

一站式刷遍各大互联网公司人工智能笔试面试题

涵盖所有考点 精准详尽解析



七月在线APP

笔试题库

面试大题

错题集锦

我的收藏

< 返回

计算广告

收藏

1、了解CTR预估么，请详细说说它的原理

🔒 隐藏解析

💬 1条讨论

📖 1 / 8

↓ 下一题

解析：

我们在刷微信和微博的时候，有时候会惊叹广告非常精准——“就是我要的！”，有时候又会感觉广告极其不靠谱——“我怎么可能点呢？”，所有的这些背后，广告的工程师们已经做了他们最大的努力——

即便在我们看来并不精准，但在所有定向给你的广告中，他们已经挑出机器认为你最可能点的广告，而这背后依赖的就是CTR预估技术。

CTR预估是竞价广告核心的逻辑之一，很多人认为CTR预估非常高深，根本搞不清其中的逻辑，的确，CTR预估作为一个纯技术的模块，要做一个简单的科普并不容易。

致力于用简洁语言描述复杂问题的卫夕今天尝试用人话来解释CTR预估的来龙去脉，放心，本文不会涉及到任何高深难懂的公式和技术逻辑，看得懂中文的同学都能明白其中的逻辑。

一、CTR在竞价广告系统中起神马作用？

CTR (Click-Through-Rate) 为点击率，它是互联网广告中最基本的概念，我们先来看看点击率对于一个竞价广告系统而言意味着神马？

我们先来做一道简单的题目——阿迪和耐克作为广告主来竞价，阿迪出2块钱一个点击，耐克出1块钱一个点击，假如微信广告平台有100次曝光，它应该给谁？阿迪还是耐克？有人说，当然是给阿迪啊，它出的价钱高。




1元一个点击

2元一个点击

我们有100次曝光，我应该给谁？

点击率5%

点击率1%

5个点击，收入5块

1个点击，收入2块

这个答案是不对的，如果阿迪的点击率为1%，耐克的点击率为5%，那么给阿迪能产生1个点击，每个点击1块钱，微信能赚2块；给耐克则产生5个点击，每个点击1块，微信能赚5块，广告平台不傻，当然要给耐克！

好，从这个简单的案例中，我们可以看出几点：

- ①广告平台关心自己的流量价值——即自己的流量曝光卖的贵还是便宜，一般用ECPM(earning cost per mille)这个指标来衡量，即每1000次曝光带来收入。
- ②广告主一般按点击进行扣费，即广告主通常关心结果，出价原则就看一个点击需要花多少钱。

- ③广告平台需要把点击出价转化成ECPM进行扣费和排序。
- ④CTR架起了从点击到曝光的一座桥梁，为排序提供基础。

总结起来就是一个简单的公式—— **$ECPM = 1000 * CTR * \text{点击出价}$** ，在上面这个案例中阿迪的 $ECPM = 1000 * 1\% * 2 = 20$ ，而耐克的 $ECPM = 1000 * 5\% * 1 = 50$ ，显然耐克的ECPM大于阿迪的，因此耐克会在这次竞价中胜出。

从这个案例中我们可以看到CTR是为广告排序用的，而排序本身就是竞价广告的核心，因此CTR预估也是竞价广告的核心技术之一。

不仅广告系统需要CTR预估来排序，推荐系统也需要CTR预估来排序，因此CTR预估应用其实比我们想象的要更加广泛：

YouTube、亚马逊、今日头条等等涉及到个性化推荐的系统都会应用到CTR预估，即他们要把你最可能点的内容推荐给你。

而广告系统的CTR预估在具体的数值上比推荐系统要求更高，比如推荐系统可能只需要知道A的CTR比B大就可以排序了，而广告由于不是直接用CTR进行排序，还加上了出价，因此广告系统不仅要知道A的CTR比B大，而且还需要知道A的CTR比B的CTR大多少。

二、为什么要进行CTR预估？

到这里大家已经明白CTR的重要性了，即CTR是广告排序的前提，它连接了点击和曝光，这时候大家又会有一个新的问题：那CTR为什么要预估呢？多少投放一点不就知道了它的CTR了吗？

