

推荐系统中的冷启动问题

原创 小军 算法猿的日常 2020-09-16

收录于话题

#推荐系统

3个

推荐系统中有两个重要的问题，**EE**和冷启动，上一篇文章中我们介绍了**EE**，本篇文章给大家介绍一下推荐系统中冷启动(cold start)问题。

冷启动的定义

什么是冷启动？首先我们要理解推荐系统的本质，不管是信息流、电商、短视频本质上我们是基于用户行为建模，通过用户或者物品的行为建立联系，从而相应推荐合适的用户或者物品。那么，冷启动即是在没有大量用户或者物品行为下，如何给新用户/新物品推荐相应的item/user。这就是冷启动。

冷启动的分类

冷启动问题主要分为以下三类：

- 1)用户冷启动:如何给新用户推荐他感兴趣的物品。
- 2)物品冷启动:如何将新的物品推荐给可能对它感兴趣的用户。
- 3)系统冷启动:如何给一个新的系统(新的app、网站，没有用户和用户行为)设计个性化推荐系统。

下面主要来讲讲冷启动问题的解决办法。

冷启动的解决办法

1、热门推荐

对于新用户来说，一般最常见的做法就是提供非个性化的方式去探索用户兴趣，即热门推荐，把最近一周或者一个月的热门推荐给新用户，事实上很多公司比如Netflix做过用户冷启动相关的研究，得出的结论就是，新用户刚开始并没有明确的喜好，更多的是参考大多数人的行为，长时间的用户行为积累之后才会产生个性化的兴趣。即新用户更倾向于热门，老用户则更偏长尾。

2、用户注册信息

当新用户新注册一个app时，我们不知道他喜欢什么物品，于是只能给他推荐一些热门的商品。但如果我们知道她是一位女性，那么可以给她推荐女性都喜欢的热门商品。这也是一种个性化的推荐。当然这个个性化的粒度很粗，因为所有刚注册的女性看到的都是同样的结果，但相对于不区分男女的方式，这种推荐的精度已经大大提高了。因此，利用用户的注册信息可以很好地解决注册用户的冷启动问题。

用户的注册信息主要分为3种：

- 1)人口统计学信息，包括年龄、性别、职业、民族、学历和居住地。
- 2)用户兴趣的描述，部分网站会让用户用文字来描述兴趣。
- 3)从其他网站导入的用户站外行为，比如用户利用社交网站账号登录，就可以在获得用户授权的情况下导入用户在该社交网站的部分行为数据和社交网络数据。

人口统计学特征包括年龄、性别、工作、学历、居住地、国籍、民族等，这些特征对预测用户的兴趣有很重要的作用，比如男性和女性的兴趣不同，不同年龄的人兴趣也不同。

Register for free!

Create an account to save your stations and access Pandora anywhere.

Your email

Password

Birth year

Why?

US zip code

US residents only

Gender

☐ Male

☐ Female

Why?

☒ Send me personalized recommendations and tips.

