

基础岛任务笔记6-OpenCompass 评测 InternLM-1.8B 实践

- 基础任务（完成此任务即完成闯关）

- 使用 OpenCompass 评测 internlm2-chat-1.8b 模型在 ceval 数据集上的性能，记录复现过程并截图。

- 进阶任务（闯关不要求完成此任务）

- 使用 OpenCompass 进行主观评测（选做）
- 使用 OpenCompass 评测 InternLM2-Chat-1.8B 模型使用 LMDeploy部署后在 ceval 数据集上的性能（选做）
- 使用 OpenCompass 进行调用API评测（优秀学员必做）

环境配置

首先创建开发机并创建好conda环境，其次面向GPU环境安装好

```
1 conda create -n opencompass python=3.10
2 conda activate opencompass
3 conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pytorch-
  cuda=12.1 -c pytorch -c nvidia -y
4
5 # 注意：一定要先 cd /root, git clone opencompass 时一定要将 opencompass clone 到
  /root 路径下
6 cd /root
7 git clone -b 0.2.4 https://github.com/open-compass/opencompass
8 cd opencompass
9 pip install -e .
10
11
12 apt-get update
13 apt-get install cmake
14 pip install -r requirements.txt
15 pip install protobuf
```

CPU 5.83% GPU-1: 10% Nvidia A100 0%
内存 0.70 / 24 GB 2.92% 显存 0 / 8182 MiB 0%

```
#
# To activate this environment, use
#
#     $ conda activate opencompass
#
# To deactivate an active environment, use
#
#     $ conda deactivate

(base) root@intern-studio-40077780: ~ # conda activate opencompass
(opencompass) root@intern-studio-40077780: ~ # conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2
.1.2 pytorch-cuda=12.1 -c pytorch -c nvidia -y
Collecting package metadata (current_repodata.json): done
Solving environment: unsuccessful initial attempt using frozen solve. Retrying with flexible solve.
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

environment location: /root/.conda/envs/opencompass

added / updated specs:
- pytorch-cuda=12.1
- pytorch==2.1.2
- torchaudio==2.1.2
- torchvision==0.16.2

The following packages will be downloaded:

package | build
-----|-----
networkx-3.2.1 | py310h06a4308_0 2.4 MB defaults
```

```
CPU 6.99% GPU-1: 10% Nvidia A100 0% 内存 0.72 / 24 GB 3.02% 显存 0 / 8182 MiB 0%
Successfully installed Levenshtein-0.25.1 OpenCC-1.1.9 absl-py-2.1.0 accelerate-0.34.0 addict-2.4.0 aiohttp-3.10.5
aiohttp-3.10.5 aiosignal-1.3.1 annotated-types-0.7.0 anyio-4.4.0 async-timeout-4.0.3 attrs-24.2.0 boto3-1.35.12
botocore-1.35.12 click-8.1.7 cn2an-0.5.22 colorama-0.4.6 contourpy-1.3.0 cpm_kernels-1.0.11 cycycler-0.12.1 datasets-2.19.1
dill-0.3.8 distro-1.9.0 einops-0.5.0 evaluate-0.4.2 exceptiongroup-1.2.2 fairscale-0.4.13 fonttools-4.53.1 frozenlist-1.4.1
fsspec-2024.3.1 func_timeout-4.3.5 fuzzywuzzy-0.18.0 h11-0.14.0 httpcore-1.0.5 httpx-0.27.2 huggingface-hub-0.24.6
immutabledict-4.2.0 importlib-metadata-8.4.0 jieba-0.42.1 jiter-0.5.0 jmespath-1.0.1 joblib-1.4.2 kiwisolver-1.4.7 langdetect-1.0.9
ltp-4.2.14 ltp-core-0.1.4 ltp-extension-0.1.13 lxml-5.3.0 markdown-it-py-3.0.0 matplotlib-3.9.2 mdurl-0.1.2 mmengine-lite-0.10.4
multidict-6.0.5 multiprocessing-0.70.16 nltk-3.8 openai-1.43.0 opencv-python-headless-4.10.0.84 packaging-24.1 pandas-1.5.3
platformdirs-4.2.2 portalocker-2.10.1 prettytable-3.11.0 proces-0.1.7 psutil-6.0.0 pyarrow-17.0.0 pyarrow-hotfix-0.6
pydantic-2.8.2 pydantic-core-2.20.1 pyext-0.7 pygments-2.18.0 pyparsing-3.1.4 pyproject-0.52.0 python-Levenshtein-0.25.1
python-dateutil-2.9.0.post0 pytz-2024.1 rank_bm25-0.2.2 rapidfuzz-3.9.7 regex-2024.7.24 requests-2.31.0 rich-13.8.0 rouge-1.0.1
rouge_chinese rouge_score-0.1.2 s3transfer-0.10.2 sacrebleu-2.4.3 safetensors-0.4.4 scikit_learn-1.2.1 scipy-1.14.1 seaborn-0.13.2
sentence-transformers-2.2.2 sentencepiece-0.2.0 shellingham-1.5.4 six-1.16.0 sniffio-1.3.1 tabulate-0.9.0 termcolor-2.4.0 threadpoolctl-3.5.0
tiktoken-0.7.0 timeout_decorator-0.5.0 tokenizers-0.19.1 tomli-2.0.1 tqdm-4.64.1 transformers-4.44.2 typer-0.12.5 wcwidth-0.2.13
xxhash-3.5.0 yapf-0.40.2 yarl-1.9.11 zipp-3.20.1
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager, possibly rendering your system unusable. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv. Use the --root-user-action option if you know what you are doing and want to suppress this warning.
(opencompass) root@intern-studio-40077780: ~/opencompass# pip install protobuf
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting protobuf
Using cached https://pypi.tuna.tsinghua.edu.cn/packages/19/15/da43113361db20f2d521bc38d92549edbe06856aeec085c420b2b8af5751/protobuf-5.28.0-cp38-abi3-manylinux2014_x86_64.whl (316 kB)
Installing collected packages: protobuf
Successfully installed protobuf-5.28.0
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager, possibly rendering your system unusable. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv. Use the --root-user-action option if you know what you are doing and want to suppress this warning.
(opencompass) root@intern-studio-40077780: ~/opencompass#
```

