



全球

World Of Tech 2017

2017年12月1日-2日 • 深圳中洲万豪酒店

软件开发技术峰会

DEVELOPMENT



赋能提效，防微杜渐

基于数据和机器学习的经纪人管理

冯扬

链家网 - 数据策略部 | 高级技术总监

自我介绍



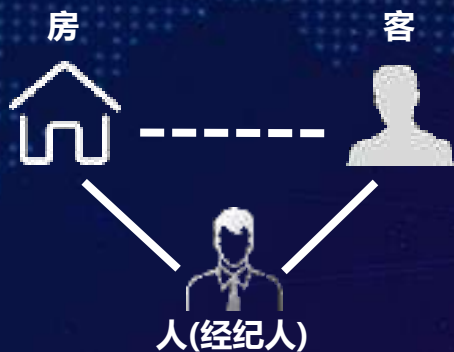
冯扬 (Young)

- 2006/09~2010/12: 北京理工大学攻读博士学位
- 2011/03~2013/05: 新浪微博-搜索-推荐, 推荐算法设计以及推荐系统研发
- 2013/05~2014/09: 腾讯SOSO & 搜狗-社区搜索部, 问答社区推荐系统研发
- 2014/09~2017/07: 新浪微博-推荐产品部, 微博推荐平台设计及研发, 视频推荐
- 2017/07~: 链家网-产品技术中心-数据策略部, 高级技术总监