☐ I have read, understood, and agreed to the Terms of Use and the Privacy Policy

Register

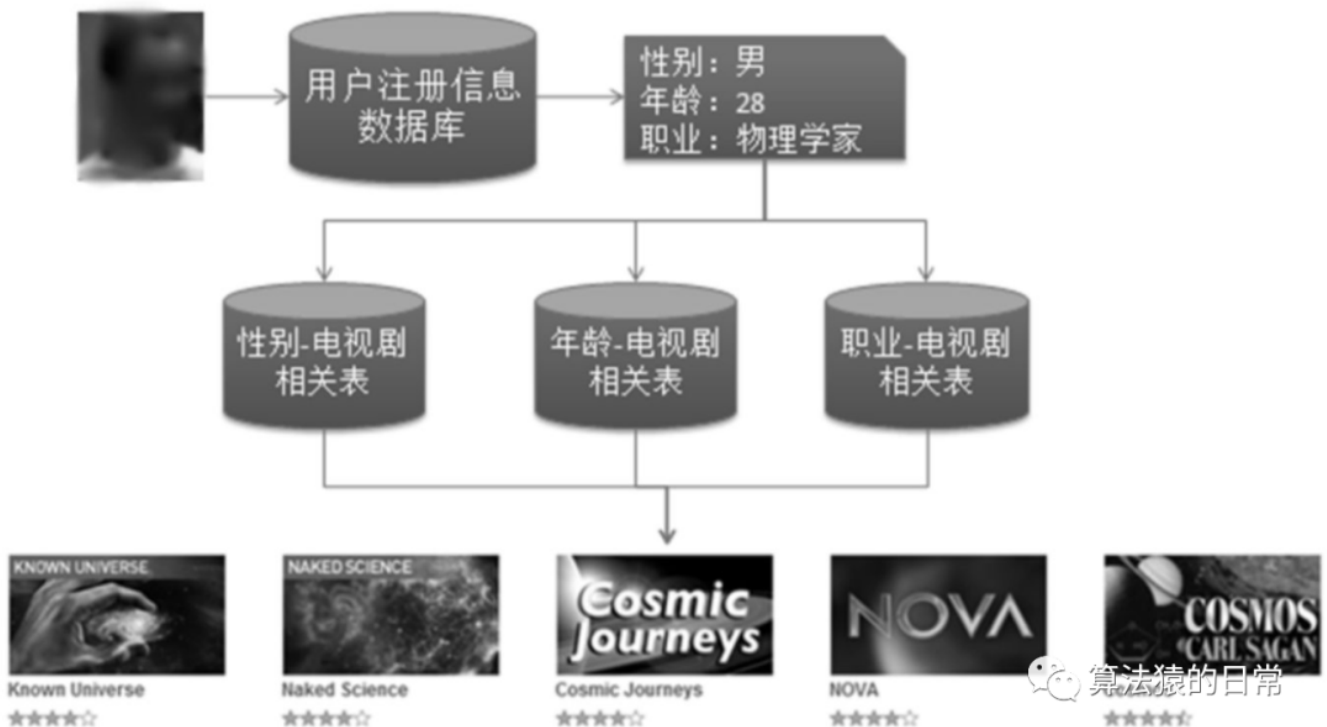
Go back

Already registered? Sign in.

算法猿的日常

基于用户注册信息的个性化推荐流程基本如下：

- 获取用户的注册信息
- 根据用户的注册信息对用户分类
- 给用户推荐他所属分类中用户喜欢的物品



当一个新的注册用户访问推荐系统时，我们首先从用户注册信息数据库中查询他的注册信息。比如上图中的用户，我们查到他是一位28岁的男性，是一位物理学家。然后，查询3张离线计算好的相关表：

- 1) 性别-电视剧相关表，从中可以查询男性最喜欢的电视剧；
- 2) 年龄-电视剧相关表，从中可以查询到28岁用户最喜欢的电视剧；
- 3) 职业-电视剧相关表，可以查询到物理学家最喜欢的电视剧。

我们可以将用这3张相关表查询出的电视剧列表按照一定权重相加，得到给用户的最终推荐列表。基于用户注册信息的推荐算法其核心问题是计算每种特征的用户喜欢的物品。也就是说，对于每种特征 f ，计算具有这种特征的用户对各个物品的喜好程度 $p(f,i)$ 。

3、利用用户反馈

用户在登录时对一些物品进行反馈，收集用户对这些物品的兴趣信息，然后给用户推荐那些和这些物品相似的物品。

一般来说，能够用来启动用户兴趣的物品需要具有以下特点：

- 比较热门，如果要让用户对物品进行反馈，前提是用户得知道这是什么东西。
- 具有代表性和区分性，启动用户兴趣的物品不能是大众化或老少咸宜的，因为这样的物品对用户的兴趣没有区分性。
- 启动物品集合需要有多多样性，在冷启动时，我们不知道用户的兴趣，而用户兴趣的可能性非常多，为了匹配多样的兴趣，我们需要提供具有很高覆盖率的启动物品集合，这些物品能覆盖几乎所有主流的用户兴趣。

这种类型的冷启动也非常常见，比如用户初次登陆某一个音乐app时，界面开始会让你选择你喜欢的音乐类型，比如rock、乡村音乐等。

4、利用物品的内容信息

主要用来解决物品的冷启动问题，即如何将新加入的物品推荐给对它感兴趣的用戶。物品冷启动问题在新闻网站等时效性很强的网站中非常重要，因为这些网站时时刻刻都有新物品加入，而且每个物品必须能够在第一时间展现给用户，否则经过一段时间后，物品的价值就大大降低了。针对协同过滤的两种推荐算法——userCF算法、itemCF算法来分别了解一下物品冷启动的问题。