数据准备

- 解压评测数据集，将在OpenCompass下看到data文件夹

```
1 cp /share/temp/datasets/OpenCompassData-core-20231110.zip /root/opencompass/
2 unzip OpenCompassData-core-20231110.zip
```

```
CPU 6.38% GPU-1: 10% Nvidia A100 0%
内存 0.71 / 24 GB 2.94% 显存 0 / 8182 MiB 0%

inflating: data/mmlu/dev/miscellaneous_dev.csv
inflating: data/mmlu/dev/moral_disputes_dev.csv
inflating: data/mmlu/dev/nutrition_dev.csv
inflating: data/mmlu/dev/philosophy_dev.csv
inflating: data/mmlu/dev/professional_psychology_dev.csv
inflating: data/mmlu/dev/astronomy_dev.csv
inflating: data/mmlu/dev/college_computer_science_dev.csv
inflating: data/mmlu/dev/high_school_biology_dev.csv
inflating: data/mmlu/dev/high_school_physics_dev.csv
inflating: data/mmlu/dev/high_school_psychology_dev.csv
inflating: data/mmlu/dev/high_school_statistics_dev.csv
inflating: data/mmlu/dev/professional_law_dev.csv
inflating: data/mmlu/dev/sociology_dev.csv
creating: data/siqa/
inflating: data/siqa/dev.jsonl
inflating: data/siqa/train-labels.lst
inflating: data/siqa/dev-labels.lst
inflating: data/siqa/train.jsonl
creating: data/TheoremQA/
inflating: data/TheoremQA/test.csv
creating: data/triviaqa/
inflating: data/triviaqa/trivia-dev.qa.csv
inflating: data/triviaqa/trivia-test.qa.csv
creating: data/lambada/
inflating: data/lambada/test.jsonl
creating: data/strategyqa/
inflating: data/strategyqa/strategyQA_train.json
(opencompass) root@intern-studio-40077780: ~/opencompass# ls
LICENSE          configs          opencompass.egg-info  setup.py
OpenCompassData-core-20231110.zip  data            requirements           src
README.md        docs            requirements.txt       tests
README_zh-CN.md  opencompass    run.py                tools
(opencompass) root@intern-studio-40077780: ~/opencompass# python
```