内容提要

- ◆ 01 背景、问题、挑战
- ◆ 02 解决方案：数据 + 机器学习
- ◆ 03 应用案例：提升效率，防范风险

房产交易领域中“人”的因素



- 通过“人”连接“房”与“客”
- 经纪人的质量决定了服务的质量
- 经纪人管理的成本是企业成本中的大头

15万+ 经纪人分布于32个城市



房产经纪人管理中的挑战



- 数量多、分布广
- 个体、地域、类型差异大
- 培养成本高，流失率高
- 利益巨大，违规风险高



- 认：客观性评估和刻画
- 帮：提升经纪人工作、成长效率
- 防：风险预测及防范



数据的问题

- 经纪人信息、行为的采集
- 数据规范化 & 再组织



挖掘的问题

- 发掘最大模式
- 寻找显著特性



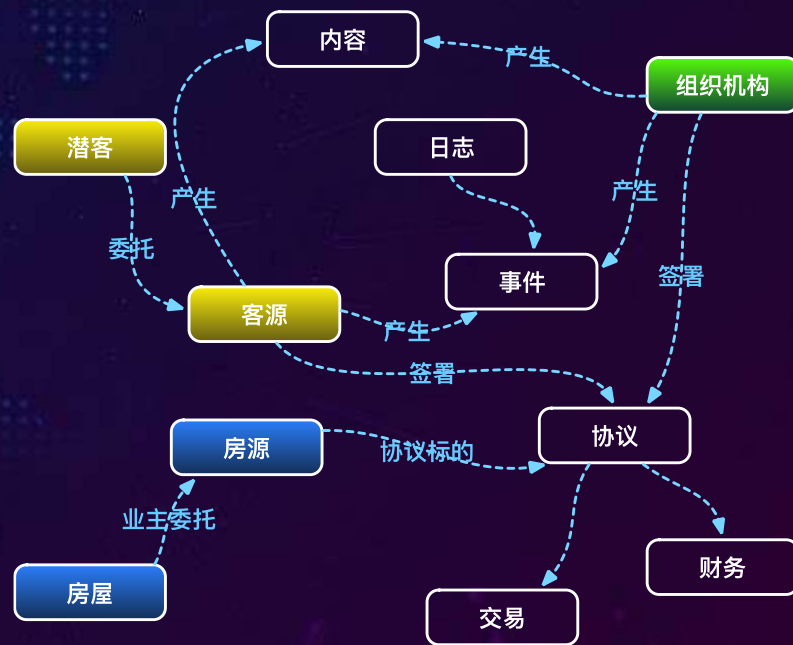
预测的问题

- 对模式进行判定
- 对风险进行评估

经纪人评价及管理平台技术架构



解决数据的问题：数据建设



链家大数据平台



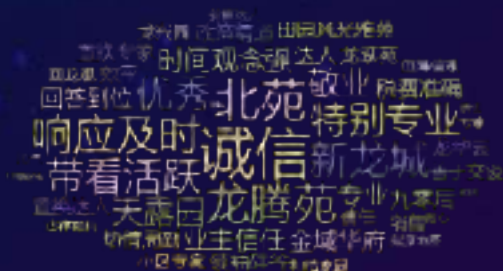
解决挖掘的问题：经纪人画像



基本信息 & 业绩统计

性别：女
 出生日期：1990-11-09
 岗位：综合经纪人
 级别：M10
 运营区域：京北区
 房源实勘业绩：10589
 房源速销业绩：27065
 客源成交业绩：753271