UserCF对物品冷启动问题并不非常敏感，因为，**UserCF**在给用户进行推荐时，会首先找到和用户兴趣相似的一群用户，然后给用户推荐这一群用户喜欢的物品。在很多网站中，推荐列表并不是给用户展示内容的唯一列表，那么当一个新物品加入时，总会有用户从某些途径看到这些物品，对这些物品产生反馈。那么，当一个用户对某个物品产生反馈后，和他历史兴趣相似的其他用户的推荐列表中就有可能出现这一物品，从而更多的人就会对这个物品产生反馈，导致更多的人的推荐列表中会出现这一物品，因此该物品就能不断地扩散开来，从而逐步展示到对它感兴趣用户的推荐列表中。有些网站中推荐列表可能是用户获取信息的主要途径，比如豆瓣网络电台。那么对于**UserCF**算法就需要解决第一推动力的问题，即第一个用户从哪儿发现新的物品。只要有一小部分人能够发现并喜欢新的物品，**UserCF**算法就能将这些物品扩散到更多的用户中。解决第一推动力最简单的方法是将新的物品随机展示给用户，但这样显然不太个性化，因此可以考虑利用物品的内容信息，将新物品先投放给曾经喜欢过和它内容相似的其他物品的用户。

对于**ItemCF**算法来说，物品冷启动就是一个严重的问题了。因为**ItemCF**算法的原理是给用户推荐和他之前喜欢的物品相似的物品。**ItemCF**算法会每隔一段时间利用用户行为计算物品相似度表(一般一天计算一次)，在线服务时**ItemCF**算法会将之前计算好的物品相关度矩阵放在内存中。因此，当新物品加入时，内存中的物品相关表中不会存在这个物品，从而**ItemCF**算法无法

推荐新的物品。解决这一问题的办法是频繁更新物品相似度表，但基于用户行为计算物品相似度是非常耗时的事情，主要原因是用户行为日志非常庞大。而且，新物品如果不展示给用户，用户就无法对它产生行为，通过行为日志计算是计算不出包含新物品的相关矩阵的。为此，我们只能利用物品的内容信息计算物品相关表，并且频繁地更新相关表(比如半小时计算一次)。

5、利用用户的手机等兴趣偏好进行冷启动

用户安装的手机app侧面说明了用户的兴趣或者状态，比如某一个用户安装了虎牙、斗鱼等直播app是不是说明该用户可能喜欢游戏相关的内容。如果一个用户安装了宝宝树、大姨妈相关的app，是不是也可以说明该用户大概率是女性，可能还是一个孕妈等等。因此通过用户的手机的兴趣偏好也可以进行用户冷启动。

6、专家标注

很多推荐系统在建立时，既没有用户的行为数据，也没有充足的物品内容信息来计算准确的物品相似度。那么，为了在推荐系统建立时就让用户得到比较好的体验，很多系统都利用专家进行标注。这方面的代表系统是个性化网络电台Pandora和电影推荐网站Jinni。

Pandora是一个给用户播放音乐的个性化电台应用。众所周知，计算音乐之间的相似度是比较困难的。首先，音乐是多媒体，如果从音频分析入手计算歌曲之间的相似度，则技术门槛很高，而且也很难计算得令人满意。其次，仅仅利用歌曲的专辑、歌手等属性信息很难获得令人满意的歌曲相似度表，因为一名歌手、一部专辑往往只有一两首好歌。为了解决这个问题，Pandora雇用了一批懂计算机的音乐人进行了一项称为音乐基因的项目。他们听了几万名歌手的歌，并对这些歌的各个维度进行标注。

7、基于深度学习的冷启动

基于深度学习的冷启动主要集中在物品冷启动上，主要的方式是对于新的物品，通过一些side information的embedding代替item embedding，在airbnb 2018年KDD best paper中，对于新的listing，因为host在上传每个listing时要上传一些meta-data，例如位置，价格，类型等，那么我们可以找到这个新listing的三个距离最近的，有相似价格的、房屋类型的listing，用这三个listing的embedding平均值作为listing的冷启动。

好了，关于冷启动的内容介绍完了，感兴趣或者有疑问的朋友可以在下方留言。

-END-

喜欢推荐广告的朋友可以三连+关注，一起学习推荐系统，成为一个优秀的人。