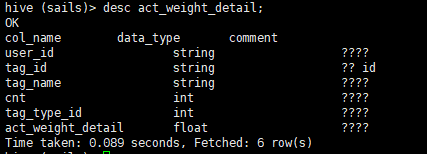
实验 4-2 用户行为画像分析

建议课时： 60 分钟

一、 实验目的  
λ 了解用户行为标签权重的计算方法；  
λ 了解用户的各标签表的关联合并；  
λ 了解用户行为权重标记对营销的意义；  
二、 实验环境  
Dsight 智慧实验室的 hadoop 环境  
三、 实验步骤  
本节实验所做内容如下红色标注：  
本节实验主要是对用户行为标签表加工生成用户行为权重表， 并将该表与用  
户属性标签表进行关联形成宽表供后续分析使用。  
具体实验步骤如下：  
1. 用户行为权重值的计算  
1.1 行为权重介绍  
一个用户标签表里面包括常见的字段如： 用户 id、 用户姓名、 标签  
id、 标签名称、 用户与该标签发生行为的次数（如搜索了两次“大数据”  
这个关键词） 、 行为类型（不同的行为类型对应用户对商品不同的意愿  
强度， 如购买某商品>收藏某商品>关注某商品>点击某商品） ， 行为时  
间（越久远的时间对用户当前的影响越小， 如 5 年前你会点击某三星手  
机， 而现在你会点击华为或苹果） 。 最后非常重要的一个字段是标签权  
重， 该权重影响着对用户属性的归类， 属性归类不准确， 接下来基于画  
像对用户进行推荐、 营销的准确性也就无从谈起了。  
1.2 权重的计算  
用户标签权重 = 行为类型权重 \* 时间衰减 \* TF-IDF 计算得到每个用  
户身上的标签权重 \* 行为次数  
ϖ 行为类型权重： 一般而言操作复杂度越高的行为权重越大。 该权重  
值一般由运营人员或数据分析人员主观给出；  
自定义购买权重为 5， 加入购物车权重为 4， 关注权重为 3， 点击权  
重为 2  
ϖ 时间衰减： 时间衰减是指用户的行为会随着时间的过去， 历史行为  
和当前的相关性不断减弱。  
套用牛顿冷却定律数学模型：  
F(t)=初始温度×exp(-冷却系数×间隔的时间)"冷却系数"是一个你自己决定的值。如果假定手机的的初始点击次数  
是 5 次， 24 小时之后"冷却"为 1 次， 那么可以计算得到"冷却系数  
"约等于0.067。 如果你想放 慢更新率， "冷却系数"就取一个较小  
的值， 否则就取一个较大的值。  
ϖ TF-IDF 计算标签权重：  
根据 TF \* IDF 即可得到该用户该标签的权重值

2. 新建用户行为权重表并加载数据  
在用户行为标签表的基础上， 新建用户行为权重表用户行为标签权重表结构设计：

drop table if exists dwd act\_weight\_detail;  
create table dwd act\_weight\_detail  
( u  
ser\_id string comment '用户编码',  
tag\_id string comment '标签 id',  
tag\_name string comment '标签名称',  
cnt int comment '行为次数',  
tag\_type\_id int comment '标签类型',  
act\_weight\_detail float comment '行为权重'  
) comment '用户画像-用户行为标签权重表';

  
加载数据： 同步用户行为标签表中的数据以及添加权重值到该表中。

insert into table act\_weight\_detail

select user\_id,tag\_id,tag\_name,action\_count,tag\_type,

case

when tag\_id="B211U001\_001" then 2\*action\_count\*100\*exp(0.067)

when tag\_id="B211U001\_002" then 3\*action\_count\*100\*exp(0.067)

when tag\_id="B211U001\_003" then 4\*action\_count\*100\*exp(0.067)

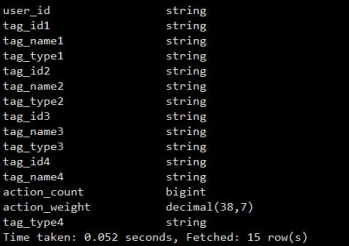
when tag\_id="B211U001\_004" then 5\*action\_count\*100\*exp(0.067)

end

from person\_user\_tag\_action

limit 5

3. 生成宽表  
根据用户 id 关联用户所有属性标签表和用户行为权重表生成新的宽表并  
加载数据  
表结构设计：  
drop table if exists profile\_user\_tb;  
create table profile\_user\_tb  
( user\_id string comment '用户编码',  
tag\_id1 string comment '标签 1 ID',  
tag\_name1 string comment '标签 1 名称',  
tag\_type1 string comment '标签 1 类型',  
tag\_id2 string comment '标签 2 ID',  
tag\_name2 string comment '标签 2 名称',  
tag\_type2 string comment '标签 2 类型',  
tag\_id3 float comment '标签 3 ID',  
tag\_name3 string comment '标签 3 名称',  
tag\_type3 string comment '标签 3 类型',  
tag\_id4 string comment '标签 4 ID',  
tag\_name4 string comment '标签 4 名称',  
action\_count bigint comment '标签 4 行为次数',  
action\_weight decimal(38,7) comment '标签 4 权重',  
tag\_type4 string comment '标签 4 类型'  
) comment '用户画像-用户标签宽表';



查询语句：

