推荐系统无处不在

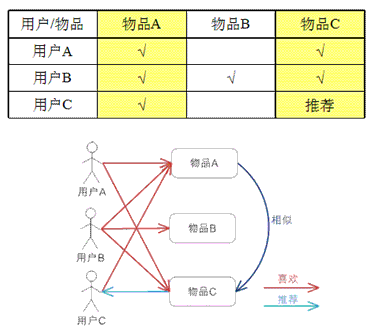
比如：

购买成功后：购买了该商品的其他用户购买了以下商品

搜索成功后：您可能感兴趣的以下商品

主页或广告：您可能感兴趣的以下商品

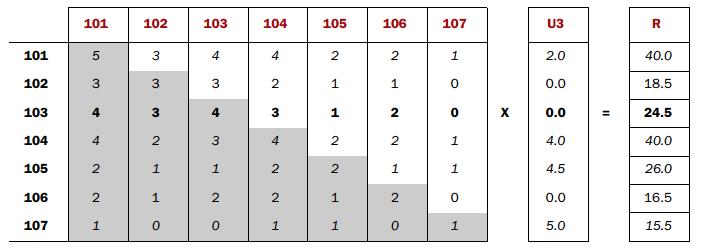
如下图，用户A购买物品AC;用户B购买物品ABC;当用户C购买物品A后，根据数据分析发现，购买物品A的用户大多数购买了物品C,所以提醒用户C”购买了该商品的其他用户购买了以下商品:**物品C**”。 下面我们就模拟实现这个功能



这里有一堆数据，记录的是用户购物记录；要求用户推荐和他之前喜欢的物品相似的物品。

|  |
| --- |
| item\_id,user\_id,action,vtime  i161,u2625,click,2014/9/18 15:03  i161,u2626,click,2014/9/23 22:40  i161,u2627,click,2014/9/25 19:09  i161,u2628,click,2014/9/28 21:35  i161,u2629,click,2014/9/27 16:33  i161,u2630,click,2014/9/5 18:45  i161,u2631,click,2014/9/29 16:57  i161,u2632,click,2014/9/24 21:58  i161,u2633,click,2014/9/25 22:41  i161,u2634,click,2014/9/16 13:30  i161,u2635,click,2014/9/20 9:23  i161,u2636,click,2014/9/21 1:00  i161,u2637,click,2014/9/24 22:51  i161,u2638,click,2014/9/27 22:40 |

**数据分析：**

****

101…107指商品分数 U3指用户评分

R指矩阵乘积,如 40=2\*(5+3+4+4+2+2+1)

R最高的代表推荐商品

**分析mapReduce处理过程**

第一次mapReduce做的事情？去重复数据；除空行；排序

第一次map

|  |  |  |
| --- | --- | --- |
| 入参 | i1,u2723,click,2014/9/1 18:31  i1,u2723,click,2014/9/14 9:31  i1,u2724,click,2014/9/18 6:33  i1,u2729,click,2014/9/19 6:48  i1,u2732,click,2014/9/17 1:00  i1,u2735,click,2014/9/3 16:23 | |
| 出参 | key | value |
| i1,u2723,click,2014/9/1 18:31  i1,u2723,click,2014/9/14 9:31  i1,u2724,click,2014/9/18 6:33  i1,u2729,click,2014/9/19 6:48  i1,u2732,click,2014/9/17 1:00  i1,u2735,click,2014/9/3 16:23 |  |
| 备注 |  | |

第一次reduce

|  |  |  |  |
| --- | --- | --- | --- |
| 入参 | key | value | |
| i1,u2723,click,2014/9/1 18:31  i1,u2723,click,2014/9/14 9:31  i1,u2724,click,2014/9/18 6:33  i1,u2729,click,2014/9/19 6:48  i1,u2732,click,2014/9/17 1:00  i1,u2735,click,2014/9/3 16:23 |  | |
| 出参 | key | | value |
| i1,u2723,click,2014/9/1 18:31  i1,u2723,click,2014/9/14 9:31  i1,u2724,click,2014/9/18 6:33  i1,u2729,click,2014/9/19 6:48  i1,u2732,click,2014/9/17 1:00  i1,u2735,click,2014/9/3 16:23 | |  |
| 备注 |  | | |

