

用户重复行为分析

赵海臣

用户重复行为分析

- 复购行为：
 - 考察时间段内所购买的商品在之前相邻时间内同一用户购买的比例；
- 重复浏览行为：
 - 考察时间段内所浏览的商品在之前相邻时间内同一用户浏览的比例；
- 浏览购买行为：
 - 考察时间段内所购买的商品在之前相邻时间内同一用户浏览过的比例

用户重复行为分析

▶ 复购行为：

- >>> a =
sqlContext.table("recommend.userordertableuserfiltered_a
lsfortest1")
- >>> b = a.filter((a["data_date"] >= "2017-03-16") &
(a["data_date"] <= "2017-03-22"))
- >>> c = a.filter((a["data_date"] >= "2017-03-09") &
(a["data_date"] <= "2017-03-15"))
- >>> b.join(c, ["userId", "productId"], "left_semi").count()
- >>> b.count()

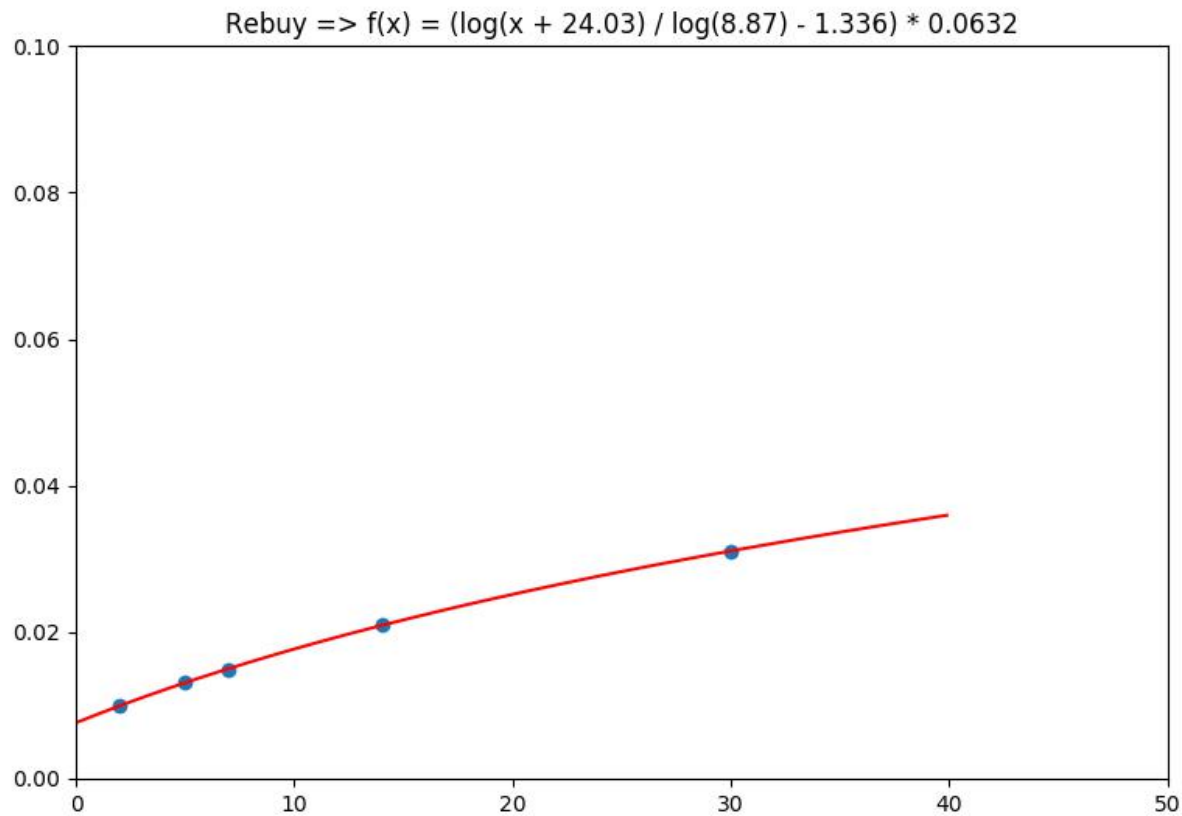
复购行为

- 用户重复购买同一商品的比例(相邻7天vs7天)
 - ([2017-03-09, 2017-03-15]与[2017-03-16, 2017-03-22]): $14128/1651568 = 0.855\%$
- 用户重复购买同一商品的比例(相邻14天vs7天)
 - ([2017-03-02, 2017-03-15]与[2017-03-16, 2017-03-22]): $23534/1651568 = 1.4249\%$

