

商业智能技术驱动业务增长

张重阳

微信小程序商业技术负责人



极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

人工智能基础课

“通俗易懂的人工智能入门课”

王天一
博士 副教授



扫一扫，免费试读

AI技术内参

你的360度人工智能信息助理

洪亮劼
EasyV数据科学主管



扫一扫，免费试读



关注落地技术，探寻AI应用场景

- 14万AI领域垂直用户
- 8000+社群技术交流人员，不乏行业内顶级技术专家
- 每周一节干货技术分享课
- AI一线领军人物的访谈
- AI大会的专家干货演讲整理
- 《AI前线》月刊
- AI技能图谱
- 线下沙龙



扫码关注带你涨姿势

QCon

全球软件开发大会

成为软件技术专家的 必经之路

[北京站] 2018

会议：2018年4月20-22日 / 培训：2018年4月18-19日
北京·国际会议中心

8折 购票中, 每张立减1360元
团购享受更多优惠



识别二维码了解更多

ArchSummit

全球架构师峰会

2018 · 深圳站

从2012年开始算起，InfoQ已经举办了9场ArchSummit全球架构师峰会，有来自Microsoft、Google、Facebook、Twitter、LinkedIn、阿里巴巴、腾讯、百度等技术专家分享过他们的实践经验，至今累计已经为中国技术人奉上了近千场精彩演讲。

限时7折报名中，名额有限，速速报名吧！

● 2012.08.10-12 深圳站

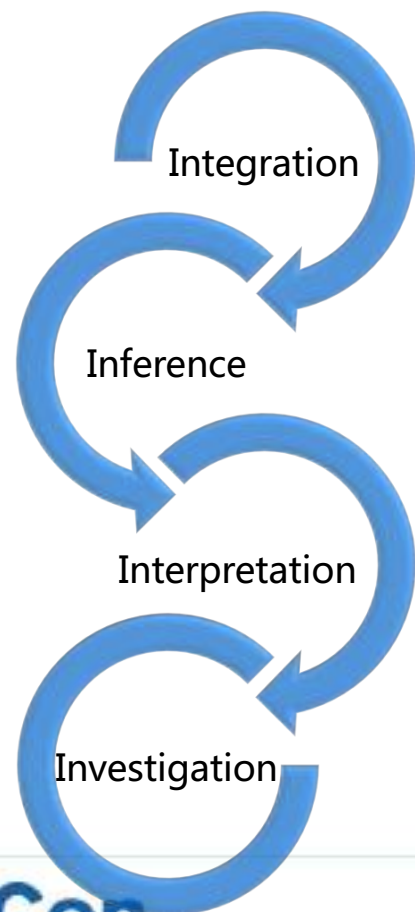
● 2018.07.06-09 深圳站

会议：07.06-07.07

培训：07.08-07.09



智能技术在商业过程中应用的4个环节



- 整合: 针对问题收集并整合数据
 - 如何在计算机中表示数据便于存储和计算
 - 如何处理保密数据, 如何在处理用户数据时保护用户隐私
- 推断: 使用统计和机器学习的方法求解问题的最优解
 - 如何结合多个模型优点
- 解释: 结合数据对模型的推断结果进行分析和解释
 - 如何对黑盒模型的结果进行解释
- 调查: 利用人工知识制定策略并使用自动化的方法验证效果
 - 如何在实际环境下验证效果

微信商业智能系统

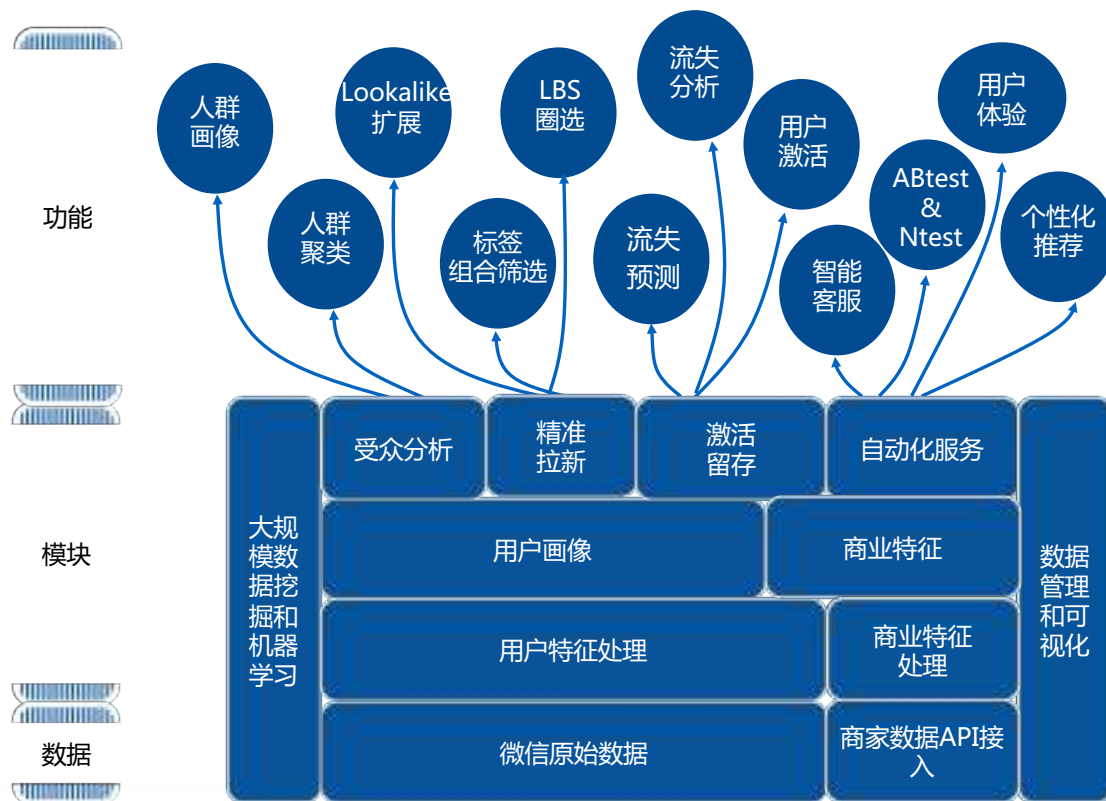


TABLE OF CONTENTES

整合 Integration

- 实例: 用户画像

推断

解释

调查

总结

用户画像

用户画像



存储结构

	特征a	特征b	特征c	特征d	...
用户1					
用户2					
...					
用户n					

距离度量

测量用户之间的相似度

应用场景:

- 用户分类, 聚类, 对相似用户推荐相同商品

用户A: x 用户B: y

相似度函数: $F(x, y)$

- 实际使用中常根据不同的应用场景定义不同的相似度函数或是用机器学习的方法在数据上拟合这个函数

多源异构



可标签化

80后 已购房 健身
有车 招行 喜欢外卖
白领 男性 网球 游泳
单身 IT行业
炒股 居住中档小区
北京海淀



异构信息

社交网络



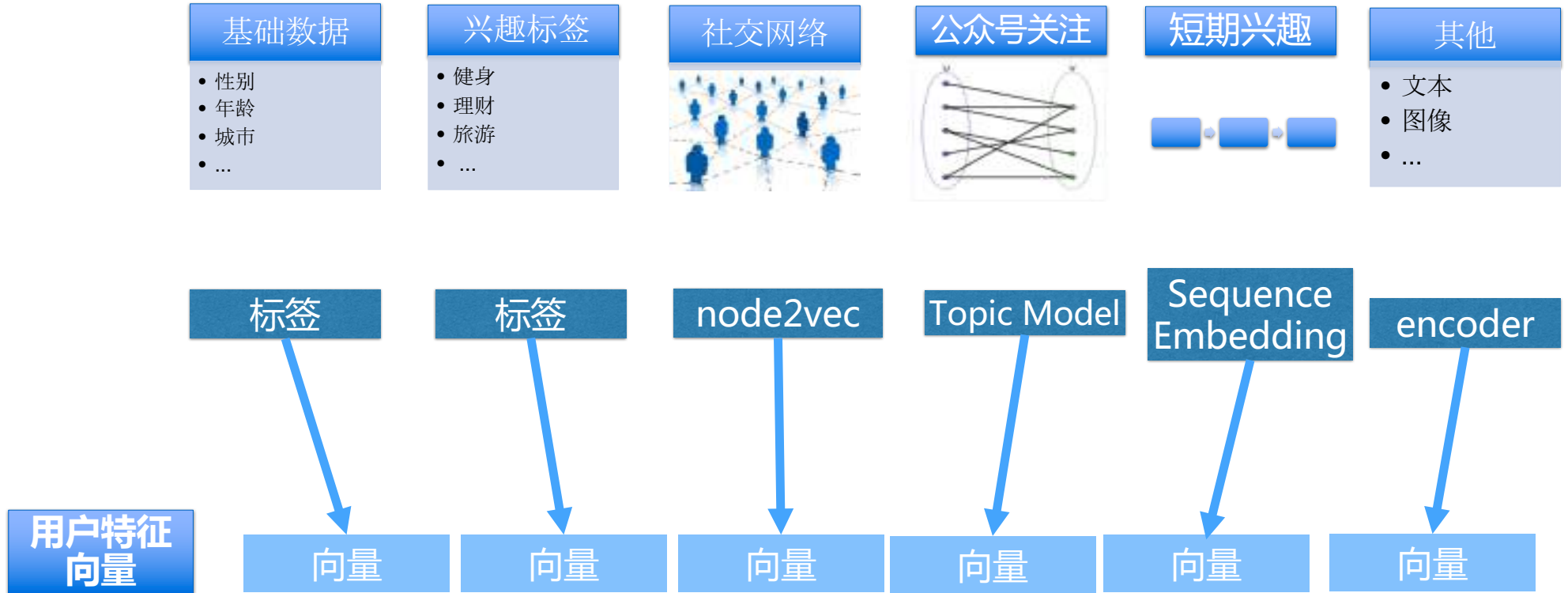
公众号关注



短期兴趣



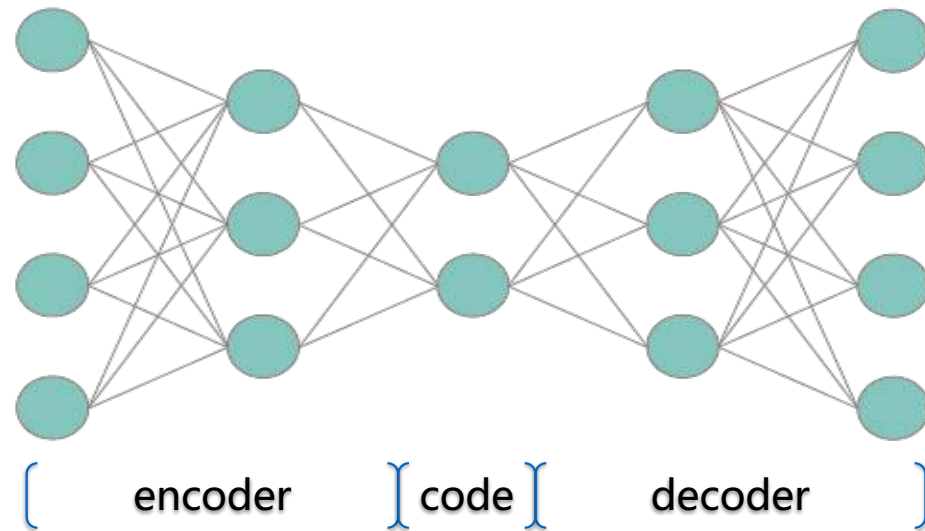
数据编码



降维

Auto-encoder

- 降维可以使实际应用的机器学习任务支持更小的样本输入
- 降维后的特征向量对用户之间的距离度量影响不大，但相当于原始特征能更好的保护用户隐私



用户隐私保护遵守的原则

1. 分析一群人而不分析一个人。
2. 不使用个人可辨识信息 (**Personal Identifiable Information**)，如：姓名，身份证号，手机号等。（我们数据处理时使用无任何物理含义的**User ID**作为各个数据中的统一标识）
3. 通讯和聊天内容神圣不可侵犯，不保存和使用任何通讯和聊天内容。
4. 控制精度，这里的精度并不是指准确度，比如我们在分析用户住址时，只定位到小区，而不再做楼栋和楼层的定位。
5. 只保留和使用一年以内的数据。
6. 所有的标签都由算法自动化生成而不使用人工标注，工程师只负责设计算法。

