

推荐系统常用特征

特征选取是每一个算法工程师都需要认真对待的事情，如果特征选取的得当往往能起到事半功倍的效果。

对于推荐系统而言，所需的特征总体而言可以分为用户特征和物品特征。

在这里我会和同学们一起总结一下推荐系统常用的特征，希望能够给同学们提供参考。

一、用户特征

用户是所有产品的根基，准确地抓住用户特征往往能够使得我们在后续的分析、建模过程中更加得心应手。

对于用户特征我想可以从以下三点着手分析：

1. 自然属性特征：

自然属性特征包括用户的姓名、性别、年龄、地域等特征。这一类特征是用户自身的固有属性，一旦获取到就不会发生变化。

在推荐系统中我们最常用的自然属性特征就是用户的性别和年龄。我们都知道不同性别、不同年龄的人所喜好的事物是明显不同的。而且有时候往往根据目标用户的性别和年龄，就足以给他们进行基础的推荐了。以电影推荐为例：年轻的男性可能更适合推荐科幻动作片，而对于年龄较大的女性则更有可能喜欢文艺片。

因此同学们在搭建推荐系统的时候，可以从产品的角度去考虑一下如何能够收集到更多的用户自然属性特征，相信一定会有所帮助。

2. 画像特征：

画像特征描述了用户的兴趣点。具体而言，就是通过用户的行为等数据，推断出的用户对于系统内物品的喜好。

这类特征是推荐系统得以做到“千人千面”的重点。因为即使是相同年龄、相同性别的两个人也很有可能有着完全不同的喜好，那么是否能准确的捕获到这些喜好就至关重要。

更具体而言，画像特征可以分为两类：隐式反馈和显式反馈。显式反馈顾名思义就是用户明确表达出对一个物品的喜好，比如对电影的打分，或者对商品的评价。而相反的，隐式反馈则是由用户的行为间接推断出他的喜好，比如用户在一个页面的停留时间，是否点击了某个内容等等。

3. 关系特征：

关系特征即用户的人群属性。比如关注关系、好友关系等等都可以作为推荐系统的特征之一。良好的运用关系特征可以有效的帮我们解决诸如冷启动等问题，同时也便于发觉用户的隐藏喜好。

二、物品特征

和用户特征类似的，物品的特征也可以分为如下几类：

1. 静态特征：

静态特征包括诸如物品的分类、价格、标签等信息。类似于用户的自然属性，这类特征往往很少会发生变动。如果产品和运营完善的话，物品的静态特征是非常容易收集的，因此大家务必要对其善加利用。

2. 动态特征：

动态特征主要包括物品的热度等信息。这类特征会随时间变化，此外还可以根据不同的需要对各个维度、层面上的热度信息进行聚合。比如对于商品而言，每个sku可以有自己热度，而对于某个品牌、某个店铺，又可以聚合它下面的所有商品从而得出品牌或者店铺的热度。因此在具体使用的时候需要大家针对不同场景对动态特征进行取舍。

3. 相关性特征：

相关性特征指的是物品之间的相似（匹配）程度。和用户不同，对于绝大部分产品而言都可以由运营团队对物品进行打标。比如大部分电商平台都会给服装打上款式、季节、品牌等等的标签。有了这些标签我们就可以很轻松的计算两个物品之间的相似度并将其用在推荐系统当中。

以上我们总结了推荐系统当中用户和物品各自常用的几种特征，在后续的课程中我们会专门针对这些特征进行处理和分析。同学们在实际工作中也可以参考以上内容，审视一下自己的特征选择是否得当。