

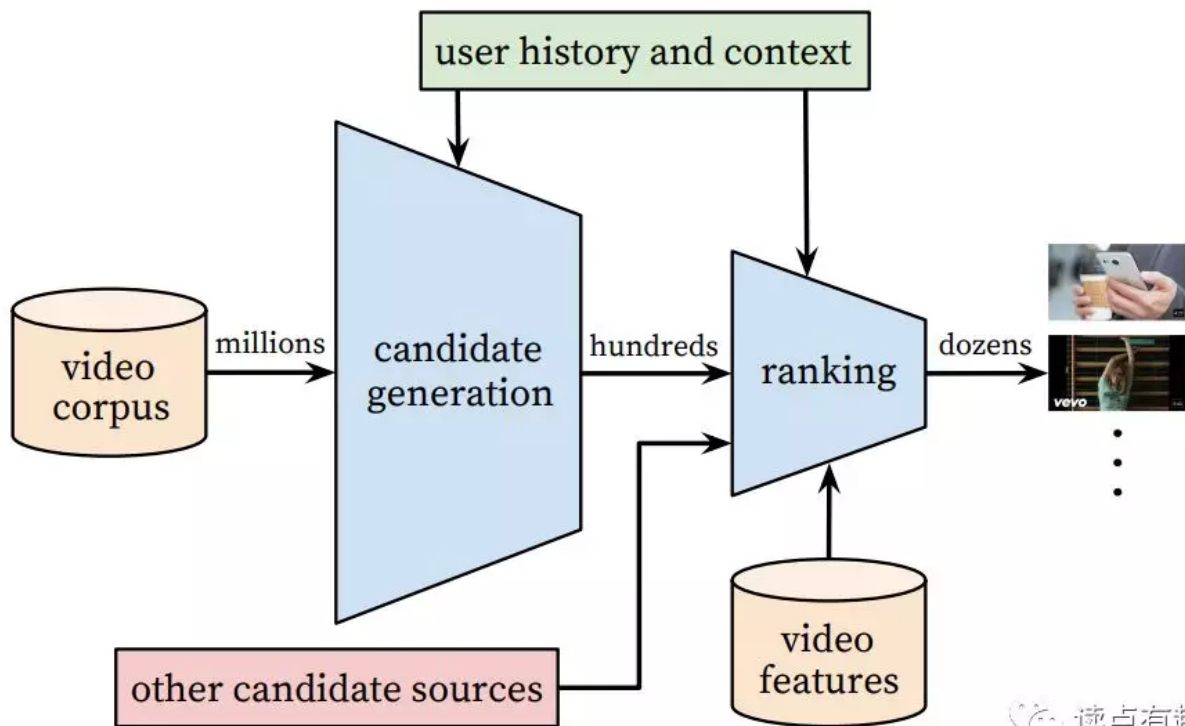
推荐系统落地 101：召回怎么做？

原创 yyzhang 读点有趣的 2018-11-07

在对物品进行推荐时，如果出现物品特别多（例如有很多的文章、很多的商品），在用户访问页面无法在有限的时间内，加载物品和用户的全部特征用算法进行排序，现在业界通用的做法会先进行召回，减小需要排序的物品量，再用算法进行排序（精排），达到个性化推荐的目的。物品特别少的情况下，不需要做召回的操作，直接对所有物品排序。召回看起来非常非常简单，但是召回却非常重要：所谓巧妇难为无米之炊，召回结果的好坏直接决定了后面精排算法效果的下限，召回做的好，后面精排的效果也不会太差。并且相比优化精排算法、建立新的数据模型、或者开发实验新的特征，优化召回操作简单、方便实验，有四两拨千斤的效果。

