

个性化推荐之召回的方法

原创 策小编 神策数据 2019-03-12

收录于话题

#神策智能推荐

14个

点击上方蓝字可关注 个



个性化推荐之召回的方法

在这个信息爆炸的时代，每个人都是内容的消费者，也轻而易举地可以成为内容的创造者。在海量的信息中，用户找到所需、所想、所爱的优质内容变得异常困难，个性化推荐也因此应运而生。

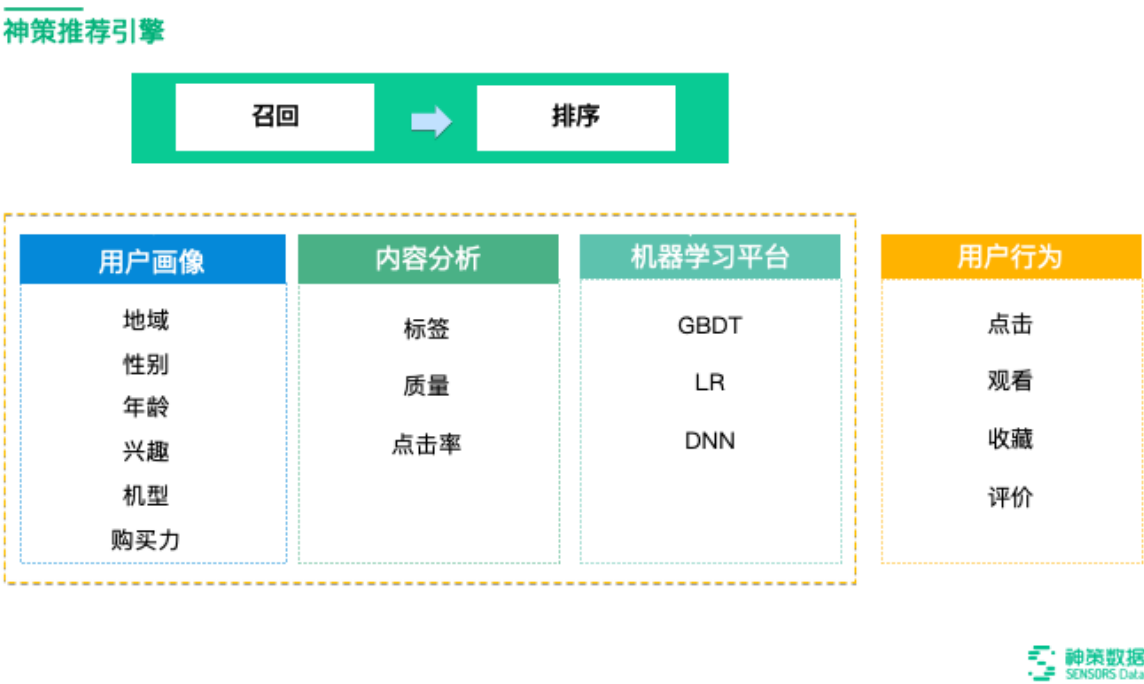
个性化推荐是在大数据分析和人工智能技术的基础上，通过用户行为分析，进行深度计算，从而给用户提供高质量的个性化内容，起到**降低信息过载、发掘长尾、提高转化率**的作用，最终更好更快地满足用户需求。

神策智能推荐是神策数据打造的一款基于用户行为分析的全流程智能推荐系统。该系统主要分为两个阶段：召回阶段和排序阶段。因此，在详细讲述召回的方法前，我们先来看看召回阶段的作用和目的。

一、召回阶段的作用和目的

召回阶段可以理解为根据用户的历史行为数据，为用户在海量的信息中粗选一批待推荐的内容，挑选出一个小的候选集，相当于粗排序。排序阶段再在此基础上进行更精准的计算，做到给每一个内容进行精确打分，相当于精排序。

