

# Item-Based Collaborative Filtering

---

Yang LI

2017 年 10 月 19 日

School of Software Engineering, Tongji University

# Visualization

---

# Item-Count Chart

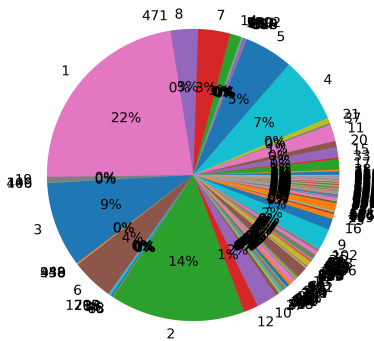


图 2: train\_item

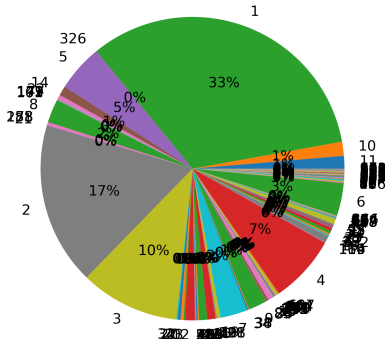


图 3: test\_item

# User-Count Chart

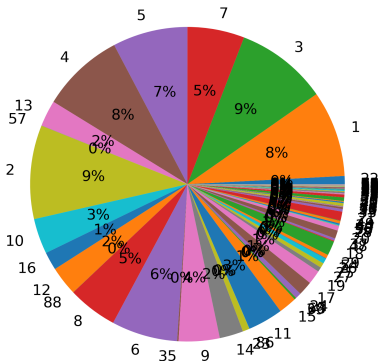


图 4: train\_user

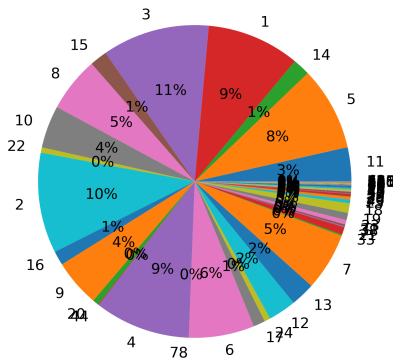


图 5: test\_user

# Improvement

---

- 更新了数据集, 使用了三个月的训练数据
  - 相较于上次实验结果, 本次测试结果出现下降
- 变化了测试时用户购买商品的种数的阈值
- 对于训练集未出现的商品, 在测试集中予以删除
  - 可预料地, 提高了 recall, 而 precision 不变
- 根据用户测试集中购买的商品种数, 进行推荐 (上限为 20)
  - 可预料地, 提高了 precision, 而 recall 不变

# Result

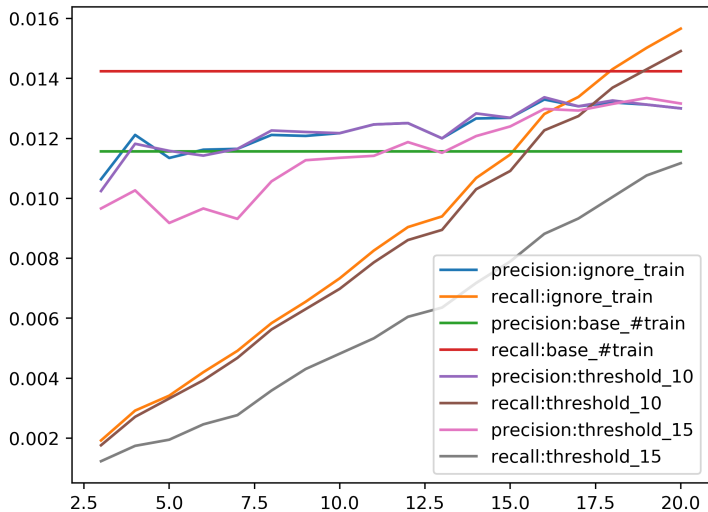


图 5: precision&recall

## Dive into individual

Take User 13021568889 for example

- In the Training

item	item_id	count	amt	price
伊利 lifeup 红枣风味酸奶 260g	15114035	2	2.0	13.8
花色小面包	23110004	2	7.98	7.98
鲜到家黑椒牛排	24401037	1	0.196	23.48

- In the Recommendation

item	item_id
片仔癀牙火清牙膏 (甄选留兰香)145g	11301138
益达纤柔焕彩牙刷 (情侣装)	11303197
海太菠萝果粒果汁饮料 238ml	10119128



■ In the Test

item	item_id	count	amt	price
七度空间公主系列卫生巾 5 片	11501062	1	1.0	9.9
苏菲裸感 S 夜用 29cm12 片	11501116	1	1.0	15.8
格力高百利滋蓝莓芝士 45g	14050013	1	1.0	3.9
达利园瑞士卷巧克力味 22g*10	14070003	1	1.0	6.9
盼盼小面包奶香味 440g	14072017	1	1.0	10.9
甘汁园阿胶红糖 350g	14863006	2	2.0	19.6
蒙牛冠益乳原味/草莓果粒 450g	15115013	1	2.0	27.8
蓝莓 (大果/甜)	22007031	1	0.64	12.8
鱿鱼串	24220050	1	0.264	10.51
儿童牛排	24401031	1	0.108	12.92
鲜到家黑椒牛排	24401037	1	0.108	12.94
包菜肉包	25111045	1	4.0	4.0
酱菜系列	27100542	2	0.498	8.48
山楂糕	27410044	1	0.244	4.83 <sub>6</sub>