第二次mapReduce做的事情？

按用户分组，计算所有物品出现的组合列表，得到用户对物品的喜爱度得分矩阵

u13 i160:1,

u14 i25:1,i223:1,

u16 i252:1,

u21 i266:1,

u24 i64:1,i218:1,i185:1,

u26 i276:1,i201:1,i348:1,i321:1,i136:1,

第二次map

|  |  |  |
| --- | --- | --- |
| 入参 | i1,u2723,click,2014/9/1 18:31  i1,u2723,collect,2014/9/14 9:31 | |
| 出参 | key | value |
| u2723 u2723 | i1:1 i1:2 |
| 备注 | Value中：后的数字时指用户操作所对应的分数  Click打1分collect打2分cart打3分alipay打4分 | |

第二次reduce

|  |  |  |
| --- | --- | --- |
| 入参 | key | value |
| u2723 u2723 | i1:1 i1:2 |
| 出参 | key | value |
| u2723 | i1:1, i1:2, |
| 备注 |  | |

第三次mapReduce做的事情？

对物品组合列表进行计数，建立物品的同现矩阵

i100:i100 3

i100:i105 1

i100:i106 1

i100:i109 1

i100:i114 1

i100:i124 1

第三次map

|  |  |  |
| --- | --- | --- |
| 入参 | u2723 i1:1, i2:2,  u13 i160:1,i1:1 | |
| 出参 | key | value |
| i1:i160  i1:i1  i2:i160  i2:i1 i1:i160  i1:i1  i2:i160  i2:i1 | 1  1  2  2 2  2  2  2 |
| 备注 |  | |

第三次reduce

|  |  |  |
| --- | --- | --- |
| 入参 | key | value |
| i1:i160  i1:i1  i2:i160  i2:i1 i1:i160  i1:i1  i2:i160  i2:i1 | 1  1  2  2 2  2  2  2 |
| 出参 | key | value |
| i1:i160  i1:i1  i2:i160  i2:i1 | 3  3  4  4 |
| 备注 |  | |

第四次mapReduce做的事情？

把同现矩阵和得分矩阵相乘

第四次map 之同现矩阵

|  |  |  |
| --- | --- | --- |
| 入参 | i100:i100 3  i100:i105 1  i100:i106 1  i100:i109 1  i100:i114 1  i100:i124 1 | |
| 出参 | key | value |
| i100  i100  i100  i100  i100  i100 | A:i100,3  A:i105,1  A:i106,1  A:i109,1  A:i114,1  A:i124,1 |
| 备注 |  | |

第四次map 之得分矩阵

|  |  |  |
| --- | --- | --- |
| 入参 | u13 i160:1,  u14 i25:1,  u16 i252:1,  u21 i266:1, | |
| 出参 | key | value |
| u13  u14  u16  u21 | B:i160,1  B:i25:1  B:i252,1  B:i266,1 |
| 备注 |  | |

第四次reduce

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 入参 | key | | value | |
| i100  i100  i100  i100  i100  i100 | | A:i100,3  A:i105,1  A:i106,1  A:i109,1  A:i114,1  A:i124,1 | |
| map | mapA同现 | | mapB得分 | |
| key | value | key | value |
| i100 | 3 | u13 | 1 |
| i105 | 1 | u14 | 1 |
| 出参 | key | | value | |
| u13  u13  u14  u14 | | i100,3  i105,1  i100,3  i105,1 | |
| 备注 | u13. i100.出参.value= u13.mapB. value\* i100. mapA.value | | | |

第五次mapReduce做的事情？

把相乘之后的矩阵相加获得结果矩阵

第五次map

|  |  |  |
| --- | --- | --- |
| 入参 | u3010 i405,2.0  u3010 i405,3.0 | |
| 出参 | key | value |
| u3010  u3010 | i405,2.0  i405,3.0 |
| 备注 |  | |

第五次reduce

|  |  |  |
| --- | --- | --- |
| 入参 | key | value |
| u3010  u3010 | i405,2.0  i405,3.0 |
| 出参 | key | value |
| u3010 | i405,5.0 |
| 备注 |  | |

第六次mapReduce做的事情？

按照推荐得分降序排序，每个用户列出10个推荐物品

第六次map

|  |  |  |
| --- | --- | --- |
| 入参 | u13 i405,1.0  u13 i406,3.0 | |
| 出参 | key | value |
| u13:1  u13:1 | i405:1  i406:3 |
| 备注 |  | |

第六次reduce

|  |  |  |
| --- | --- | --- |
| 入参 | key | value |
| u13:1  u13:1 | i405:1  i406:3 |
| 出参 | key | value |
| u13 | i405:1,i406:3 |
| 备注 | 排序后取前10 | |