

PR #24652 完整报告

sgl-project/sglang

fix(mimo_v2): auto-disable multimodal when vision/audio configs are absent

合并时间: 2026-05-08 13:40

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/24652>

执行摘要

- 一句话: 修复纯文本 MiMoV2 启动崩溃
- 推荐动作: 值得合并, 修复了一个清晰的配置 - 模型不一致 bug, 改动小且安全。

功能与动机

纯文本 MiMoV2 checkpoint 启动时因缺少 `vision_config` 而抛出 `AttributeError` (PR body 中贴出了完整栈跟踪)。模型层 `mimo_v2.py:1038-1040` 的 `_is_multimodal` 已要求同时存在 `vision_config` 和 `audio_config` 才视为多模态, 但配置层未做对应检查, 导致两处判定不一致。

实现拆解

1. 定位问题: 在 `python/sglang/srt/configs/model_config.py` 的 `enable_multimodal` 自动判定分支中, 增加针对 MiMoV2 的特殊逻辑。
2. 新增判断条件: 当 `architectures[0]` 在 `MIMO_V2_MULTIMODAL_ARCHS` 中, 且 HF 配置不包含 `vision_config` 或 `audio_config` 任一属性时, 将 `enable_multimodal` 设为 `False`。
3. 日志记录: 当禁用多模态时, 打印一条 `info` 日志说明原因。
4. 无其他改动: 变更仅 10 行, 不涉及测试、配置或部署配套设施。

关键文件:

- `python/sglang/srt/configs/model_config.py` (模块 配置层; 类别 `source`; 类型 `data-contract`): 配置层入口, 新增逻辑使配置级多模态开关与模型级检测一致。

关键符号: 未识别

关键源码片段

`python/sglang/srt/configs/model_config.py`

配置层入口, 新增逻辑使配置级多模态开关与模型级检测一致。

```
# python/sglang/srt/configs/model_config.py 中 enable_multimodal 自动判定逻辑
if enable_multimodal is None:
    mm_disabled_models = [
        "Gemma3ForConditionalGeneration",
        "Llama4ForConditionalGeneration",
        "Step3VLForConditionalGeneration",
```

```

]
if (
    self.hf_config.architectures[0] in mm_disabled_models
    and self.model_impl != ModelImpl.TRANSFORMERS
):
    enable_multimodal = False
    logger.info(
        f"Multimodal is disabled for {self.hf_config.model_type}. "
        "To enable it, set --enable-multimodal."
    )
elif self.hf_config.architectures[0] in MIMO_V2_MULTIMODAL_ARCHS and not (
    hasattr(self.hf_config, "vision_config")
    and hasattr(self.hf_config, "audio_config")
):
    # 新增分支：当架构是 MiMoV2 但 HF 配置缺少 vision_config 或 audio_config 时，
    # 视为纯文本变体，禁用多模态以避免后续多模态处理器初始化崩溃。
    enable_multimodal = False
    logger.info(
        "Multimodal is disabled for this MiMoV2 checkpoint: "
        "vision_config/audio_config not found in the model config "
        "(likely a text-only MiMoV2 variant)."
    )
else:
    enable_multimodal = True

```

评论区精华

审核者 ShangmingCai 批准了 PR，无进一步讨论。CI 测试也已通过。

- 暂无高价值评论线程

风险与影响

- 风险：风险极低。变更仅增加一个 elif 分支，且检查条件与模型层已有的 `_is_multimodal` 逻辑完全一致。多模态 MiMoV2 变体不受影响，因为其 HF 配置包含 `vision_config` 和 `audio_config`，不会进入该分支。
- 影响：只有纯文本 MiMoV2 用户受到影响——他们原本无法启动服务，现在可以正常启动。多模态 MiMoV2 用户无感知。团队无需额外维护成本。
- 风险标记：暂无

关联脉络

- 暂无明显关联 PR