: 0
```

Downloading and Extracting Packages:

Downloading and Extracting Packages:

```
Preparing transaction: done
Verifying transaction: done
Executing transaction: done
#
# To activate this environment, use
#
#     $ conda activate opencompass
#
# To deactivate an active environment, use
#
#     $ conda deactivate
```

```
(opencompass) root@intern-studio-40069509:~# cd opencompass
bash: cd: opencompass: No such file or directory
(opencompass) root@intern-studio-40069509:~# git clone https://github.com/open-compass/opencompass
Cloning into 'opencompass'...
remote: Enumerating objects: 8422, done.
remote: Counting objects: 100% (989/989), done.
remote: Compressing objects: 100% (578/578), done.
remote: Total 8422 (delta 580), reused 672 (delta 401), pack-reused 7433
Receiving objects: 100% (8422/8422), 3.27 MiB | 5.14 MiB/s, done.
Resolving deltas: 100% (4890/4890), done.
Updating files: 100% (1847/1847), done.
(opencompass) root@intern-studio-40069509:~# cd opencompass
(opencompass) root@intern-studio-40069509:~/opencompass# pip install -e .
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Obtaining file:///root/opencompass
Preparing metadata (setup.py) ... done
```

```
(opencompass) root@intern-studio-40069509:~/opencompass# python tools/list_configs.py internlm ceval
```

Model	Config Path
hf_internlm2_1_8b	configs/models/hf_internlm/hf_internlm2_1_8b.py
hf_internlm2_20b	configs/models/hf_internlm/hf_internlm2_20b.py
hf_internlm2_7b	configs/models/hf_internlm/hf_internlm2_7b.py
hf_internlm2_base_20b	configs/models/hf_internlm/hf_internlm2_base_20b.py
hf_internlm2_base_7b	configs/models/hf_internlm/hf_internlm2_base_7b.py
hf_internlm2_chat_1_8b	configs/models/hf_internlm/hf_internlm2_chat_1_8b.py
hf_internlm2_chat_1_8b_sft	configs/models/hf_internlm/hf_internlm2_chat_1_8b_sft.py
hf_internlm2_chat_20b	configs/models/hf_internlm/hf_internlm2_chat_20b.py
hf_internlm2_chat_20b_sft	configs/models/hf_internlm/hf_internlm2_chat_20b_sft.py
hf_internlm2_chat_20b_with_system	configs/models/hf_internlm/hf_internlm2_chat_20b_with_system.py
hf_internlm2_chat_7b	configs/models/hf_internlm/hf_internlm2_chat_7b.py
hf_internlm2_chat_7b_sft	configs/models/hf_internlm/hf_internlm2_chat_7b_sft.py
hf_internlm2_chat_7b_with_system	configs/models/hf_internlm/hf_internlm2_chat_7b_with_system.py
hf_internlm2_chat_math_20b	configs/models/hf_internlm/hf_internlm2_chat_math_20b.py
hf_internlm2_chat_math_20b_with_system	configs/models/hf_internlm/hf_internlm2_chat_math_20b_with_system.py
hf_internlm2_chat_math_7b	configs/models/hf_internlm/hf_internlm2_chat_math_7b.py
hf_internlm2_chat_math_7b_with_system	configs/models/hf_internlm/hf_internlm2_chat_math_7b_with_system.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
lmdeploy_internlm2_chat_20b	configs/models/hf_internlm/lmdeploy_internlm2_chat_20b.py
lmdeploy_internlm2_chat_7b	configs/models/hf_internlm/lmdeploy_internlm2_chat_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_contamination_ppl_810ec6	configs/datasets/contamination/ceval_contamination_ppl_810ec6.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_internal_ppl_1cd8bf	configs/datasets/ceval/ceval_internal_ppl_1cd8bf.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

```
(opencompass) root@intern-studio-40069509:~/opencompass# python run.py --datasets ceval_gen --hf-path /share/temp/model_repos/internlm-chat-7b/ --tokenizer-path /share
/temp/model_repos/internlm-chat-7b/ --tokenizer-kwags padding_side='left' truncation='left' trust_remote_code=True --model-kwags trust_remote_code=True device_map='a
uto' --max-seq-len 2048 --max-out-len 16 --batch-size 4 --num-gpus 1 --debug
04/13 11:08:24 - OpenCompass - Loading ceval_gen: configs/datasets/ceval/ceval_gen.py
04/13 11:08:24 - OpenCompass - Loading example: configs/summarizers/example.py
04/13 11:08:24 - OpenCompass - WARNING - SlurmRunner is not used, so the partition argument is ignored.
