

×

9月17日, Gitee Xtreme 极智AI重磅发布,来Gitee直播间一起探索AI时代的软件研发新模式





〈/〉代码

រា Pull Requests 0

II Wiki

辿 统计

₩ 流水线

该仓库未声明开源许可证文件(LICENSE),使用请关注具体项目描述及其代码上游依赖。 3

Q + \Box

🌉 姜卓 update README.md. 1ecd94e 2个月前

% 分支 2 ♡ 标签 0

■ README.md

update README.md.

• TeleChat2.5

■ README

- 目录
- 模型介绍
 - 训练策略
 - 数据
 - 基础模型训练
 - 后训练阶段
 - 模型下载
- 效果评测
- 模型推理
- ModelScope
- vLLM 推理
- 离线推理
- OpenAI 兼容的 AP...
- 国产化适配
- 声明、引用
 - 声明
 - 引用

TeleChat2.5

🤴 github • 🔗 Hugging Face • 🖶 ModelScope • 🐾 gitee • 🗩 WeChat

目录

- 模型介绍
- 效果评测
- 模型推理
- 国产化适配
- 声明、协议、引用

模型介绍

TeleChat2.5 是 TeleChat 系列新版通用问答模型,由中国电信人工 智能研究院(TeleAI)基于国产算力研发训练,包括了 TeleChat2.5-35B 与 TeleChat2.5-115B。TeleChat2.5 基于最新强化 的 TeleBase2.5 系列模型进行训练,在理科、通用问答、Function Call等任务上有显著的效果提升。TeleChat2.5 的微调方法延续了 TeleChat2 系列, 具体请参考 TeleChat2。

训练策略

数据

• 为了提高模型训练数据的数量和质量, TeleChat2.5 在训练过 程中采用了大量理科学科和编程领域的合成数据。在合成过 程中, 为了减少错误信息的引入, 主要以基于知识点或知识 片段的教育类知识合成为主。

基础模型训练

- 交到 master 分支, 4a f1abf...1ecd94e
- 姜 4个月前推送了新 交到 master 分支 04 de5a5...4af1abf
- 姜 4个月前同步了行 ☑
- 姜 4个月前推送了新 交到 master 分支 9cd81...04de5a5
- 4个月前推送了新PI m aster 分支

加载更多 ▼



模型优化。

 在最终训练阶段,为了平衡模型在各个维度的能力表现,我们选取 优的多个模型,并基于各模型的综合表现进行参数加权融合,其中 正相关。

后训练阶段

我们采用分阶段优化的模型训练策略:

- 融合优化阶段:整合复杂推理与通用问答能力,针对语言理解、数解构重组。通过重构任务框架并融合多维度解题思路,生成优化后答案长度会适度增加,并基于优化数据实施微调训练。
- 能力强化阶段: 针对数理逻辑与编程类任务,通过注入结构化解题 强化学习奖励机制,显著提升模型对复杂任务的理解与处理能力。
- 泛化提升阶段:面向安全合规、指令响应、函数调用、数学推理、 类型进行系统性强化学习增强,全面提升模型的通用任务处理能力。

模型下载

模型版本	下载链接
TeleChat2.5-35B	nıodelscope
TeleChat2.5-115B	nıodelscope

效果评测

模型	MATH-500	AlignBench	BFCL(avg v1
Qwen2.5-32B	82	7.39	81.11
Qwen2.5-72B	82	7.62	79.15
Qwen3-32B (通用)	83	8.23	81.84
GPT-4o-1120	75	7.49	78.66
TeleChat2-35B	65	6.97	75.32
TeleChat2-115B	75	7.56	77.47
TeleChat2.5-35B	85	7.73	78.28
TeleChat2.5-115B	87	7.93	83.39

模型推理

TeleChat2.5 系列模型支持使用 transformers 库进行推理,示例如下:

import torch

from transformers import AutoModelForCausalLM, AutoTokenizer,

tokenizer = AutoTokenizer.from_pretrained("TeleChat2.5/TeleCha
model = AutoModelFor lausalLM.from_pretrained(

"TeleChat2.5/TeleChat2.5-35B",

?

