

李言迪

(+86) 18401596880 | liyandi007@163.com | yandili | Yandi | yandi.space
北京, 中国

工作经历

- 新浪微博** 北京, 中国 11/2014 - 12/2025
- 算法技术负责人 / 资深算法工程师
 - 历经初级、高级研发, 2019年起担任核心业务技术负责人 (后转为核心技术专注架构演进)。
 - 主导构建微博基础画像体系, 输出通用特征库, 支撑全站多个核心业务线。
 - 负责大图页推荐等场景的系统架构上下游打通与最终业务效果把控。
- Yahoo! 研究院** 巴塞罗那, 西班牙 02/2014 - 07/2014
- 实习研究员
 - 以第一作者身份发表论文: Applying space syntax to online mapping tools (推荐系统顶会 WSDM 2017)。

工作项目

- 微博全站用户画像与特征工程** 资深算法工程师 05/2022 - 12/2025
- 基础画像体系构建:** 主导构建微博通用的基础画像体系 (涵盖人口属性、兴趣标签、行为偏好、博主亲密度等), 为推荐、广告、Feed、超话等多个核心业务线提供底层特征支持, 并全量输出至在线特征库。
 - 深度兴趣建模:** 将多类用户特征与行为日志映射为统一稠密向量空间, 构建支持十万级词表的多目标预测系统。引入 SENet 动态学习特征权重与 Caser 提取局部交叉特征, 结合 RQ-VAE 提升长尾数据泛化能力。
 - 试探策略输出:** 在模型训练中引入“意外度”作为动态目标, 按需打压历史同质化高频行为。输出“试探兴趣”特征, 供下游推荐策略按需进行探索性分发。
 - 博主运营标签:** 利用大语言模型进行少样本分类打标。在 Hive 中完成博文、分组标签等特征的拼接, 并将 Prompt 配置化存储, 支持运营团队无代码修改。配合基础的监控与对账脚本记录 API 失败及特征缺失情况, 最终结果写入 MySQL 供下游系统调用。
 - 技术栈:** Hive, Spark, Fuxi-CTR, PyTorch
- 长短期兴趣计算与商业定向** 技术负责人 05/2019 - 05/2022
- 短期兴趣实时化:** 兴趣计算由离线 Hive 升级为 Flink 流计算, 更新延迟降至 10 分钟内, 接入行为日志增加 3 倍, 实现推荐“刷内更新”。优化时间衰减与平滑策略, 提升兴趣灵敏度与退场效率 (获专利CN115827966A)。推荐流整体: 千曝点击 +5%, 时长 +3.4%
 - 长期兴趣去偏:** 针对马太效应设计“人群归一化”特征提取算法, 通过计算点击强度与大盘均值偏差剥离热门噪音。落地后头部标签基尼系数由 0.58 降至 0.45, 单路千曝点击 +6%, 千曝互动 +2.5%。
 - 商业定向重构:** 广告定向由统计模型升级为监督式算法 (GBDT)。优化后兴趣标签财务消耗占比由 8% 提升至 10% (重点行业达 12%), 核心行业覆盖 UV 提升 6 倍。
 - 技术栈:** Flink, Hivemall, LightGBM, Grafana, Clickhouse
- 微博图片后推荐** 技术负责人 08/2017 - 04/2019
- 架构打通与效果把控:** 负责图片右滑推荐场景的整体统筹。对接客户端团队, 带领算法团队在现有 C++ 引擎中实现基于标签、博主、热门及图片语义向量的多路召回, 并落地 LR/FM 排序模型。
 - 高并发网关开发:** 个人负责开发大图页推荐前端机, 基于 Sanic 构建高性能异步网关, 对接渲染层与后端推荐引擎。实现多页并发缓存及自动降级保护, 支撑 5000+ QPS, 在 100ms 内稳定响应。
 - 技术栈:** asyncio, Sanic, PyTorch, Spark, Faiss, Redis
- 图片特征挖掘与内容理解** 08/2016 - 10/2017
- 质量与风控:** 研发图片分类、美学质量及敏感度识别模型, 为推荐流与大图页提供底层特征支撑。研发**作弊拦截模型**, 打击灰黑产行为。
 - 智能裁剪:** 结合人脸识别和显著性目标检测算法, 为微博信息流图片选取最适合的缩略图裁剪位置。
 - 技术栈:** Tensorflow, Keras, Docker, CNN, Faiss, Nginx, MTCNN

- **长文本分类**: 基于标题、摘要、正文等多维度, 构建 TextCNN, 多模型 Stacking。
- **标题党识别**: 研发并上线基于 Bi-LSTM 的**标题党识别模型**。
- **系统优化**: 构建相似图片排重系统 (基于 phash)。
- **技术栈**: Keras, scikit-learn, Elasticsearch, MySQL

个人项目

Cyber Dingding (多智能体协作系统)

12/2025 - 02/2026

- 独立开发的基于 WebSocket 的多智能体协作系统, 支持人类在环干预。
- **状态与通信**: 基于 WebSocket 构建通信网关, 使用 SQLite 持久化消息, 管理多 Agent 生命周期与任务队列。
- **协同机制**: 构建多角色聊天室模式 (如架构师与程序员), Agent 通过 @ 机制自由交互、分配任务及共享上下文, 同时实现私有上下文隔离。
- **代码执行**: 深度接入 OpenInterpreter, 赋予底层 Agent 真实的计算机操作与代码执行能力。
- **技术点**: WebSocket, SQLite, Agent, OpenInterpreter

KDD Cup(Authorship Disambiguation)

里昂, 法国

03/2013 - 06/2013

- 对微软学术论文 5000 万数据库中 1900 万作者名字进行实体甄别, [\[结题报告\]](#)
- **主要技术**: R, PostgreSQL, Python, LaTeX

教育经历

加泰罗尼亚理工(UPC) & 里昂二大

巴塞罗那, 西班牙 & 里昂, 法国

09/2012 - 09/2014

- 硕士, 数据挖掘与知识管理专业 (DMKM)
 - Erasmus Mundus 奖学金
 - 课程包括: Kernel-based Learning, Statistical NLP, Advanced Statistical Modeling, Bioinformatics

浙江大学

杭州, 中国

09/2008 - 06/2012

- 本科, 数学与应用数学
 - GPA: 3.87/4.0, 学院排名前 5%
 - 荣誉: 浙江大学 ACM 程序设计竞赛三等奖、省级大学生高等数学竞赛一等奖, 多年一等奖学金。

加州大学洛杉矶分校 (UCLA)

洛杉矶, 美国

08/2009 - 09/2009

- 暑期学校, 金融、英文写作

论文研究

[P1] Apply Space Syntax to Online Mapping Tools, *WSDM 2017*

- **Yandi Li**, Nicola Barbieri(Tumblr), Daniele Quercia(Bell Labs)
- 研究生毕设, 基于时间、天气信息, 推荐有趣导航路线的系统。
- 技术点: Factorization Machine, BPR, PostgreSQL, PostGIS, QGIS, A*, Space Syntax

[P2] Chebyshev-Tao方法的不同实现及其比较

- 本科毕设, [\[论文部分\]](#), [\[文献综述\]](#), [\[答辩\]](#)

技能和其他

- **编程语言**: Python, SQL
- **工具框架**: PyTorch, 大模型部署 (Ollama, Huggingface), Flink, Spark, Hive, Docker
- **爱好**: 网球🎾, 乒乓🏓, 太极🧘, 滑雪🎿, 修理🔧