

# PR #6358 完整报告

verl-project/verl

[doc] chore: OPD docs

合并时间: 2026-05-15 12:53

原文链接: <http://prhub.com.cn/verl-project/verl/pull/6358>

## 执行摘要

- 一句话: 新增 OPD 算法文档与多教师示例脚本
- 推荐动作: 建议团队尽快修复 review 中指出的教师池配置问题, 避免用户踩坑。OPD 文档内容详实, 可作为算法参考和用户教程。对于想了解蒸馏进阶用法的开发者, 推荐精读 docs/algos/opd.md 中的 MOPD 和 Policy Gradient 部分。

## 功能与动机

PR body 明确说明: Adds docs for OPD. Also adds back a missing example script for multiple teachers. 此前 OPD 缺乏正式文档, 多教师脚本在重构中被遗漏, 本 PR 填补了这些空白。

## 实现拆解

1. 创建 OPD 算法文档 (docs/algos/opd.md): 完整介绍 OPD 的背景 (对比 SFT/KD/RLVR)、数学形式化、学生与教师模型的配置参数、蒸馏损失函数 (K1/K3/Forward KL TopK)、Policy Gradient 模式、多教师扩展 (MOPD)、训练流程架构图、实用提示 (性能、调试) 等。
2. 新增多教师启动脚本 (examples/on\_policy\_distillation\_trainer/run\_qwen3\_8b\_mopd\_fsdp.sh): 支持同时使用两个教师模型 (GSM8K 文本 + Geo3K 视觉) 进行蒸馏, 通过环境变量控制教师重复数和 TP 大小, 脚本内使用 Hydra 多值参数列表合并训练数据集。
3. 更新示例 README (examples/on\_policy\_distillation\_trainer/README.md): 修正表格增加多教师行, 补充多教师环境变量和配置键说明。
4. 注册文档索引 (docs/index.rst): 将 algos/opd.md 加入 toctree, 使其在官方文档中可访问。

关键文件:

- docs/algos/opd.md (模块 算法文档; 类别 docs; 类型 documentation): OPD 算法核心文档, 包含完整原理、配置和使用指南, 是本次 PR 的主要变更。
- examples/on\_policy\_distillation\_trainer/run\_qwen3\_8b\_mopd\_fsdp.sh (模块 示例脚本; 类别 other; 类型 configuration): 多教师蒸馏的启动脚本, 被 review 指出配置问题, 是用户直接使用的入口。
- examples/on\_policy\_distillation\_trainer/README.md (模块 示例说明; 类别 docs; 类型 documentation): 更新示例说明文档, 增加多教师脚本条目。

- docs/index.rst (模块 文档索引; 类别 docs; 类型 documentation) : 将 OPD 文档加入索引树, 使其在文档站可见。

关键符号: 未识别

## 关键源码片段

[examples/on\\_policy\\_distillation\\_trainer/run\\_qwen3\\_8b\\_mopd\\_fsdp.sh](#)

多教师蒸馏的启动脚本, 被 review 指出配置问题, 是用户直接使用的入口。

```
#!/usr/bin/env bash
# On-policy distillation | multi-teacher (gsm8k text + geo3k VL) | vLLM rollout | FSDP training |
NVIDIA GPUs

set -xeuo pipefail

# ---- user-adjustable ----
STUDENT_MODEL=${STUDENT_MODEL:-Qwen/Qwen3-VL-8B-Instruct}
GSM8K_TEACHER_MODEL=${GSM8K_TEACHER_MODEL:-Qwen/Qwen3-32B}
GEO3K_TEACHER_MODEL=${GEO3K_TEACHER_MODEL:-Qwen/Qwen3-VL-32B-Instruct}

# 注意: 以下 NNODES 控制训练器节点数, 但教师池总大小需独立计算
NNODES=${NNODES:-1}
NGPUS_PER_NODE=${NGPUS_PER_NODE:-8}

# 每个教师的副本数, 总教师 GPU 数 = sum(num_replicas) * teacher_tp
TEACHER_NNODES=${TEACHER_NNODES:-1}
TEACHER_NUM_REPLICAS_GSM8K=${TEACHER_NUM_REPLICAS_GSM8K:-1}
TEACHER_NUM_REPLICAS_GEO3K=${TEACHER_NUM_REPLICAS_GEO3K:-1}
teacher_tp=${TEACHER_TP:-2}
TEACHER_WORLD_SIZE=$(( (TEACHER_NUM_REPLICAS_GSM8K + TEACHER_NUM_REPLICAS_GEO3K) * teacher_tp ))

# ... 后续配置
```

## 评论区精华

仅有一条来自 [gemini-code-assist\[bot\]](#) 的 review 评论 (high 优先级), 指出多教师脚本中教师资源池配置错误地与训练器节点数 **NNODES** 耦合。在 ver1 中, 教师池是独立分配, 总大小必须等于所有教师 world size 之和。若用户设置 **NNODES > 1**, 该脚本会错误地将教师池大小乘以 **NNODES**, 导致验证时 **ValueError**。评论建议解耦教师池配置并确保 Hydra 覆盖语法正确。该问题在合并前未得到作者明确回复, 需后续跟进。

- 多教师脚本教师池资源配置错误 (correctness): 未在 PR 中修复, 需后续关注。

## 风险与影响

- 风险: 主要风险在新增的多教师脚本中教师资源池配置与训练器节点数耦合, 用户按需扩展时可能遇到验证失败。文档本身不存在功能风险, 但 OPD 算法配置项较多, 用户可能因理

解偏差导致蒸馏效果不佳。此外，文档中引用的 `teacher_logprobs.py` 文件尚有一个 TODO，表明该组件仍在开发中。

- 影响：对用户：提供了 OPD 从理论到实践的完整指南，降低了使用门槛；多教师脚本填补了多蒸馏场景的空缺。对系统：无代码逻辑变更，不影响已有功能。对团队：规范了蒸馏相关的文档，有利于功能推广和协作。
- 风险标记：教师池配置耦合，遗留 TODO 待完成

## 关联脉络

- PR #6350 [fsdp] fix: emit distillation outputs in use\_remove\_padding=False path (#6293): 蒸馏功能线的 bugfix 补丁，本 PR 补充了该功能的文档和示例。
- PR #6360 Revert "[fsdp] fix: emit distillation outputs in use\_remove\_padding=False path (#6293)": 回退了蒸馏输出的修复，表明该功能线存在不确定性，文档和脚本可作为参考。