



POLIXIR
南栖仙策



2022

强化学习控制白皮书

2022- RL-Control White Paper

前言：时代背景和行业趋势

能源与流程工业制造是国民经济必不可少的基础产业，是支撑经济持续增长和世界经济的重要力量。经过几十年的发展，我国的工业生产已经有了跨越式的进步，整体行业规模迅速提高，整体实力不断变强。与此同时，我国工业生产面临的能耗高、资源消耗高、产品附加值低、环境污染严重等问题越来越突出。

当前，随着人工智能、移动互联网、云计算、工业互联网等技术的快速发展，智能制造已成为提升能源企业生产整体竞争力的核心技术，人工智能技术已成为生产工艺发展的重要趋势。人工智能技术加速了智能制造的发展。以机械技术为基础的蒸汽机和反馈调速器的出现引发了第一次工业革命；电力和基于电气技术的控制系统的出现引发了第二次工业革命；可编程逻辑控制器（programmable logic controller, PLC）和集散控制系统（distributed control system, DCS）的出现引发了第三次工业革命；基于数据驱动，将人工智能、移动互联网、云计算、建模、控制、优化等信息技术与流程工业物理资源深度融合，第四次工业智能革命正在袭来。

人工智能赋能时代已经到来，南栖仙策正在创造新一代控制方法与产品。我们期望您的加入与合作，共同谱写未来华章。

目录

CONTENT

1. 流程工业生产控制的难题	1
2. 南栖强化学习智能控制	2
2.1 Polixir REVIVE 工业软件	4
(一) Polixir REVIVE 工业软件的特色	4
(二) Polixir REVIVE 工业软件设计四步法	4
2.2 ReinOptima 工业边缘控制器	6
2.3 云边协同架构	7
3. 技术优势与案例	9
3.1 高精度控制	9
3.2 高稳定控制	12
3.3 大滞后控制	16
3.4 全局目标控制	18



1. 流程工业生产控制的难题

当前能源与流程制造企业的生产控制过程一般可以分为三个层面：

- 管理人员得到生产过程各设备的参数，然后根据自身积累的经验和知识，对产品综合生产指标（产品质量、能耗和成本等）的目标值范围做出决定。
- 生产部门经理根据得到的生产信息，通过自己积累的专家经验来决定生产制造全流程的生产指标目标。
- 运行管理者和工艺工程师针对生产指标目标，根据工作人员的经验做出决策，设置控制系统。

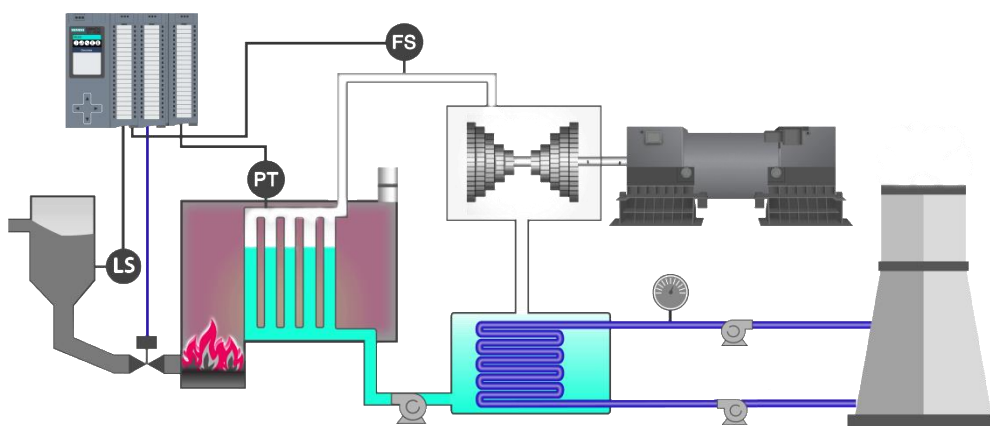


图 1. 流程制造示意图

在现有的生产管理流程中，多个环节由人工把控，依赖经验知识且响应速度慢；过程控制基于经典控制算法，难以满足大范围动态条件复杂系统的优化控制需求；系统运维缺乏数据支撑，故障难预警、风险不可控。以上流程控制中的不足，常造成企业**产品质不够稳定、生产线持续性不够高、产量波动的应对能力不足、生产损耗过大成本难以控制**等问题，已成为现代能源企业生产控制过程中的关键难题。



2. 南栖强化学习智能控制

Polixir REVIVE¹ 是南栖仙策自主研发的数据驱动强化学习技术，攻克理论局限，创新基于数据驱动环境虚拟强化学习算法，打破了经典强化学习技术无法突破封闭环境的屏障，实现了实际业务中智能决策的落地，并在多种业务场景中得以验证。

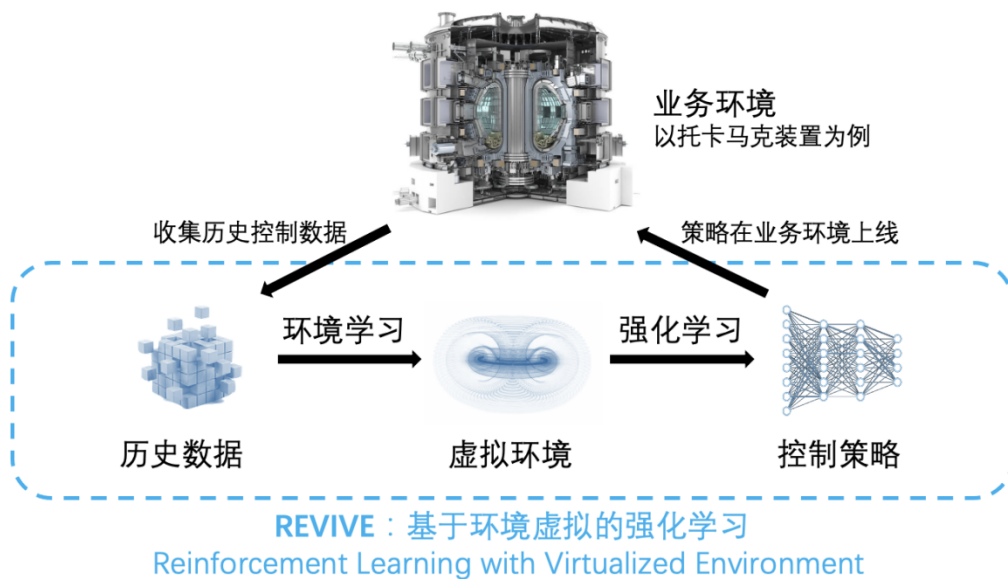


图 2. Polixir REVIVE 主要流程

Polixir REVIVE 在流程工业控制领域，已应用于火力发电、化工生产、自来水/污水生产处理等工业场景，形成了一套完整的流程工业控制优化、预测性维护、工业生产节能降碳解决方案。方案能够满足智能运行优化、动态条件适应、风险早期识别等智能自主工业 4.0 需求需要，具备如下特性：