复购行为(考察2天)

- 用户重复购买同一商品的比例(相邻30天vs2天)
 - ([2017-02-14, 2017-03-15]与[2017-03-16, 2017-03-17]):
 $16299/524815 = 3.1057\%$
- 用户重复购买同一商品的比例(相邻14天vs2天)
 - ([2017-03-02, 2017-03-15]与[2017-03-16, 2017-03-17]):
 $10969/524815 = 2.0901\%$
- 用户重复购买同一商品的比例(相邻7天vs2天)
 - ([2017-03-09, 2017-03-15]与[2017-03-16, 2017-03-17]):
 $7836/524815 = 1.4931\%$
- 用户重复购买同一商品的比例(相邻5天vs2天)
 - ([2017-03-14, 2017-03-15]与[2017-03-16, 2017-03-17]):
 $6894/524815 = 1.3136\%$
- 用户重复购买同一商品的比例(相邻2天vs2天)
 - ([2017-03-14, 2017-03-15]与[2017-03-16, 2017-03-17]):
 $5192/524815 = 0.9893\%$

复购行为回归曲线



用户重复行为分析

▶ 重复浏览行为:

- `>>> a = sqlContext.table("recommend.browsedatatableuserfiltered_alsfortest1")`
- `>>> b = a.filter((a["data_date"] >= "2017-03-16") & (a["data_date"] <= "2017-03-22"))`
- `>>> c = a.filter((a["data_date"] >= "2017-03-09") & (a["data_date"] <= "2017-03-15"))`
- `>>> b.join(c, ["userId", "productId"], "left_semi").count()`
- `>>> b.count()`