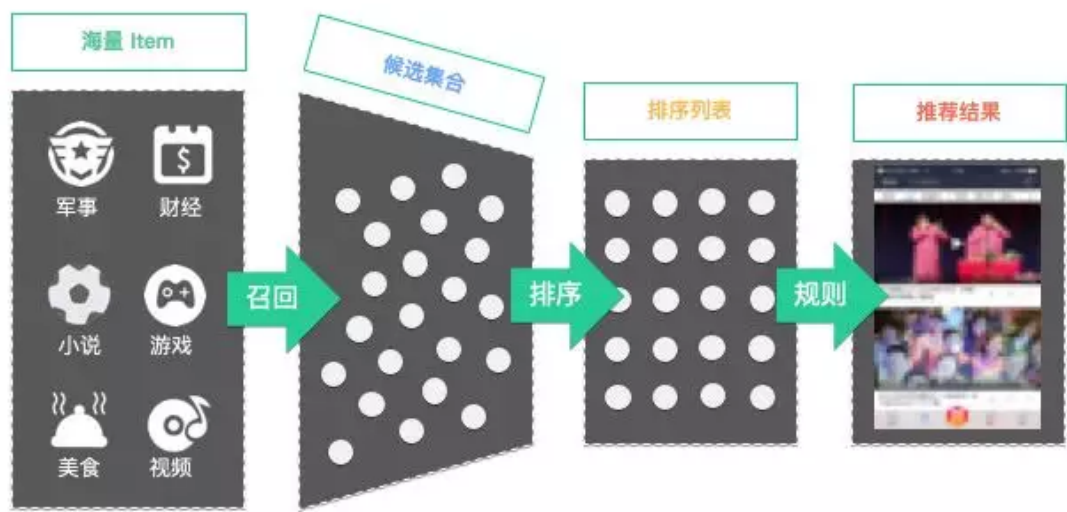
如下图，用户画像、内容分析、机器学习平台是支持召回和排序的三个独立组件，另外，通过点击、观看、收藏、评价等用户行为数据作为数据输入的主要来源。



举个例子，神策的某个资讯类客户，每天的内容量为十万以上，在召回阶段根据用户的历史行为数据分析出其兴趣和偏好，再在海量的政治、历史、体育、娱乐等内容中，挑选出一个小的候选集，如用户主要对历史和体育感兴趣，候选集中就主要是历史和体育的内容，娱乐、政治等其他内容就会被过滤掉，排序阶段会以此为基础进行更精准地计算，通过机器学习预估用户的喜欢程度对使用不同的召回方法获得的候选集的内容进行更精细化地打分，使其变得有序，经过这两个阶段后，最终为用户精挑细选出用户最感兴趣的高质量内容。

综上所述，个性化推荐系统是一个流程：海量 Item——召回（粗排）——候选集合——排序（精排）——排序列表——规则（多样化推荐）——推荐结果。

神策推荐引擎



事实上，作为信息过滤的第一环，召回对于个性化推荐系统极其重要，神策智能推荐在保障数据源的准确性基础上，使用了多种召回方法加强个性化推荐的推荐精准性。下面将为你详解：

召回的方法



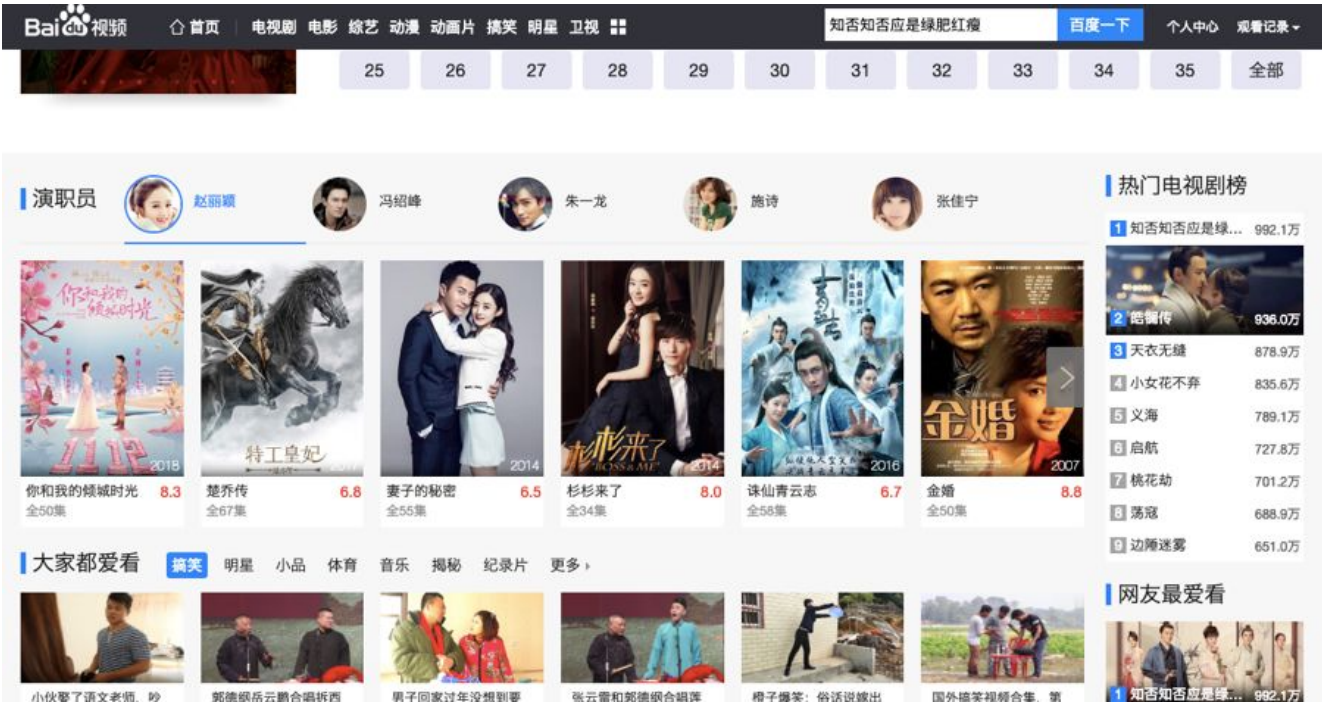
二、基于深度用户行为分析的召回

基于深度用户行为分析的召回，能更全面且精细化地解读用户真实需求，并为其推荐。市场上熟知的基于用户行为分析的召回主要通过以下两种思路：

1.基于内容的协同过滤

基于内容的协同过滤，简而言之是根据用户喜欢的内容推荐与之相似的内容。比如某个用户喜欢 A 内容，算法就会为用户推荐 A1、A2 等与 A 相似的内容，而不是推荐 Z9 等相关性小的内容。

举个例子，百度视频会根据用户正在观看或者历史偏好的视频，进行个性化的推荐。如在百度视频首页搜索《知否知否应是绿肥红瘦》就会出现根据主演赵丽颖推荐的相关电视剧。

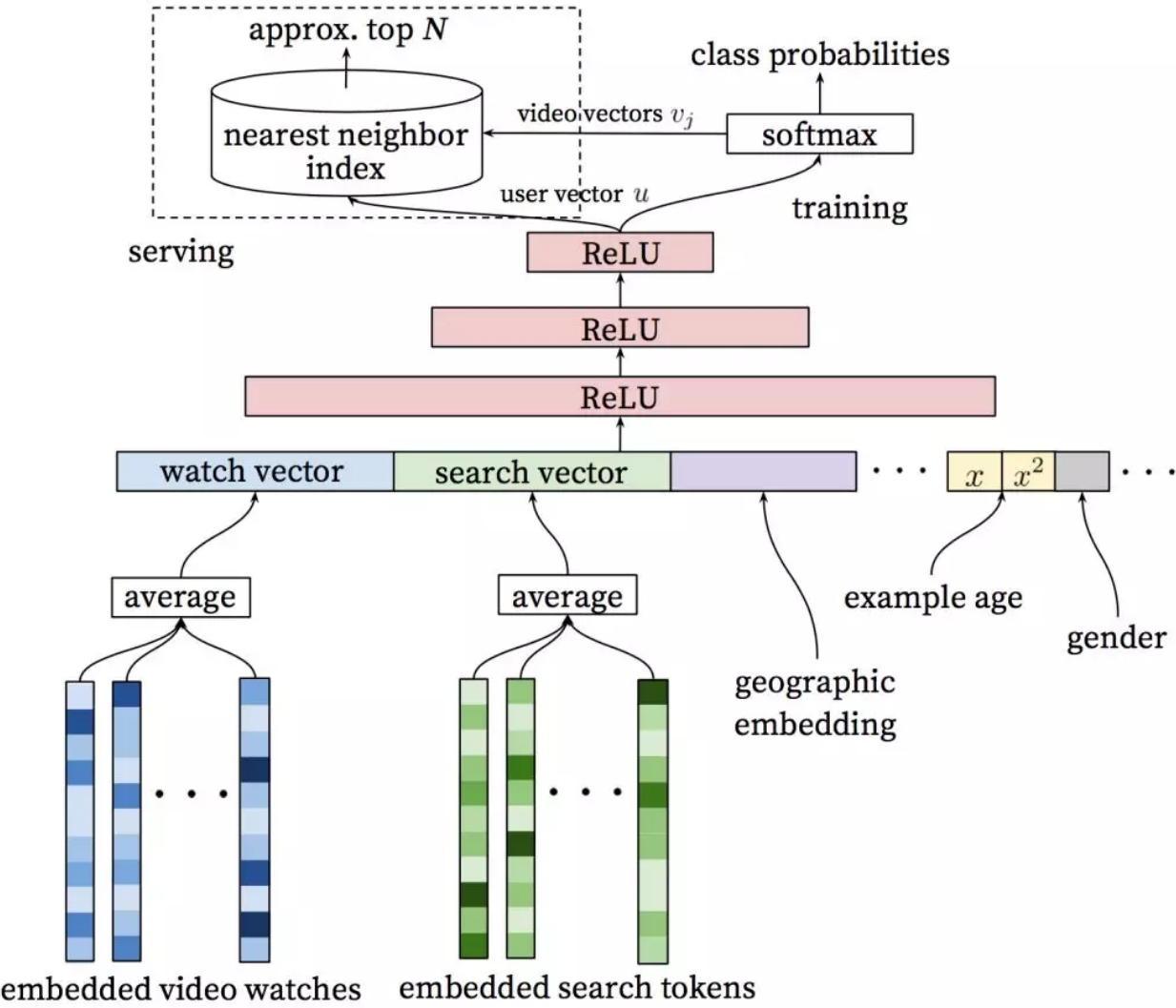


2.基于用户的协同过滤

基于用户的协同过滤，简而言之，推荐相似用户喜欢的内容。比如，A 用户和 B 用户喜欢的内容一致，这种方式就会向 A 用户推荐 B 用户喜欢但是 A 用户没有看的内容。

举个例子，某咨询类 App 有娱乐、奇趣、健康、美食等内容分类，对于喜欢娱乐新闻的用户可以将所有同样喜欢娱乐内容的用户最近最爱看的一些资讯放在有娱乐新闻偏好的用户的热门流中。

神策智能推荐在具备多种协同过滤推荐机制的基础上，采用基于行为的深度学习召回模型，再根据用户行为数据训练模型，加强了推荐系统推荐的智能性和准确性。该深度学习召回模型是借鉴谷歌的个性化推荐思路（如下图），并汲取用户行为分析实践经验结合企业的最新实践需求进行了个性化的优化迭代。



图片来源：谷歌发布的《Deep Neural Networks for YouTube Recommendations》论文

相比基于矩阵分解的协同过滤方式，深度学习召回模型有如下优势：

对比协同过滤，深度学习召回模型的优势

更全面的行为表达	在模型中结合点击，收藏，搜索等多种行为，更全面地表示用户行为偏好
可添加画像特征	可加入性别，地域等用户画像相关的特征
包含行为顺序	模型中考虑用户的行为顺序
组合复杂特征	神经网络进行更复杂的特征组合，挖掘更深层次的关联关系



第一，更全面的行为表达。在模型中结合点击、收藏、搜索等多种行为，能更全面地表示用户行为偏好。

第二，可添加画像特征，可加入性别、地域等用户画像相关的特征。如果你有额外的一些标签或发生的信息，这个模型是可以兼容的，它可以把所有信息糅杂在同一模型里面去做，而在协同过滤模型里面是完全无法引入的。

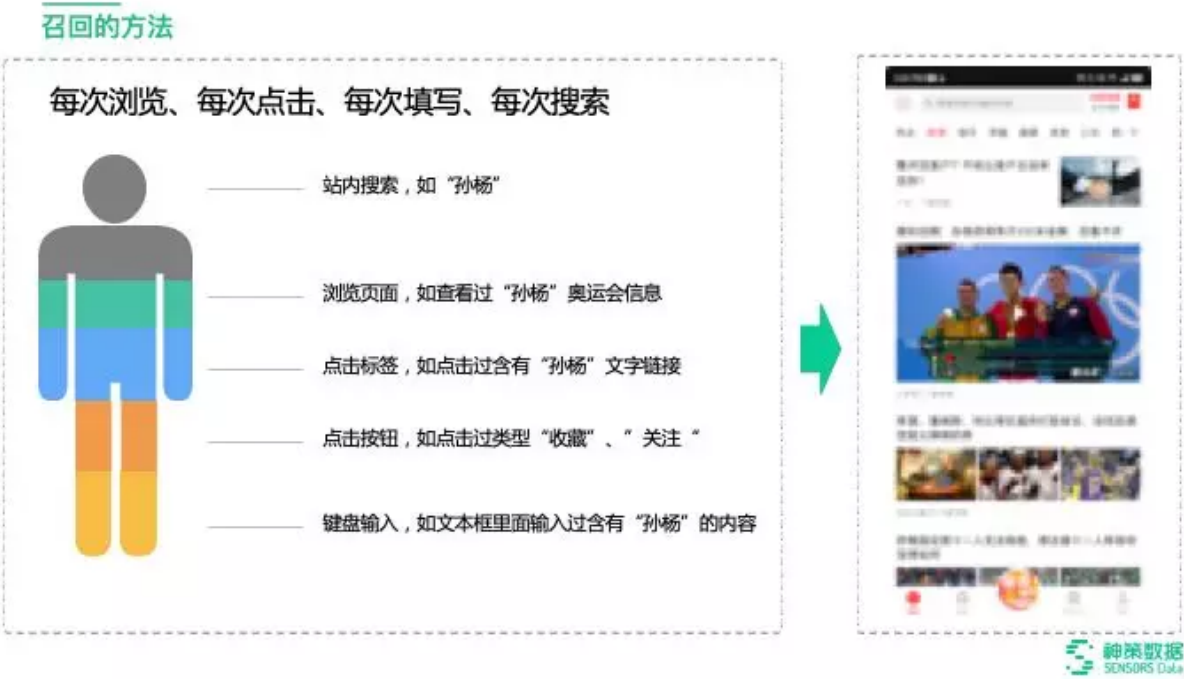
第三，考虑用户的行为顺序。比如用户通常的行为顺序是，先买一个手机，然后再去买一个手机壳；买了一个汽车后可能会买个汽车坐垫。

第四，组合复杂特征。神经网络可以进行更复杂的特征组合，挖掘更深层次的关联关系。

三、基于用户画像的召回

1.利用已有的关注关系

关注关系就是用户显性化的喜爱偏好。用户的每次浏览、每次点击、每次填写、每次搜索都隐藏了你的用户偏好。如下图，当用户的站内搜索、浏览页面、点击标签、点击按钮等行为的数据都能抓到并进行分析，就能做到给用户的内容更逼近其心中所想和心中所爱。



再比如，淘宝也会根据用户收藏和历史的搜索行为进行推荐，如下图淘宝的收藏中有运动裤，下面的猜你喜欢就会推荐相关的运动裤。





¥146.00

找相似

— 你可能还喜欢 —



dcw宽松运动裤女高腰收口小脚夜跑步长裤速干透



软性亲水接触镜



注册证号：国食药监械（注）字2014第3221594号
生产企业名称：甘肃康视科技发展有限公司
生产许可证号：甘食药监械生产许20120033号
请仔细阅读产品说明书或在医务人员指导下购买和使用
禁忌内容或注意事项详见说明书
甘肃康视（文）第 号

博士伦买1送1非离子混血学生年抛女2片装

2. 基于标签构建用户画像

不同性别、年龄、职业、地区的用户对内容的兴趣偏好有所不同，即使同一性别、年龄、职业的偏好也有很大差异，所谓千人千面，每个用户都有其特征和偏好。因此，很多企业都有自身的标签平台，如纵横小说会根据用户行为数据结合内容标签，构建用户画像标签，通过这种方式召回用户感兴趣的内容。

四、保障推荐的有效冷启动和推荐多样性的方法

推荐系统的冷启动场景主要分为三类：

用户冷启动：即如何给新用户做个性化推荐，事实上，第一次展现给用户的 item 极其重要，决定了用户的第一印象；

内容冷启动：即如何将新的内容推荐给潜在对它感兴趣的用户；

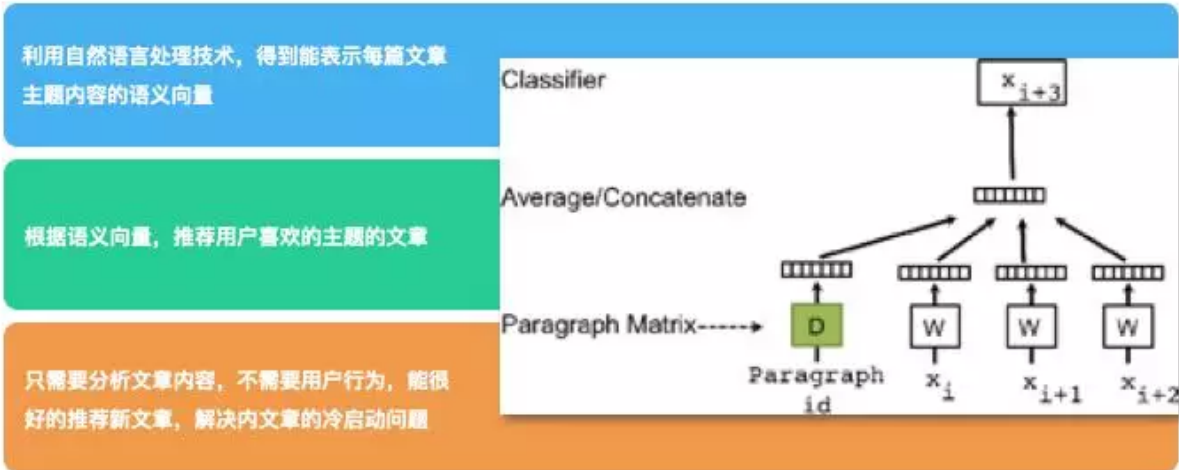
系统冷启动：即如何在一个新开发的产品中（无用户、无用户行为，只有一些内容）设计个性化推荐，从而在产品刚发布就让用户体验到个性化推荐服务。

