

Machine Learning Lab

重庆邮电大学机器学习研究所

淘宝穿衣搭配-挑战Baseline

邵帅

2016级硕士研究生

s160231053@stu.edu.com

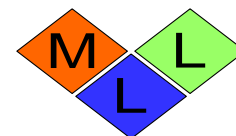




最终成绩

最终成绩：

| | | | | |
|----|-----------------|------------|-------|------------|
| 8 | X-Data-4 | 重庆邮电大学 | 6.11% | 2017-03-21 |
| 9 | X-Data-5 | 重庆邮电大学 | 6.10% | 2017-03-22 |
| 10 | kekedan | 电子科技大学 | 5.95% | 2017-03-09 |
| 11 | X-Data-3 | 重庆邮电大学 | 5.93% | 2017-03-24 |
| 12 | 就是来逛一逛的0 | 重庆邮电大学 | 5.87% | 2017-03-26 |
| 13 | x-data-12 | 重庆邮电大学 | 5.49% | 2017-03-28 |
| 14 | 罗曼罗迪 | 重庆大学 | 5.39% | 2017-03-22 |
| 15 | X-Data-8 | 重庆邮电大学 | 4.58% | 2017-03-24 |
| 16 | x-data-1 | 重庆邮电大学 | 4.35% | 2017-03-20 |
| 17 | Chiangshm | 西安外国语大学 | 3.34% | 2016-06-22 |
| 18 | FNo0 | 吉林农大 | 2.64% | 2017-03-16 |
| 19 | TiranoGreatLand | 哈尔滨工业大学 | 2.36% | 2017-02-25 |
| 20 | 欢茄炒蛋 | 其它-想花哄胆信给你 | 2.16% | 2017-03-20 |

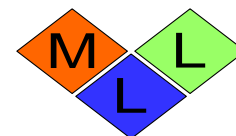
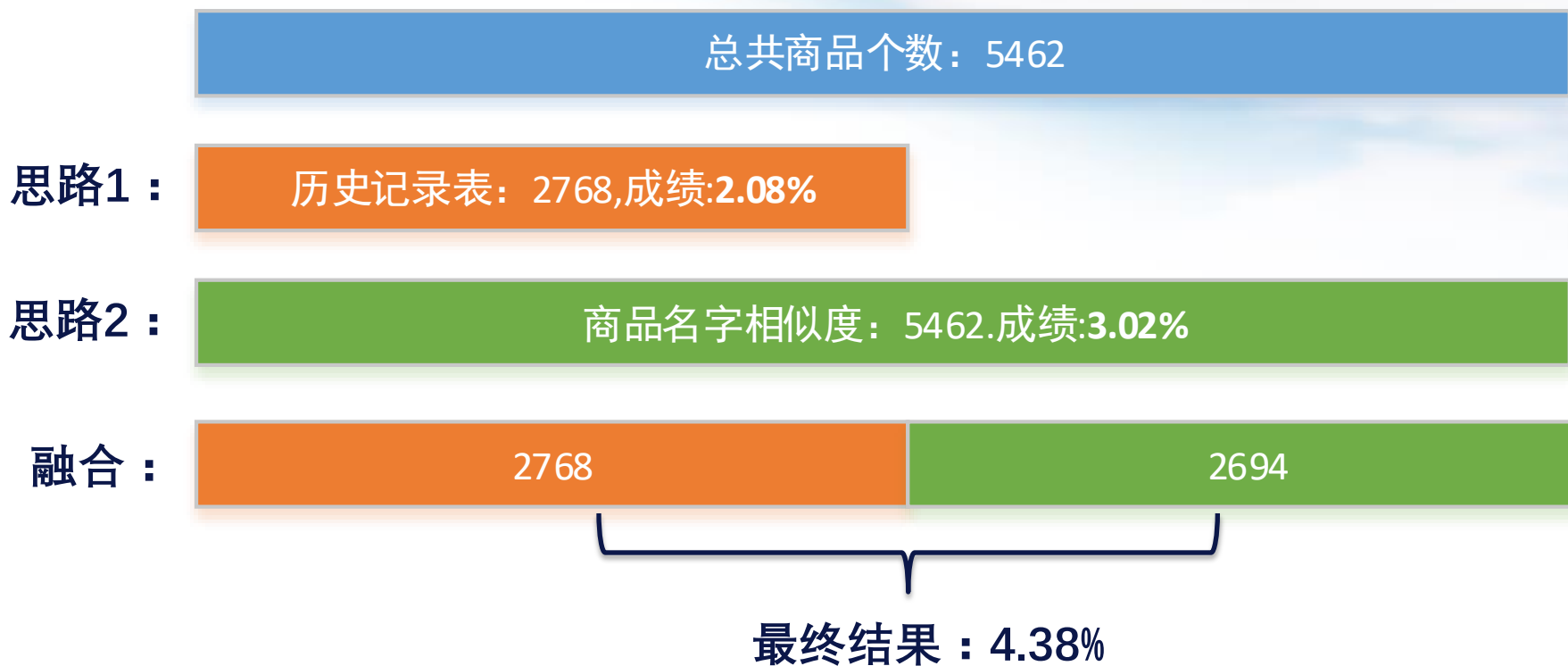