“多少投放一点不就知道了它的CTR了吗？”这句话大体上没错，但事情远没有这么简单，我举一个例子你就知道了。

假如广告平台有四个广告主来投广告，这个广告平台总共有4000个用户，这四个广告主分别为卖布娃娃的、卖游戏机的、卖西装的和卖高跟鞋的，出价都是1块钱一个点击。

这时候我们不知道CTR的情况下，我们应该怎么办，于是我们试投了一下，每个广告投了100个曝光，最后发现他们的CTR都是25%，出价一样、CTR一样。

按照我们上面的公式，ECPM就一样，而我们按ECPM来排序，因此我们就没办法排序——每个人ECPM都相同，因此最终就只能在广告平台随机出这四个广告了。



$ECPM = 1000 * CTR * \text{点击出价}$

但我们回去看一看实际情况，这个25%的点击率是如何发生的我们就能明白总体CTR是不靠谱的，真相是平台总共有4000个用户，而他们均匀分成了四组：

分别是1000为成熟男人；1000成熟女人、1000少女、1000少男，这四个人群只点击他们喜欢的东西，如成熟男人只点击西装，他们对自己喜欢的东西的点击率为100%，对自己不喜欢的点击率为0%，当我们随机投放的时候，每个人的点击率都是25%。



好，现在你看到了，直接拿试投这个点击率去预估导致的结果就是用群体数据代表了个体差异，即当一位成熟男人来请求广告的时候，我们会认为投给他一个布娃娃和投给他一套西装没有任何区别。

因为在试投阶段布娃娃和西装的整体点击率都是25%，尽管实际上他们的点击率有天壤之别，一个是100%、另一个是0%。

因此，当我们通过特征差异能正确地个性化预估CTR的时候，我们才能正确地排序，当一位成熟男人来请求广告的时候，西装这个广告主就会在竞价中胜出，因为我们知道它的CTR100%高于布娃娃、高跟鞋、游戏机的0%。

对的，针对不同的人群投放针对性的不同广告，这就是**个性化预估**、**个性化推荐**。

所以从这个案例中我们就能明白，不能直接笼统地看试投的点击率，而应该有针对性地针对个体单独预估，试投这个阶段依然存在，但我们会通过某个模型用交叉特征地看每一个个体的点击率，从而下一个拥有该特征的人群来访问的时候，我们就能相对更加准确地预估了。

同时，我们也从上面的案例中发现，正确预估提升了CTR，从原来的25%提升到了100%，同样广告平台的收入也会提升，这就是CTR预估的意义。

三、CTR预估是如何进行的？

通过上面那个案例我们看到，至少有三个特征决定了一个广告的**点击率**——广告主行业、用户的年龄、用户的性别，事实上，在实际的广告系统中，有无数的因素决定了广告的点击率，我们把这些因素主要分成3类——

第一是**广告主侧**，比如广告创意、广告的表现形式、广告主行业等，一个劳斯莱斯的广告和一个可口可乐的广告点击率肯定有天壤之别。

第二是**用户侧**，如人群属性，以上的案例是年龄和性别，事实上，决定创意的因素人群属性极其多——年龄、性别、地域、手机型号、WiFi环境、兴趣……

第三是**广告平台侧**，比如不同的广告位、投放时间、流量分配机制、频次控制策略等。



广告侧：

广告主行业、广告文案、图片、广告形式.....



用户侧：

年龄、性别、地域、手机、兴趣、使用环境、活跃度.....



平台侧：

频次控制、时间、竞价策略、流量分配策略、历史CTR.....

这些决定因素在CTR预估中被称之为特征，而CTR预估的第一步就是“特征工程”，即把这些特征找到并数据化。

特征工程是一件复杂的工程，光判断不同特征能否对CTR产生影响就是一个浩大的工程，特征工程的攻城狮们第一步得列出来不同的特征可能对CTR产生影响，这有时候靠直觉，有时候靠经验，

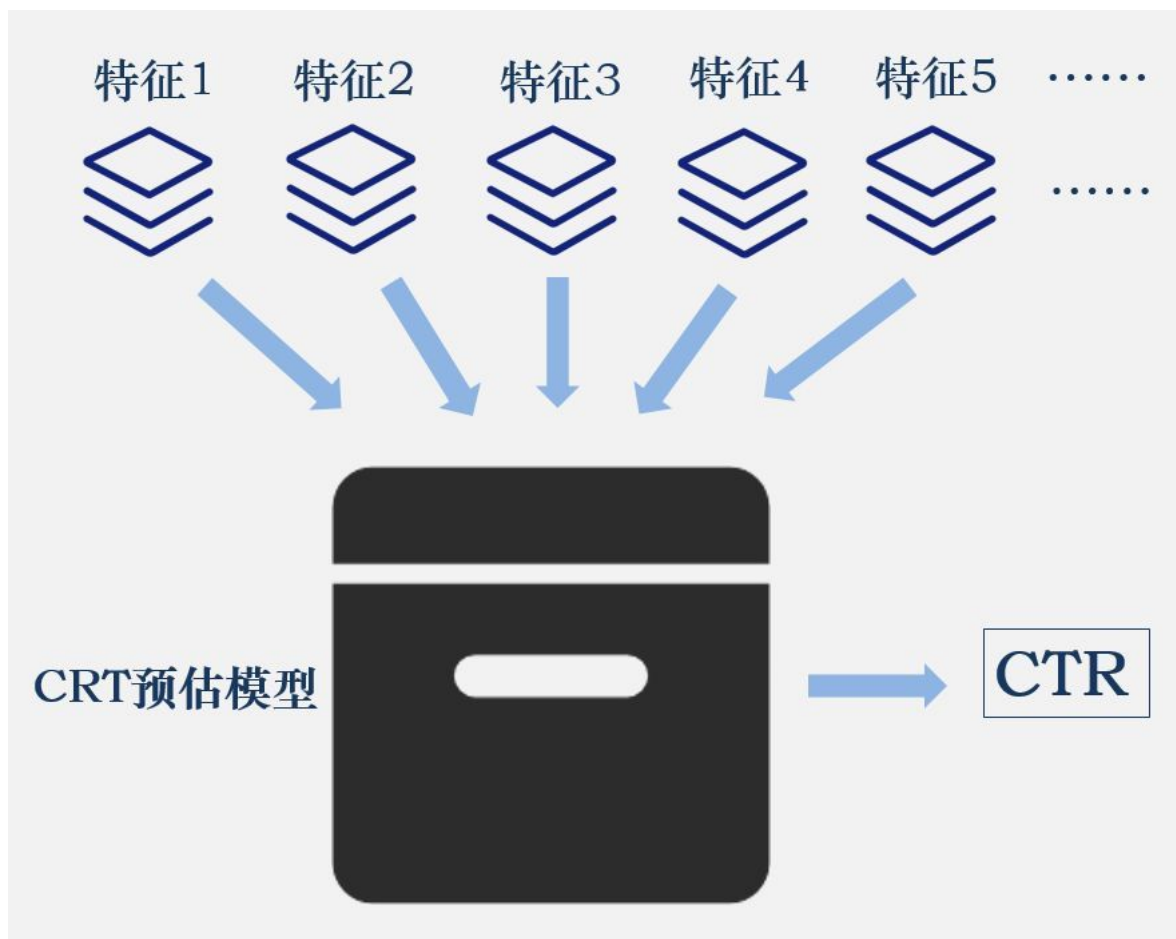
确定了特征之后就需要对这些特征进行处理——即把**特征数据化**，比如把所有的特征变成0和1的二值化，把连续的特征离散化，把特征的值平滑化、把多个特征向量化.....

以上这一段看不懂？看不懂没关系，你只需要明白，所有的这些特征都会被攻城狮们编码变成一串可计算的数组就行，特征工程是一项重要的基础工作，像今日头条、百度这样的广告平台都有一个庞大的团队来做特征工程的工作。

好，特征工程完成之后就开始建立模型了，很多同学看到“模型”二字就开始皱眉头了，没关系，你可以这样理解模型：

模型就是一个黑盒子，在盒子的一方我们输入一大堆参数，盒子的另一端就会输入一个**CTR的值**。

比如我们输入这样的参数：一个高跟鞋广告主、投放给北京地区、年龄20-30岁、对财经感兴趣的女性用户，在周末投放在微信的朋友圈第三位.....请给出一个CTR的值。模型就会根据不同特征的值计算出一个值。