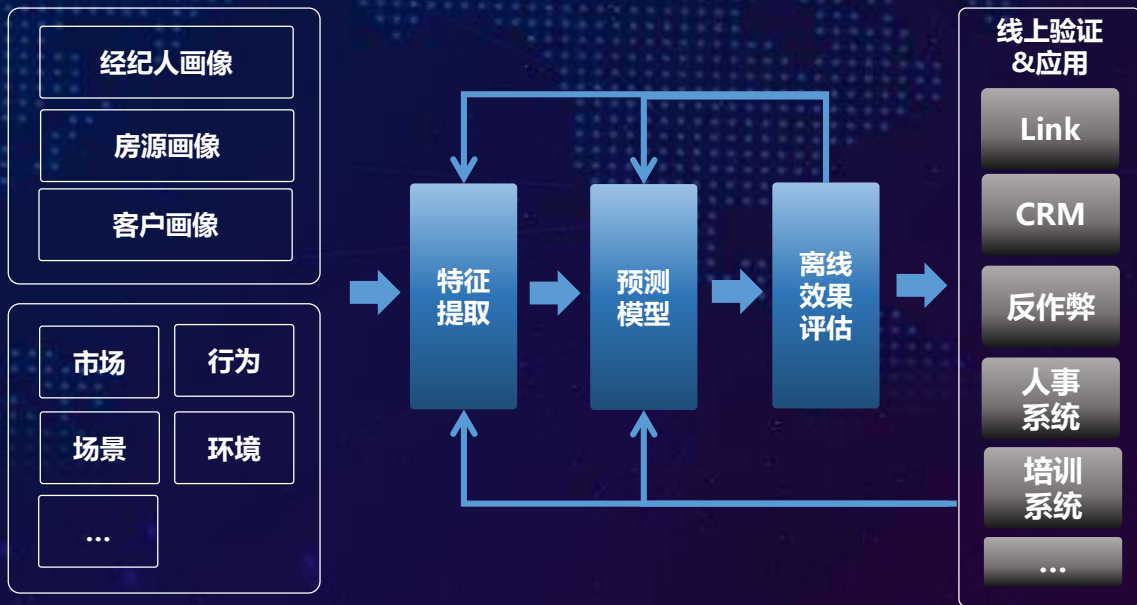
标签信息



从个体刻画到群体刻画



解决预测的问题：机器学习



应用方向：

- 状态判定、成长预测
- 赋能提效
- 违规风险预测

状态判定实例：经纪人离职风险预测

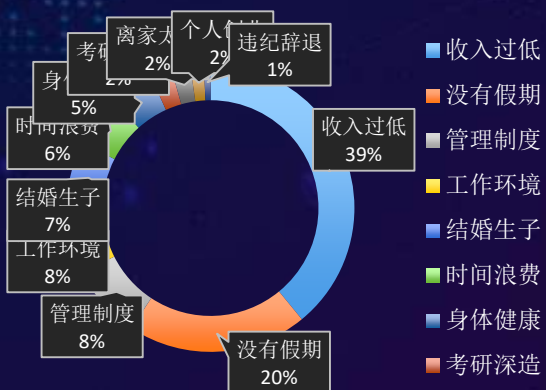
北京离职人数



特征

个体	画像特征	行为特征	业绩特征	展位特征
群体	门店特征	区域特征	转正期特征	试用期特征
市场	市场特征			

离职原因调研

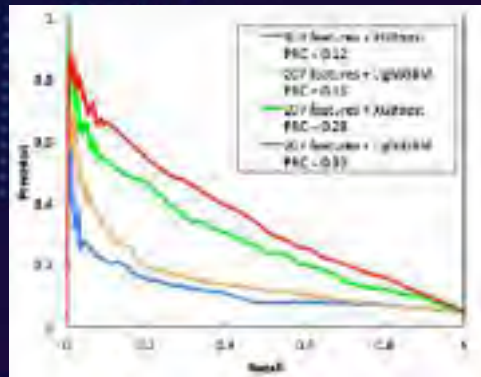
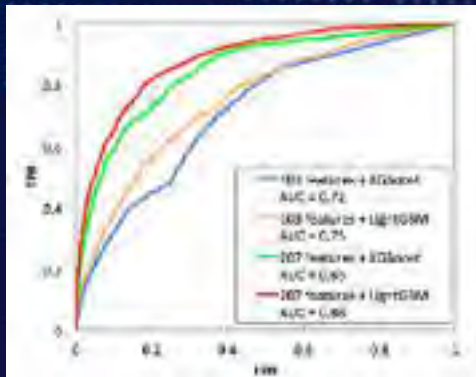


离职风险 => 未来60天离职判定



经纪人离职风险预测效果

离线评估



实际验证

北京地区评估（北京8万经纪人）

被识别出来的经纪人三分之二以上会在60天内提出离职（准确率 $\approx 69\%$ ）

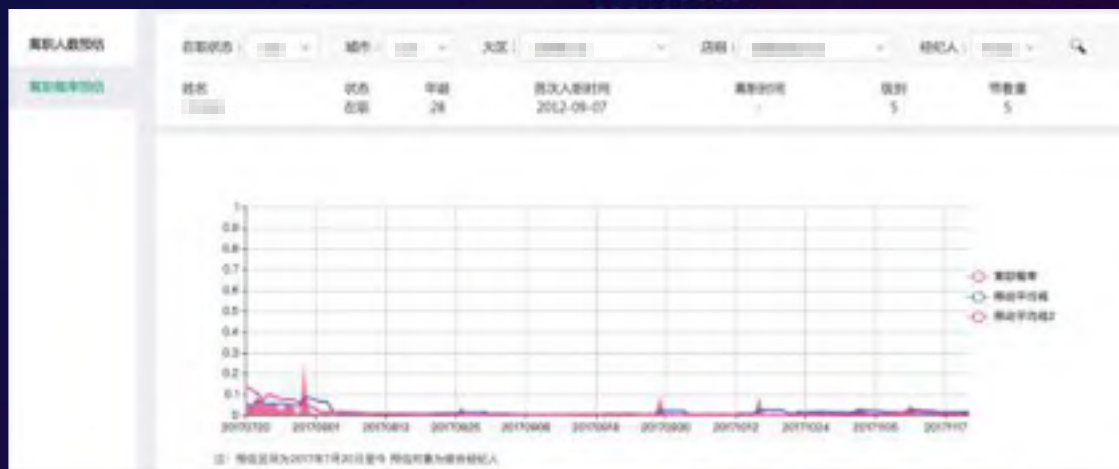
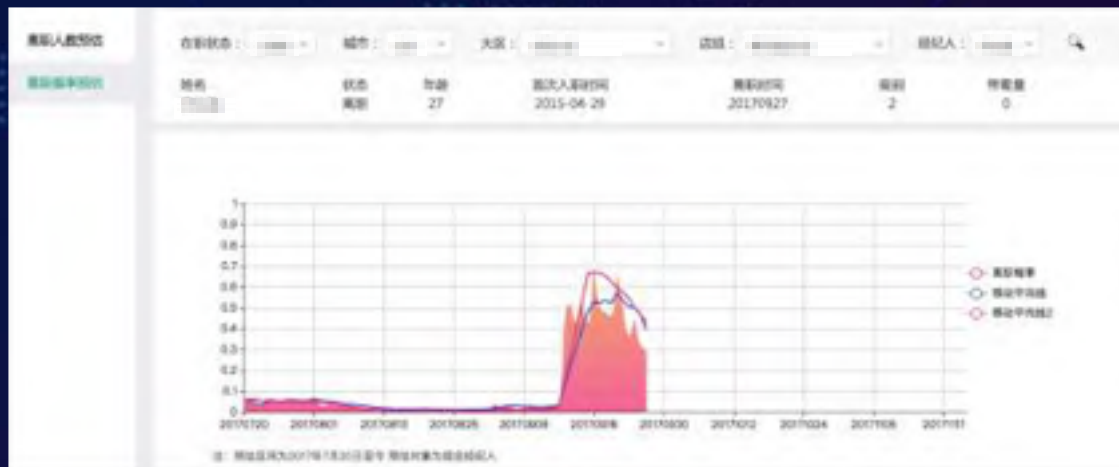
有五分之一以上的离职经纪人在预测列表中（召回率 $> 20\%$ ）