TABLE OF CONTENTES

整合

推断 Inference

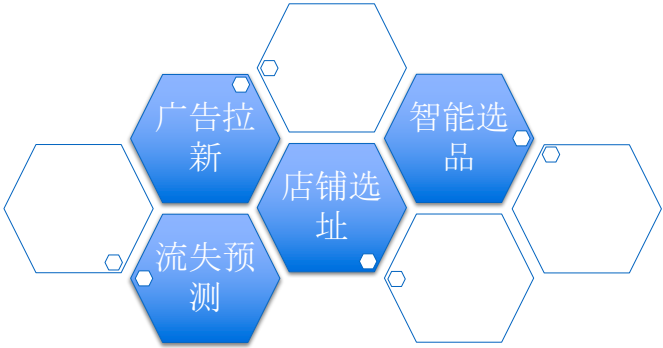
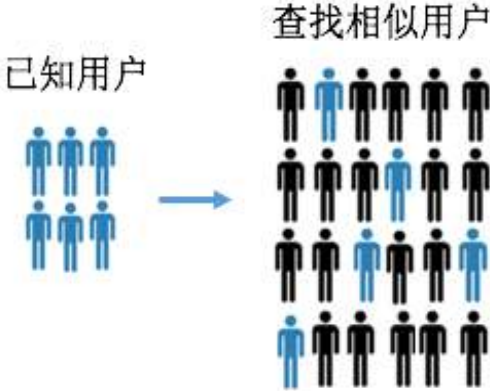
- 实例: Lookalike人群定向与流失预测

解释

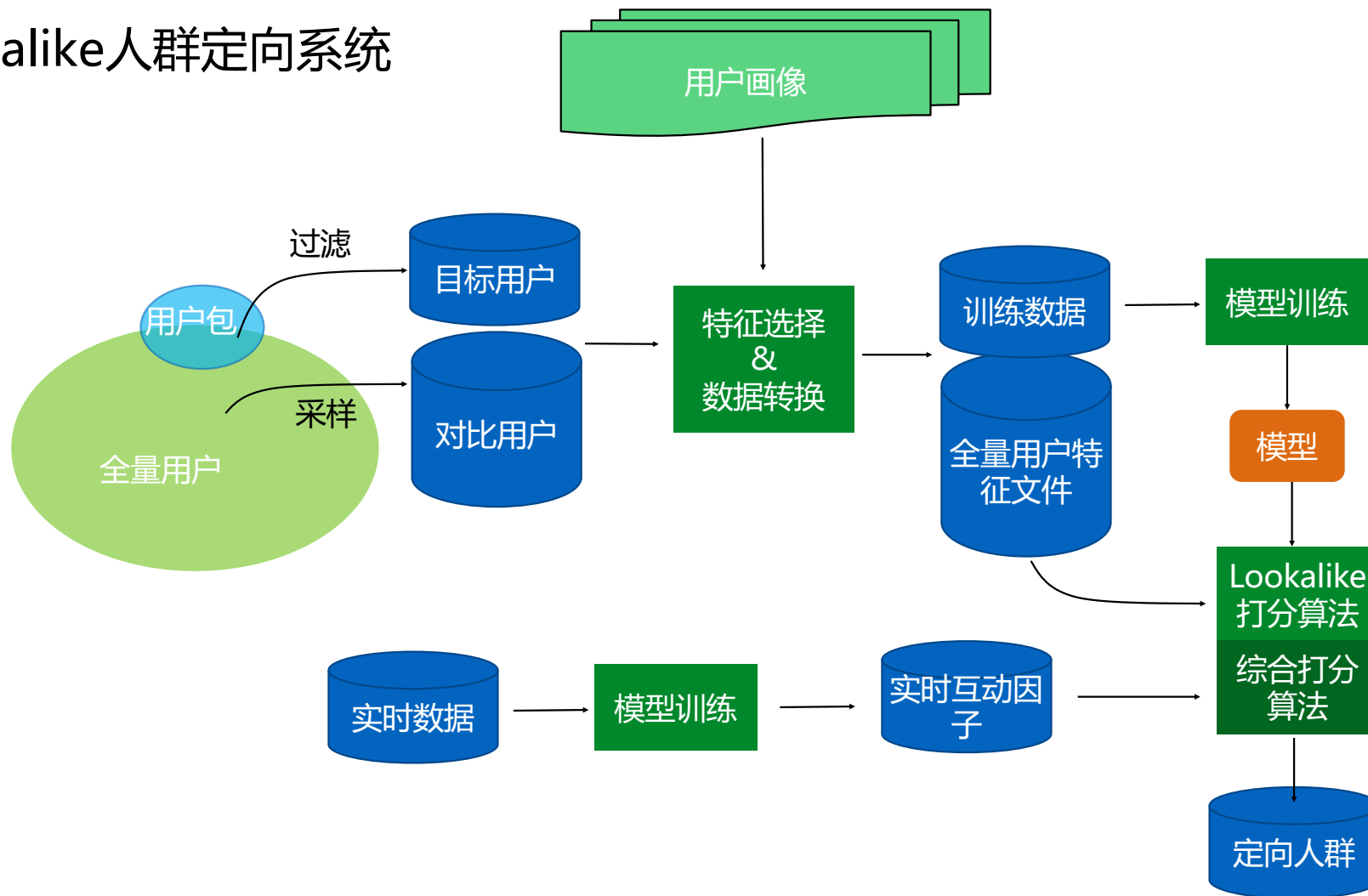
调查

总结

Lookalike潜在应用场景



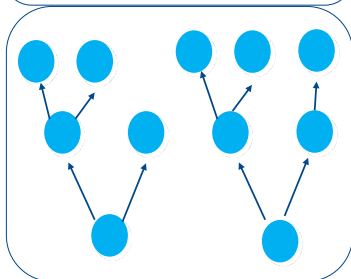
Lookalike人群定向系统



模型整合 (JointTrain vs Ensemble)

