



## 扒一扒京东推荐系统里的商品知识图谱 | 知识整理



道比伯尔

同济大学 工学硕士

关注他

10 人赞同了该文章

全文近2500字，阅读时长应该很快吧。

电商推荐系统都是分成捕捉用户兴趣、满足用户兴趣和用户兴趣的拓展，其中用户兴趣的拓展相对他类型的推荐系统会更重要。因为只有不断探索用户对商品的兴趣并形成购买，才能打破用户整体价值的上限、打破电商系统的成长空间天花板。

### 电商推荐系统的特点

用户的角度

- 跟资讯类的、信息流的推荐系统相比，电商推荐系统的用户行为相对少
- 资讯类的、信息流的系统，用户消费的内容就在线上，系统可以完整地记录到这个用户观看或消费平台内容的过程；而对电商系统来说，它只是交易的过程产生在线上，事前的考虑和事后的使

## 商品的角度

- 商品的规模特别大，电商系统它后台真正的候选级物品都是数十亿SKU的候选，也是因为现在这个时代，已经进入了一个供过于求的状态，所以它更依赖于这些中长尾的商品，更依赖于推荐系统，去找到合适的用户
- 大量的热点是随着一些节日或者平台自己做的一些临时性的活动带来的用户，那么从逻辑上来说这些商品其实它们的前续的用户行为一段时间内也是缺乏的

总结，第一、兴趣拓展对于电商推荐系统来说特别重要。第二、传统的协同过滤，依赖于用户行为的这种方式是不够的。

- 推荐系统围绕用户兴趣的挖掘和扩展 (EE)
- 用户兴趣拓展能力决定推荐系统上限



## 京东推荐系统的创新点

推荐系统的工作流程有三个环节，第一个环节是要从刚刚所说的数十亿的SKU集合里，召回出来大概数千级别的候选SKU。第二个环节是要对召回出来的数千级别的SKU去做CTR，去做点击率或者转化率的这个预估，就是对每一个后选的商品，跟用户打一个预估的CTR的分，或者预估的转化率的分。第三个环节是拿到打完分的这一千个候选之后，怎么推出来最后返回给用户看到的那几条，或者几十条的候选的过程。

那么最后 呈现给用户的商品包括两部分：第一部分 其实已经知道它对这些商品是感兴趣的

- 1. 召回：基于商品知识图谱的兴趣召回
- 2. CTR预估：基于维度关系建模的泛化兴趣排序
- 3. 重排：基于Listwise全局建模的兴趣重排

我们这里只讨论“基于商品知识图谱的兴趣召回”。



基于商品知识图谱的兴趣召回

做商品知识图谱是因为原先的系统里存在大量的用户行为缺失，商品也缺少对应的行为。它依赖于传统行为的这种方法，不管是CF也好，还是基于用户行为去训练各种向量模型，或者其他各种模型化的召回方法，在行为缺失的时候都失效了。

造成的原因包括但不限于：

新品上架：新品上来的时候，其实是完全没有用户行为的，这个时候如果依赖于行为的模式去召回，都会面临商品本来很热，但几乎没有用户行为，推荐系统是推不出去的

时效活动：参与活动的商品，也是突然地从一个常规的销售要变成一个突然热门的销售，去依赖于传统的行为也是推不出来的

时效热点：今年的疫情期间，比如说口罩、方便面、洗手液这些东西，正常的情况来看它们其实不会建立关联的，但是因为疫情，它们之间建立了关联关系

精准搭配：正常来说用户买完手机，推手机壳这是没有问题的，但是如果在电商场景时候，你推的这两个东西不是搭配的，比如说你推的是P40手机，你推出来是某个iPhone的保护壳，这对用户的



用户做兴趣的拓展，而且要保证你拓展出来的东西是精准的，是满足用户需求的。

- 工作背景
- 电商场景，新品多，活动多，活动周期短，传统基于用户行为建模方法无法覆盖
  - 基于商品知识图谱有效建模新品，活动，并基于商品关系完成精准搭配推荐，捕获时效热点



这个商品知识图谱的构建是多层的，包括商品SKU层、实体层、概念层和产品分类层。

最底下这层是基础的商品，是京东场景下数十亿的SKU的这些基础的商品，各个商品都是各异的。

再上一层，是我们需要算法能够抽取出来的，就是最主要是通过商品的一些的标题、描述，以及一些商详页一些属性，甚至包括用户评论的一些属性，一些信息，去抽取出来上一层的实体。这个实体的概念其实主要是在用户感知的概念里头，它是唯一的一个商品的集合。所以它的基础过程，第一步是基于标题、上下文的文本描述抽取出的这个过程，第二步是基于一些用户行为所做的一些具促之类或者是说一些数据清洗之类的逻辑去做的一个实体。

