

QwQ-32B Qwen推理大模型解读

📅 2025年3月6日 ⌚ 2 分钟阅读

#AI #深度思考 #QwQ-32B #大模型 #Qwen

本文介绍了深度求索（DeepSeek）公司推出的新一代推理模型QwQ-32B，并对其技术原理、主要贡献、论文方法、评估结果和局限性进行了详细解读。

模型介绍

这是一款拥有 320 亿参数的模型，其性能可与具备 6710 亿参数（其中 370 亿被激活）的 DeepSeek-R1 媲美。这一成果突显了将强化学习应用于经过大规模预训练的强大基础模型的有效性。QwQ32B在推理模型中集成了与 Agent 相关的能力，使其能够在使用工具的同时进行批判性思考，并根据环境反馈调整推理过程。通过这种方式，QwQ32B 能够执行复杂的推理任务，如数学问题解决和编程挑战。这个模型于2025年3月6日发布。

技术指标

类型：因果语言模型

训练阶段：训练前和训练后（监督微调和强化学习）

架构：具有RoPE、SwiGLU、RMSNorm和注意力QKV偏置的变压器

参数数：32.5B

参数数量（非嵌入）：31.0B

层数：64

注意头数（GQA）：Q 40个，KV 8个

上下文长度：完整的131,072个令牌

评测结果

QwQ-32B 在一系列基准测试中进行了评估，测试了数学推理、编程能力和通用能力。以下结果展示了 QwQ-32B 与其他领先模型的性能对比，包括 DeepSeek-R1-Distilled-Qwen-32B、DeepSeek-R1-Distilled-Llama-70B、o1-mini 以及原始的 DeepSeek-R1。

目录

文章信息

字数

阅读时间

发布时间

更新时间

标签

#AI #深度思考 #QwQ-32B #Qwen

训练细节

QwQ基于Qwen2.5，在冷启动的基础上开展了大规模强化学习，这个和DeepSeek R1的训练方式类似。在初始阶段，我们特别针对数学和编程任务进行了RL训练。与依赖传统的奖励模型（reward model）不同：

通过校验生成答案的正确性来为数学问题提供反馈，

通过代码执行服务器评估生成的代码是否成功通过测试用例来提供代码的反馈。

随着训练轮次的推进，这两个领域中的性能均表现出持续提升。

在第一阶段的RL过后，我们增加了另一个针对通用能力的RL。

此阶段使用通用奖励模型和一些基于规则的验证器进行训练。

通过少量步骤的通用RL，可以提升其他通用能力，同时在数学和编程任务上的性能没有显著下降。

代码调用

其代码已在最新的HuggingFace transformers中（v4.37）。

```

1
2 from openai import OpenAI
3 import os
4
5 # Initialize OpenAI client
6 client = OpenAI(
7     # If the environment variable is not configured,
8     # replace with your API Key: api_key="sk-xxx"
9     # How to get an API Key:
10    https://help.aliyun.com/zh/model-studio/developer-
11    reference/get-api-key
12    api_key=os.getenv("DASHSCOPE_API_KEY"),
13
14    base_url="https://dashscope.aliyuncs.com/compatible-
15    mode/v1"
16)
17
18 reasoning_content = ""
19 content = ""
20
21 is_answering = False
22
23 completion = client.chat.completions.create(
24     model="qwq-32b",
25     messages=[
26         {"role": "user", "content": "Which is
27         larger, 9.9 or 9.11?"}
28     ],
29     stream=True,
30     # Uncomment the following line to return token
31     # usage in the last chunk
32     # stream_options={
33     #     "include_usage": True
34     # }
35)
36
37 print("\n" + "=" * 20 + "reasoning content" + "=" *
38 20 + "\n")
39
40 for chunk in completion:
41     # If chunk.choices is empty, print usage
42     if not chunk.choices:
43         print("\nUsage:")
44         print(chunk.usage)
45     else:
46         delta = chunk.choices[0].delta
47         # Print reasoning content
48         if hasattr(delta, 'reasoning_content') and
49         delta.reasoning_content is not None:
50             print(delta.reasoning_content, end='',
51                   flush=True)
52             reasoning_content +=
53             delta.reasoning_content
54         else:
55             if delta.content != "" and is_answering
56             is False:
57                 print("\n" + "=" * 20 + "content" +
58                       "=" * 20 + "\n")
59                 is_answering = True
60                 # Print content
61                 print(delta.content, end='', flush=True)
62                 content += delta.content
63
64 ]

```

未来工作

未来，我们将继续探索将智能体与RL集成，以实现长时推理，目标是通过推理时间扩展来释放更高的智能。这是Qwen在大规模强化学习（RL）以增强推理能力方面的第一步。通过这一旅程，Qwen认识到预训练语言模型中尚未开发的可能性。更强大的基础模型与依托规模化计算资源的RL相结合，将会使Qwen更接近实现人工通用智能（AGI）。

参考链接

Demo: <https://huggingface.co/spaces/Qwen/QwQ-32B-Demo>

Qwen 官方文档:

<https://qwen.readthedocs.io/en/latest/>

Qwen 博客

<https://qwenlm.github.io/zh/blog/qwq-32b/>

大模型访问和下载:

<https://modelscope.cn/models/Qwen/QwQ-32B>

<https://huggingface.co/Qwen/QwQ-32B>

分享这篇文章



相关文章推荐

DeepSeek R1 论文解读

本文介绍了深度求索（DeepSeek）公司推出...

Chain of Draft 论文解读

本文介绍了Chain of Draft（CoD）论文，并对其技...

DeepSeek FlashMLA 代码解读

本文介绍了深度求索
(DeepSeek) 公 ...