- (1) 适用于高动态性系统，能稳定控制运行指标；
- (2) 生产条件动态变化下，能敏捷感知快速适配；

¹ REVIVE = Reinforcement Learning with Virtualized Environment, 产品主页请访问 <http://REVIVE.cn>



- (3) 面向运行指标最优化，能自主求解最优控制；
- (4) 针对全系统总体目标，能全局多点协同控制；
- (5) 可实时推演系统状态，能早期预测防范风险。

◇ Polixir REVIVE 与 MPC 的对比

MPC (Model Predictive Control) 控制技术由两部分组成：系统模型用于预测未来状态，优化器用于基于未来状态求解最优控制。与 MPC 相似，REVIVE 强化学习控制首先学习虚拟环境模型（即系统模型），并基于虚拟环境求解最优控制策略。两者在结构上相似，但技术路线的不同带来诸多关键差异。

表 1. Polixir REVIVE 强化学习控制与 MPC 控制的差异

	MPC 控制	Polixir REVIVE 强化学习控制
系统建模方法不同	<ul style="list-style-type: none"> ● 主要基于人工动力学建模, 严重依赖专家经验, 建模耗时长、成本高 ● 系统模型误差的大小直接决定了最终控制的性能, 且难以自动更新减小误差 	<ul style="list-style-type: none"> ● 主要基于数据驱动, 无需动力学建模, 可选择性的注入专家经验, 建模耗时短、成本低 ● 虚拟环境模型可自然的随数据的积累而更新, 贴近真实系统
系统模拟能力不同	<ul style="list-style-type: none"> ● 大量采用了近似技术, 不可避免的建模误差导致系统模型只能进行短期预测 ● 受限于凸优化技术, 难以对多环节的复杂系统进行建模 	<ul style="list-style-type: none"> ● 基于虚拟环境学习的数学理论突破, 可进行数千步以上的长期预测 ● 基于神经网络模块的流程建模方式, 可以对大至城市级的系统进行建模
控制策略能力不同	<ul style="list-style-type: none"> ● 基于系统模型和优化器进行实时求解, 在线计算开销 ● 受限于凸优化技术, 控制策略的代价函数不可随意设置 	<ul style="list-style-type: none"> ● 得益于深度学习与强化学习的长足发展, 具备高度的非线性和自适应能力, 且神经网络推理计算开销小 ● 在虚拟环境中接收对应业务需求设计的奖励, 可面向全局目标求解有全局约束的问题

表 1 列举的差异，使得 Polixir REVIVE 强化学习控制具有实施快、成本低、建模灵活、适用面广等特点，更具实用性。



2.1 Polixir REVIVE 工业软件

Polixir REVIVE 工业软件是 REVIVE 强化学习控制的训练系统，提供了包括工业数据接入、工艺决策流设计、工业仿真环境搭建、控制策略优化、一体化智能应用运维服务等功能。

(一) Polixir REVIVE 工业软件的特色

1. 多源数据接入能力：支持 ERP、MES、DCS 等不同平台不同系统的数据接入服务、支持实时数据、离线数据的采集、支持结构化和非结构化数据的传输，支持 https、MQ 消息、webservers、TCP 等多种数据接口协议。
2. 工业数理仿真环境构建能力：支持基于历史数据搭建工业仿真环境，满足大型、高危场景推演需要，基于实际数据的仿真环境更贴近实际生产环境。
3. 一键工艺智能优化建模能力：支持用户对生产环境的多维度分析，从能源利用效率、产品质量分析、故障预警预测角度一键自动生成智能控制策略，帮助企业提高能源利用效率、提升产品质量。
4. 自动化运维监控服务能力：支持企业自主多维度智能化应用在线自动化运维和监控服务，保障企业 7*24 小时安全、稳定、可靠运行。

(二) Polixir REVIVE 工业软件设计四步法



在使用 REVIVE 的预备阶段，需采集实际工业制造过程中涉及到的人、机、料、法、环对应的性能状态数据；进行简单的数据整理工作，包括数据整理、数据清洗、特征生成等。整体过程操作可分四步完成。

第一步，构建业务模型：根据业务控制基础逻辑，确定数据各个维度在控制逻辑图中各个模块的输入输出位置；

第二步，虚拟环境学习：导入历史生产数据，使用 REVIVE 中的环境模型训练功能，训练控制逻辑图中所有的神经网络模块，得到可运行的数据仿真环境；可视化验证虚拟环境与真实环境的一致性，便于评价虚拟环境模型的性能表现；

第三步，控制模型优化：设置奖励函数（即代价函数、目标函数），使用 REVIVE 的强化学习算法中的策略模型训练功能，得到优化后的控制模型；

第四步，控制模型验证：将优化后的控制模型在虚拟环境中进行验证，评估有效后，进行实际生产验证实验。若达到预期，优化结束，否则加入新的业务验证数据后返回第二阶段继续。



图 3. Polixir REVIVE 实施流程图



2.2 ReinOptima 工业边缘控制器

基于 Polixir REVIVE 得到的控制模型可载入到「ReinOptima 睿优」工业边缘控制器进行部署实施，另外 ReinOptima 还可同时加载 REVIVE 产生的虚拟环境模型，从而具备预测性维护的功能。ReinOptima 工业边缘控制器具备低延迟、低数据通信和存储成本、高网络可用性，能够在企业现有控制系统或单元基础上实现不停产智能化改造升级，大大降低改造成本。ReinOptima 工业边缘控制器具备的能力包括：

1. 多源数据实时自动采集能力：支持 OPC、modbus 等主流工业接口协议，满足主流工业传感器、控制器如 PLC（西门子、施耐德、AB 等）、DCS 系统的数据接入服务。
2. 自动化部署调试能力：支持用户自动化部署服务，支持在线实时数据调试服务，支持手自投切服务。
3. 实时监测和安全保障能力：支持用户投产后实时监控 AI 模型运行态势；自定义安全边界和告警，告警自动切换并声光报警；具备实时心跳监测、数据校验服务。

