PR #25956 完整报告

sgl-project/sglang

Avoiding the problem of printing a large number of compatibility warn···

合并时间：2026-05-21 22:10

原文链接：http://prhub.com.cn/sgl-project/sglang/pull/25956

执行摘要

- 一句话：抑制 Transformers 兼容性警告日志
- 推荐动作：PR 改动简单，无需深入精读。但 review 中关于环境变量覆盖的讨论值得注意：对于测试脚本，应优先使用 os.getenv("VAR", default) 模式以保留开发者调试的可能性。

功能与动机

当 pipeline 执行测试用例时，Transformers 版本更新导致相关方法调用时打印大量兼容性警告日志，影响执行时间和失败用例的定位。

实现拆解

1. 在 python/sglang/test/ascend/vlm_utils.py 的 setUpClass 方法中，设置环境变量 TRANSFORMERS_VERBOSITY，值取自系统环境变量，若未设置则默认为 error。
2. 在 python/sglang/test/ascend/gsm8k_ascend_mixin.py 的 env 字典中，添加 TRANSFORMERS_VERBOSITY 键，值同样采用 os.getenv("TRANSFORMERS_VERBOSITY", "error")，允许开发者通过系统环境变量覆盖。

关键文件：

- python/sglang/test/ascend/vlm_utils.py（模块 VLM 测试；类别 test；类型 test-coverage）：在 VLM 测试的 setUpClass 中设置 TRANSFORMERS_VERBOSITY 环境变量，抑制 Transformers 兼容性警告日志。
- python/sglang/test/ascend/gsm8k_ascend_mixin.py（模块 GSM8K 测试；类别 test；类型 test-coverage）：在 GSM8K 测试的 env 字典中添加 TRANSFORMERS_VERBOSITY，采用相同策略，避免硬编码。

关键符号：未识别

关键源码片段

python/sglang/test/ascend/vlm_utils.py

在 VLM 测试的 setUpClass 中设置 TRANSFORMERS_VERBOSITY 环境变量，抑制 Transformers 兼容性警告日志。

```
# python/sglang/test/ascend/vlm_utils.py
# 在 setUpClass 末尾设置 Transformers 日志级别
```

```
    @classmethod
    def setUpClass(cls):
        # ... 原有设置 ...
        os.environ["OPENAI_API_KEY"] = cls.api_key
        os.environ["OPENAI_API_BASE"] = f"{cls.base_url}/v1"
        # 设置 Transformers 日志级别为 error 以屏蔽兼容性警告
        # 允许通过系统环境变量覆盖，默认值为 error
        os.environ["TRANSFORMERS_VERBOSITY"] = os.getenv(
            "TRANSFORMERS_VERBOSITY", "error"
        )
```

python/sglang/test/ascend/gsm8k_ascend_mixin.py

在 GSM8K 测试的 env 字典中添加 TRANSFORMERS_VERBOSITY，采用相同策略，避免硬编码。

```
# python/sglang/test/ascend/gsm8k_ascend_mixin.py
# 在 env 字典中添加 Transformers 日志级别控制
class GSM8KAscendMixin(ABC):
    # ... 其他属性 ...
    env = {
        **os.environ,
        "PYTORCH_NPU_ALLOC_CONF": "expandable_segments:True",
        # ... 其他环境变量 ...
        "P2P_HCCL_BUFFSIZE": "20",
        # 设置 Transformers 日志级别为 error 以屏蔽兼容性警告
        # 使用 os.getenv 允许开发者通过系统环境变量覆盖默认值
        "TRANSFORMERS_VERBOSITY": os.getenv("TRANSFORMERS_VERBOSITY", "error"),
    }
```

评论区精华

Review 中 gemini-code-assist[bot] 指出初始硬编码 "error" 会覆盖系统环境变量，不利于调试；同时提醒该方案仅覆盖两个测试文件，未解决其他测试套件中的日志噪声。作者随后改为 os.getenv("TRANSFORMERS_VERBOSITY", "error") 以允许覆盖。

- TRANSFORMERS_VERBOSITY 硬编码与覆盖问题 (design): 作者改为使用 os.getenv 读取系统环境变量，若未设置则默认为 error，允许开发者覆盖。

风险与影响

- 风险：风险极低：变更仅作用于测试环境的环境变量，不影响生产逻辑。但如果其他测试套件未做类似处理，仍可能存在日志噪声。需注意 gsm8k_ascend_mixin.py 中设置的位置在 **os.environ 之后，会覆盖系统环境变量（当前版本已使用 os.getenv 解决）。
- 影响：影响范围限于 Ascend NPU 上的两个测试套件：VLM 测试和 GSM8K 测试。用户不会感知到任何行为变化。CI 日志将更清晰，测试执行时间可能略有缩短。其他测试套件（如 CUDA 端）未覆盖。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR