12段话，了解今日头条算法的秘密

今日头条
12小时前 · 今日头条官方头条号

▲资深算法架构师曹欢欢讲解今日头条算法原理

1月11日，一场问诊算法、建言算法的 “让算法公开透明” 分享交流，在北京今日头条总部举办。资深算法架构师、中国科学技术大学曹欢欢博士，介绍了今日头条的推荐算法原理。同时，解答大家对算法的疑问，接受大家对算法的建议。

中央电视台、新华社、人民日报等媒体机构从业者，和阿里、腾讯、百度、美团、新浪、网易等科技公司的算法工程师、产品经理等100多人，参加了活动。

1

资讯推荐系统本质上要解决用户、环境和资讯的匹配。

今日头条算法推荐系统，主要输入三个维度的变量。

一是内容特征，图文、视频、UGC小视频、问答、微头条等，每种内容有很多自己的特征，需要分别提取。二是用户特征，包括兴趣标签、职业、年龄、性别、机型等，以及很多模型刻画出的用户隐藏兴趣。三是环境特征，不同的时间不同的地点不同的场景（工作/通勤/旅游等），用户对信息的偏好有所不同。结合这三方面纬度，今日头条的推荐模型做预估，这个内容在这个场景下对这个用户是否合适。

2

点击率、阅读时间、点赞、评论、转发，这些都是可以量化的。但一个大体量的推荐系统，服务用户众多，不能完全由指标评估，引入数据以外的要素，也很重要。有些算法可以完成，有些算法还做不到、做的不好，这就需要内容干预。

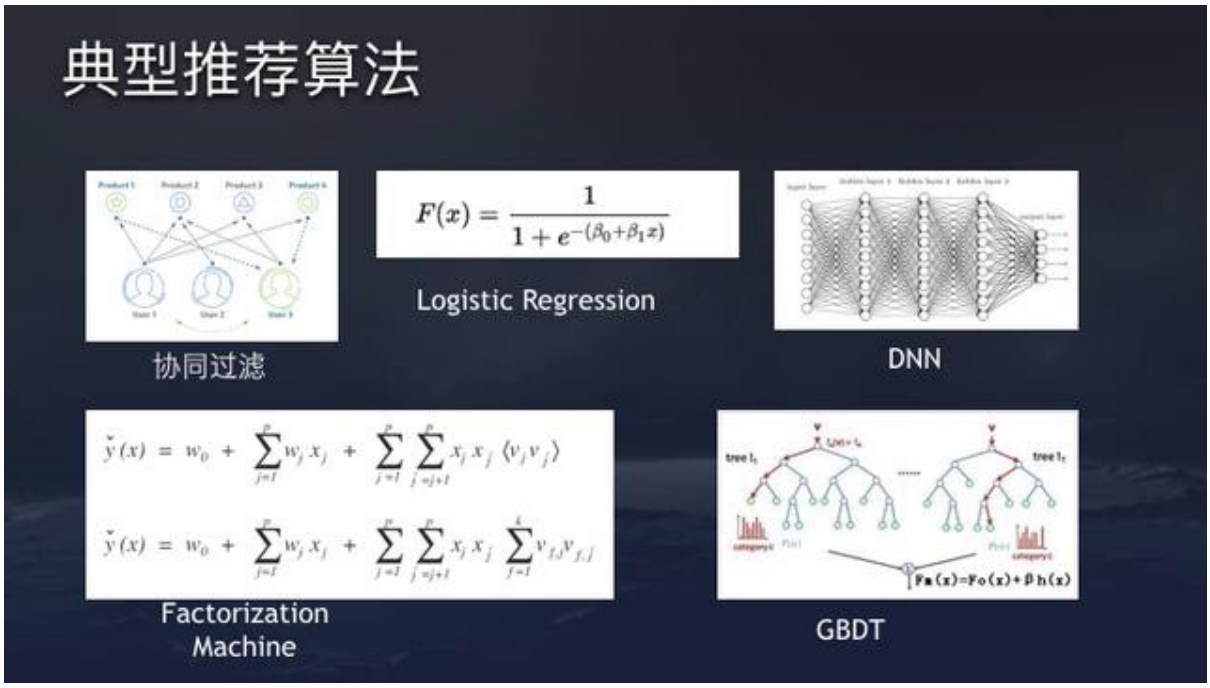
3

没有一套通用的模型架构，适用所有的推荐场景。我们需要一个非常灵活的算法实验平台，这个算法不行，马上试另一个算法，实际上是各种算法的一个复杂组合。西瓜视频、火山小视频、抖音短视频、悟空问答，都在用头条这一套推荐系统，但具体到每套系统，架构都不一样，需要不断去试。

4

算法推荐要达到不错的效果，需要解决好这四类特征：相关性特征、环境特征、热度特征和协同特征。

相关性特征，解决内容和用户的匹配。环境特征，解决基础特征和匹配。热度特征，在冷启动上很有效。协同特征，考虑相似用户的兴趣，在一定程度上解决所谓算法越推越窄的问题。



5

今日头条有一个世界范围内比较大的在线训练推荐模型，包括几百亿特征和几十亿的向量特征。

完全依赖模型推荐成本过高，因此有了简化策略的召回模型。基于召回策略，把一个海量、无法把握的内容库，变成一个相对小、可以把握的内容库，再进入推荐模型。这样有效平衡了计算成本和效果。

6

在今日头条工作前三年，我收到用户反馈最大的一个问题，就是，“怎么老给我推重复的？”

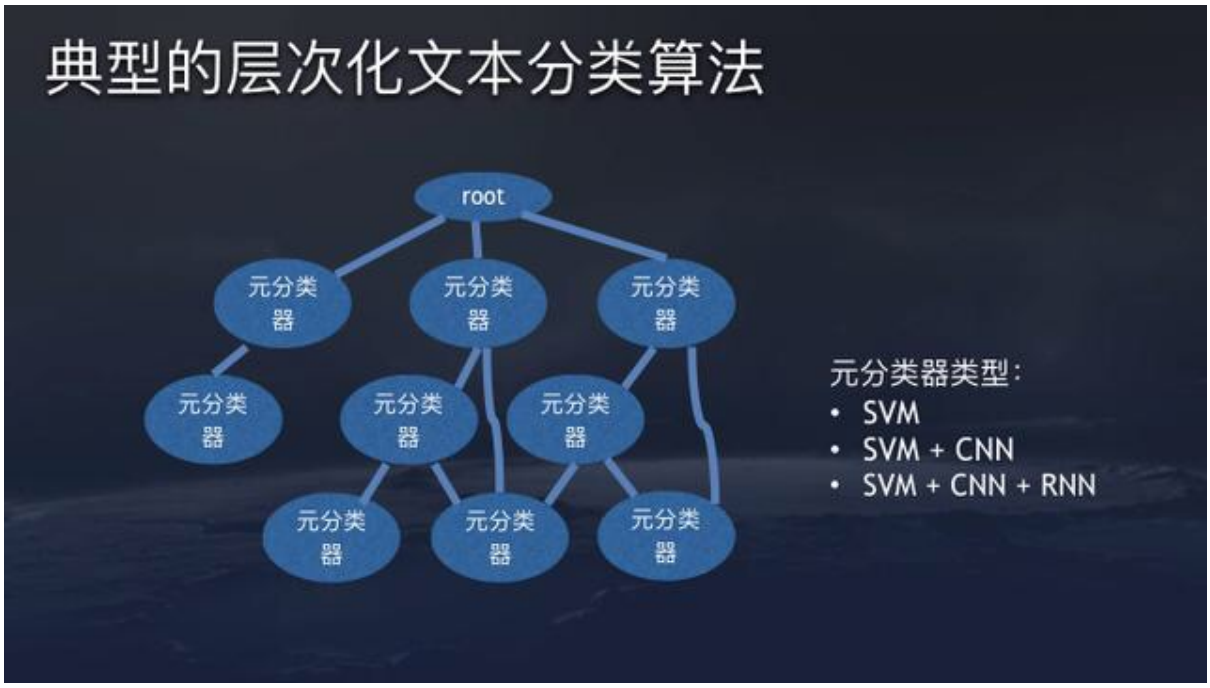
其实，每个人对重复的定义不一样。有人昨天看到一篇讲巴萨的文章，今天又看到两篇，可能就觉得烦了。但对于一个重度球迷来讲，比如巴萨的球迷，可能恨不得所有的报道都看一遍。解决这个，实际上需要精确抽取文本特征，比如哪些文章说的是一个事儿，哪些文章基本一样等等。

文本特征对于推荐的独特价值在于，没有文本特征，推荐引擎无法工作，同时，文本特征颗粒度越细，冷启动能力越强。

7

语义标签的效果，是检查一个公司NLP（自然语言处理）的试金石。

频道、兴趣表达等重要产品功能，需要一个有明确定义、容易理解的文本标签体系。所以，在隐式语义特征已经可以很好地帮助推荐，且做好语义标签需要投入远大于隐式语义特征的情况下，我们仍然需要做好语义标签。



8

除了用户的自然标签，推荐还需要考虑很多复杂的情况：

1）过滤噪声：过滤停留时间短的点击，打击标题党；2）惩罚热点：用户在热门文章上的动作做降权处理；3）时间衰减：随着用户动作的增加，老的特征权重会随时间衰减，新动作贡献的特征权重会更大；4）惩罚展现：如果一篇推荐给用户的文章没有被点击，相关特征

（类别、关键词、来源）权重会被惩罚；5）考虑全局背景：考虑给定特征的人均点击比例。

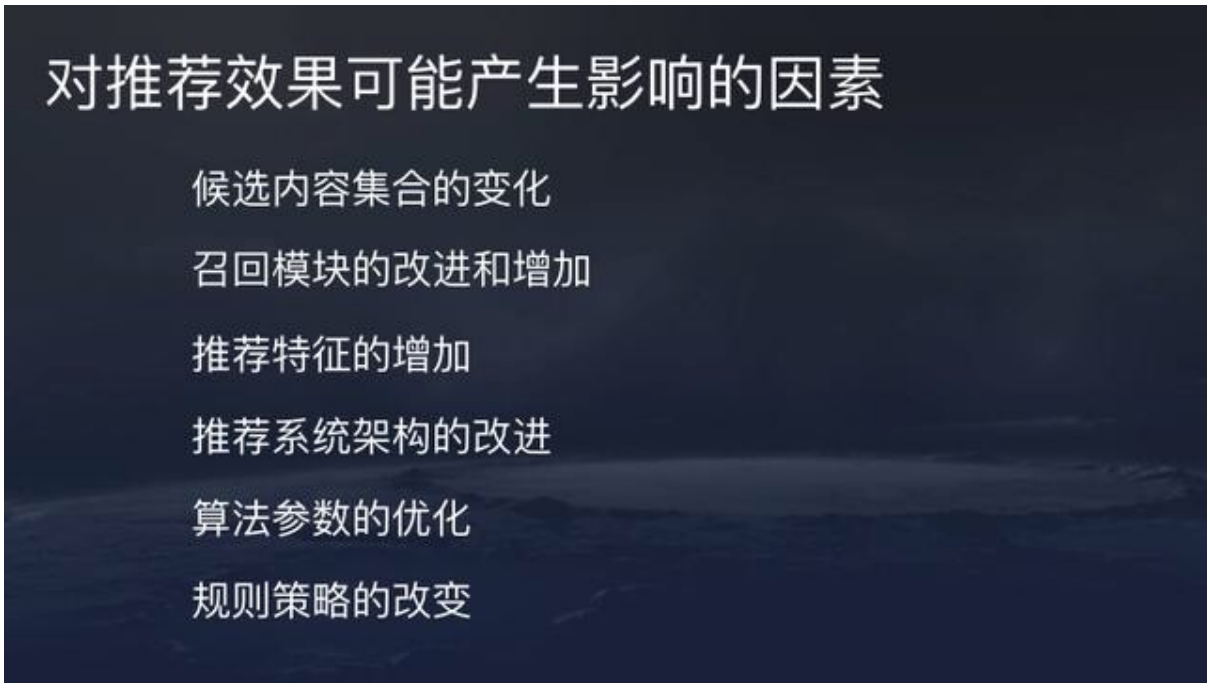
9

比起批量计算用户标签，采用流式计算框架，可以大大节省计算机资源，可以准实时完成用户兴趣模型的更新。几十台机器就可以支撑每天数千万用户的兴趣模型更新，99%的用户可以实现发生动作后10分钟模型更新。

10

影响推荐效果的因素有很多，我们需要一个完备的评估体系，不能只看单一指标，点击率、留存、收入或是互动，我们需要看很多指标，做综合评估：兼顾短期指标和长期指标，兼顾用户指标和生态指标，注意协同效应的影响，有时候需要做彻底的统计隔离等。

有人问，所有的这些指标，能合成唯一的一个公式吗？我们苦苦探索了几年，目前还没有做到。



11

很多公司的算法做得不好，不是人的问题，是实验平台的问题。

如果A/B Test，每次数据都是错的，不是这儿错就是那儿错，总上不了线，这个事就废了。而一个强大的实验平台，可以实现每天数百个实验同时在线，高效管理和分配实验流量，降低实验分析成本，提高算法迭代效率。

12

头条现在拥有健全的内容安全机制。除了人工审核团队，我们还有技术识别。包括风险内容识别技术，构建千万张图片样本集的鉴黄模型，超过百万样本库的低俗模型和谩骂模型等，以及泛低质内容识别技术。

我们一直按行业最高的标准要求自己。