



招商银行信用卡中心金融数据大赛

成 绩：A榜第二，B榜第三

团队名称：我只吃两个剩下的都给你



赛题分析

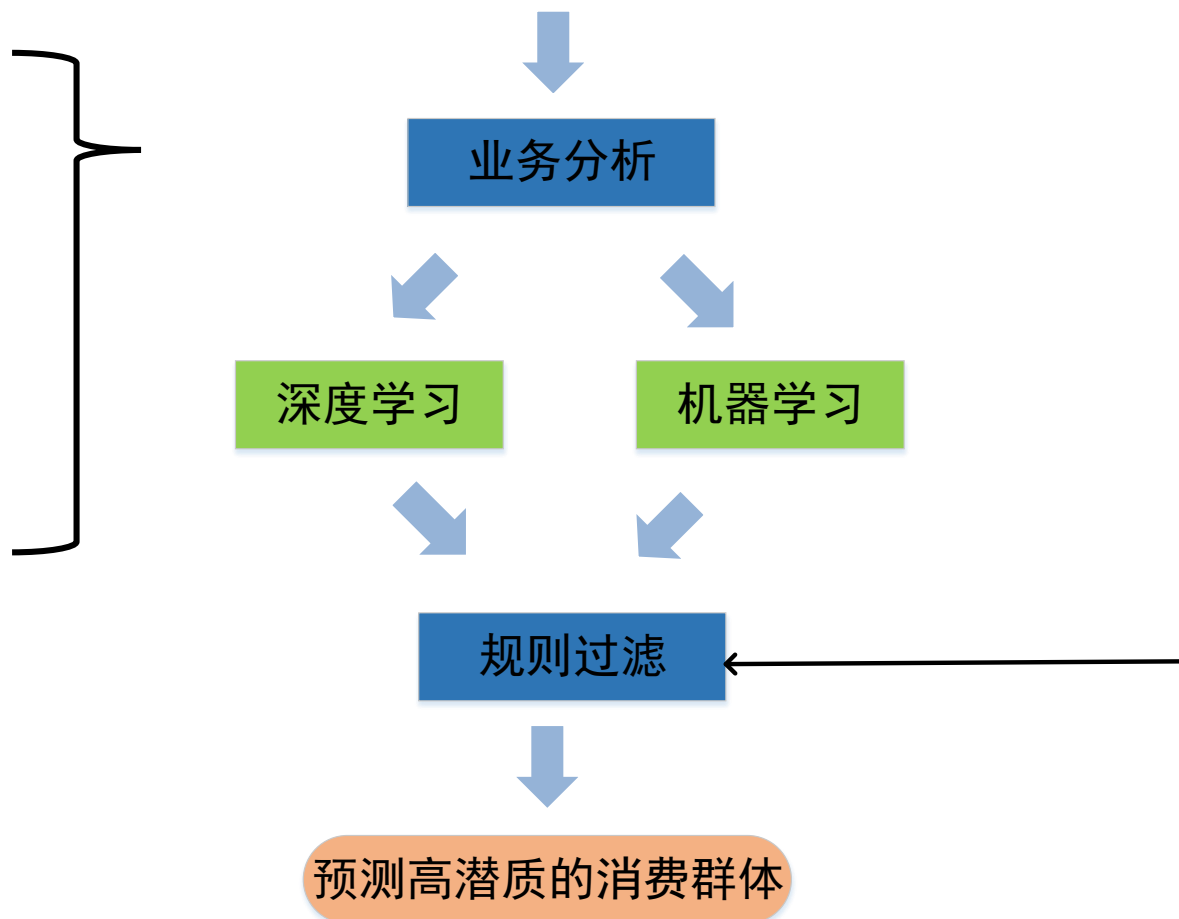
有监督学习

标签

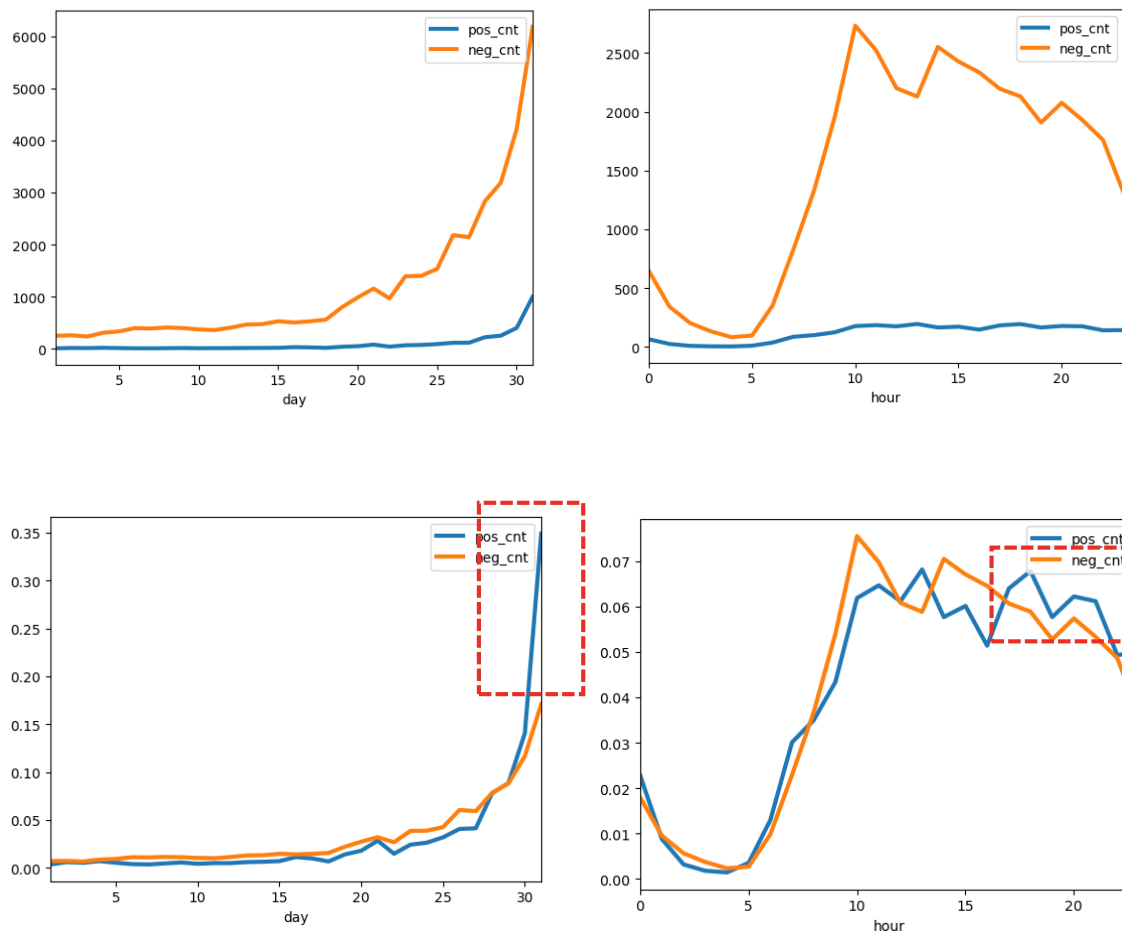
个人属性与信用卡消费agg表

APP操作行为日志log表

- 随着人工智能和大数据等技术不断渗透，依靠金融科技主动收集、分析、整理各类金融数据，为细分人群提供更为精准的消费需求服务，成为解决消费金融精确服务的有效途径。
- 识别出高潜质的消费群体，成为金融领域提供更为精准的推荐消费服务的关键。



特征工程——时序关系



用户消费和未消费时序关系 (day/hour)

用户消费和未消费趋势有规律性

统计用户每天消费
和未消费出现次数

统计用户小时消费
和未消费出现次数

用户 (每天/小时) 消费和未消费概率趋势

提取关于day时序
特征

提取关于hour时序
特征

特征工程——特征扩展

官方未告知agg表里面各个属性的含义，很难提取到agg表相关的特征。但可以通过一些合理的猜测发掘相关信息

train_agg	column	V1	V2	V3	V4	V5	...	V29	V30
	cnt	335	2	20	2	2		14293	68

test_agg	column	V1	V2	V3	V4	V5	...	V29	V30
	cnt	256	2	20	2	2		7644	41

猜测：V2、V4、V5为离散属性

假设：V2表示性别男女

V4表示职业1、2

V5表示有无不良记录



V2_V4、V2_V5、V4_V5、V2_V4_V5：
Labelencoder编码进行特征扩展

column	V1	V2	V3	V4	V5	...	V29	V30	V2V4	V2V5	V4V5	V2V4V5
--------	----	----	----	----	----	-----	-----	-----	------	------	------	--------