- InternLM和ceval相关的配置文件

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

```
README_zh-CN.md opencompass run.py tools
(opencompass) root@intern-studio-40077780: ~/opencompass# python tools/list_configs.py internlm ceval

A module that was compiled using NumPy 1.x cannot be run in
NumPy 2.0.1 as it may crash. To support both 1.x and 2.x
versions of NumPy, modules must be compiled with NumPy 2.0.
Some module may need to rebuild instead e.g. with 'pybind11>=2.12'.

If you are a user of the module, the easiest solution will be to
downgrade to 'numpy<2' or try to upgrade the affected module.
We expect that some modules will need time to support NumPy 2.

Traceback (most recent call last): File "/root/opencompass/tools/list_configs.py", line 5, in <module>
  from opencompass.utils import match_files
File "/root/opencompass/opencompass/utils/__init__.py", line 4, in <module>
  from .collect_env import * # noqa
File "/root/opencompass/opencompass/utils/collect_env.py", line 2, in <module>
  from mmengine.utils.dl_utils import collect_env as collect_base_env
File "/root/.conda/envs/opencompass/lib/python3.10/site-packages/mmengine/utils/dl_utils/__init__.py", li
ne 3, in <module>
  from .collect_env import collect_env
File "/root/.conda/envs/opencompass/lib/python3.10/site-packages/mmengine/utils/dl_utils/collect_env.py",
line 10, in <module>
  import torch
File "/root/.conda/envs/opencompass/lib/python3.10/site-packages/torch/__init__.py", line 1382, in <modul
e>
  from .functional import * # noqa: F403
File "/root/.conda/envs/opencompass/lib/python3.10/site-packages/torch/functional.py", line 7, in <module
>
  import torch.nn.functional as F
File "/root/.conda/envs/opencompass/lib/python3.10/site-packages/torch/nn/__init__.py", line 1, in <modul
e>
  from .modules import * # noqa: F403
```

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	configs/datasets/ceval/ceval_ppl.py
ceval_ppl_1cd8bf	configs/datasets/ceval/ceval_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

启动评测

- 使用命令行配置参数法进行评测

```
1 #打开 opencompass文件夹下configs/models/hf_internlm/的
   hf_internlm2_chat_1_8b.py ,贴入以下代码
2 from opencompass.models import HuggingFaceCausalLM
3
4
5 models = [
6     dict(
```

```

7         type=HuggingFaceCausalLM,
8         abbr='internlm2-1.8b-hf',
9         path="/share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b",
10        tokenizer_path='/share/new_models/Shanghai_AI_Laboratory/internlm2-
chat-1_8b',
11        model_kwargs=dict(
12            trust_remote_code=True,
13            device_map='auto',
14        ),
15        tokenizer_kwargs=dict(
16            padding_side='left',
17            truncation_side='left',
18            use_fast=False,
19            trust_remote_code=True,
20        ),
21        max_out_len=100,
22        min_out_len=1,
23        max_seq_len=2048,
24        batch_size=8,
25        run_cfg=dict(num_gpus=1, num_procs=1),
26    )
27 ]

```

The screenshot shows a VS Code editor with the following details:

- System Bar:** CPU 5.9%, GPU-1: 10% Nvidia A100 0%, 内存 0.71 / 24 GB, 2.95%, 显存 0 / 8182 MiB, 0%.
- Explorer:**
 - ROOT
 - falcon
 - gemini
 - gemma
 - hf_internlm
 - hf_internlm_7b.py
 - hf_internlm_20b.py
 - hf_internlm_chat_7b_8k.py
 - hf_internlm_chat_7b_v1_1.py
 - hf_internlm_chat_7b.py
 - hf_internlm_chat_20b.py
 - hf_internlm2_1_8b.py
 - hf_internlm2_7b.py
 - hf_internlm2_20b.py
 - hf_internlm2_base_7b.py
 - hf_internlm2_base_20b.py
 - hf_internlm2_chat_1_8b_sft.py
 - hf_internlm2_chat_1_8b.py**
 - hf_internlm2_chat_7b_sft.py
 - hf_internlm2_chat_7b_with_system.py
 - hf_internlm2_chat_7b.py

- Editor:**
- File: `xtuner_streamlit_demo.py` M
- File: `hf_internlm2_chat_1_8b.py` X
- Content of `hf_internlm2_chat_1_8b.py`:

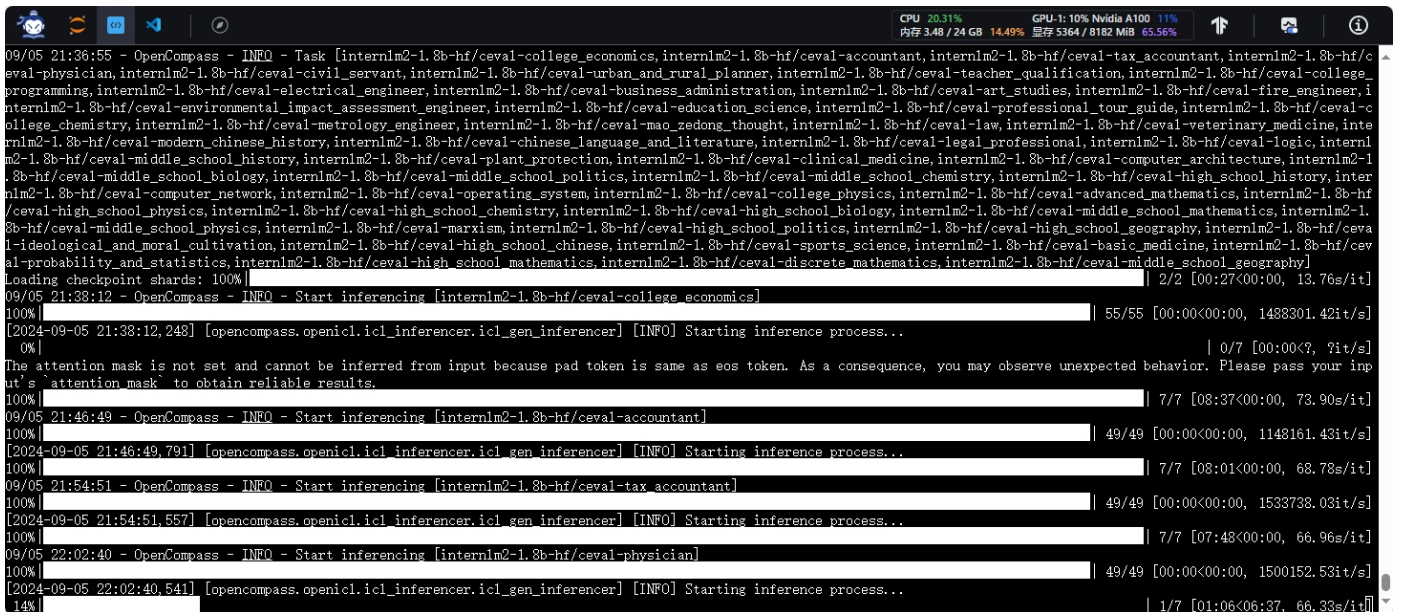
```

1 from opencompass.models import HuggingFaceCausalLM
2
3
4 _meta_template = dict(
5     round=[
6         dict(role='HUMAN', begin='<|im_start|>user\n',
7             dict(role='BOT', begin='<|im_start|>assistant\n',
8                 eos_token_id=92542
9             ]
10 )
11
12 models = [
13     dict(
14         type=HuggingFaceCausalLM,
15         abbr='internlm2-chat-1.8b-hf',
16         path="internlm/internlm2-chat-1_8b",
17         tokenizer_path='internlm/internlm2-chat-1_8b',
18         model_kwargs=dict(
19             trust_remote_code=True,
20             device_map='auto',
21         ),
22         tokenizer_kwargs=dict(
23             padding_side='left',
24             truncation_side='left',
25             use_fast=False,

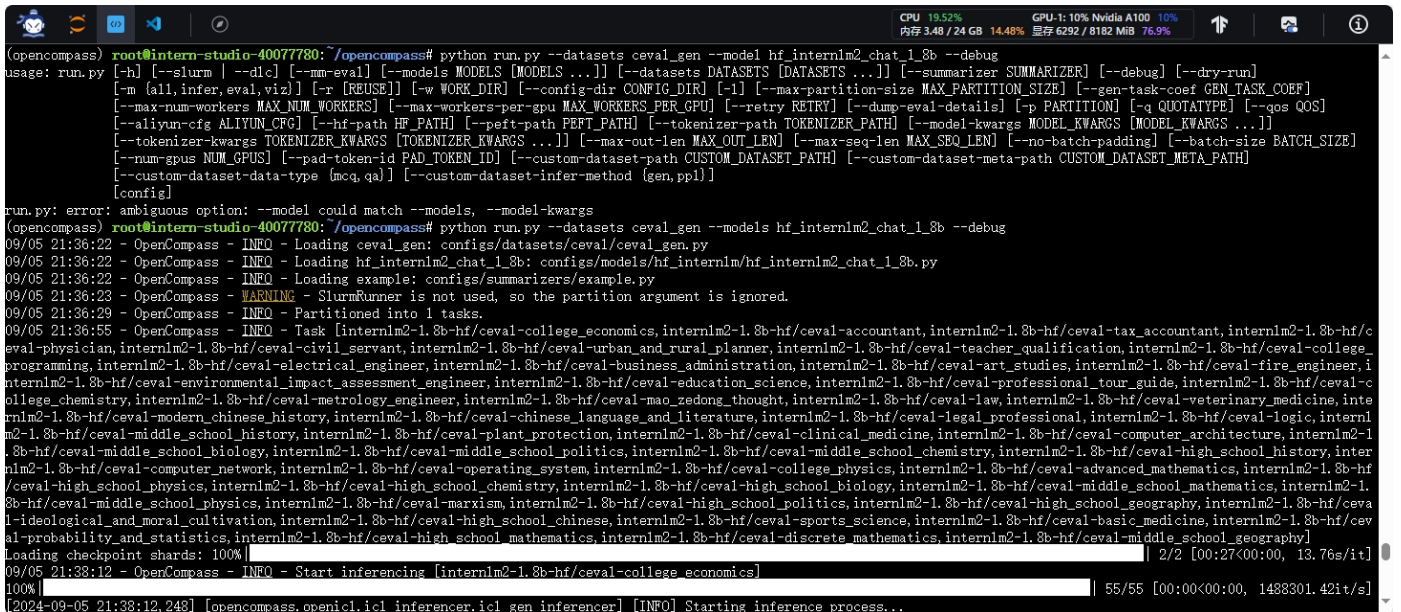
```
- Status Bar:** 0.2.4, 0 0 0, 8, Git Graph, Ln 1, Col 1, Spaces: 4, UTF-8, LF, Python 3.11.5 ('base': conda)

