

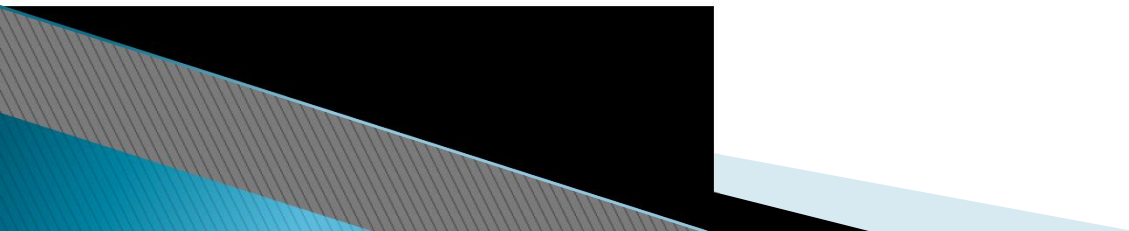
# 用户流失模型

赵海臣

# 建立用户流失模型的必要性

- ▶ 老客户可以为企业带来更多的利润，老客户为企业所带来的利润比新客户高20%-85%，Reichheld（1996）的研究表明：如果忠诚顾客每增长5%，那么企业利润将增加25%~95%。，所以预防客户流失是企业的重要任务。

- (<http://finance.qq.com/a/20120705/006578.htm>)



# 用户流失模型两种主流方法

- ▶ 方式1：把用户流失看作二元分类问题——流失/不流失
  - 选取一段时间，通过模型得到客户在该段时间内流失概率
- ▶ 方式2：评估用户的剩余生存期
  - 预测用户还能继续交易多久，通常是建立在生存分析的原理上进行研究的

# CRM客户价值

- ▶ 用户价值是企业在选择个性化服务对象时参考依据：
  - 当前价值：
    - 客户的当前关系，这种价值可在一段时间内持续。
  - 潜在价值：
    - 如果客户没有流失，其在将来的购买行为中将给企业带来的价值。
  - 影响价值：
    - 现有的每个客户会有自身的关系网，企业可以利用客户本身的网络影响和吸收其他新客户来企业进行交易，这种方式为企业带来价值就是客户的影响价值
- ▶ 每类价值客户的需求都不相同，企业需要针对不同类型客户的需求来采取相应的措施，从而从整体上提高客户的忠诚度，防止客户大面积的流失。

# 用户流失分类

## ▶ 客户主动流失：

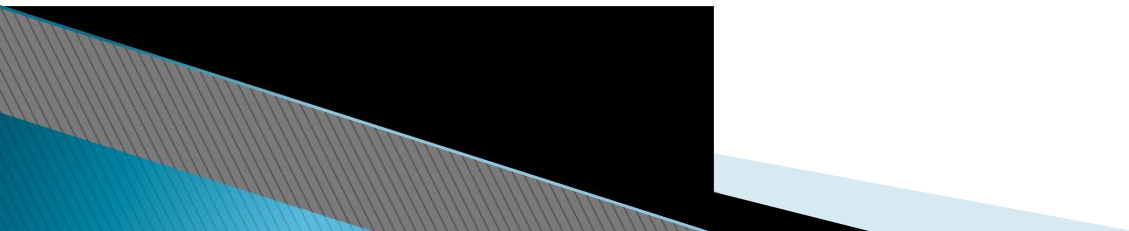
- 客户不再选择该电子商务企业进行交易，而是选择了别家企业

## ▶ 客户交易量大幅下降流失：

- 企业根据自身产品的周期规定某个时间段内，若在这段时间内客户交易相对减少，减少量超过规定数值

## ▶ 零次消费流失：

- 客户在一段时间内交易次数为零



# 流失模型分析指标(维度)

- ▶ 性别(男女对化妆品的行为区别)
- ▶ 年龄(不同年龄的行为区别)
- ▶ 地区(地区经济差异)
- ▶ 用户等级画像(类似于之前优质用户筛选)
- ▶ 总交易次数
- ▶ 总交易金额
- ▶ 近期交易次数(RFM – Frequency)
- ▶ 近期交易金额(RFM – Monetary)
- ▶ 购物评价均值
- ▶ 近期购物评价均值\*
  - (\*重要\*用户对聚美假货问题怀疑成为用户流失主因，可以根据评价获得)
- ▶ 最晚购买时间(RFM – Recency)

# 用户行为与流失关系猜测

## ▶ 交易总次数

- 未流失客户的总交易次数均值比流失客户的总交易次数均值多：交易次数越多，客户越不容易流失。

## ▶ 交易总金额

- 未流失客户总交易金额均值比流失客户总交易金额均值明显大很多：总交易金额越大，客户越不容易流失。

## ▶ 交易评价均值

- 未流失客户的评分均值比流失客户的评分均值高：客户对交易评价分值越高，客户对本次交易越满意，客户将会继续进行交易，客户流失率越低。

## ▶ 最后购买时间点

- 未流失客户的最后购买时间的均值比流失客户的最后购买时间均值短：客户最后购买时间离观测点时间越长，那么该客户流失的概率就越大。

# 用户流失模型所适用机器学习算法

- ▶ 用户流失模型实质上是对用户的分类，可以适用大部分分类算法：
  - 逻辑回归
  - 决策树
  - SVM
  - 神经网络
  - ...



# 流失模型训练

## ▶ 1. 时间段的选择

- 选择可以明确流失用户的时间段：目前定义用户120天时间段确定是否流失，选择120天前的时间点作为参考时间点

## ▶ 2. 参考时间点用户状态(维度)

- 计算出训练集用户参考时间点的流失分析指标，作为用户当前的状态

## ▶ 3. 参考时间点后用户流失情况(反馈-2值分类)

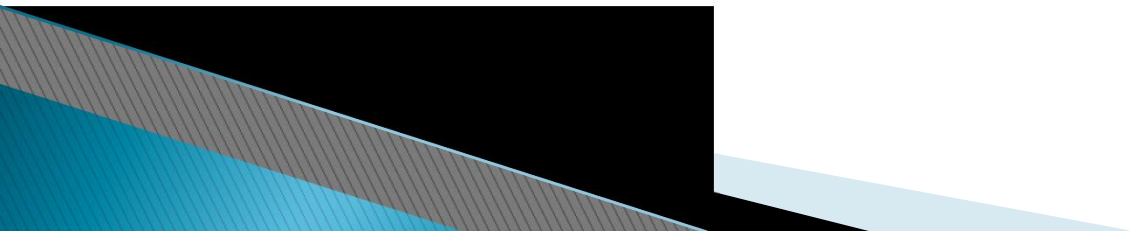
- 对每个训练集用户标记出参考时间点之后的流失情况

# 流失模型训练

- ▶ 4. 通过逻辑回归/决策树/SVM/神经网络对训练数据进行建模
  - 模型输入：用户流失分析维度
  - 模型输出：流失/非流失概率
- ▶ 5. 通过模型预测测试用户，获得模型性能数据，进行模型选择，调参
- ▶ \*性别，年龄，地区可以分别单独建模，相互之间会有行为差异。

# SMC用户潜在价值模型

- ▶ Schmittlien. Morrison and Colombo于1987年提出了如何预测客户交易行为的模型组，简称为SMC模型
  - SMC模型的主要思想是：用客户过去的历史交易数据来预测客户将来的交易行为。



# SMC用户潜在价值模型

## ▶ 五个假设

### ◦ 活跃客户的交易：

- 泊松分布的参数 $\lambda$ 是单位时间(或单位面积)内随机事件的平均发生率。泊松分布适合于描述单位时间内随机事件发生的次数。
- 活跃客户与企业的交易率概率满足泊松分布，用 $\lambda$ 表示用户长期以来单位时间交易率，当客户活跃时，客户在时间 $t$ 内交易总次数为 $x$ 的概率为：

$$P[X = x | \lambda(t), t] = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, x = 0, 1, 2, \dots$$

- \*已经有很多研究已经证实了可以用泊松随机过程来描述个体客户的购买行为

# SMC用户潜在价值模型

- 个体客户的死亡率：
  - 假设每个客户的个体生存时间都服从指数分布。用 $\mu$ 来表示客户的流失率，那么其密度函数可以表示为：

$$f[t | \mu] = \mu e^{-\mu t}, t > 0$$

- 不同客户购买率：
  - 有的客户购买频繁，有的确很少购买，假设不同客户的购买率 $\lambda$ 服从 $\Gamma$ （Gamma）分布
- 不同客户的流失率
  - 服从 $\Gamma$ （Gamma）分布

The end