集成学习

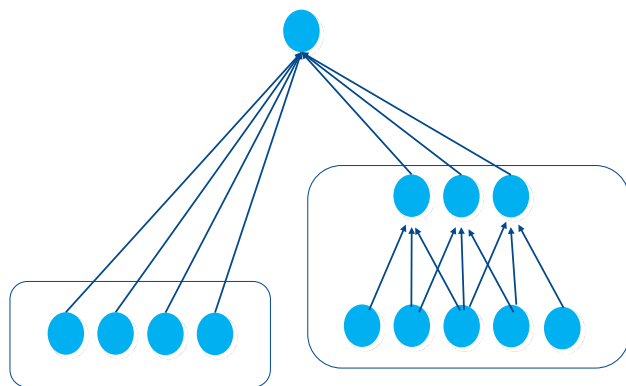
LR/FM/DNN



GBDT2LR
GBDT2FM
GBDT2DNN

•通用方法

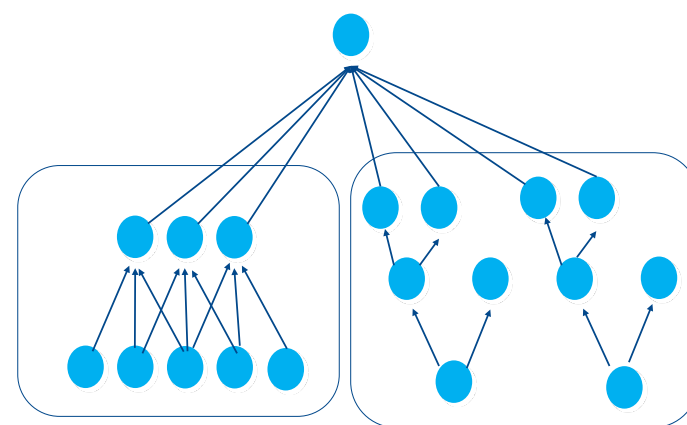
联合训练



Wide & Deep

•可以结合人工定义的特征

集成学习



DNN + GBDT

我们系统实际采用的模型

➢ 不需要加入人工特征，且实际效果更好

应用实例：流失预测与流失分析

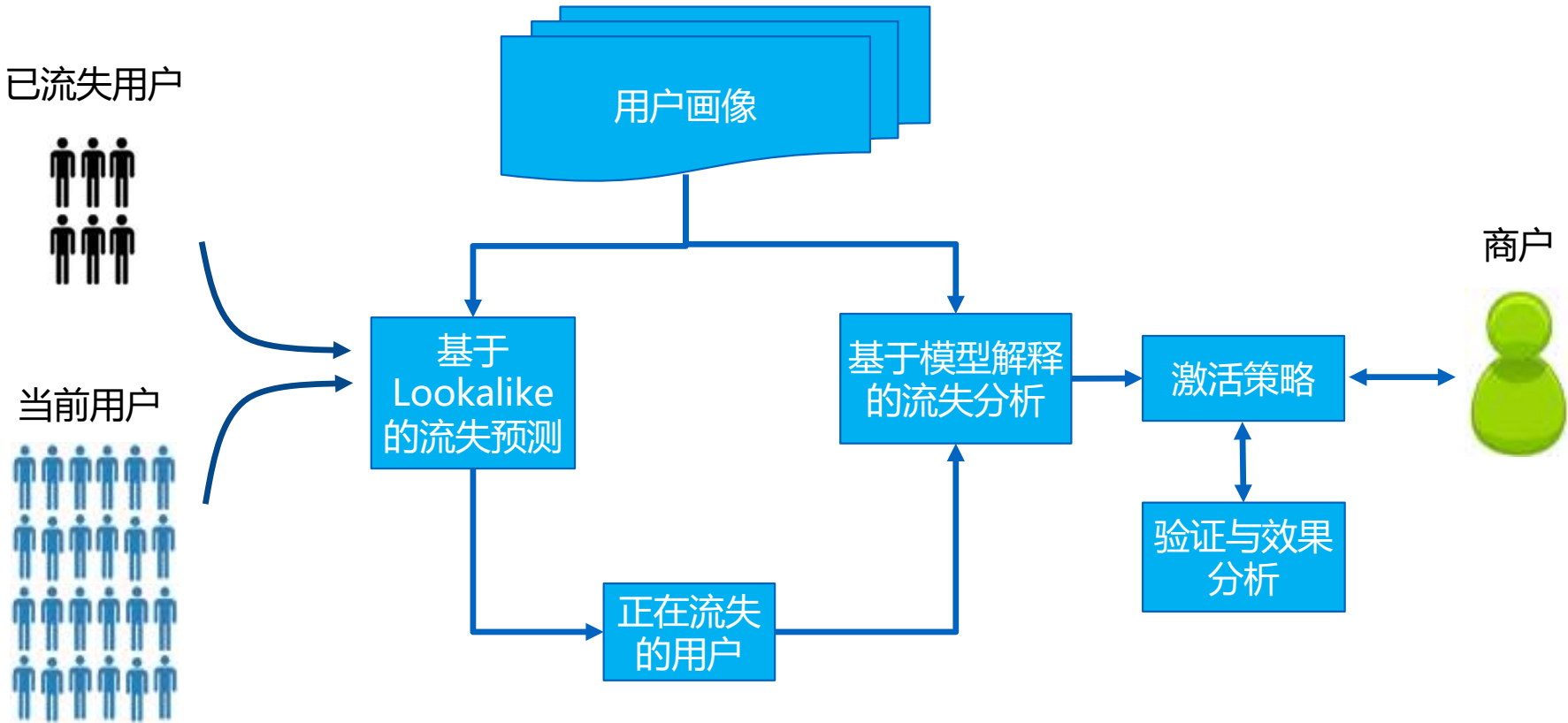


TABLE OF CONTENTES

整合

推断

解释 Interpretation

- 实例: 流失分析

调查

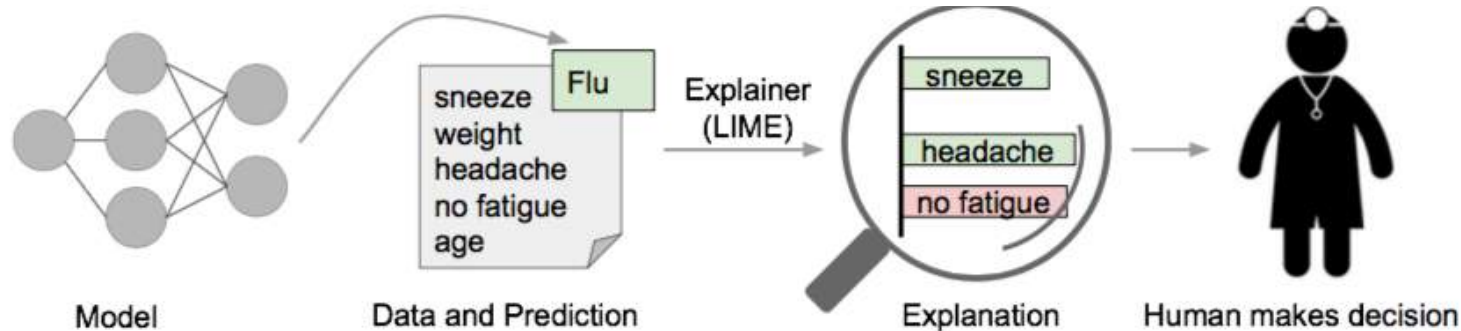
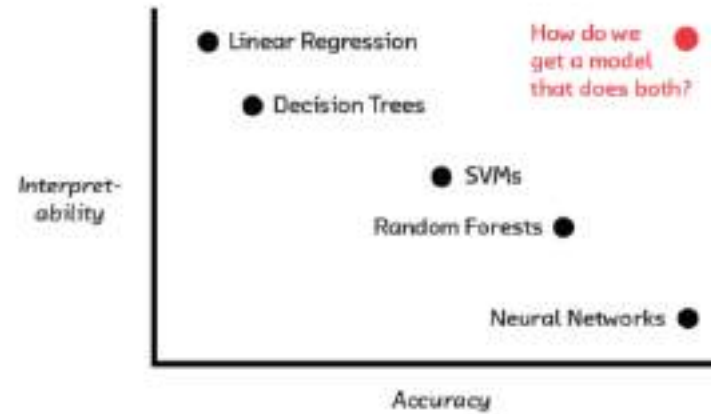