通过以下命令评测 InternLM2-Chat-1.8B 模型在 C-Eval 数据集上的性能

```
1 #环境变量配置
2 export MKL_SERVICE_FORCE_INTEL=1
3 #或
4 export MKL_THREADING_LAYER=GNU
5
6 python run.py --datasets ceval_gen --models hf_internlm2_chat_1_8b --debug
7
8 #命令解析
9 python run.py
10 --datasets ceval_gen \ # 数据集准备
11 --models hf_internlm2_chat_1_8b \ # 模型准备
12 --debug
```



```
09/05 21:36:55 - OpenCompass - INFO - Task [internlm2-1.8b-hf/ceval-college_economics,internlm2-1.8b-hf/ceval-accountant,internlm2-1.8b-hf/ceval-tax_accountant,internlm2-1.8b-hf/ceval-physician,internlm2-1.8b-hf/ceval-civil_servant,internlm2-1.8b-hf/ceval-urban_and_rural_planner,internlm2-1.8b-hf/ceval-teacher_qualification,internlm2-1.8b-hf/ceval-college_programming,internlm2-1.8b-hf/ceval-electrical_engineer,internlm2-1.8b-hf/ceval-business_administration,internlm2-1.8b-hf/ceval-art_studies,internlm2-1.8b-hf/ceval-fire_engineer,internlm2-1.8b-hf/ceval-environmental_impact_assessment_engineer,internlm2-1.8b-hf/ceval-education_science,internlm2-1.8b-hf/ceval-professional_tour_guide,internlm2-1.8b-hf/ceval-college_chemistry,internlm2-1.8b-hf/ceval-metrology_engineer,internlm2-1.8b-hf/ceval-mao_zedong_thought,internlm2-1.8b-hf/ceval-law,internlm2-1.8b-hf/ceval-veterinary_medicine,internlm2-1.8b-hf/ceval-modern_chinese_history,internlm2-1.8b-hf/ceval-chinese_language_and_literature,internlm2-1.8b-hf/ceval-legal_professional,internlm2-1.8b-hf/ceval-logic,internlm2-1.8b-hf/ceval-middle_school_history,internlm2-1.8b-hf/ceval-plant_protection,internlm2-1.8b-hf/ceval-clinical_medicine,internlm2-1.8b-hf/ceval-computer_architecture,internlm2-1.8b-hf/ceval-middle_school_biology,internlm2-1.8b-hf/ceval-middle_school_politics,internlm2-1.8b-hf/ceval-middle_school_chemistry,internlm2-1.8b-hf/ceval-high_school_history,internlm2-1.8b-hf/ceval-computer_network,internlm2-1.8b-hf/ceval-operating_system,internlm2-1.8b-hf/ceval-college_physics,internlm2-1.8b-hf/ceval-advanced_mathematics,internlm2-1.8b-hf/ceval-high_school_physics,internlm2-1.8b-hf/ceval-high_school_chemistry,internlm2-1.8b-hf/ceval-high_school_biology,internlm2-1.8b-hf/ceval-middle_school_mathematics,internlm2-1.8b-hf/ceval-middle_school_physics,internlm2-1.8b-hf/ceval-marxism,internlm2-1.8b-hf/ceval-high_school_politics,internlm2-1.8b-hf/ceval-high_school_geography,internlm2-1.8b-hf/ceval-1-ideological_and_moral_cultivation,internlm2-1.8b-hf/ceval-high_school_chinese,internlm2-1.8b-hf/ceval-sports_science,internlm2-1.8b-hf/ceval-basic_medicine,internlm2-1.8b-hf/ceval-al-probability_and_statistics,internlm2-1.8b-hf/ceval-high_school_mathematics,internlm2-1.8b-hf/ceval-discrete_mathematics,internlm2-1.8b-hf/ceval-middle_school_geography]
Loading checkpoint shards: 100% | 2/2 [00:27<00:00, 13.76s/it]
09/05 21:38:12 - OpenCompass - INFO - Start inferencing [internlm2-1.8b-hf/ceval-college_economics]