04/13 11:08:24 - OpenCompass - Partitioned into 1 tasks.
04/13 11:08:43 - OpenCompass - Task [opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-college_economics, opencompass.models.huggingf
ace.HuggingFace_model_repos_internlm-chat-7b/ceval-accountant, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-tax_accountant, opencompa
ss.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-physician, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-civil_servant
, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-urban_and_rural_planner, opencompass.models.huggingface.HuggingFace_model_repos_internlm
-chat-7b/ceval-teacher_qualification, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-college_programming, opencompass.models.huggingface.Hu
gingFace_model_repos_internlm-chat-7b/ceval-electrical_engineer, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-business_administration,
opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-art_studies, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval
-fire_engineer, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-environmental_impact_assessment_engineer, opencompass.models.huggingface.Hu
gingFace_model_repos_internlm-chat-7b/ceval-education_science, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-professional_tour_guide, op
encompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-college_chemistry, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/c
eval-metrology_engineer, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-mao_zedong_thought, opencompass.models.huggingface.HuggingFace_mod
el_repos_internlm-chat-7b/ceval-law, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-veterinary_medicine, opencompass.models.huggingface.Hu
gingFace_model_repos_internlm-chat-7b/ceval-modern_chinese_history, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-chinese_language_and_
literature, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-legal_professional, opencompass.models.huggingface.HuggingFace_model_repos_inte
rnlm-chat-7b/ceval-logic, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-middle_school_history, opencompass.models.huggingface.HuggingFace
_model_repos_internlm-chat-7b/ceval-plant_protection, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-clinical_medicine, opencompass.models
.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-computer_architecture, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-middle_
school_biology, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-middle_school_politics, opencompass.models.huggingface.HuggingFace_model_re
pos_internlm-chat-7b/ceval-middle_school_chemistry, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_history, opencompass.models
.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-computer_network, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-operating_sy
stem, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-college_physics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat
-7b/ceval-advanced_mathematics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_physics, opencompass.models.huggingface.Hugging
Face_model_repos_internlm-chat-7b/ceval-high_school_chemistry, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-marxism, opencompass.models.h
uggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_politics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_geography, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_chinese, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-basic_medicine, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-probability_and_statistics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-high_school_mathematics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-discrete_mathematics, opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b/ceval-middle_school_geography]