总结

模型解释

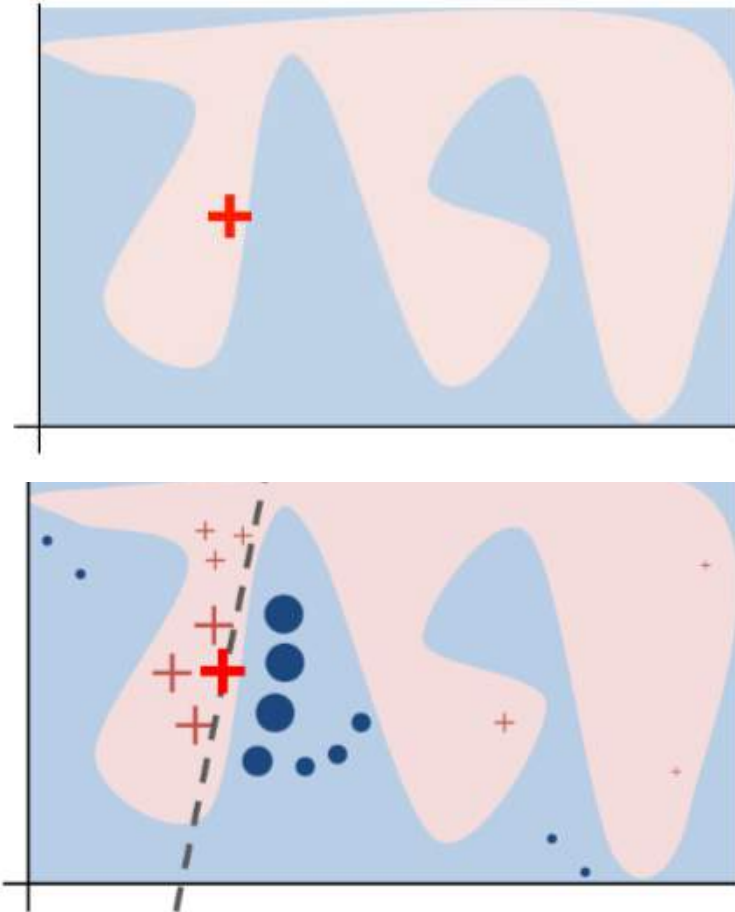
模型的可解释性和精度同等重要，然而模型的能力越强就越复杂越不易解释

对模型的有效解释可以：

- 发现训练数据和模型中的问题
- 更容易说服模型输出结果的反馈对象
- 从解释结果中获取更多信息从而制定策略解决问题



模型无关的局部解释算法LIME



KDD 2016 : “Why Should I Trust You?”
Explaining the Predictions of Any Classifier

Algorithm 1 Sparse Linear Explanations using LIME

Require: Classifier f , Number of samples N

Require: Instance x , and its interpretable version x'