特征工程——特征群

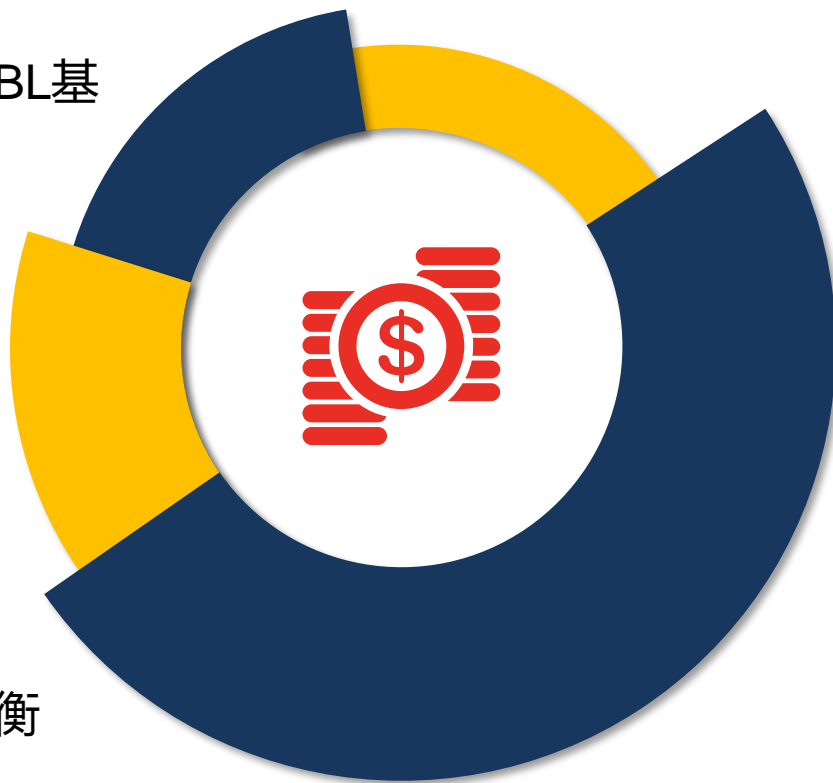
用户行为特征群:

用户行为TCH_TYP、EVT_LBL基础统计特征, 保证特征规模

- 最小值、均值
- 方差、中位数
- 标准差、众数

信息完整度特征群:

我们使用访问app的次数作为衡量一个用户信息的完整程度



用户属性特征群:

agg表的属性应直接作为特征, 但还可利用Labelencoder进行特征扩展。

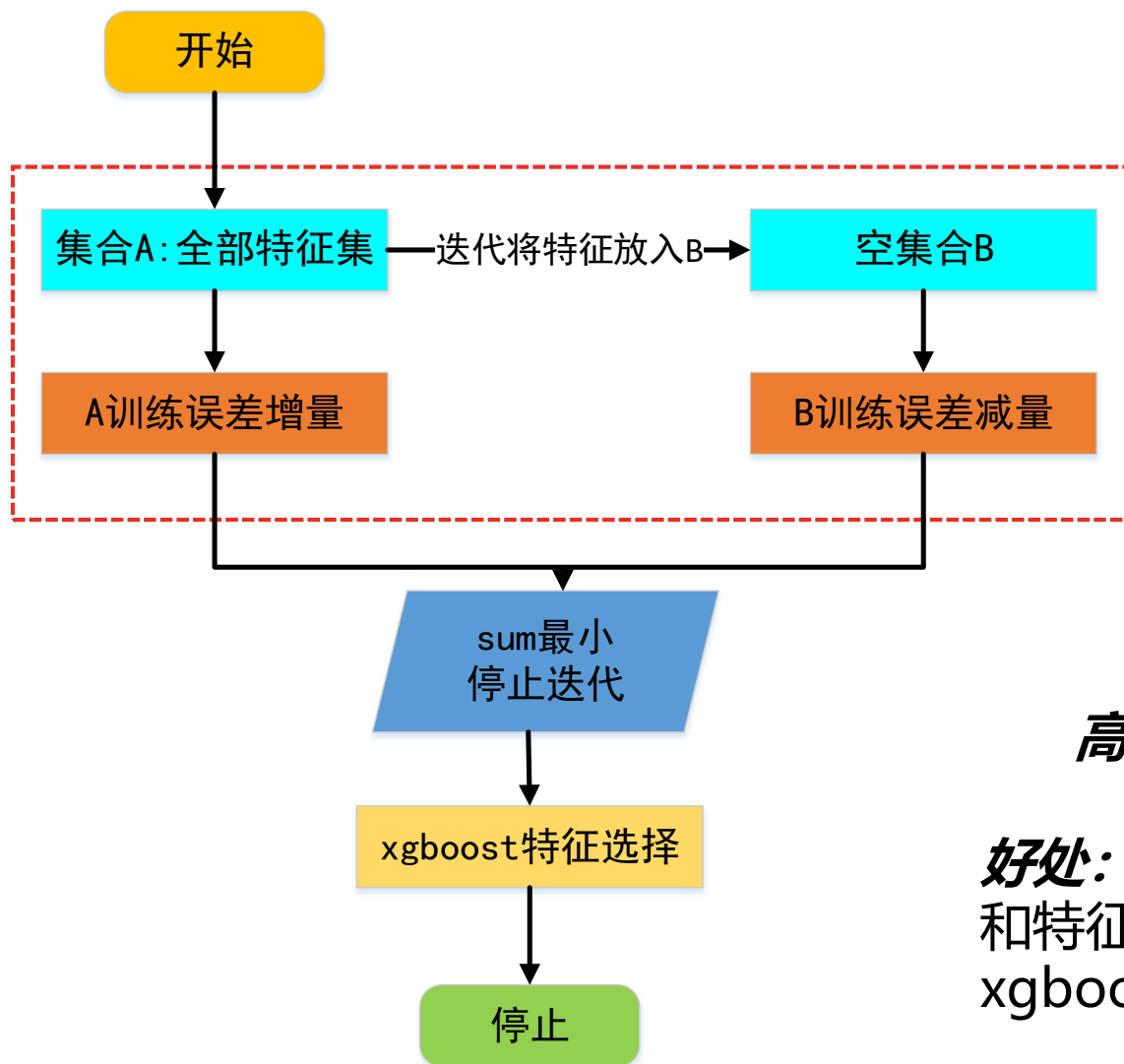
- V2_V4
- V2_V5
- V4_V5
- V2_V4_V5等进行特征扩展

业务时序特征群:

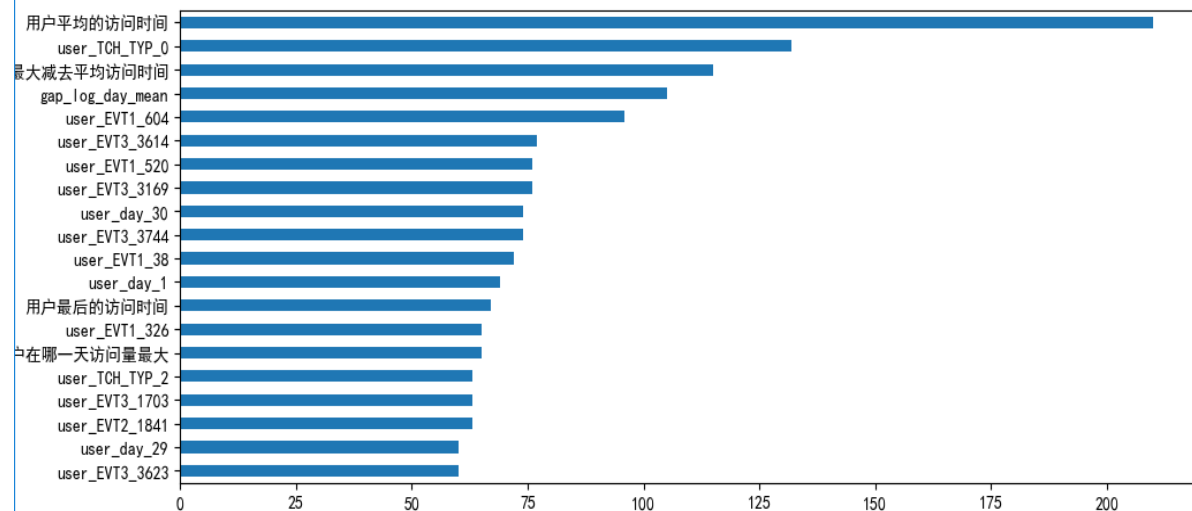
基于业务逻辑提取的特征, 主要基于log表提取关于时序特征

- 用户访问app的最后时间
- 用户访问app的平均时间
- 用户访问app时间间隔的最小值... ..