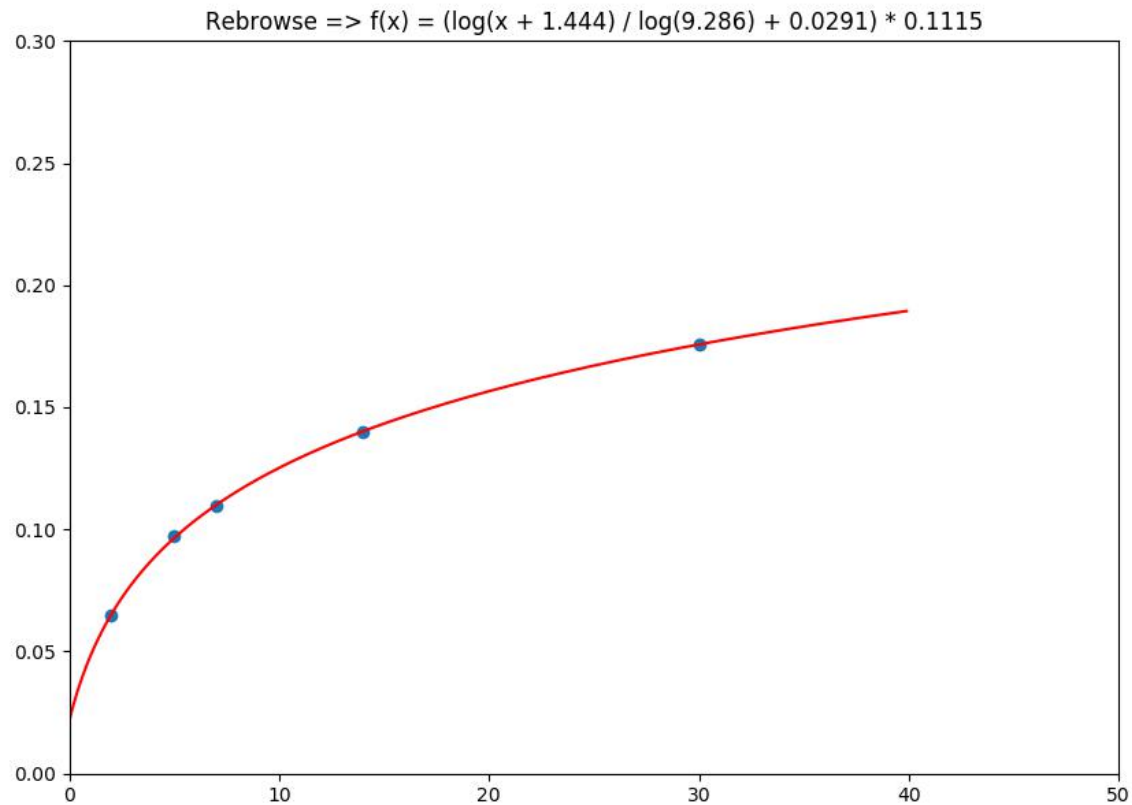
重复浏览行为

- 用户重复浏览同一商品的比例(相邻7天vs7天)
 - ([2017-03-09, 2017-03-15]与[2017-03-16, 2017-03-22]): $4814519 / 58444978 = 8.2377\%$
- 用户重复浏览同一商品的比例(相邻14天vs7天)
 - ([2017-03-02, 2017-03-15]与[2017-03-16, 2017-03-22]): $6741107 / 58444978 = 11.5341\%$

重复浏览行为(考察2天)

- 用户重复浏览同一商品的比例(相邻30天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-02-19, 2017-03-20]):
 $2651957/15076031 = 17.5905\%$
- 用户重复浏览同一商品的比例(相邻14天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-19, 2017-03-20]):
 $2107954/15076031 = 13.9821\%$
- 用户重复浏览同一商品的比例(相邻7天vs2天)
 - ([2016-03-21, 2017-03-22]与[2017-03-14, 2017-03-20]):
 $1653337/15076031 = 10.9666\%$
- 用户重复浏览同一商品的比例(相邻5天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-19, 2017-03-20]):
 $1462858/15076031 = 9.7032\%$
- 用户重复浏览同一商品的比例(相邻2天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-19, 2017-03-20]):
 $979321/15076031 = 6.49588\%$

重复浏览行为回归曲线



用户浏览购买行为分析

▶ 重复浏览行为:

- >>> a = sqlContext.table("recommend.userordertableuserfiltered_alsfortest1")
- >>> b = sqlContext.table("recommend.browsedatatableuserfiltered_alsfortest1")
- >>> c = a.filter((a["data_date"] >= "2017-03-16") & (a["data_date"] <= "2017-03-22"))
- >>> d = b.filter((b["data_date"] >= "2017-03-09") & (b["data_date"] <= "2017-03-15"))
- >>> c.join(d, ["userId", "productId"], "left_semi").count()
- >>> c.count()