Require: Similarity kernel π_x , Length of explanation K

$\mathcal{Z} \leftarrow \{\}$

for $i \in \{1, 2, 3, \dots, N\}$ **do**

$z'_i \leftarrow \text{sample_around}(x')$

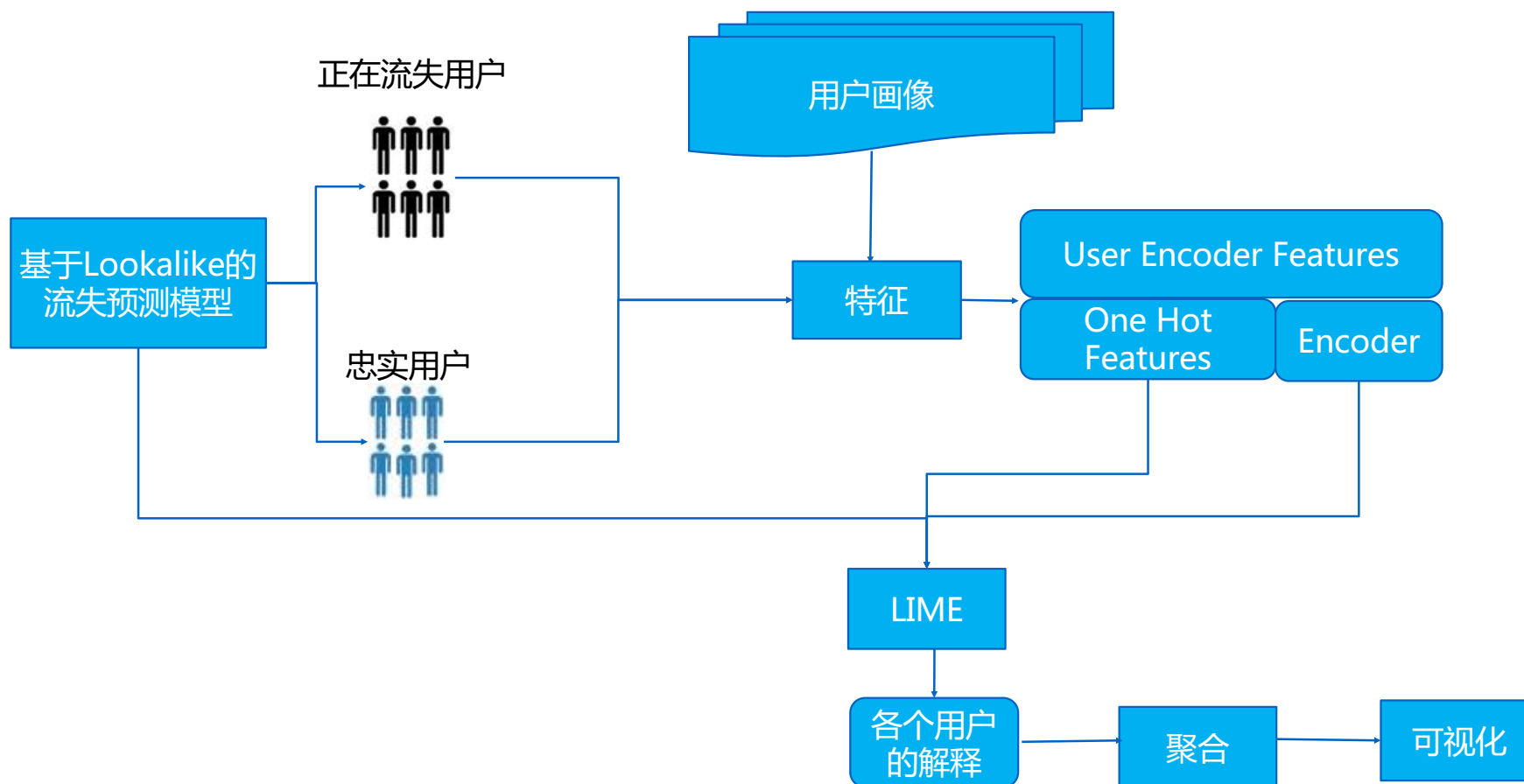
$\mathcal{Z} \leftarrow \mathcal{Z} \cup \langle z'_i, f(z_i), \pi_x(z_i) \rangle$

end for

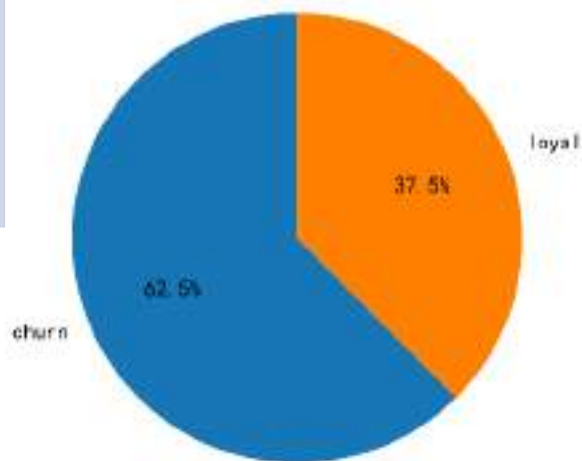
$w \leftarrow \text{K-Lasso}(\mathcal{Z}, K)$ \triangleright with z'_i as features, $f(z)$ as target