Machine Learning Lab

重庆邮电大学机器学习研究所



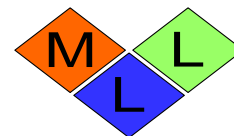
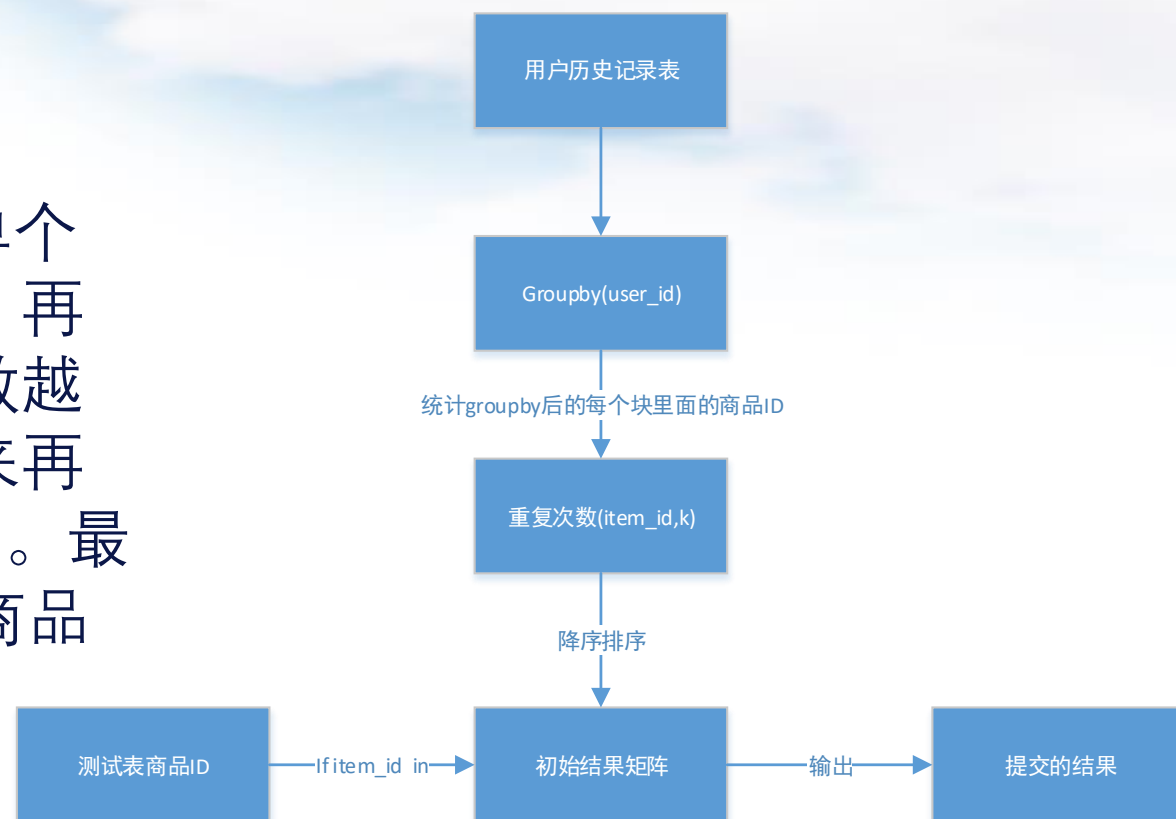
淘宝穿衣搭配





思路1：

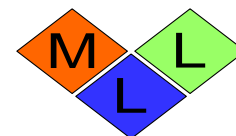
❖ 用所有用户的历史记录表，统计单个用户一年内的购买商品的历史记录，再统计这些记录的重复次数，重复次数越多，排名越往前。然后去重。接下来再用这些记录去merge测试表的商品ID。最终形成搭配记录。这里只有2768个商品ID有历史记录。线上成绩：**2.08%**。





思路2：

- ❖ 通过对商品表里面的商品名称分词的列表作余弦相似度匹配。找出测试表商品ID相似度最高的前200个ID作为最终的搭配记录。线上成绩**3.02%**。
- ❖ 将这两个成绩融合。最终**4.35%**。

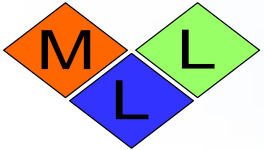




总结：

- ❖ 只用到了两个表，其他表的数据基本没动。
- ❖ 历史记录表的结果，还可以根据搭配表去过滤。
- ❖ 历史记录表的结果很多都未满足200个，所以成绩还可以提高。
- ❖ 只用了余弦相似度计算，还可以尝试下TDIDF，相同个数等方法。





Machine Learning Lab

重庆邮电大学机器学习研究所

Thank you!