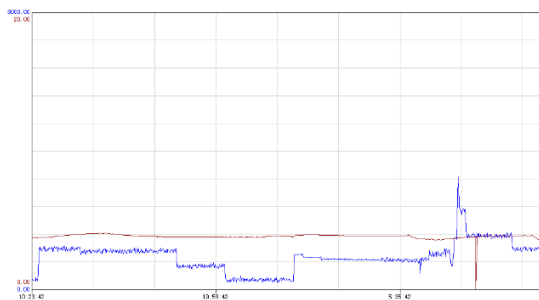


图 4. 「ReinOptima 睿优」工业边缘器运行监视示意图



2.3 云边协同架构

南栖的 Polixir REVIVE 工业软件和 ReinOptima 工业边缘控制器并非割裂的系统。通过云边协同架构，可实现 REVIVE 训练系统与 ReinOptima 边缘控制器的联动，当 ReinOptima 的预测性维护模块识别出系统发生偏移时，自动将 ReinOptima 中的历史数据传到 REVIVE 系统，从而更新虚拟环境模型和控制策略，实现对环境变化的自动适应和持续自我进化。

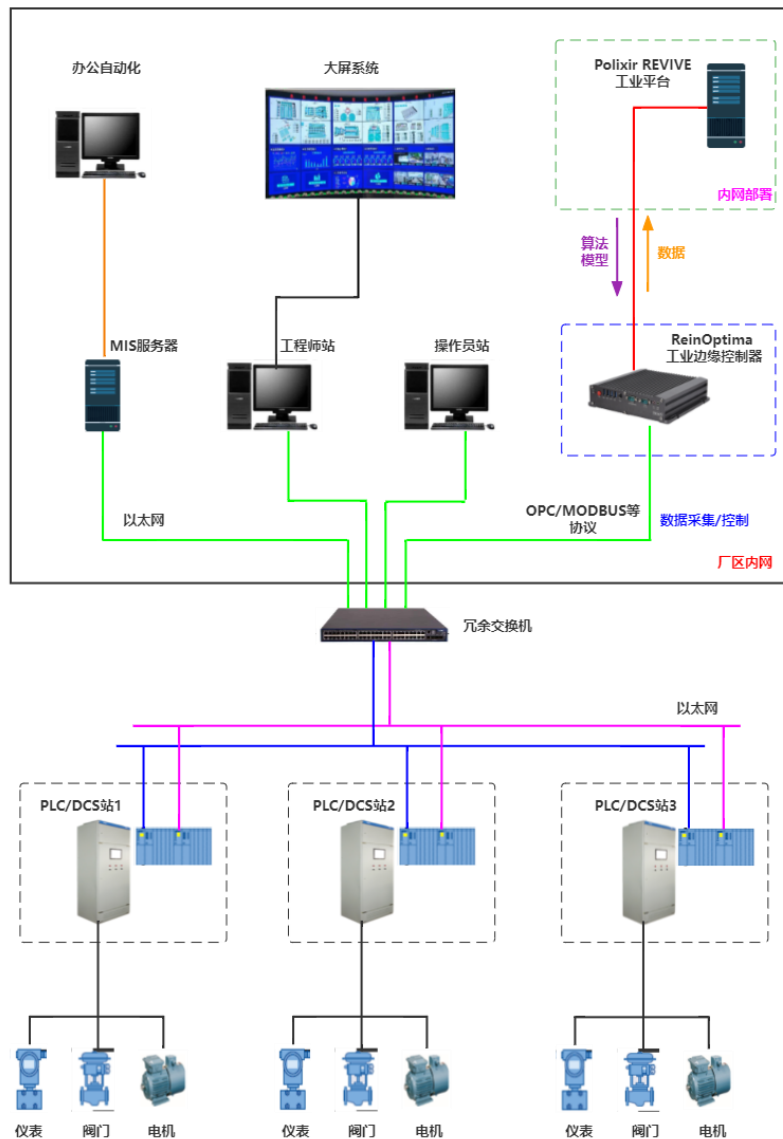


图 5.南栖云边协同解决方案示意图



南栖云边协同架构在工业客户场景中具有如下突出优势：

（一）场景适配性强、改造成本低

该方案充分考虑当前工业用户现有系统的建设情况，采用标准的工业 modbus、opc 协议接口，能够对接客户现有的 PLC/DCS 柜等控制单元，无需进行大规模的改造升级。仅需要在边缘侧增加 ReinOptima 工业边缘控制器即可实现模型部署、调试和投生产。

（二）状态自监测、模型自学习

Polixir REVIVE 工业软件支持私有化部署，能够与部署在边缘侧的 ReinOptima 工业边缘控制器进行协同服务，实时接收到边缘侧返回的模型运行状态信息并根据需要完成模型的自学习和更新。依托 ReinOptima 工业边缘控制器自带的状态实时监测能力，当工况发生变化时会自动将状态变化信息协同传递给 Polixir REVIVE 工业软件系统，系统依据反馈的状态信息和最新的生产数据进行模型的重新训练，并生成新的环境模型和决策模型，由运营人员下达指令将全新的模型下发到边缘设备完成模型的自主学习和自主更新。

（三）模型维护成本低、安全可靠性强

Polixir REVIVE 工业软件系统支持多模型的在线运维管理服务，运营人员可实现在线可视化运维管理服务，支持拖拉拽式模型的自主开发服务；支持模型手动/自动一键更新服务；支持模型快速部署、在线调试、一键投切服务。同时，为了确保安全生产需要，ReinOptima 工业边缘控制器支持自定义阈值设置、异常告警自动



切换、可视化在线监控以及实时心跳波监测等安全保障能力，能够与现有控制系统实现无扰动切换。

3. 技术优势与案例

3.1 高精度控制

场景描述

增压泵房是给水系统中的重要组成部分，为城市供水管网提供所需的水量和压力，是整个供水系统中能耗组成的最大单元。传统人工控制靠操作人员经验，存在精度不足和非最优策略等问题，导致水量和能耗出现过高或过低等情况，影响供水系统的整体效率。人工控制的难点主要体现在：