算法模型——特征选择



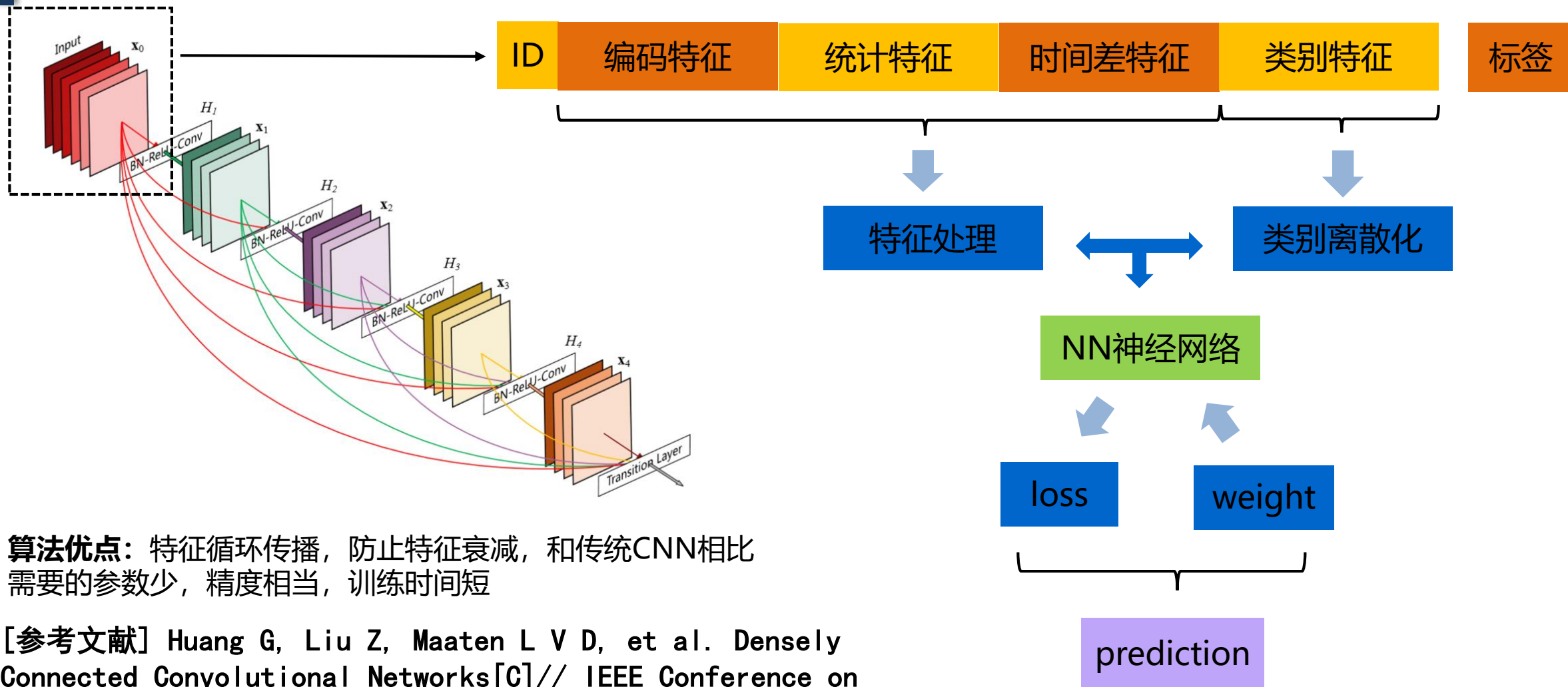
XGBOOST特征选择



高效的特性选择方案：基于视图+嵌入式的特征排序方法

好处：基于视图聚类特征选择将训练误差最小化，模型的学习和特征选择是同时进行的，基于 xgboost 来做特征选择，xgboost 模型训练完成后可以输出特征的重要性。

