

PR #22704 完整报告

sgl-project/sglang

[NPU] --attn-cp-size --init-expert-location --eplb-algorithm parameter docs update

合并时间: 2026-04-13 22:42

原文链接: <http://prhub.com.cn/sgl-project/sglang/pull/22704>

执行摘要

本次 PR 更新了 Ascend NPU 支持特性文档，澄清了 `--attn-cp-size`、`--init-expert-location` 和 `--eplb-algorithm` 三个参数的约束和有效值，属于低风险、小范围的文档维护，旨在提升用户配置的清晰度，无需深入技术审查。

功能与动机

根据 PR 标题和 body（含截图，但内容未提供），动机是更新 Ascend NPU 相关参数的文档，以澄清其约束和有效值。这延续了近期多个 NPU 文档 PR（如 #22705、#22687、#22700）的维护趋势，旨在持续改进平台文档的准确性和可用性，帮助用户避免配置错误。

实现拆解

仅修改了 `docs/platforms/ascend/ascend_npu_support_features.md` 文件，具体变更如下：

| 参数 | 变更前 | 变更后 | 说明 | |-----|-----|-----|-----| | `--attn-cp-size` | `Type: int` | `Type: int; must be equal to --tp-size` | 添加了必须等于 `--tp-size` 的约束 | | `--init-expert-location` | `Type: str` | `trivial, <path.pt>, <path.json>, <json_string>` | 明确了有效值格式，包括字面值和文件路径占位符 | | `--eplb-algorithm` | `Type: str` | `auto, deepseek` | 修正为具体选项列表，移除了可能引起混淆的尖括号 |

评论区精华

review 中只有 `gemini-code-assist[bot]` 提出了两条格式建议：

"For consistency with the formatting in line 89, consider using a semicolon instead of parentheses to describe the constraint." "The use of angle brackets (<...>) typically denotes a placeholder for a user-provided value. Since `deepseek` is a literal value for this parameter, the angle brackets should be removed to avoid confusion."
作者在后续提交中采纳了这些建议，确保了文档格式的一致性和清晰性。

风险与影响

- 风险：极低，纯文档变更无代码逻辑影响；唯一潜在风险是文档准确性，但 review 已纠正格式问题。
- 影响：帮助 Ascend NPU 用户更清晰地理解参数用法，减少配置错误；对系统无运行时影响；属于常规文档维护，延续了近期 NPU 文档更新节奏。

关联脉络

本次 PR 是近期 Ascend NPU 文档维护序列的一部分，与 #22705、#22700、#22697 等 PR 类似，都修改了 `ascend_npu_support_features.md` 文件，反映了团队对 NPU 平台文档持续优化的重视。这些变更共同提升了文档的完整性和用户友好性，支持 SGLang 在异构硬件（如 Ascend NPU）上的部署体验。