(1) **控制精度要求高**：当水务集团下发出水流量指令时，操作人员依据调控经验，始终无法突破 1000 吨水的调度精度，使出水总量波动大。

(2) **多个设备存在差异**：由于不同水泵的品牌和传感器老化程度不同，导致各水泵配水电耗和供给流量能力不同，人工难以精准感知，难以给出最优泵频组合。

实施过程

第一步：根据增压泵房业务逻辑，在 Polixir REVIVE 中为每个水泵构建决策流图。

第二步：基于过去一年的泵频控制历史数据和决策流图，对每个水泵进行虚拟水泵控制环境构建。



第三步：根据目标流量、调节池水位和管网状态，在虚拟水泵控制环境中学习优化泵频控制策略。

第四步：控制策略线下验证。

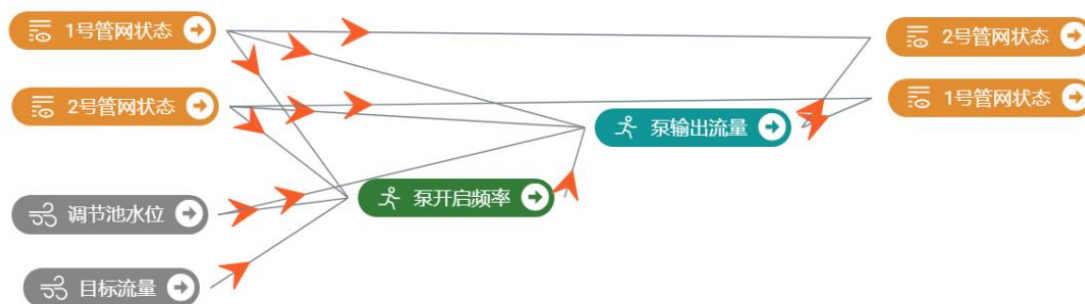


图 6. 水泵场景与 Polixir REVIVE 水泵供水决策流图

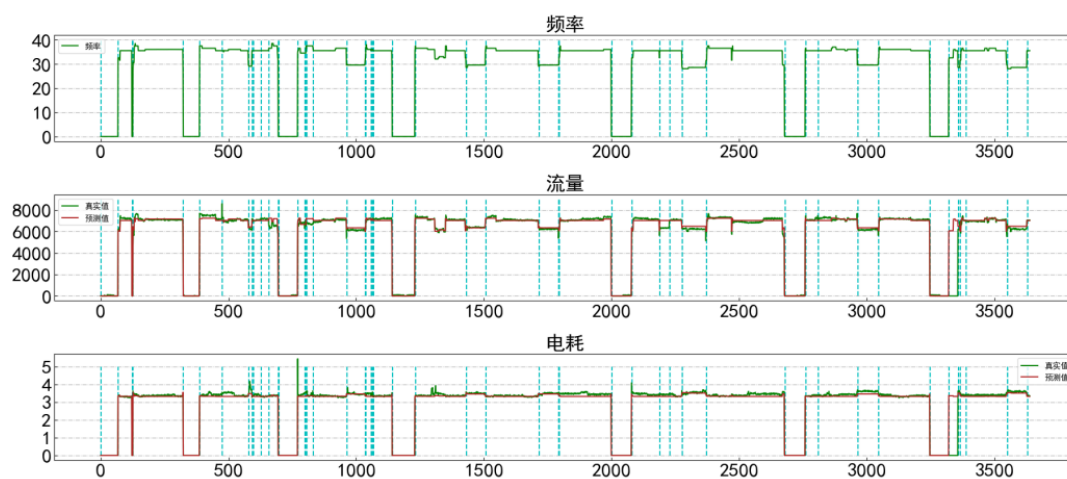


图 7. Polixir REVIVE 中随机单泵的虚拟水泵控制环境学习评估，由测试集中的真实泵频数据进行回放，在流量和电耗上观察到真实值和环境预测值近乎重合

实测结果与价值



在 24 小时的连续线下测试过程中，使用优化后的泵频控制策略，出口流量与目标流量差值上下浮动不超过 350m³/h，达到了系统所期望的精准控制。并与历史同时期千吨水电耗相比，节省电耗约 7.16%，有效提高了泵房的经济效益。