结果应用

对大区、门店生成离职预估报告，辅助线下管理，降低离职率
针对城市分析离职原因，改善经纪人招聘、培养环节潜在问题



经纪人离职风险预测效果



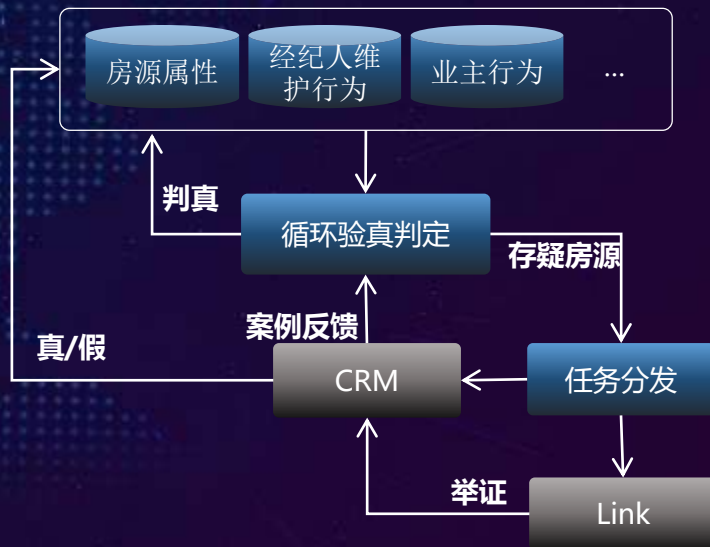


附能提效案例：房源验真辅助



纯人工回访的问题

- 监管覆盖 **10%** 房源
- 300套/人/天，无效回访 **35%**
- 超过 **87%** 的假房源举报来自监管未覆盖房源



假房源识别 => 房源存疑风险排序

对Top N%的存疑房源进行举证及回访，提升人效

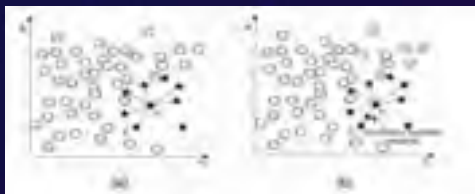
存疑房源识别关键问题解决

- 样本分布极度不平衡，正（假房源）负样本比例 1 : 22
- 排序问题中 Top N% 的选择

案例反馈



数据合成



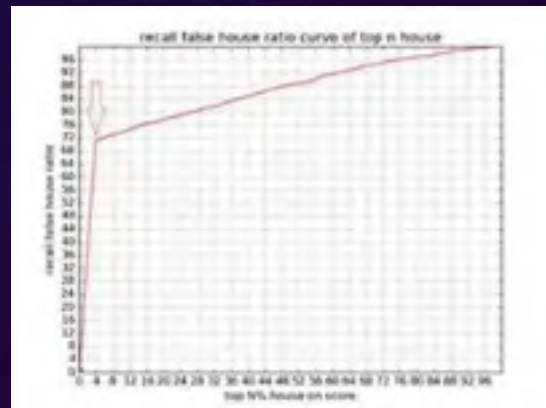
$$x_{new} = x + rand(0,1) \times (\bar{x} - x)$$

