

范叶亮

✉ leo-van@hotmail.com 📞 (+86) 186-3159-3756 🏠 <https://leovan.me>

🎓 教育背景

河北工业大学 硕士 (保送) 信息管理 GPA: 85/100 Top 2% 2012 - 2015

- 二等奖学金 (Top 5%), 三好学生 (Top 5%), 优秀学生干部 (Top 5%)
- 河北省普通高等学校优秀毕业研究生 (Top 2%)
- “创青春”全国大学生创业大赛河北省一等奖 (团队负责人)

河北工业大学 学士 工商管理 GPA: 87/100 Top 2% 2008 - 2012

- 一等奖学金 (Top 2%), 三好学生 (Top 5%), 优秀学生干部 (Top 5%)
- 经济管理学院学生会主席

👤 工作经历

美团 👤 风险数据挖掘专家 2020年08月 - 至今

安全和风险 风险数据挖掘专家 2020年08月 - 至今

京东 👤 高级算法工程师 2015年04月 - 2020年08月

- 京东集团 HIPO 高潜力员工 (2019)
- 京东数字科技 (京东金融) 荣誉讲师 (2017, 2018), 京东集团铜牌认证讲师 (2019)
- 京东集团京芽杯创新大赛“创新种子奖” (2017, 全视之眼项目, 排名 20/378)

智慧农业 算法 & 产品负责人 2019年07月 - 2020年08月

- **智能种植和智能禽类解决方案**: 主导构思和设计了智能种植和智能禽类商业模式和技术解决方案。带领算法和产品团队设计和实现了整个解决方案中的相关数据和算法模型, 完成 SaaS 和 APP 原型设计, **从 0 到 1** 地实现了整体解决方案从 MVP 到真实生产环境的落地。
- **智能环控**: 设计和实现了基于时间序列分析, 深度学习和强化学习的智能环控算法和解决方案。将智能环控算法拆解为环境模型和控制模型两部分, 实现了同一作物和家畜在不同环境下种养殖模型的复用。通过专家知识引擎和机器学习算法实现了在保持现有产量不变的前提下各项环境指标控制误差相比于人工降低了 **50%+**, 整体平均成本 (水, 电, 肥等) 降低 **20%+**。参加 2019 年国际智慧温室种植挑战赛, 在 24 小时 Hackathon 模拟挑战赛中, 获得人工智能策略方法 **4/21** 名, 虚拟西红柿种植净利润 **9/21** 名。
- **智能集蛋**: 设计和实现了一套基于计算机视觉和 IoT 传感器的智能集蛋装置和算法。在笼养模式集蛋过程中, 利用摄像头和 IoT 传感器采集的数据实现了鸡蛋计数和隶属笼体的识别, 准确率 **99%+**。通过鸡蛋隶属笼体识别可以精准分析不同笼体内的料蛋比, 为淘汰鸡提供强有力的数据支持, 同时也为蛋品的溯源提供了更细力度的数据。

Daat (复杂网络和知识图谱) 项目负责人 2018年04月 - 2019年06月

- **数据知识工程和知识问答系统**: 设计和构建数据仓库, 数据集市和数据工具领域的知识本体, 并据此进行知识抽取和知识库构建。设计和实现了数据知识问答系统, 包括: 用户问题意图识别, 槽填充, 查询改写, 结果排序以及基于深度语义模型的问题匹配。问答系统主要用于辅助数据使用者对于仓库和集市数据的查询使用, 以及相关概念、流程、工具等方面问题的解答。系统服务内部 **3000+** 用户, 数据相关问题咨询人工解答量降低 **50%+**。

- **自动化敏感信息识别**: 用于数据仓库在数据接入过程中的表的敏感信息识别, 进一步辅助数据加密策略实施。根据数据的元信息 (例如: 表名, 表注释, 字段名, 字段注释等) 和值信息 (即字段存储的数据值), 利用 Wide & Deep 网络构建识别模型。提取传统特征构建 Wide 网络, 针对文本特征, 利用 Char Embedding + CNN 构建 Deep 网络, 模型测试数据的 F1-Score 为 **0.95+**。
- **大规模异构网络嵌入**: 实现针对**千万级别顶点**和**亿级别边**的异构网络嵌入算法。通过 Alias 方法进行高效地加权节点采样, 实现了基于具有丰富业务含义元路径 (Meta-Path) 和其他算法的网络节点嵌入。利用异构网络节点嵌入结果作为节点特征 (例如: 用户, 商品等) 为风控, 营销和推荐等业务模型和系统提供服务。
- **基于用户网络 and 用户行为的推荐和营销**: 通过用户的历史购买记录构建包含用户, 商品 (实物, 理财等) 等多类型节点和边的异构网络, 并利用异构网络嵌入算法构建不同类型节点的向量表示。结合用户传统特征和用户在线行为, 构建推荐候选结果生成和营销人群生成模型, 结合最终业务模型, 平均辅助提升效果 **20%+**。

全视之眼 (中文地址分析项目) 项目负责人

2015 年 04 月 - 2018 年 04 月

- **地址分析算法引擎**: 设计和开发中文地址分词, 分类, 完整度, POI 识别, 相似度等基础算法和服务, 算法平均准确率 **90%+**。
- **地址画像系统**: 依托地址分析算法引擎, 设计地址 POI 画像相关指标体系, 存储和服务框架。对内, 服务于线下支付等场景下的营销和运营, 平均提升用户转化率 **30%+**; 对外, 接入京东征信等多个平台, 提供精确查询, 模糊查询和经纬度范围查询等多种查询方式, 实现金融科技的对外输出。
- **授信和反欺诈模型**: 依据地址画像构建授信策略, 辅筛选和授信用户超过 **1000 万人**。依托地址分析算法和地址画像, 构建中介套现识别等反欺诈模型, 识别套现订单金额约 **20 万元/日**。
- **企业地址画像系统**: 构建企业、地址和用户关系, 设计和开发企业地址画像相关指标体系, 服务于京东企业信用 (<https://icredit.jd.com>), 对内外部提供离线数据和实时查询服务。
- **农村金融小站选址模型**: 根据地址画像系统, 结合农村金融相关业务, 构建基于地址画像的农村金融小站选址模型和系统, 为农村金融小站的线下选址提供辅助决策。

用户行为分析项目 算法设计和开发

2017 年 10 月 - 2017 年 12 月

- 设计开发用户在线行为表示方法, 基于层次聚类和深度搜索, 构建了一个用于识别用户异常行为的混合模型。在校园白条中介激活和账号盗用问题上, 相比传统的 Bag of Words 和 N-GRAM 等表示方法, 在保证准确率 **80%+** 的前提下, 提高异常用户识别数量 **3+ 倍**。

动产融资项目 算法设计和开发

2015 年 04 月 - 2015 年 10 月

- 设计开发一种基于 Bass 扩散模型, 优化的时间序列相似度和聚类算法的产品生命周期识别模型, 尾货识别准确率 **95%+**, 辅助贷款方商品质押准入决策和质押率的制定。
- 设计开发不同数据源的商品信息融合模型, 相同商品识别准确率 **90%+**, 配合 Elasticsearch 开发商品模糊匹配查询服务, 为商品质押业务提供精准的价格等相关信息。

🔧 专业技能

- 开发语言: Python: ★★★★★☆ R: ★★★★★☆ SQL: ★★★★★☆ JavaScript/TypeScript: ★★★★★☆
- 平台框架: Qt: ★★★★★☆ PyTorch: ★★★★★☆ Tensorflow: ★★★★★☆ React: ★★★★★☆
- 工具: Axure: ★★★★★☆ Sketch: ★★★★★☆ Omnigraffle: ★★★★★☆
- 语言能力: CET-6: 518, 熟练运用英语进行口语交流, 阅读资料, 撰写文档

📄 论文

- Zhou, F., Yin, H., Zhan, L., Li, H., **Fan, Y.**, & Jiang, L. (2018) . A Novel Ensemble Strategy Combining Gradient Boosted Decision Trees and Factorization Machine Based Neural Network for Clicks Prediction.

In *2018 International Conference on Big Data and Artificial Intelligence (BDAI)* (pp. 29-33) . IEEE.

- Li, J., **Fan, Y.***, Xu, Y., & Feng, H. (2013) . An Improved Forecasting Algorithm for Spare Parts of Short Life Cycle Products Based on EMD-SVM. In *Information Science and Cloud Computing Companion (ISCC-C)* , *2013 International Conference on* (pp. 722-727) . IEEE.
- **Fan, Y.**, Li, J., Chu, C. (2014) . IEAF: A Hybrid Method for Forecasting Short Life Cycle Spare Parts. *Unpublished*.

📄 专利

-
- 一种中文地址分词方法及系统 (发明, CN 105159949, **授权**, 第一发明人) 2015
 - 一种产品库存预测方法及装置 (发明, CN 106056239, **授权**, 第一发明人) 2016
 - 数据仓库信息处理方法、装置、系统、介质 (发明, CN 109388637, **授权**, 第一发明人) 2018
 - 一种数据处理方法、装置、设备及介质 (发明, CN 110309235, **授权**, 第一发明人) 2019
 - 用于生成信息的方法和装置 (发明, CN 112395490, **实审**, 第一发明人) 2019

📁 开源项目

-
- Github: <https://github.com/leovan> 👤 Followers: 208
 - Python 数据科学导论: 一份以 Python 为基础的数据科学入门教程。🌟 Stars: 57
 - R 语言数据科学导论: 一份以 R 语言为基础的数据科学入门教程。🌟 Stars: 98
 - Sci-Hub EVA: 一个跨平台的 Sci-Hub 界面化应用。🌟 Stars: 878
 - Hive 函数: 实用自定义 Hive 函数集。🌟 Stars: 3
 - Cytoscape 手册: Cytoscape 中文手册。🌟 Stars: 11