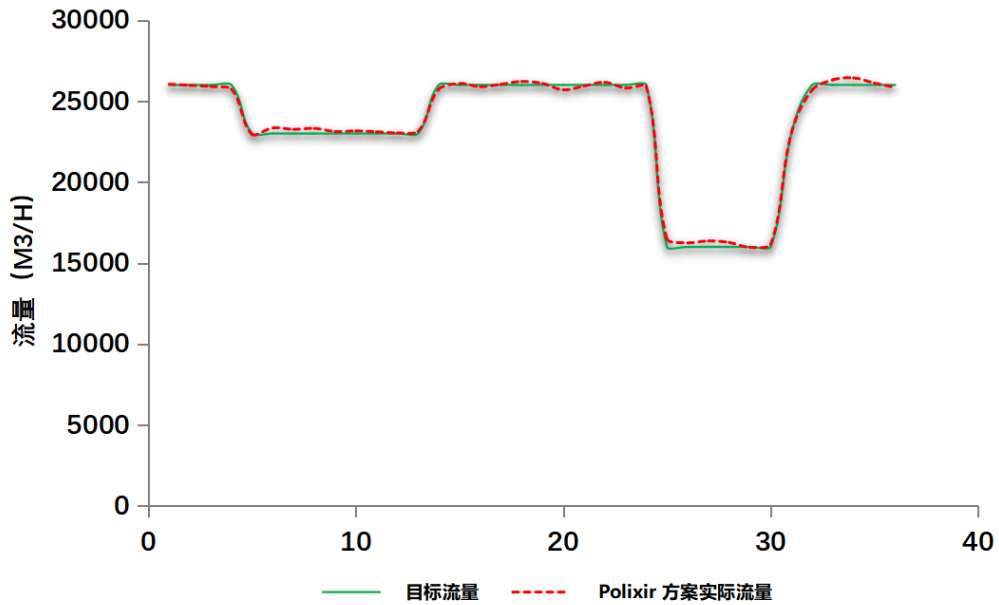


图 8：控制流量与目标流量对比，所控流量可快速跟踪响应目标流量。

同时期千吨水电耗对比

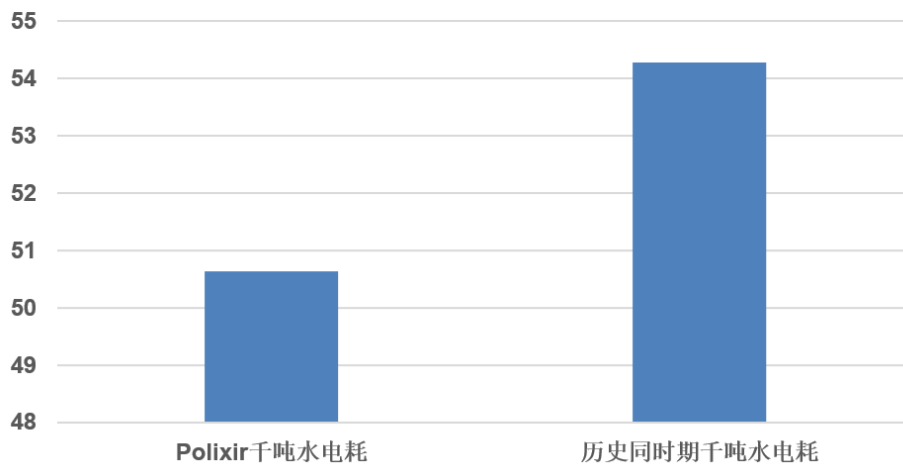


图 9. 千吨水电耗与历史同时期千吨水电耗对比，节省电耗约 7.16%



3.2 高稳定控制

场景描述

在燃煤火电机组控制中，主蒸汽温度是一个非常重要的被控参数，是提高电厂经济效益，保证机组安全运行不可缺少的环节。然而，由于外部扰动（如主蒸汽流量波动、主汽压力波动、汽水分离器水位波动、给水温度及流量变化、蒸汽吹灰投入等）变化频繁且扰动量较大，致使**主蒸汽温度长期不能稳定**。能否对主蒸汽温度进行稳定有效控制，使主蒸汽温度快速收敛平稳到目标温度，是 PID 现行控制的痛点难点，体现为 PID 控制存在严重的过调、延迟等问题，导致超温、温度过低等情况出现，影响发电效率等。

实施过程

第一步：根据主蒸汽温度控制业务逻辑，在 Polixir REVIVE 中构建主蒸汽温度控制的决策流图。

第二步：基于过去半年的多工况的主蒸汽温度控制历史数据和决策流图，进行虚拟主蒸汽温度控制环境构建。

第三步：根据目标温度值和控制反馈预测，在虚拟环境中控制 A/B 侧一二级减温水调节阀开度，不断学习优化减温水阀开度控制策略。

第四步：控制策略线下验证。



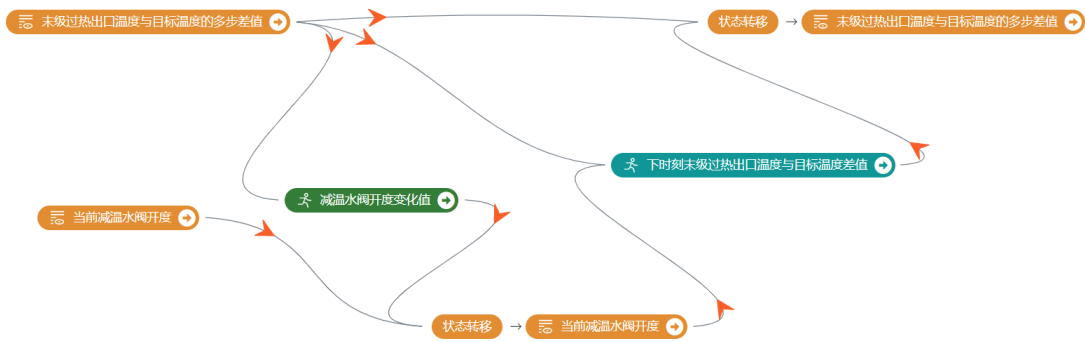


图 9. 主蒸汽场景与 Polixir REVIVE 主蒸汽温度控制决策流程图

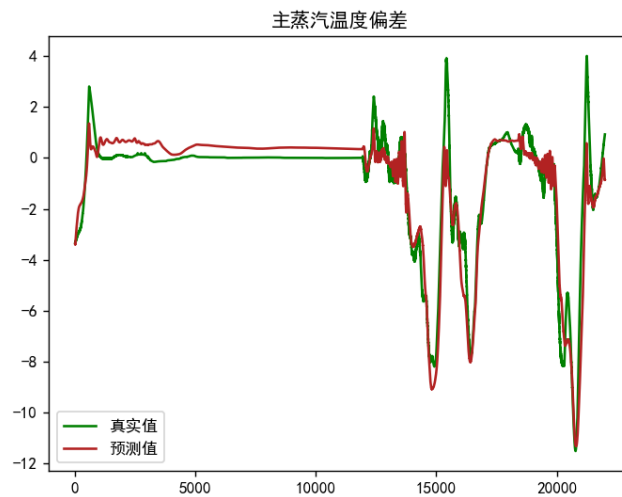


图 10. Polixir REVIVE 中的虚拟主蒸汽温度控制环境学习评估，在测试集中，将真实减温水阀开度进行连续回放，虚拟环境中的主蒸汽温度偏差和真实温度偏差趋势一致，幅度相近



实测结果与价值

在 4 小时未预见工况大幅波动的情况下, 二级出口温度始终稳定在 520°C 左右, 末级出口温度稳定在 570°C 左右, 上下浮动不超过 5°C。与此同时, 通过该控制点共节省燃煤 0.06%, 保障稳定的同时延长了设备的使用寿命。在相同的变负荷工况条件下, 相比于现运行的 PID 控制器, Polixir REVIVE 控制使得温度分布显著更集中在目标温度附近, 更稳定。