100% | 55/55 [00:00<00:00, 1488301.42it/s]
[2024-09-05 21:38:12,248] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
0% | 0/7 [00:00<?, ?it/s]
The attention mask is not set and cannot be inferred from input because pad token is same as eos token. As a consequence, you may observe unexpected behavior. Please pass your input's attention_mask to obtain reliable results.
100% | 7/7 [08:37<00:00, 73.90s/it]
09/05 21:46:49 - OpenCompass - INFO - Start inferencing [internlm2-1.8b-hf/ceval-accountant]
100% | 49/49 [00:00<00:00, 1148161.43it/s]
[2024-09-05 21:46:49,791] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 7/7 [08:01<00:00, 68.78s/it]
09/05 21:54:51 - OpenCompass - INFO - Start inferencing [internlm2-1.8b-hf/ceval-tax_accountant]
100% | 49/49 [00:00<00:00, 1533738.03it/s]
[2024-09-05 21:54:51,557] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 7/7 [07:48<00:00, 66.96s/it]
09/05 22:02:40 - OpenCompass - INFO - Start inferencing [internlm2-1.8b-hf/ceval-physician]
100% | 49/49 [00:00<00:00, 1500152.53it/s]
[2024-09-05 22:02:40,541] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
14% | 1/7 [01:06<06:37, 66.33s/it]
```



```
(opencompass) root@intern-studio-40077780: /opencompass# python run.py --datasets ceval_gen --model hf_internlm2_chat_1_8b --debug
usage: run.py [-h] [--slurm] [--dlc] [--mm-eval] [--models MODELS [MODELS ...]] [--datasets DATASETS [DATASETS ...]] [--summarizer SUMMARIZER] [--debug] [--dry-run]
               [-m [all,infer,eval,viz]] [-r [REUSE]] [-w WORK_DIR] [--config-dir CONFIG_DIR] [-i] [--max-partition-size MAX_PARTITION_SIZE] [--gen-task-coef GEN_TASK_COEF]
               [--max-num-workers MAX_NUM_WORKERS] [--max-workers-per-gpu MAX_WORKERS_PER_GPU] [--retry RETRY] [--dump-eval-details] [-p PARTITION] [-q QUOTATYPE] [--qos QOS]
               [--aliyun-cfg ALIYUN_CFG] [--hf-path HF_PATH] [--peft-path PEFT_PATH] [--tokenizer-path TOKENIZER_PATH] [--model-kwags MODEL_KWARGS [MODEL_KWARGS ...]]
               [--tokenizer-kwags TOKENIZER_KWARGS [TOKENIZER_KWARGS ...]] [--max-out-len MAX_OUT_LEN] [--max-seq-len MAX_SEQ_LEN] [--no-batch-padding] [--batch-size BATCH_SIZE]
               [--num-gpus NUM_GPUS] [--pad-token-id PAD_TOKEN_ID] [--custom-dataset-path CUSTOM_DATASET_PATH] [--custom-dataset-meta-path CUSTOM_DATASET_META_PATH]
               [--custom-dataset-data-type {mcq,qa}] [--custom-dataset-infer-method {gen,pp1}]
               [config]
run.py: error: ambiguous option: --model could match --models, --model-kwags
(opencompass) root@intern-studio-40077780: /opencompass# python run.py --datasets ceval_gen --models hf_internlm2_chat_1_8b --debug
09/05 21:36:22 - OpenCompass - INFO - Loading ceval_gen: configs/datasets/ceval/ceval_gen.py
09/05 21:36:22 - OpenCompass - INFO - Loading hf_internlm2_chat_1_8b: configs/models/hf_internlm2_chat_1_8b.py