Ľ





 $\overline{\uparrow}$

)





```
prompt = "生抽和酱油的区别是什么?"
  messages = [{"role": "user", "content": prompt}]
  text = tokenizer.apply_chat_template(messages,
     tokenize=False,
     add_generation_p rompt=True
  model_inputs = tokenizer([text], return_tensors="pt").to(model
  generated_ids = model.generate(
     **model inputs
  generated ids = [
     output ids[len(i put ids):] for input ids, output ids in z
  response = tokenizer.batch_decode(generated_ids, skip_special_
  print(response)
ModelScope
TeleChat2.5 系列模型支持使用 ModelScope 推理,示例如下:
  import os
  import torch
  from modelscope impo t AutoModelForCausalLM, AutoTokenizer, Ge
  tokenizer = AutoToke nizer.from pretrained('TeleChat2.5/TeleCha
  model = AutoModelFor lausalLM.from_pretrained('TeleChat2.5/Tele
                                                  torch_dtype=
  prompt = "生抽与老抽的区别?"
  messages = [{"role": "user", "content": prompt}]
  text = tokenizer.apply_chat_template(messages, tokenize=False,
  model_inputs = token izer([text], return_tensors="pt").to(model
  generated_ids = model.generate(
      **model_inputs
  )
  generated_ids = [
     output_ids[len(i put_ids):] for input_ids, output_ids in z
  response = tokenizer batch_decode(generated_ids, skip_special_
  print(response)
vLLM 推理
TeleChat2.5 支持使用 vLLM 进行部署与推理加速,示例如下:
离线推理
  from transformers import AutoTokenizer
  from vllm import LLM, SamplingParams
  tokenizer = AutoToke izer.from_pretrained("TeleChat2.5/TeleCha
  sampling_params = SamplingParams(temperature=0.0, repetition_r
  11m = LLM(model="TeleChat2.5/TeleChat2.5-35B", trust_remote_cc
  prompt = "生抽和酱油的区别是什么?"
  messages = [{"role": "user", "content": prompt}]
  text = tokenizer.apply_chat_template(
     messages.
      tokenize=False,
      add_generation_p rompt=True
  )
```















```
prompt = output.prompt
      generated_text = output.outputs[0].text
      print(f"Prompt: [prompt!r], Generated text: {generated_tex
OpenAI 兼容的 API 服务
您可以借助 vLLM,构建一个与 OpenAI API 兼容的 API 服务。请按照以
  vllm serve TeleChat2.5/TeleChat2.5-35B \
      --trust-remote-code \
      --dtype bfloat16 \
      --disable-custom-all-reduce
然后,您可以与 TeleChat 2.5 进行对话:
  from openai import OpenAI
  openai api key = "EM'TY"
  openai api base = "http://localhost:8000/v1"
  client = OpenAI(api_key=openai_api_key, base_url=openai_api_ba
  chat response = client.chat.completions.create(
     model="TeleChat2.5/TeleChat2.5-35B",
     messages=[
         {"role": "us ar", "content": "生抽和酱油的区别是什么?"},
      ],
      temperature=0.0,
     max_tokens=8192,
     extra_body={
         "repetition_penalty": 1.01,
         "skip_special_tokens": False,
         "spaces_between_special_tokens": False,
     },
  print("Chat response:", chat_response)
```

国产化适配

TeleChat2.5系列模型均进行了**国产化算力适配**,具体信息可见

- 1. MindSpore-Lab/Tele Chat2.5-35B
- 2. MindSpore-Lab/Tele Chat2.5-115B

声明、引用

声明

我们在此声明,不要使用 TeleChat2.5 系列模型及其衍生模型进行任何; 法的活动。同时,我们也要求使用者不要将 TeleChat2.5 系列模型用于 互联网服务。我们希望所有使用者遵守上述原则,确保科技发展在合法

我们已经尽我们所能,来确保模型训练过程中使用的数据的合规性。然了巨大的努力,但由于模型和数据的复杂性,仍有可能存在一些无法预由于使用 TeleChat2.5 系列开源模型而导致的任何问题,包括但不限于论风险,或模型被误导、滥用、传播或不当利用所带来的任何风险和问责任。

?

E







 $\overline{\Lambda}$





```
אוואנחון דערוואנרווער וואנרווער האוא אוויים וויים וויי
               @misc{wang2025techni:alreporttelechat2telechat25,
                                                           title={Technic | Report of TeleChat2, TeleChat2.5 and T1
                                                           author={Zihan | Vang and Xinzhang Liu and Yitong Yao and C
                                                         year={2025},
                                                         eprint={2507.13013},
                                                           archivePrefix=[arXiv],
                                                         primaryClass={:s.CL},
                                                         url={https://arxiv.org/abs/2507.18013},
               }
```

6 gitee

载

北京奥思研工智能科技有限公司版权所有

Git 大全 Gitee 封面人 OpenAPI 关于我们 Git 命令学习 MCP Server 加入我们 GVP 项目 CopyCat 代码 帮助文档 使用条款 克隆检测 Gitee 博客 在线自助服务 意见建议 APP与插件下 更新日志 合作伙伴

Gitee 公益计 划

Gitee 持续集

client@oschina.cn

企业版在线使用: 400-606-0201

专业版私有部署: 13670252304

13352947997



技术交流QQ群 微信服务号

成

开放原子开源基金会 合作代码托管平台

◆ 违法和不良信息举报中心

京ICP备2025119063号

●简体/繁體/[≦