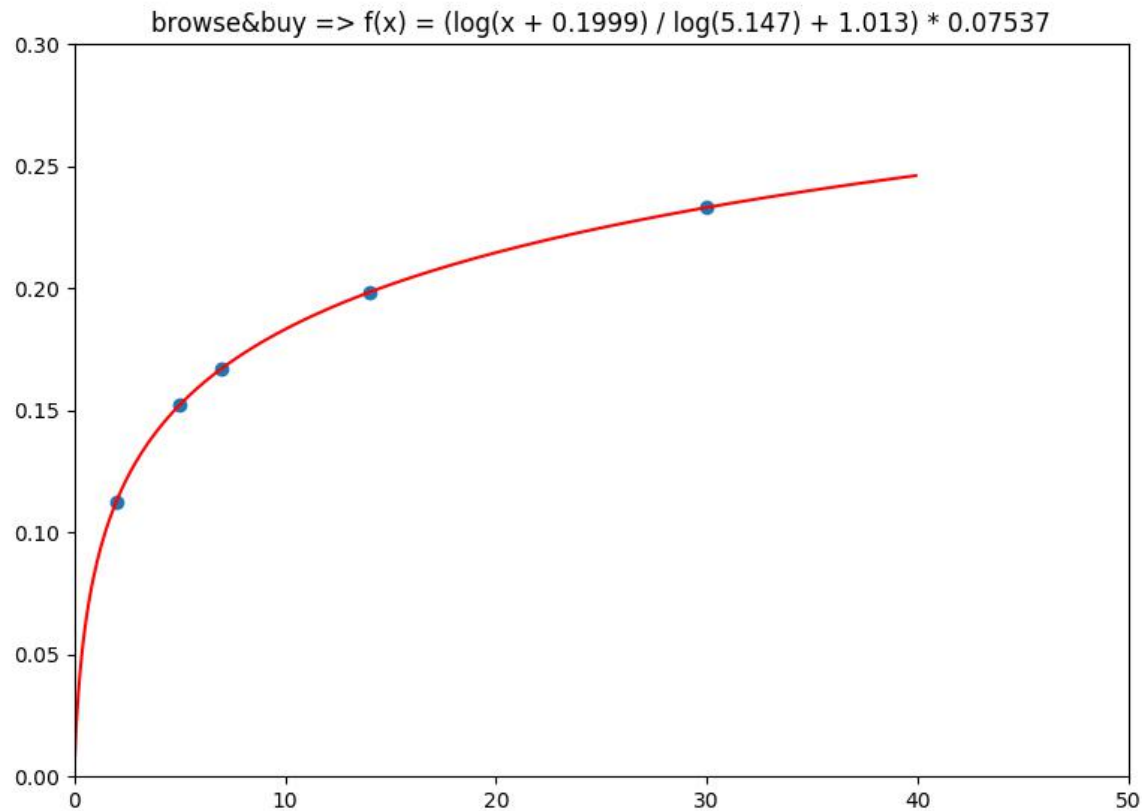
用户浏览购买行为分析

- 用户浏览购买同一商品的比例(相邻7天vs7天)
 - ([2017-03-09, 2017-03-15]与[2017-03-16, 2017-03-22]): $202742/1651568 = 12.2757\%$
- 用户浏览购买同一商品的比例(相邻14天vs7天)
 - ([2017-03-02, 2017-03-15]与[2017-03-16, 2017-03-22]): $269264/1651568 = 16.3035\%$

用户浏览购买行为分析(考察2天)

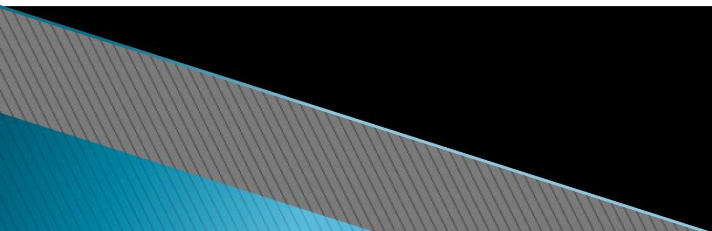
- 用户浏览购买同一商品的比例(相邻30天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-02-19, 2017-03-20]):
 $106228 / 455407 = 23.326\%$
- 用户浏览购买同一商品的比例(相邻14天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-07, 2017-03-20]):
 $90194 / 455407 = 19.805\%$
- 用户浏览购买同一商品的比例(相邻7天vs2天)
 - ([2016-03-21, 2017-03-22]与[2017-03-14, 2017-03-20]):
 $76089 / 455407 = 16.7079\%$
- 用户浏览购买同一商品的比例(相邻5天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-16, 2017-03-20]):
 $69429 / 455407 = 15.2455\%$
- 用户浏览购买同一商品的比例(相邻2天vs2天)
 - ([2017-03-21, 2017-03-22]与[2017-03-19, 2017-03-20]):
 $51250 / 455407 = 11.2537\%$

浏览购买行为回归曲线



结论

- ▶ 可以看出，用户的重复行为在2周内呈现较大的发生率，超过2周的时间后重复行为曲线增长趋于平缓；
- ▶ 用户的短期兴趣在2周的时间范围内稳定性较高。
- ▶ 短期行为可以侧重于最近2周的行为。



The end