至于它是如何计算的，非技术同学理解起来难度会非常大，我们可以简单理解为CTR是无数特征的一个函数， $CTR=f(x_1, x_2, x_3, x_4, x_5, \dots)$ ，而模型就是选择神马样的函数来进行预测。

我举个大家最容易理解的线性函数： $CTR=ax_1+bx_2+cx_3+dx_4+ex_5+\dots$ （这只是一个假设，实际上模型要比这个复杂N多倍）。

每一个函数都有相应的参数（比如刚刚案例中的a、b、c、d），而这些参数是如何确定的呢，就需要历史数据进行训练，就是把已经知道的一些数据去喂这个模型，模型根据这些数据不断地调整参数，喂的越多调教的就越准确，最终预测的效果也就越好。

在所有的这些特征中，有些特征影响因子会大一些，有一些会小一些，比如搜索广告中，搜索关键词和广告关键词匹配程度就是一个影响因子极大的因子、历史CTR也是一个影响因子极大的因子，其他因子会有不同程度上的影响。

四、如何评估CTR预估的效果？

好，接下来的一个问题，如何评估CTR预估的效果？工程师们会有特定的指标去衡量效果，对于非技术同学，有人会直观地说，那当然是看CTR有没有变高啊：

正确预估CTR不就是为了把真正高CTR的广告挑出并展示出来么，错误地预估——把高的CTR低估或把低的CTR高估都会让高的CTR不会排在最前面，从而会降低CTR。

这个说法在推荐系统中是成立的，但在广告系统里是不准确的，因为广告的排序不仅仅是CTR排序，它还综合了出价进行排序，即ECPM进行排序，有时候CTR预估准确，反而（实际）CTR会降，我们看一个实际的例子：

广告主	实际CTR	预测CTR	出价	预测ECPM	实际ECPM	结果
阿迪	1%	1%	3	30	30	胜出
耐克	2%	2%	1	20	20	失败

图1，CTR预测正确，CTR低

广告主	实际CTR	预测CTR	出价	预测ECPM	实际ECPM	结果
阿迪	1%	1%	3	30	30	失败
耐克	2%	4%	1	40	20	胜出

图2，CTR预测错误，CTR高

这一大堆数字大家可能看不懂，我们只需要明白以下几个逻辑就能看懂了：

- i) 广告通过预估ECPM的大小来判断谁胜出，而预估ECPM等于预估CTR*出价，在这个案例中，图一CTR预测正确，阿迪的预测CTR30高于耐克的20因而胜出。
- ii) 广告实际ECPM是胜出者的ECPM，因为广告平台把所有的曝光都给了胜出者，它等于胜出者的实际CTR*出价。
- iii) 图二里把耐克的CTR从2%错误地预估到了4%，导致其预测ECPM高而胜出，而实际上投放出来的CTR为2%。
- iv) 这个情况中，预测正确反而CTR低，预测错误反而CTR高，但是没关系图一预估正确的情况下实际的ECPM30高于图二预测错误ECPM20。

所以从这个案例中我们可以看到，从结果指标来看，**ECPM是衡量CTR预估最为重要的指标**，当然，在实际的实践过程中，**CTR预测正确通常ECPM、CTR、收入这些指标通常都会涨。**

另外一个很简单的方法就是把预估CTR和真实CTR直接进行对比，由于预估CTR针对每一次广告展示都会预估，比如这一次预估CTR为2%，但单次广告真实CTR只有两个结果，点或不点，即100%、0%，所以看单次结果是没有意义的，我们应该从一个群体来看——

将每一次广告曝光按照预测的CTR从小到大排序，然后按某个单位（比如每10000个曝光）分别统计平均预估CTR和实际CTR，就能知道预测CTR的准确程度了。

作者：卫夕，微信公众号：卫夕聊广告（ID：weixiads），人人都是产品经理专栏作家。一名兴趣广泛的广告产品经理，致力于用简单语言深度剖析互联网相关的逻辑。

链接：<http://www.woshipm.com/operate/1323824.html>

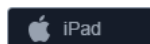


学人工智能 上七月在线

关于我们 | 加入我们

关注我们：

客户端下载：



合作伙伴：阿里云天池 腾讯云学院 讯飞AI大学 人邮异步社区 莫烦Python

联系我们：

热线：010-82712840

手机：18910848502

邮箱：support@julyedu.cc