最优化目标函数时
考虑样本权重

$$grad_i^* = grad_i \times w_i$$

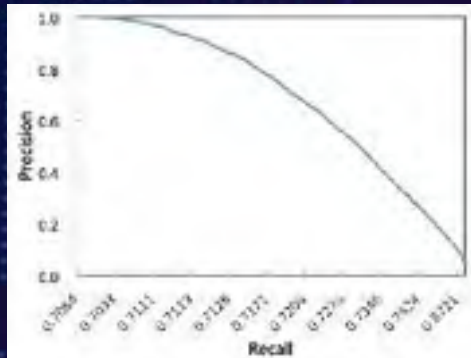
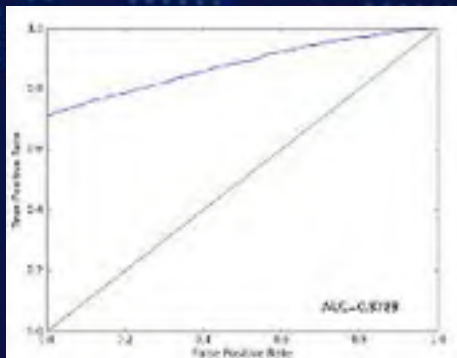
$$hess_i^* = hess_i \times w_i$$

Top N% , N=4



房源验真辅助效果

离线评估
(存疑房源识别)



结果应用

- 监管覆盖**100%**房源
- 节省人力超过**90%**
- 房源回访效率提升**4倍**



违规风险预测案例：房源泄露行为预测

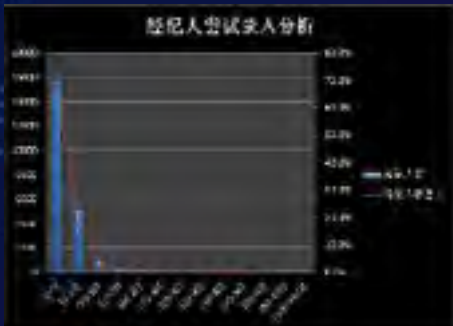
难点：

- 特征不明确
- 结果需要强解释性

房源信息试探调研



特征



经纪人属性特征

业绩特征

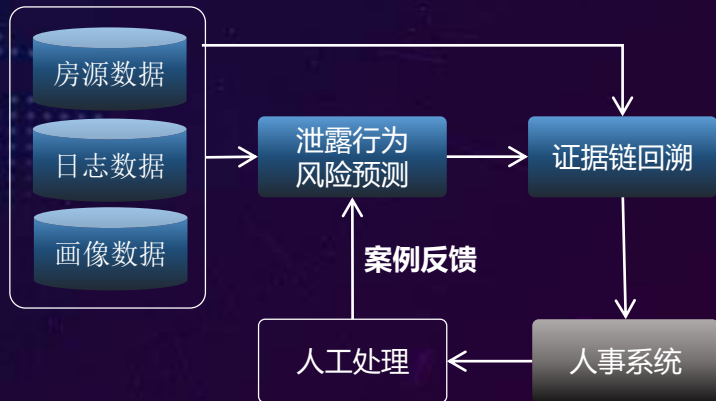
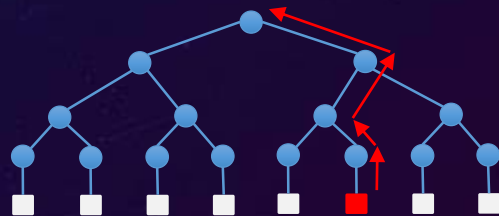
房源特征

行为特征

设备特征

环境特征

.....



对未来的思考

- 01 特征学习 vs. 特征工程？
- 02 迁移学习的可能性和成本？
- 03 统一的评价体系



Thank you!