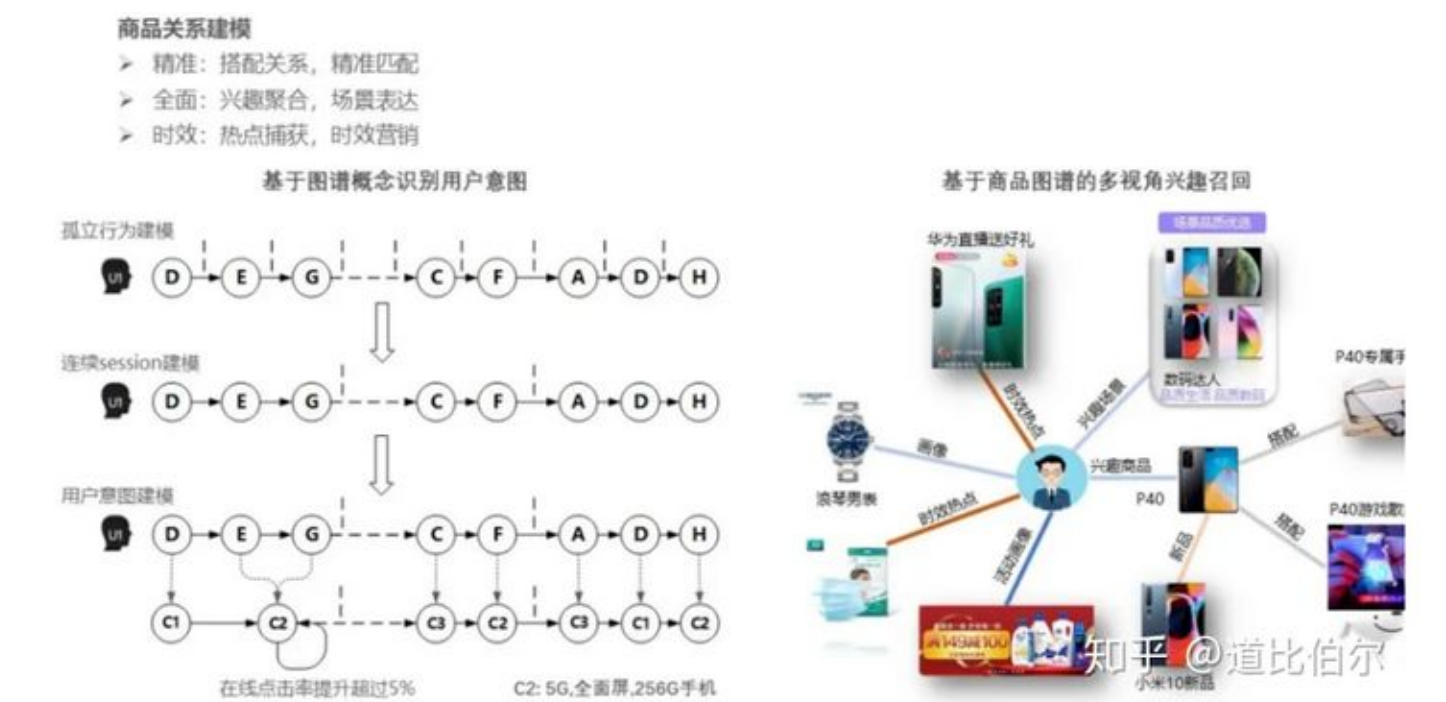
实体再上面一层是概念层，这个概念层体现出来的也是一些用户关注的商品的属性，比如说这里举个手机的例子来说，那可能它关注这个手机的主打的功能是不是拍照，是不是全面屏，是不是带5G，也就是类似于这种用户特别关注的某一个属性，叫概念层。

再上面一层才是产品分类，就是手机。



所以其实可以看到这是一个从细到粗、逐步抽象出来的知识网络的结构。所以它结合了一些商品抽取的信息, 也结合了一些用户的行为, 那么有了这个基础的商品层次的结构之后, 我们很多的关系构建就不需要基于最底下这层, 因为最底下这层的量级是特别大的, 数十亿级别的, 在这个级别下面的东西其实没有泛化能力, 如果说你想去做基于行为, 或者基于其他方式去构建, 构建出来的数据都会是特别稀疏的, 所以更多的关系是基于上面那三层, 尤其是中间那两层构建的一些商品的信息。

而且用户跟概念层某个实体形成关联之后, 我们可以基于知识图谱关联出更多实体, 因此对SKU进行召回。



构建商品知识图谱除了解决数据稀疏的问题以外, 还有助于把看似孤立的用户行为变成一个连续的

在三种不同的概念间（C1/C2/C3）跳转而已，而且是围绕C2（5G/全面屏/256G手机）这个概念在跳转，那么整个过程就等于把用户的行为降维了，基于下面这个链条去做建模，去做用户兴趣的刻画和跳转，那么召回来的东西就可以基于正常的建模的方式去做操作了。不管你是 CF 的方法也好，还是其他的模型化的方法也好，都是能够工作的。

这就是在召回方面京东所做的一个创新，其实最主要的工作就是知识图谱的构建。

\*\*\*本文是对网络上京东推荐系统知识的整理并加上了部分个人的理解，欢迎讨论\*\*\*

整理创作不易，点赞关注您觉得如何？

编辑于 2020-10-19

「真诚赞赏，手留余香」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

推荐系统    知识图谱    京东

文章被以下专栏收录

- 

道哥叨科技

关于大数据、AI及其他前沿科技

关注专栏
- 

知识图谱

知识图谱的相关知识收集与分享

关注专栏