

何星泽

15022649329 | hxz2046084122@outlook.com | GitHub

工作经历

蓝色光标 DSP 广告平台

算法工程师 | 2025.07 - 至今

算法团队核心成员。2025年10月起主导推进 DSP 平台从0到1搭建，独立负责竞价引擎、智能出价、预估模型和 AB 实验四大核心模块的架构设计与开发落地，平台日均处理 百亿级别竞价流量（ADX 侧 千亿/天）。

智能出价与策略优化体系

针对广告主成本控制严格（OCPC 成本达标率仅 68%）、平台预算利用率仅 85%、新计划冷启动成本失控三大核心痛点，设计并落地智能出价策略体系：

- 研发基于 PID 控制的预算分配算法，结合非线性减速策略实现分钟级预算动态调整，解决此前尾量集中爆发导致频繁超预算的问题。
- 设计 K 值自适应出价系数（Smart Bidding），基于后验 CTR 对 CPC/OCPC 场景动态校准出价；引入慢启动机制（Slow Start），保护新计划冷启动阶段的成本安全性。
- 建立基于保序校准的出价兜底逻辑，确保数据稀疏时期的出价合理性。

落地效果：OCPC 成本达标率 68% → 85% (+25%)，预算利用率 85% → 98%，整体 ROI 提升 22%，超预算率 3%

DSP 竞价引擎架构

独立设计投放引擎全链路架构，支撑日均 200 亿 RTB 竞价请求（峰值 QPS 50w+），覆盖流量接入、竞价决策、出价计算到归因分析。

- 基于 Go 构建高性能竞价引擎（Bidmaster Engine），实现多 ADX 协议适配与实时 RTB 请求解析，竞价响应 P99 < 100ms
- 设计模块化策略管线，将定向过滤、频控、出价等决策逻辑解耦为独立插件，支持配置热加载与灰度发布，策略上线周期从天级缩短至小时级。
- 构建基于 Kafka 的实时竞价日志管线，为模型训练、策略迭代和漏斗分析提供数据基础。

DSP-Model 广告预估模型体系

全权负责 CTR/CVR/WINR 三类预估模型从零搭建到持续迭代，直接影响竞价排序质量与广告主成本达标。

- 基于 TensorFlow + Horovod 搭建分布式训练框架，采用 TFRA 动态 Embedding 解决亿级稀疏特征的内存与更新瓶颈，支持千万级样本日级训练。
- 设计多任务学习架构（MTL）：通过 Shared Embedding + Task-Specific FM/DNN Heads，实现激活/付费/注册/唤醒多目标联合建模，改善数据稀疏目标的泛化能力。
- 引入保序校准层（Isotonic Regression），将线上 PCOC 校准至趋近 1.0，保障出价准确性。
- 建立模型全生命周期管理：自动化调度（join/update 两阶段）→ 离线评估（AUC/GAUC/Logloss/PCOC）→ SavedModel 导出 → 线上 A/B → 飞书告警，将模型迭代周期从周级缩短至天级。

落地效果：CVR 模型 AUC 从初始 0.75 持续迭代至 0.80+ (+6.7%)，带动线上转化成本达标率显著提升。

AB 实验平台

针对算法团队此前缺乏并行验证手段、模型与策略验证周期长达 2 周的痛点，设计并落地多层正交 AB 实验系统：

- 设计 CTR/CVR/策略三层正交分流架构（一致性哈希，分流耗时 < 1ms），支持 10+ 实验组并行运行。
- 建立全链路日志埋点（请求→竞价→展示→点击→转化），支持小时级指标自动聚合与可视化漏斗分析。

落地效果：实验周期从 2 周缩短至 3 天，线上问题排查时间从 1 小时缩短至 5 分钟。

快手 - 北京快手科技有限公司

IT中心 后端研发实习生 | 2025.01 - 2025.05

- 负责 KRooms 会议室系统 API 性能优化与认证模块修复，将接口响应时间从 120ms 降至 40ms (-67%)，API 错误率从 2.5% 降至 0.1%。
- 基于 Redis 设计设备健康评分算法，扩展 CPU/内存/网络延迟等多维监控指标。

项目经历

Advoo (AIMO-Tech) | 合作项目

<https://www.advoo.ai/>

2025.11 - 至今

Advoo 是面向中小品牌的 AI 营销 Agent 产品——以 \$9.9 起步价为品牌提供“AI 营销员工”，自主完成策划创意、内容生成、设计微调到跨渠道发布的完整流程，并可主动复盘与策略迭代。负责核心技术基础设施的架构设计与开发（Go）。

- **AgentKit 执行引擎**：设计并实现通用 Agent 运行时（AgentExecutor），支持多步工具编排（ToolExecutor + 串行/中断控制）、流式事件驱动架构（SpanAgentEventEmitter）、会话状态持久化及用量采集，Agent 支持最多 64 轮自主推理循环，具备用户中断恢复能力。

- **虚拟文件系统 & 工具体系**：设计 SessionSandbox 抽象层，Agent 通过标准文件操作访问品牌知识库与技能配置；实现图文编辑、视频生成、Web 搜索、品牌分析、待办管理等 10+ 工具域。

- **可观测性基础设施**：基于 SigNoz + OpenTelemetry 搭建全链路 Tracing/Logging 体系，开发飞书告警卡片适配器（webhook-adapter），支持告警关联 TraceID 快速定位。

- **LLM 服务可用性监测 (relay-pulse)**：从0搭建 LLM 通道健康监测系统（Go + React），支持配置热更新、指数退避探测、多模型父子通道继承、PostgreSQL 持久化。

竞赛与开源

Apache Higress RAG 赛道 | 决赛入围：设计“Pre-Retrieval → 多路混合检索 → CRAG 纠错闭环”全链路增强架构，支持向量 + BM25 + WebSearch 多路检索与 RRF 融合，HotpotQA 数据集 EM=0.94（10,000 样本）。

TAAC 2026 搜广推 CVR 预测竞赛 | 工业赛道：基于 HyFormer 统一架构，通过系统性消融实验将 CVR 预测 AUC 从 baseline 0.8411 提升至 0.8505 (+0.94%，10 轮有效迭代）。

核心方法：EDA 驱动特征工程（时间编码 +0.26%、噪声剔除、Null 指示器）、编码器选型消融（Longer > SwiGLU > Transformer）、训练策略优化（batch_size 隐式正则化 +0.25%、Cosine LR、Label Smoothing）。严格遵循单变量控制原则

开源贡献：Apache Higress（云原生 API 网关）Contributor，贡献 RAG 增强框架实现；Apache seata-go（分布式事务框架）Contributor，修复 insert-on-duplicate 场景空值查询 bug，添加 PR 自动化单元测试 CI。

专业技能

编程语言：精通 Go、Python，熟悉 Java、TypeScript

搜广推算法：CTR/CVR/WINR 预估模型、智能出价（PID / Smart Bidding / OCPC）、预算控制与消耗平滑、AB 实验设计与分析

机器学习工程：TensorFlow 分布式训练（Horovod）、TFRA 动态 Embedding、多任务学习（MTL）、保序校准、模型全生命周期管理

Agent & LLM：Agent 执行引擎设计、Tool-use 编排、RAG 全链路（向量+BM25+CRAG）、LLM 可用性监测

系统架构：高性能 RTB 引擎（Go, 峰值 QPS 50w+）、Kafka 实时管线、Redis 缓存与频控、微服务架构

教育经历

天津财经大学 - 本科 信息与计算科学 2021.09 - 2025.06