算法模型——神经网络



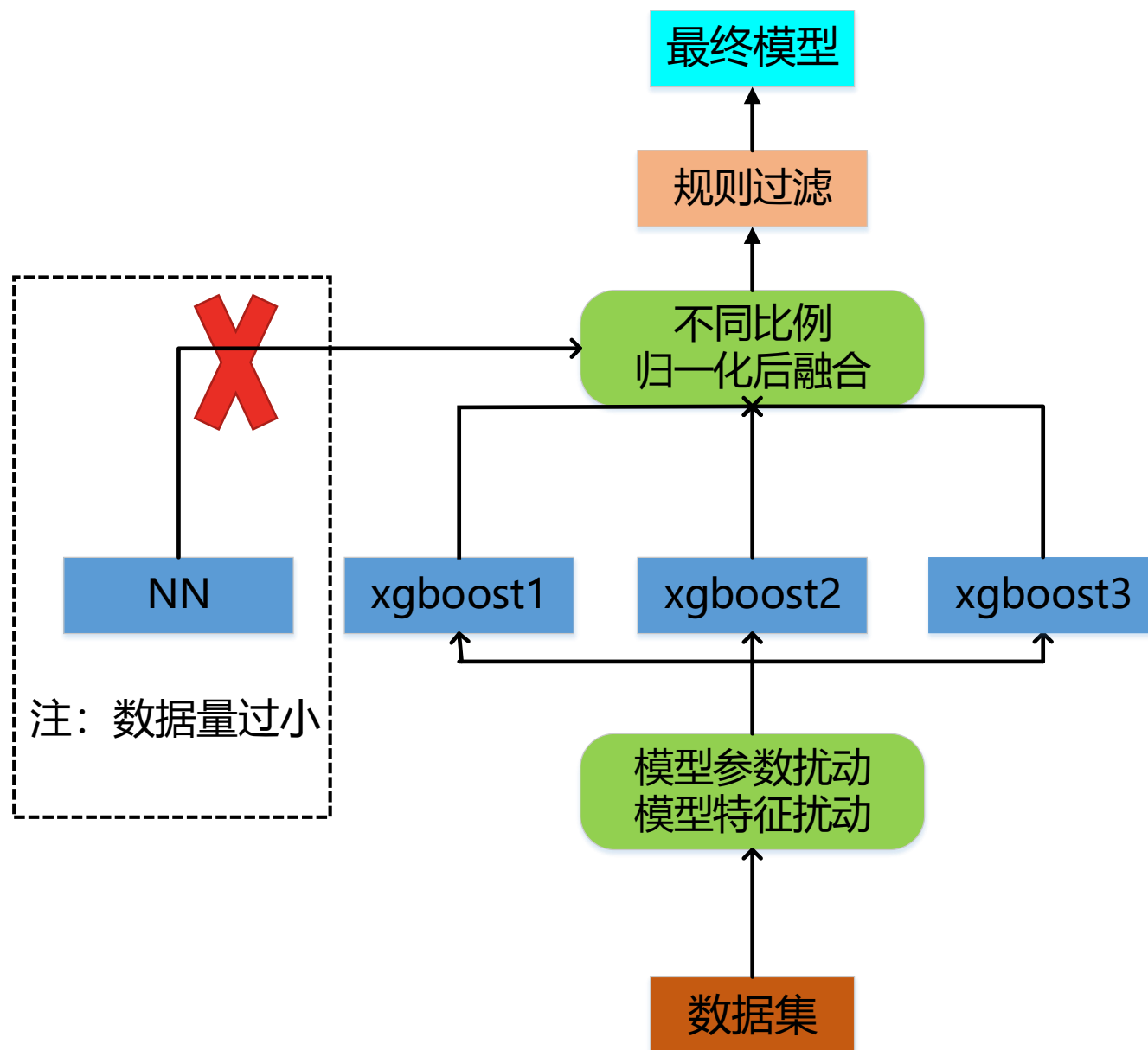
算法模型——模型融合

模型异构

- 特征工程
- 交叉验证

模型融合

- 算法层面



[注1] 交叉验证参数设置: $n_folds = 2$, $shuffle = True$, $random_state = None$

算法模型——模型融合

random_state=None



第一层:

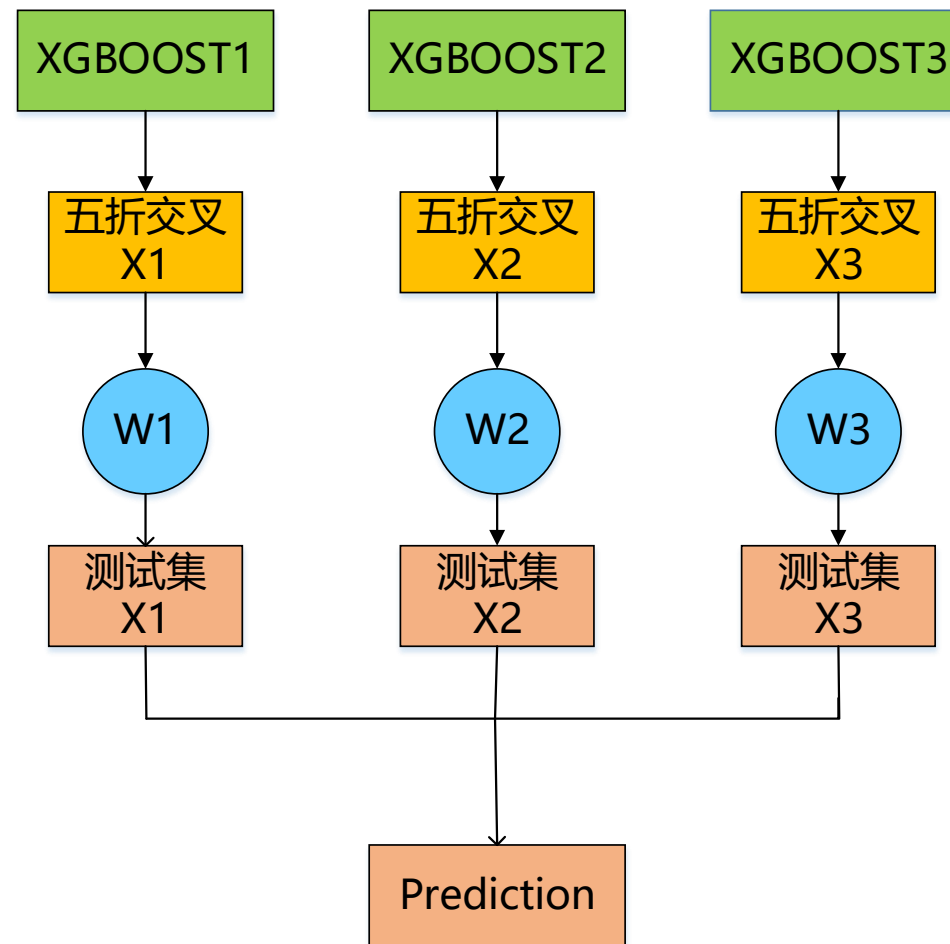
线性回归拟合验证集 $f(X_i) = x_i * w_i$

第二层:

$$Y(\text{验证集}) = x_1 * w_1 + x_2 * w_2 + x_3 * w_3$$

$$P(\text{测试集}) = X_1 * w_1 + X_2 * w_2 + X_3 * w_3$$

Stacking模型融合流程



代码亮点



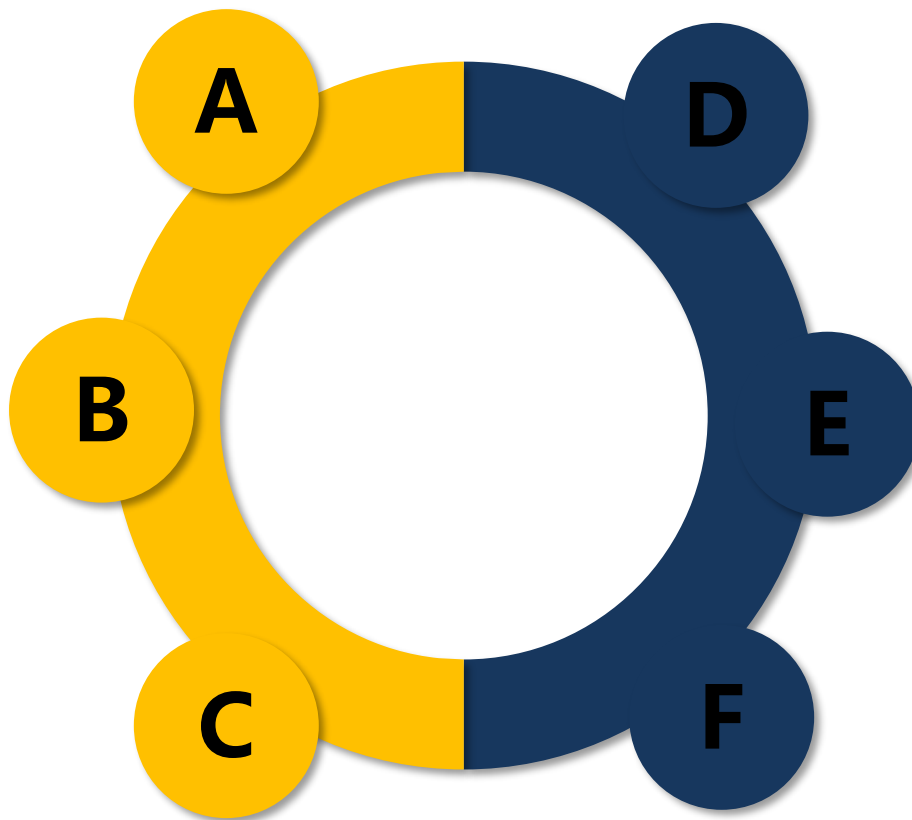
数据勘探:
通过多角度可视化探
寻数据内在联系



抗干扰性:
数据预处理, 减少数
据噪声对模型影响



特征角度:
labelencoder编码扩
展特征挖掘agg表潜
在的信息



运行效率:
采用了unstack()代替
了传统的for循环提取,
运行效率提高数倍。



模型精确度:
多个异构模型融合



模型稳定性:
模型融合、特征重组,
防止特征顺序影响模
型稳定性

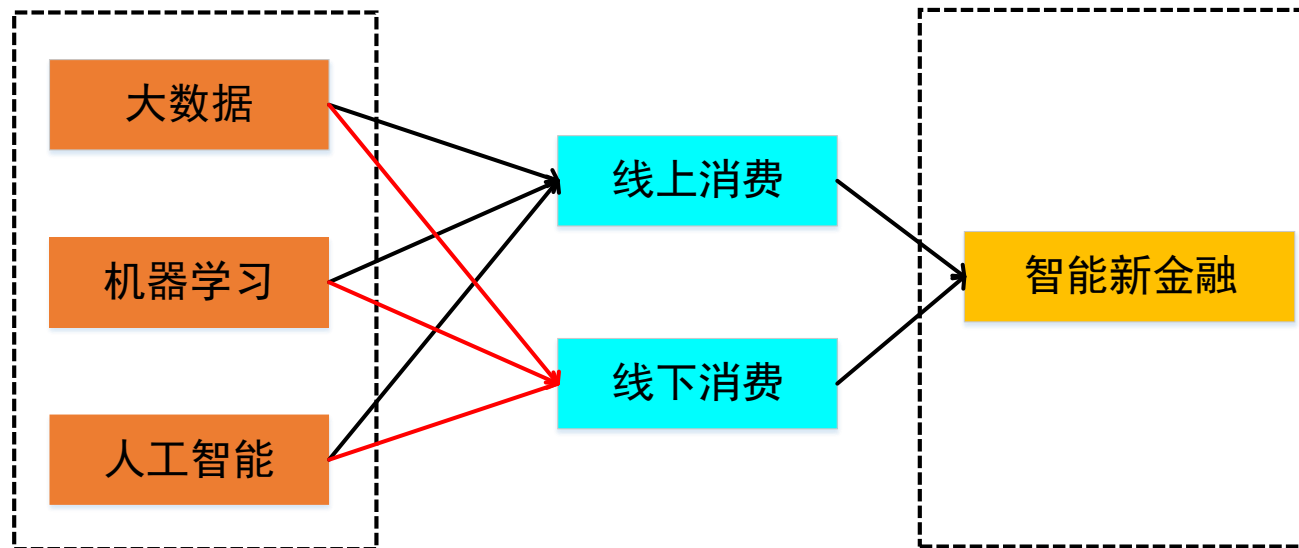
展望未来

消费场景

- 客户数据
- 信用卡消费数据
- 掌上生活APP日志

数据

金融科技



消费金融技术

智能预测助力消费金融

智能预测打造金融新概念