物品特别多的情况下，业界推荐架构举例：

Youtube：

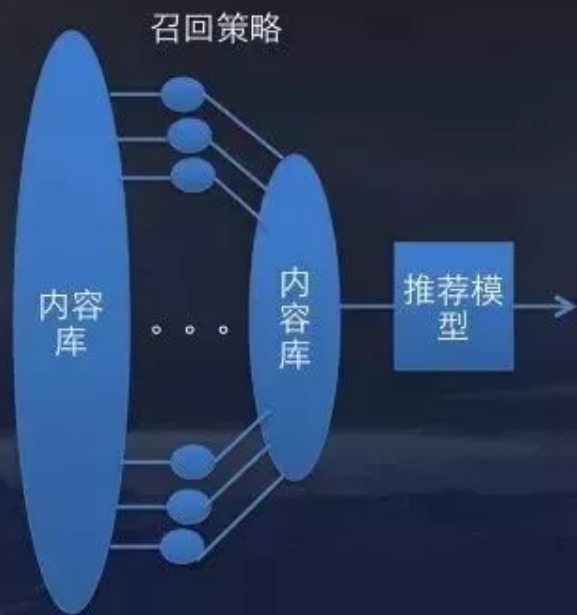


读点有趣的

今日头条：

召回策略设计

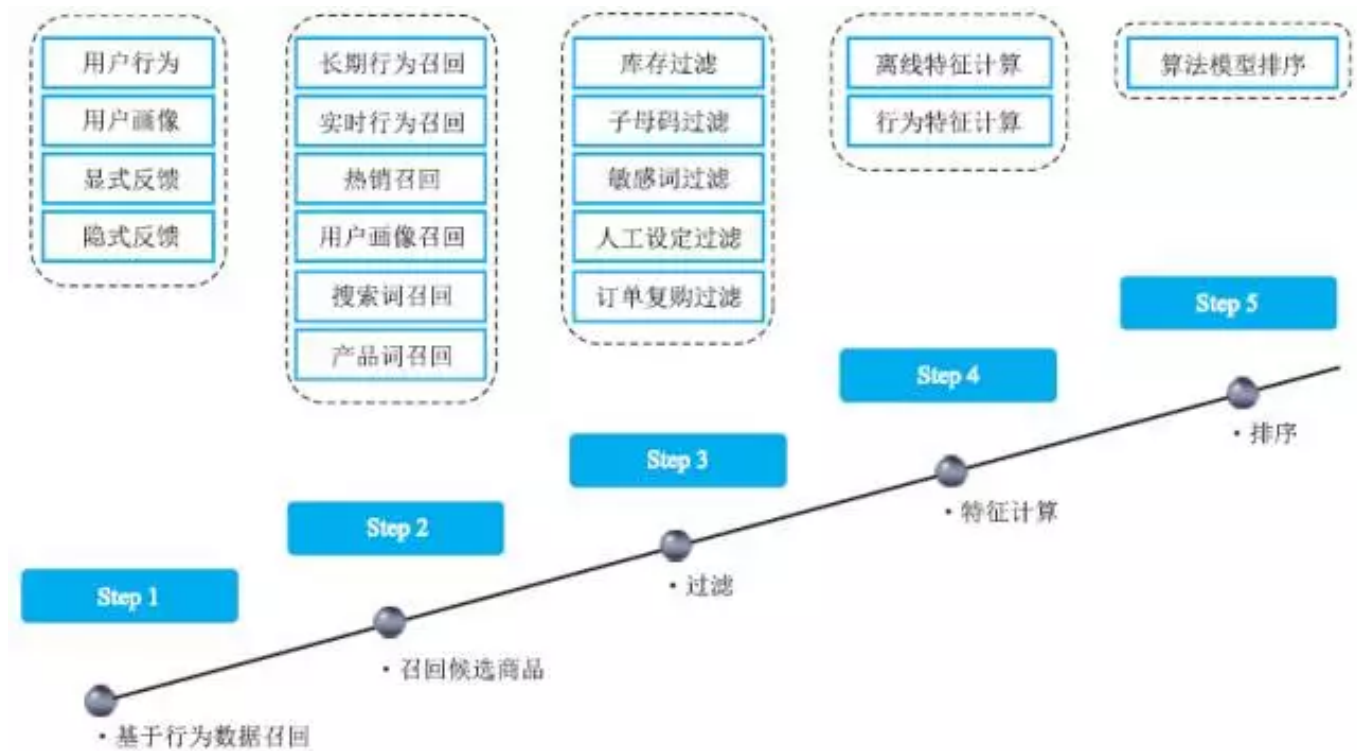
- 推荐模型计算开销相对较大，完全依赖模型推荐成本过高
- 基于简化策略的召回模块可以有效平衡计算成本和效果



典型召回策略架构



京东：



常见的召回分非个性化的召回、个性化召回。

。非个性化召回

最常见是热度召回，也就是把热门的内容召回，例如前面京东“热销召回”就属于热度召回。

也可以做一些运营召回。在笔者以往的经验中，推荐算法同学多和人工运营同学交流，运营同学指导一些内容风格，算法同学针对这些内容风格进行内容挖掘后加入到召回，对推荐效果指标的提升也非常显著。

有些被推荐的内容跟时间段敏感，例如适合早晨观看/阅读/收听的内容、突然爆发的时效性热点内容，也可以针对时间进行召回。只是这个做法不是常规做法，有些推荐系统把有时间段的内容做在推荐系统规则层进行过滤，例如现在是晚上，就在推荐时把召回的适合早晨的内容过滤掉；或者对热点内容进行投放。

。个性化召回

最常见的个性化召回是“标签定向”。这里“标签”可能是实体意义上的“标签”，或者说抽象的隐式向量，甚至一些人为定义划分的“标签”。例如今日头条用了用户兴趣标签、京东用了用户行为标签，以及用户的一些 profile 兴趣，这些属于常见的实体意义上的标签。

抽象的隐式标签，常见的用 LDA、word2vec 等算法把用户和 item 都映射成向量的方式进行召回。隐式标签的召回好处是，它有一些相似度的扩展，例如喜欢刘德华的用户，召回结果里也能返回一些周华健的结果（举例子暴露年龄了，逃...），坏处是，这种召回结果难以控制，喜欢看章子怡娱乐新闻的用户可能召回了范冰冰的结果，但是这个用户很反感范冰冰。

人为定义划分的“标签”召回，例如用RFM模型对用户进行分群，再利用分群的结果进行召回。在笔者的经验中，加入RFM模型召回的结果能有一个点左右的提升。

另外，还有一些不走寻常路的召回，例如 Youtube 用的是深度学习模型从百万的候选视频中生成几百个召回的结果。

总的说来，推荐系统是一个系统性的工程问题，细致繁杂，从用户行为产生的数据行为到机器计算推出，每个环节掉一点链子都会直接影响推荐的效果。例如，数据上报丢数据、重复上报数据，这并不会影响用户使用产品，却实实在在的影响了推荐算法计算打分的结果。正如传统互联网产品，交互、视觉设计、产品流程设计、系统响应速度、测试完备性、发布平台、安全等等，每一环都有可能影响最终产品的效果一样，在做推荐的过程中，只能尽最大可能保证每一步的效果都是最好的。

引用：

《干货 | 3 分钟了解今日头条推荐算法原理（附视频 + PPT）》

《深度解析京东个性化推荐系统演进史》

P Covington , J Adams , E Sargin. Deep Neural Networks for YouTube Recommendations. Acm Conference on Recommender Systems , 2016 :191-198