Loading checkpoint shards: 0% | 0/8 [00:00<?, ?it/s]/root/.conda/envs/opencompass/lib/python3.10/site-packages/torch/_utils.py:776: UserWarning: Type
dStorage is deprecated. It will be removed in the future and UntypedStorage will be the only storage class. This should only matter to you if you are using storages di
rectly. To access UntypedStorage directly, use tensor.untyped_storage() instead of tensor.storage()
```

dataset	version	metric	mode	opencompass.models.huggingface.HuggingFace_model_repos_internlm-chat-7b
ceval-computer_network	db9ce2	accuracy	gen	31.58
ceval-operating_system	1c2571	accuracy	gen	36.84
ceval-computer_architecture	a74dad	accuracy	gen	28.57
ceval-college_programming	4ca32a	accuracy	gen	32.43
ceval-college_physics	963fa8	accuracy	gen	26.32
ceval-college_chemistry	e78857	accuracy	gen	16.67
ceval-advanced_mathematics	ce03e2	accuracy	gen	21.05
ceval-probability_and_statistics	65e812	accuracy	gen	38.89
ceval-discrete_mathematics	e894ae	accuracy	gen	18.75
ceval-electrical_engineer	ae42b9	accuracy	gen	35.14
ceval-metrology_engineer	ee34ea	accuracy	gen	50
ceval-high_school_mathematics	1dc5bf	accuracy	gen	22.22
ceval-high_school_physics	adf25f	accuracy	gen	31.58
ceval-high_school_chemistry	2ed27f	accuracy	gen	15.79
ceval-high_school_biology	8e2b9a	accuracy	gen	36.84
ceval-middle_school_mathematics	bee8d5	accuracy	gen	26.32
ceval-middle_school_biology	86817c	accuracy	gen	61.9
ceval-middle_school_physics	8accf6	accuracy	gen	63.16
ceval-middle_school_chemistry	167a15	accuracy	gen	60
ceval-veterinary_medicine	b4e08d	accuracy	gen	47.83
ceval-college_economics	f3f4e6	accuracy	gen	41.82
ceval-business_administration	c1614e	accuracy	gen	33.33
ceval-marxism	cf874c	accuracy	gen	68.42
ceval-mao_zedong_thought	51c7a4	accuracy	gen	70.83
ceval-education_science	591fee	accuracy	gen	58.62
ceval-teacher_qualification	4e4ced	accuracy	gen	70.45
ceval-high_school_politics	5c0de2	accuracy	gen	26.32
ceval-high_school_geography	865461	accuracy	gen	47.37
ceval-middle_school_politics	5be3e7	accuracy	gen	52.38
ceval-middle_school_geography	8a63be	accuracy	gen	58.33
ceval-modern_chinese_history	fc01af	accuracy	gen	73.91
ceval-ideological_and_moral_cultivation	a2aa4a	accuracy	gen	63.16
ceval-logic	f5b022	accuracy	gen	31.82
ceval-law	a110a1	accuracy	gen	25
ceval-chinese_language_and_literature	0f8b68	accuracy	gen	30.43
ceval-art_studies	2a1300	accuracy	gen	60.61
ceval-professional_tour_guide	4e673e	accuracy	gen	62.07
ceval-legal_professional	ce8787	accuracy	gen	39.13
ceval-high_school_chinese	315705	accuracy	gen	63.16
ceval-high_school_history	7eb30a	accuracy	gen	70
ceval-middle_school_history	48ab4a	accuracy	gen	59.09
ceval-civil_servant	87d061	accuracy	gen	53.19
ceval-sports_science	70f27b	accuracy	gen	52.63
ceval-plant_protection	8941f9	accuracy	gen	59.09
ceval-basic_medicine	c409d6	accuracy	gen	47.37
ceval-clinical_medicine	49e82d	accuracy	gen	40.91
ceval-urban_and_rural_planner	95b885	accuracy	gen	45.65
ceval-accountant	002837	accuracy	gen	26.53
ceval-fire_engineer	bc23f5	accuracy	gen	22.58
ceval-environmental_impact_assessment_engineer	c64e2d	accuracy	gen	64.52
ceval-tax_accountant	3a5e3c	accuracy	gen	34.69
ceval-physician	6e277d	accuracy	gen	40.82
ceval-stem	-	naive_average	gen	35.09
ceval-social-science	-	naive_average	gen	52.79
ceval-humanities	-	naive_average	gen	52.58
ceval-other	-	naive_average	gen	44.36
ceval-hard	-	naive_average	gen	23.91
ceval	-	naive_average	gen	44.16
04/13 11:23:14 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240413_110824/summary/summary_20240413_110824.txt				
04/13 11:23:14 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240413_110824/summary/summary_20240413_110824.csv				

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