09/05 21:36:22 - OpenCompass - INFO - Loading example: configs/summarizers/example.py
09/05 21:36:23 - OpenCompass - WARNING - SlurmRunner is not used, so the partition argument is ignored.
09/05 21:36:29 - OpenCompass - INFO - Partitioned into 1 tasks.
09/05 21:36:55 - OpenCompass - INFO - Task [internlm2-1.8b-hf/ceval-college_economics,internlm2-1.8b-hf/ceval-accountant,internlm2-1.8b-hf/ceval-tax_accountant,internlm2-1.8b-hf/ceval-physician,internlm2-1.8b-hf/ceval-civil_servant,internlm2-1.8b-hf/ceval-urban_and_rural_planner,internlm2-1.8b-hf/ceval-teacher_qualification,internlm2-1.8b-hf/ceval-college_programming,internlm2-1.8b-hf/ceval-electrical_engineer,internlm2-1.8b-hf/ceval-business_administration,internlm2-1.8b-hf/ceval-art_studies,internlm2-1.8b-hf/ceval-fire_engineer,internlm2-1.8b-hf/ceval-environmental_impact_assessment_engineer,internlm2-1.8b-hf/ceval-education_science,internlm2-1.8b-hf/ceval-professional_tour_guide,internlm2-1.8b-hf/ceval-college_chemistry,internlm2-1.8b-hf/ceval-metrology_engineer,internlm2-1.8b-hf/ceval-mao_zedong_thought,internlm2-1.8b-hf/ceval-law,internlm2-1.8b-hf/ceval-veterinary_medicine,internlm2-1.8b-hf/ceval-modern_chinese_history,internlm2-1.8b-hf/ceval-chinese_language_and_literature,internlm2-1.8b-hf/ceval-legal_professional,internlm2-1.8b-hf/ceval-logic,internlm2-1.8b-hf/ceval-middle_school_history,internlm2-1.8b-hf/ceval-plant_protection,internlm2-1.8b-hf/ceval-clinical_medicine,internlm2-1.8b-hf/ceval-computer_architecture,internlm2-1.8b-hf/ceval-middle_school_biology,internlm2-1.8b-hf/ceval-middle_school_politics,internlm2-1.8b-hf/ceval-middle_school_chemistry,internlm2-1.8b-hf/ceval-high_school_history,internlm2-1.8b-hf/ceval-computer_network,internlm2-1.8b-hf/ceval-operating_system,internlm2-1.8b-hf/ceval-college_physics,internlm2-1.8b-hf/ceval-advanced_mathematics,internlm2-1.8b-hf/ceval-high_school_physics,internlm2-1.8b-hf/ceval-high_school_chemistry,internlm2-1.8b-hf/ceval-high_school_biology,internlm2-1.8b-hf/ceval-middle_school_mathematics,internlm2-1.8b-hf/ceval-middle_school_physics,internlm2-1.8b-hf/ceval-marxism,internlm2-1.8b-hf/ceval-high_school_politics,internlm2-1.8b-hf/ceval-high_school_geography,internlm2-1.8b-hf/ceval-1-ideological_and_moral_cultivation,internlm2-1.8b-hf/ceval-high_school_chinese,internlm2-1.8b-hf/ceval-sports_science,internlm2-1.8b-hf/ceval-basic_medicine,internlm2-1.8b-hf/ceval-al-probability_and_statistics,internlm2-1.8b-hf/ceval-high_school_mathematics,internlm2-1.8b-hf/ceval-discrete_mathematics,internlm2-1.8b-hf/ceval-middle_school_geography]
Loading checkpoint shards: 100% | 2/2 [00:27<00:00, 13.76s/it]
09/05 21:38:12 - OpenCompass - INFO - Start inferencing [internlm2-1.8b-hf/ceval-college_economics]
100% | 55/55 [00:00<00:00, 1488301.42it/s]
[2024-09-05 21:38:12,248] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
```

