

*Initial Algorithm*:(category based)

**类别分 0-4 共 5 类:**

- 0 - 无行为
- 1 - 点击
- 2 - 收藏
- 3 - 购物车
- 4 - 购买

**基于假设:**

1. 确定为购买行为后预测未来不会重复购买 (little sacrifice)
2. 确定为购物车行为后直接预测未来会购买 (little mistake)
3. 无论多少次点击均视为一次点击类别为 1
4. 行为类别确定原则  $0 < 1 < 2 < 3 < 4$

**推荐方法:**

暂采用基于用户和商品的协同过滤的方法 (2 种), 行为预测采用 KNN 投票的方式, 重点在于对原行为 2、1、0 的预测, 将所有预测为 4 的项推荐出来。

**相似性度量:**

采用 *Manhattan 距离*。

**Process:** (in Python)

1. 建立 user 和 brand 的索引表 (数组、列表)

```
User = ['zhangsan', 'lisi', 'wangwu', ...];
```

```
Brand = ['book', 'CD', 'pencil', ...];
```

2. 建立 user 和 brand 对应的行为矩阵

```
Behavior = [ [4, 1, 2, 3, 0, ...],  
             [1, 1, 4, 2, 0, ...],  
             [4, 1, 3, 3, 0, ...],  
             [3, 2, 4, 2, 0, ...],  
             ...  
            ];
```

3. 建立预测行为矩阵并通过算法进行填充

```
Prediction = [ [ ... ],  
               [ ... ],  
               [ ... ],  
               ...  
              ];
```

算法的填充原则: 原来为 4 的直接填充 0;

原来为 3 的直接填充 4;

原来为 2、1、0 的通过 KNN 投票预测结果填充 (k 暂取 10);