1.运用自然语言处理技术，解决内容冷启动问题

关于冷启动，神策智能推荐，会运用自然语言处理技术对新的内容进行语义分析，我们可以这样理解，市面上很多都是文本数据的一些“显式”使用方法，包括在前面介绍的标签也是，所谓显式，是指我们将可读可理解的文本本身作为了相关性计算、召回以及模型排序的特征。这样做的优势是能够清晰地看到起作用的是什么，但是其劣势是无法捕捉到隐藏在文本表面之下的深层次信息。例如，“衣服”和“上衣”指的是类似的东西，“厚外套”和“棉服”具有很强的相关性，类似这样的深层次信息，是显式的文本处理所无法捕捉的，因此我们需要一些更复杂的方法来捕捉，而自然语言处理技术就能捕捉到，运用从词下沉到主题的思路，挖掘更深层次的核心信息。

神策智能推荐运用自然语言处理技术——基于神经网络的文本语义分析模型（如下图），相比市面上通过打标签的方式推荐，可以做到更深层次的偏好挖掘推荐，举个例子，如果用户阅读了大量包含甄子丹、成龙、李小龙等关键词的文章，可以挖掘出用户对功夫类主题的内容偏好，并为其推荐。

基于神经网络的文本语义分析模型



2.运用 UCB 探索用户的潜在兴趣算法，保证推荐结果的多样性

在推荐的过程中，需要考虑给新 item 展示的机会，比如给一个喜欢历史分类资讯的用户推荐一些娱乐、政治等其他资讯，解决加强推荐多样性的问题。

大家所熟知的是通过随机分配一部分流量给新 item 曝光，得到一些反馈，然后模型才能对其有较好的建模能力，这是比较传统的冷启动套路。

神策智能推荐采用的是 upperconfidence bound(UCB) 策略: 假设有 K 个新 item 没有任何先验，每个 item 的回报也完全不知道。每个 item 的回报均值都有个置信区间，而随着试验次数增加，置信区间会变窄，对应的是最大置信边界向均值靠拢。如果每次投放时，我们选择置信区间上限最大的那个，则就是 UCB 策略。这个策略主要是通过以下两个原理达成更好地推荐：**均值差不多时，优先给统计不那么充分的资讯多些曝光；均值有差异时，优先出效果好的。**

五、根据多源数据召回，保障推荐的全面性和精准性

企业产生数据的方式多种多样，推荐系统的个性化精准推荐，离不开对业务和用户的精准把控，只有获取足够全面、颗粒度足够细的数据才能更精准的了解用户。神策智能推荐支持企业结合业务角度和时事热点等多方面，利用多种数据源的整合与综合分析，如支持将热门、编辑精选、最新、本地化内容等多类数据进行综合，吸取不同数据源的特点，并将这些内容放入到候选集里，为排序提供足够全面且详尽的内容。

比如，某个资讯类 APP 的编辑会在热门流中精选出部分内容，形成一个精选内容集，就是一个精选出来的数据源，可将其放入候选集中，再次推荐增加曝光。再比如，当出了某些热门事件，资讯类 APP 就可以通过编辑打标签、手动筛选或通过某个简单的程序抓取相关的内容，将其归为一类放入内容池，再经过一系列操作后进行推荐。

最后，我们整体上回顾一下前面介绍的召回方法（如下图）。

召回的方法



事实上，神策智能推荐是以数据为根基，4种主要召回的方法为关键，保障召回的候选集的全面、准确、个性化，为后面的排序打下良好基础。这里引用神策数据架构师房东雨的一句话，“数据和特征决定了机器学习的上限，而模型和算法只是逼近这个上限而已”，这也是神策智能推荐如此强调数据根基的原因，因为如果你的数据根基没有做好，那所有的工作都是徒劳的。事实上，为企业建立良好的数据根基也是神策数据一直在做的事，如果你有需要或疑问都可以通过 4006509827 电话联系我们，会有专业的人员为您解答。

『点击图片，解锁更多精彩』



【一文解读】
神策智能推荐对你的价值

【2019 春招】
“数”天下神人，都“据”于此