return w

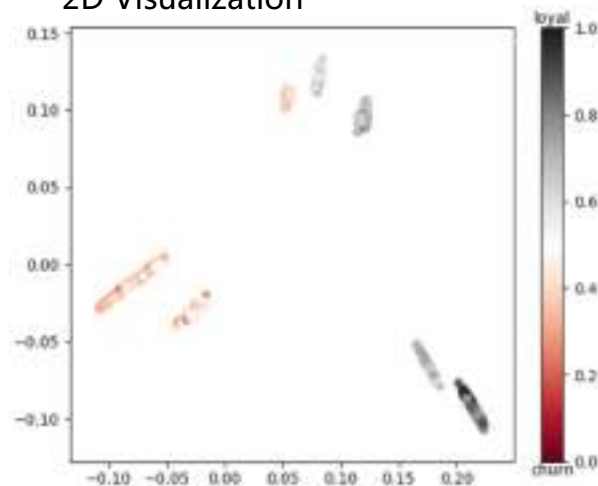
应用实例：用户流失分析



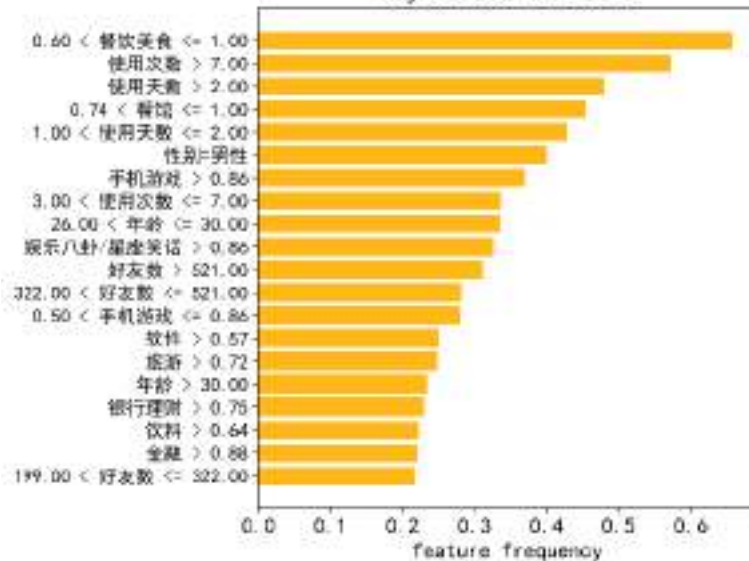
某小程序用户流失
分析结果
(流失预测AUC为
0.8)