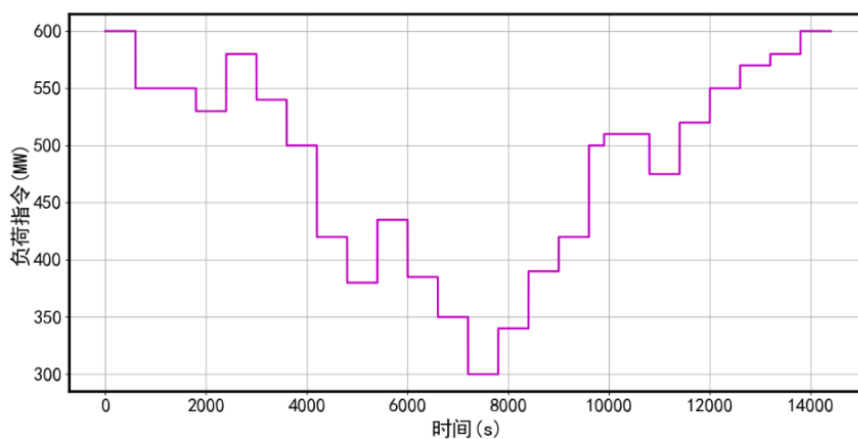


图 11. 4 小时的变负荷测试工况指令, PID 和 Polixir REVIVE 为相同测试条件

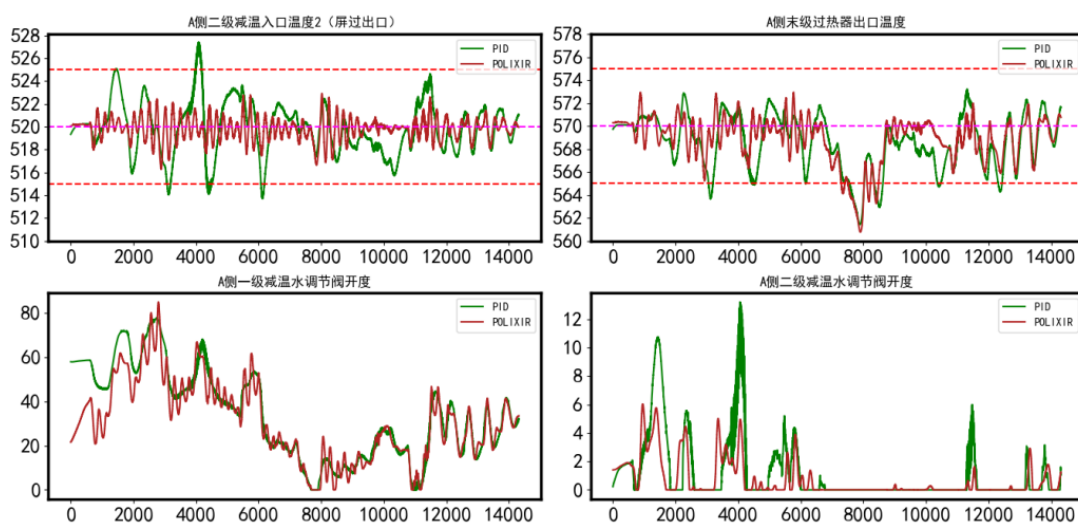


图 12. A 侧被控温度及阀门开度曲线可见, 相比于 PID, Polixir REVIVE 控制更稳定



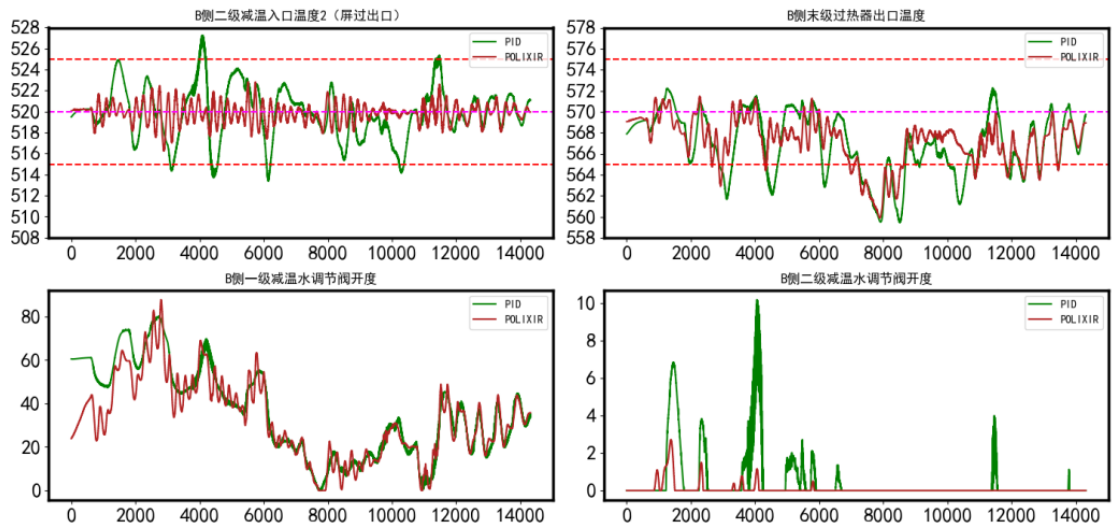


图 13. B 侧被控温度及阀门开度曲线可见，相比于 PID，Polixir REVIVE 控制更稳定

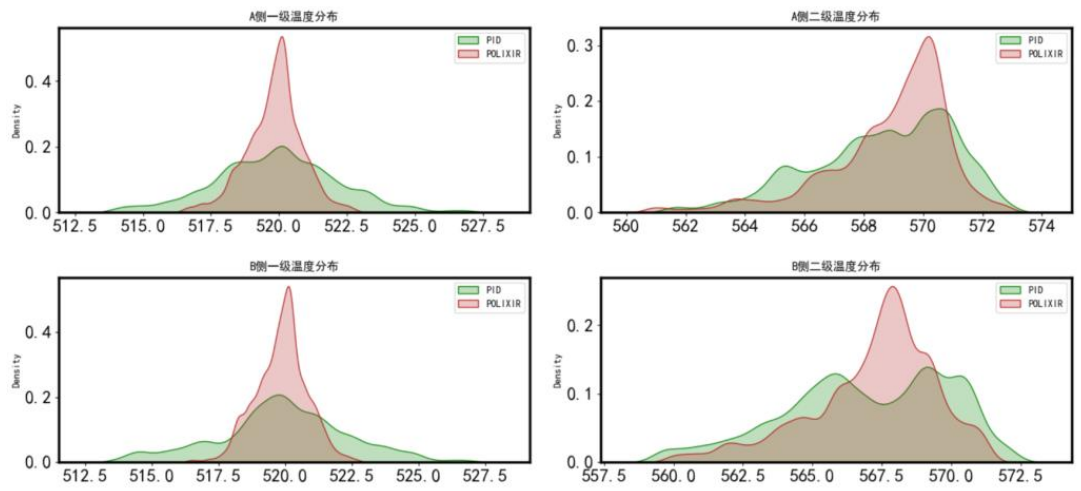


图 14. 被控温度分布图可见，相比于 PID，Polixir REVIVE 使系统一二级出口温度更集中于目标温度



3.3 大滞后控制

场景描述

在磨煤机生产过程中，对出口风速的控制决定了出口煤粉温度，进而影响下游的煤粉利用效率。尽可能根据给煤量、磨碗压差等参数，使出口煤粉温度接近目标温度，对于工艺整体的生产效率影响甚大。由于目标出口煤粉温度与控制风量之间存在**较长的时间差，整体系统存在滞后性、长距离控制**等问题，以及风速测不准导致风速控制器的 PID 控制失效，现有的自动化控制方案无法实现磨煤机保持在正常运行状态。

