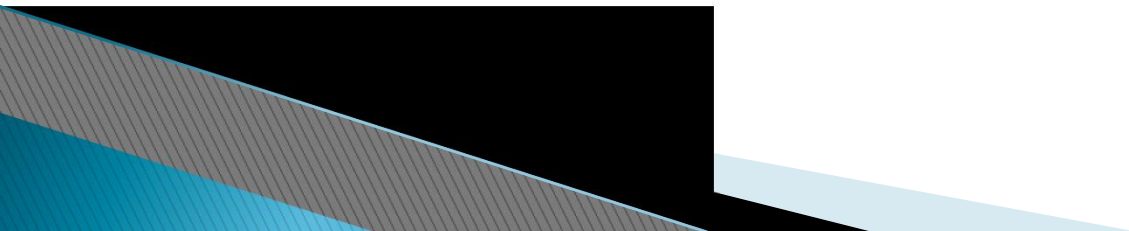


# 隐含行为分析

赵海臣

# 隐含行为的种类

- ▶ 浏览记录
- ▶ 购买记录
- ▶ 搜索记录
- ▶ 鼠标移动
- ▶ ...



# 隐含行为的特征

## ▶ 没有负反馈

- 通过观察用户的隐含行为，我们只能推断出用户可能喜欢某个东西，而不能推断出用户不喜欢某个东西。
  - 跟显式行为的区别：显式行为有明确的喜好，可以收集正面跟负面反馈，因此不分析missing data；而隐含行为没有负面反馈，大部分负面反馈隐藏在missing data中，所以missing data也需要进行分析。

## ▶ 充满噪声

- 我们只能“猜”用户的意图，例如，用户买了某个东西可能仅仅是为了做礼物而不是喜欢它；或者用户买了某个东西却讨厌它。

# 隐含行为的特征

- ▶ 隐含行为的数字分数仅仅表示Confidence信心
  - 隐含行为的数字分数通常表示用户某个行为的频率/次数，分数并不一定意味着用户的喜好，例如，最喜欢的电影可能只会看一遍，而喜欢的连续剧每天都会看，分数不同，但却都很喜欢。
  - 所以，隐含行为的数字分数仅仅表示confidence，一次性的单次行为可能跟用户的兴趣没什么关系，但是重复的行为更可能真的反应用户的兴趣。

# 显式/隐式行为ALS公式

## ▶ 显式ALS:

$$\min_{q_*, p_*} \sum_{(u, i, t) \in \mathcal{K}} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

- $r_{ui}$  是用户  $u$  对商品  $i$  的显式打分

## ▶ 隐式ALS:

$$\min_{x_*, y_*} \sum_{u, i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda \left( \sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

- $C_{ui}$  是用户  $u$  对商品  $i$  的 confidence 信心
- $P_{ui}$  是 boolean 变量值

# 隐式ALS公式

$$\min_{x_*, y_*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda \left( \sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

- $p_{ui}$ 是boolean变量值

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}$$

- 用户 $u$ 与商品 $i$ 发生过交互,  $p_{ui} = 1$
- 用户 $u$ 与商品 $i$ 未发生交互,  $p_{ui} = 0$
- $c_{ui}$ 是用户 $u$ 对商品 $i$ 的信心度confidence

$$c_{ui} = 1 + \alpha r_{ui}$$

- 当用户 $u$ 对商品 $i$ 的交互次数多, 说明用户很可能对这个商品感兴趣, 因此 $c_{ui}$ 线性增大
- 用户对所有非交互商品都有基础分1, 所有 $u-i$ 对都要计算

# 隐式公式解释

- ▶ Confidence表明了用户对商品的倾向度，而非兴趣度，因此作为“权重”乘到用户 $u$ -商品 $i$ 上，而非像显式模型那样直接作为“加数”，权重高的 $u-i$ 对将有更大的调整幅度(因为权重 $c$ 大)

The end