2D Visualization



loyal user features



churn user features

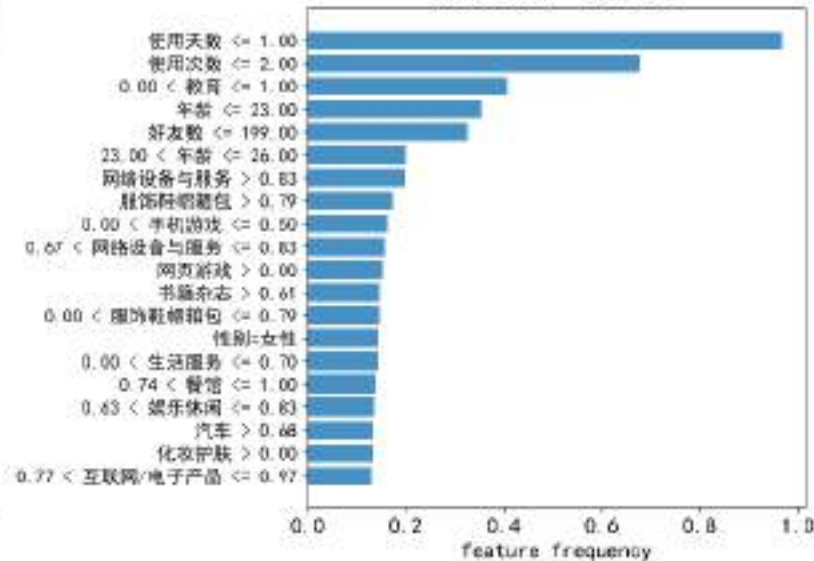


TABLE OF CONTENTES

整合

推断

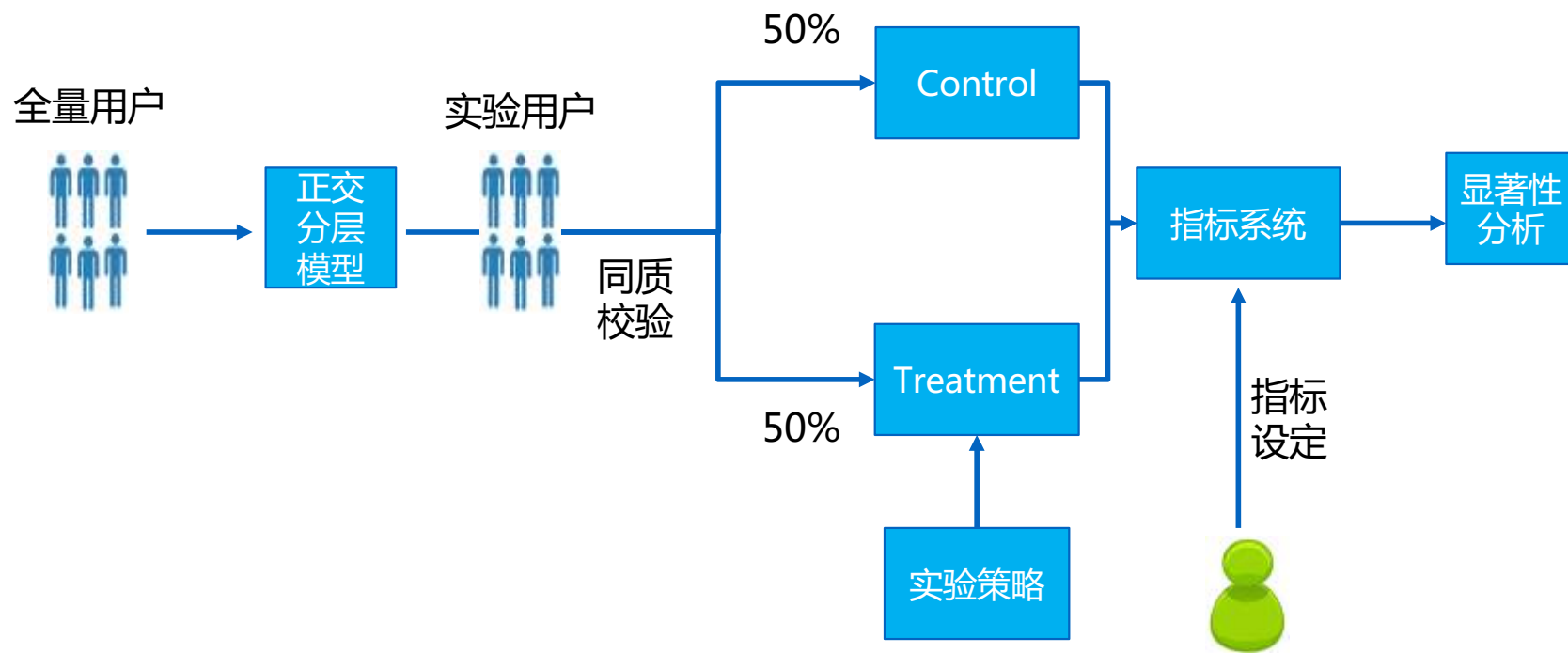
解释

调查 Investigation

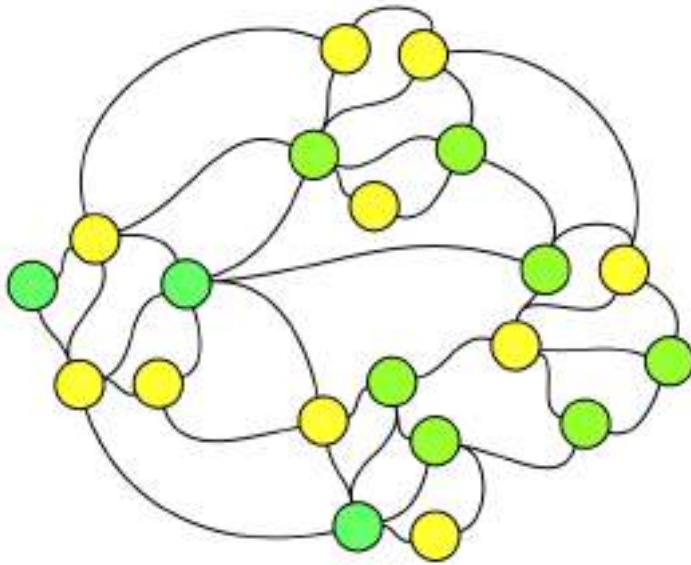
- 实例: Ntest

总结

传统ABtest系统



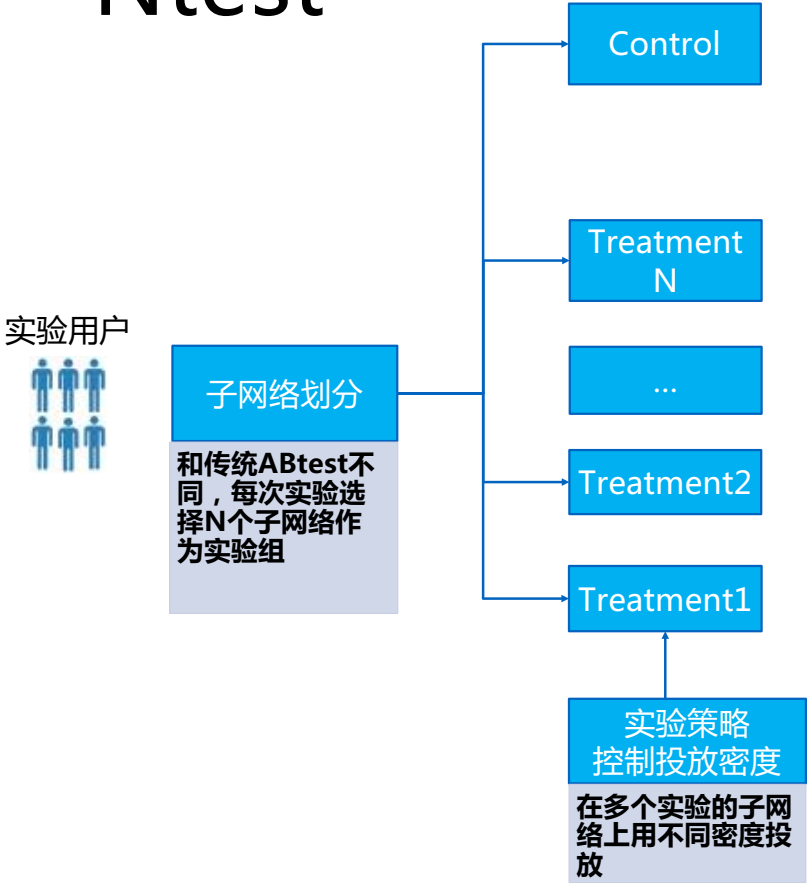
社交网络上ABtest的问题



社交网络上ABtest的问题

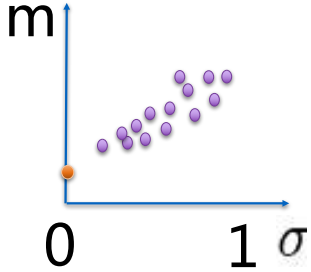
- A集合和B集合之间相互影响
- 在小样本的测试下有效，逐步加量后是否会继续有效，如何分析投放密度对效果的影响

Ntest

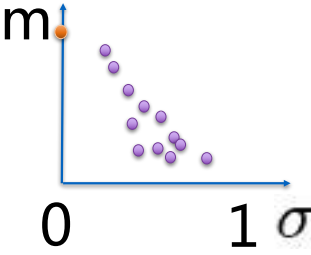


绘制各个指标实验效果和投放密度的关系

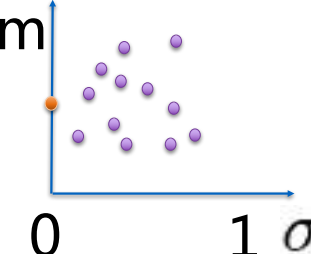
m : 指标
 σ : 密度



正相关



负相关



无关

TABLE OF CONTENTES

整合

推断

解释

调查

总结

- 人工智能和机器学习落地实践心得

总结：机器学习落地实践心得



Thanks