实施过程

第一步：根据磨煤机控制业务逻辑，在 Polixir REVIVE 中构建冷热一次风控制的决策流图。

第二步：基于过去半年的磨煤机温度、冷热一次风控制指令、出口风速、出口风粉温度和决策流图，进行虚拟冷热一次风控制环境构建。

第三步：根据目标温度设定值、风量设定值和控制反馈预测，在虚拟环境中控制冷热一次风挡板位，不断学习优化冷热风挡板的控制策略。

第四步：控制策略线下验证。



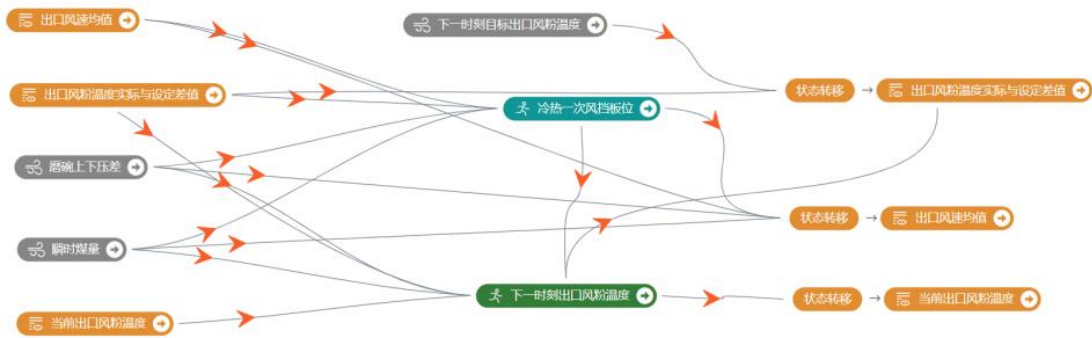


图 15. 冷热一次风场景与 Polixir REVIVE 冷热一次风控制优化决策流程图

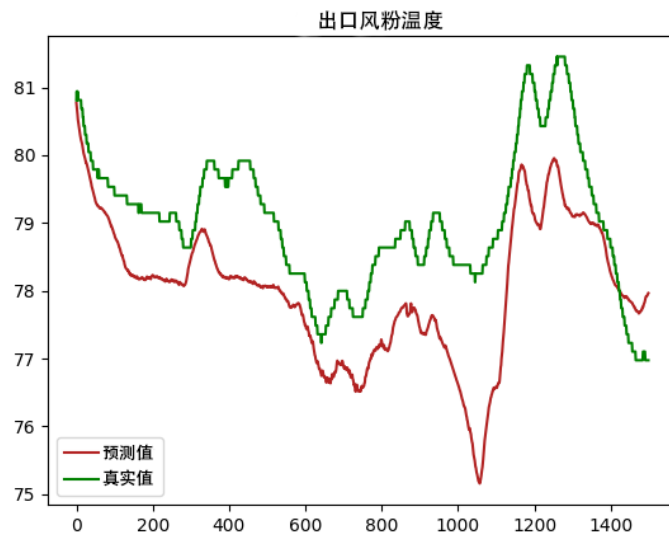


图 16. Polixir REVIVE 中的虚拟冷热一次风控制环境中的学习评估，在测试集中，将真实磨煤机温度、冷热一次风控制指令进行连续回放，虚拟环境中的出口风粉温度和真实温度趋势相近



实测结果与价值

在真实测试中，由原始温差均值 4.12，利用优化后的控制策略输出温差均值可降至 3.92，真实温度向目标温度显著靠近，提升比例约为 4.85%。相比于传统的 PID 控制器，基于 Polixir REVIVE 的冷热一次风挡板位控制策略使系统出风口风速在正常范围内 (20-35)，出口风粉温度分布更贴近目标温度，使系统运行工况更加稳定。

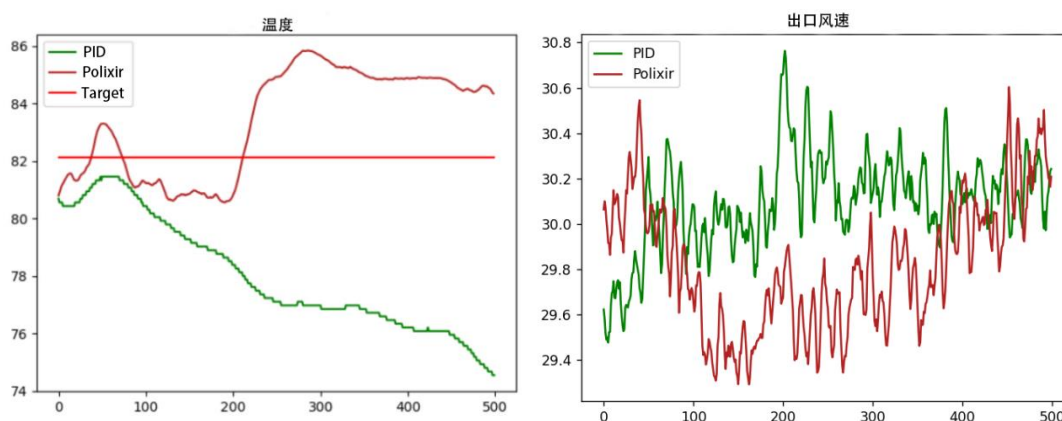


图 17. 温度控制曲线（左）与出口风速曲线（右）。相比于 PID 控制，基于 Polixir REVIVE 的冷热一次风挡板位控制策略，在使出口风速稳定在正常范围内的同时，将温差提升约 4.85%

3.4 全局目标控制

场景描述

循环水泵是火力发电设备中的重要辅机之一，其功耗占主机容量的 4% 左右。由于冷却水流量是影响汽轮机真空度的主要因素，因而循环水泵的控制历来是火电厂中的一个重要问题。然而，循环水泵转速越大需要的耗电越多，但相应的冷凝效果会更好，带来的真空度越高，能够利用气压差进行更多发电。因此，循环水泵的优化