结果出来了：从上午九点到下午7点，总算评测完成

```

cpu-computer_network,d89ce2,accuracy,gen,47.37
cpu-operating_system,1c2571,accuracy,gen,47.37
cpu-computer_architecture,a74dad,accuracy,gen,23.81
cpu-college_programming,4ca32a,accuracy,gen,27.03
cpu-college_physics,963fa8,accuracy,gen,42.11
cpu-college_chemistry,e78857,accuracy,gen,57.50
cpu-advanced_mathematics,ce03e2,accuracy,gen,26.32
cpu-probability_and_statistics,65e812,accuracy,gen,22.22
cpu-discrete_mathematics,e694ae,accuracy,gen,25.00
cpu-electrical_engineer,ae42b9,accuracy,gen,27.03
cpu-metrology_engineer,ee34ea,accuracy,gen,54.17
cpu-high_school_mathematics,1dc5bf,accuracy,gen,22.22
cpu-high_school_physics,adf25f,accuracy,gen,42.11
cpu-high_school_chemistry,2ed274,accuracy,gen,52.63
cpu-high_school_biology,8a2b9a,accuracy,gen,26.32
cpu-middle_school_mathematics,bee8d5,accuracy,gen,36.84
cpu-middle_school_biology,86817c,accuracy,gen,80.95
cpu-middle_school_physics,8accf6,accuracy,gen,47.37
cpu-middle_school_chemistry,167a15,accuracy,gen,80.00
cpu-veterinary_medicine,b4e08d,accuracy,gen,43.46
cpu-college_economics,f3f4e6,accuracy,gen,32.73
cpu-business_administration,c1614e,accuracy,gen,39.39
cpu-marxism,cf874c,accuracy,gen,68.42
cpu-mao_zedong_thought,51c7a4,accuracy,gen,70.83
cpu-education_science,591fee,accuracy,gen,55.17
cpu-teacher_qualification,fe1cec,accuracy,gen,59.09
cpu-high_school_politics,5d6de2,accuracy,gen,57.89
cpu-high_school_geography,86f461,accuracy,gen,47.37
cpu-middle_school_politics,5be3e7,accuracy,gen,76.19
cpu-middle_school_geography,8a63be,accuracy,gen,75.00
cpu-modern_chinese_history,fc01af,accuracy,gen,52.17
cpu-ideological_and_moral_cultivation,a2aa4a,accuracy,gen,73.68
cpu-logic,f9a02e,accuracy,gen,31.82
cpu-law,a110a1,accuracy,gen,29.17
cpu-chinese_language_and_literature,0f2b68,accuracy,gen,47.83
cpu-art_studies,2a1300,accuracy,gen,42.42
cpu-professional_tour_guide,4e673e,accuracy,gen,51.72
cpu-legal_professional,ce8787,accuracy,gen,34.78
cpu-high_school_chinese,315705,accuracy,gen,36.84
cpu-high_school_history,7eb30a,accuracy,gen,65.00
cpu-middle_school_history,48ab4a,accuracy,gen,36.36
cpu-civil_servant,87d961,accuracy,gen,42.55
cpu-sports_science,70f27b,accuracy,gen,52.63
cpu-plant_protection,8941f9,accuracy,gen,40.91
cpu-basic_medicine,c409d6,accuracy,gen,68.42
cpu-clinical_medicine,49e62d,accuracy,gen,36.36
cpu-urban_and_rural_planner,9c3855,accuracy,gen,52.17
cpu-accountant,002837,accuracy,gen,36.73
cpu-fire_engineer,bc23f5,accuracy,gen,38.71
cpu-environmental_impact_assessment_engineer,c64e2d,accuracy,gen,51.61
cpu-tax_accountant,3a5e3c,accuracy,gen,36.73

```