控制难点体现在需全局考虑整锅炉运行流程，寻找最优的耗电-发电平衡点，PID 控制难以实现。

实施过程

第一步：根据影响循环水泵的业务逻辑，在 Polixir REVIVE 中为循环水泵的控制构建决策流程图。

第二步：基于过去一年的循环水泵控制历史数据和决策流程图，构建虚拟循环水泵控制环境。

第三步：根据凝汽器真空度、凝水温度、循环水泵出口母管压力、凝汽器出口温度、机组符合、供热流量、主蒸汽流量和主蒸汽压力，在虚拟环境中学习优化循环水泵的控制策略。

第四步：控制策略线下验证。

实测结果与价值

在 330MW 热电联产机组循环水泵控制中，经过优化后整体指标值降低了约 0.53，相较于 PID 控制策略，节能提升 3.52%。

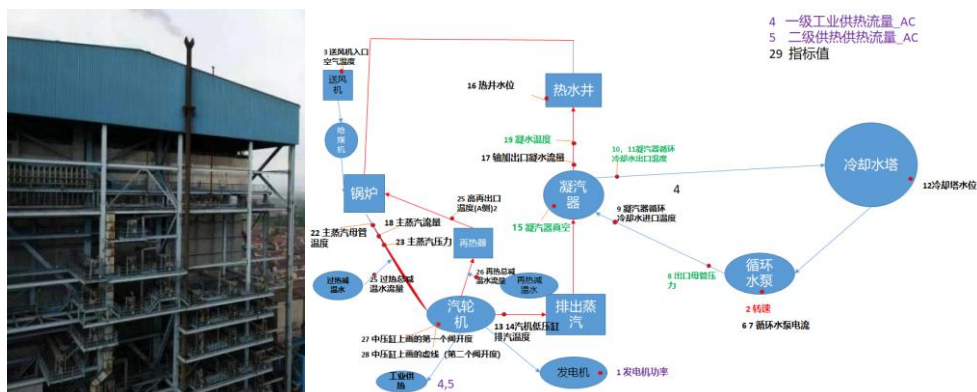


图 18. 锅炉全局业务和循环水泵子循环逻辑分析



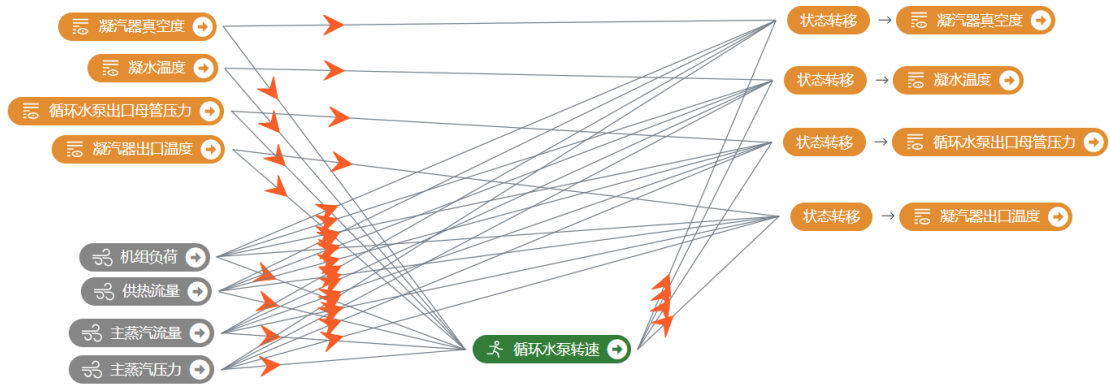


图 19. Polixir REVIVE 中抽象出的循环水泵的控制决策流图

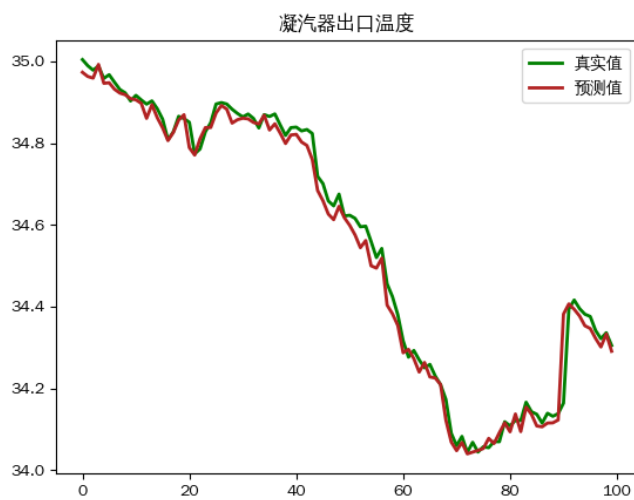


图 20. Polixir REVIVE 中循环水泵控制环境的学习评估，对测试集数据进行回放，观察到测试集中真实凝汽器出口温度值和虚拟环境中的预测值趋势一致，幅度相近

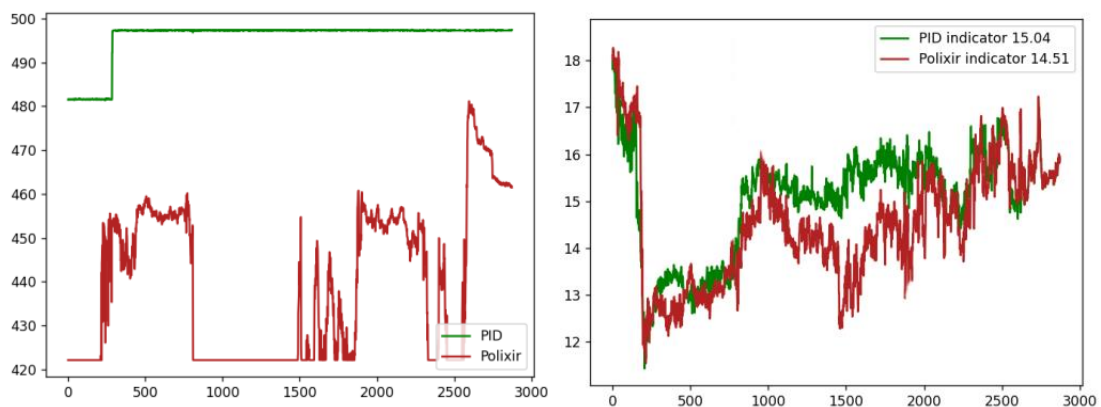


图 21. 左图为循环水泵转速值，绿色为历史数据中经 PID 控制的循环水泵转速值，红色为 Polixir REVIVE 中的控制策略转速值；右图为对应的评价指标，该值越低代表该机组煤耗越低，相较于 PID 控制策略，节能提升 3.52%



关于南栖：

南栖仙策是南京大学人工智能创新研究院技术孵化企业，专注强化学习智能决策领域，拥有深厚的原创研发能力，南栖参与的研究成果发表在 NeurIPS、ICLR 等人工智能国际顶级会议上，获得了 2021 ICAPS 决策类算法竞赛全球冠军，已拥有 30 余项知识产权。

南栖在国际上率先突破了数据驱动模拟技术，解决了近 20 年的复合误差理论障碍、首先将因果学习引入模拟环境建模，实现了“零试错”强化学习，极大地降低了智能决策技术的应用门槛。南栖独创的决策学习方式，区别于市场主流的监督学习和强化学习方式，真正能从数据中回答“怎么做”的问题，从而能迅速在实际业务中落地，并在多种业务场景中得以验证。

南栖仙策创新前沿技术，落地客户价值，致力于在广泛的业务中释放人工智能的决策力量，成为人工智能工业革命的领导者。





官方微信

电话:025-87182030

网址:polixir.ai

商务合作:business@polixir.ai

科研事务:research@polixir.ai

人力招聘:hr@polixir.ai

一般事务:contact@polixir.ai

地址:南京市江宁区总部基地32幢17层