

秋招面经 | 推荐算法工程师史上最强面经来袭

深度传送门 1周前

以下文章来源于海边的拾遗者，作者weber



海边的拾遗者

一名既爱生活也爱算法的修行者。不定期更心得，生活，学习笔记(包括 ML/GL/NLP/...

背景

作者是一名今年参加秋招的学弟，本文写在秋招结束后。

背景为985本硕&计算机科班，研究生期间方向主要为**推荐算法**，投递的方向主要为**推荐/广告/机器学习**。目前国内已**从A收割到Z**。本科做工程相关，研究生期间转算法，发表过**CCF A类**论文一作1篇，有过大厂实习经验、开源项目和机器学习比赛“划水”经历。

在这里也欢迎各位从事推荐及广告算法的小伙伴加入我们，一起交流学习，一起进步。

如下仅为部分面经，墙裂欢迎各互联网大佬来撩。

具体面经分享

百度

百度7月初就开启了AIDU计划（详情见“[百度招聘](#)”公众号），报名后可以直接在交流群联系部门负责人进行面试。

面试流程为 三轮技术面 + hr意向面 + 9月底带薪offer面。

我自主投递了商业策略平台部（凤巢），还被垂直搜索组捞了简历，两个部门同时进行面试，最终两个部门都**拿到了offer**。

AIDU计划是先约面，面完再进系统走流程，这样既可以多个部门同时面，又不用担心没过的话有面试记录，这个方式是对应届生面试比较友好的，点赞👍。但值得注意的是最终只能走一个部门的offer流程，如果拿了多个部门的offer意向需要自己取舍。

商业算法组

一面

自我介绍。

- 询问实习，很详细。

- 问2020腾讯广告算法比赛如何做的数据分析，预处理；模型后续如何提升。
- LR的损失函数，sigmoid函数由来，lr求导手推。
- 极大似然和最小二乘的区别。

答：观测值服从正态分布的情况下，最小二乘法等价于极大似然估计。（可以推导）

例子参考：<https://www.cnblogs.com/little-YTMM/p/5700226.html>

- 常见的优化器介绍；adam与adagrad的区别。
- AUC计算公式，解释意义。
- 朴素贝叶斯公式，为什么叫朴素，是生成模型还是判别模型。

答：朴素的英文是native，即“天真”，我们应用朴素贝叶斯时做了一个天真的假设：条件独立性假设。

- 生成模型和判别模型的区别。

答：举个例子：

对面一个人走过来，和你说了一句话，你需要识别出她说的到底是汉语、英语还是法语等。可以有两种方法达到这个目的：

- 学习每一种语言，你花了大量精力把汉语、英语和法语等都学会了，我指的学会是你知道什么样的语音对应什么样的语言。然后再有人过来对你说，你就可以知道他说的是什么语音。
- 不去学习每一种语言，你只学习这些语言之间的差别，然后再判断（分类）。意思是指我学会了汉语和英语等语言的发音是有差别的，我学会这种差别就好了。

那么第一种方法就是生成方法，第二种方法是判别方法。

- 生成模型： $P(y|x) = \frac{P(x,y)}{P(x)}$

$$P(y|x) = P(x,y)/P(x)$$

，先由数据学习联合概率分布和先验概率分布，然后求出条件概率分布。

- 判别模型：直接学习决策函数 $y = f(x)$

$$y = f(x)$$

- 概率题：工厂A和工厂B的产品的次品率分别为1%,2%，现从由A和B的产品分别占60%和40%的一批产品中随机抽取一件，发现是次品，则该产品属于B厂生产的概率。
- 代码题：剑指offer 面试题68 - II. 二叉树的最近公共祖先。

二面

实习介绍。

- 内积，外积，哈达玛积的区别。
- loss回传时，m个隐层单元，n个输出层单元，每个输出层单元的loss是如何回传到m个隐层单元的。
- embedding和one-hot的区别，emb的好处与坏处，做emb是否有信息缺失。
- 问贪心算法和动态规划的思路一般是什么样的。
- 代码题：二维数组最大子数组和。

三面

介绍实习，论文，很详细。

- 开放题：问搜索场景下，会提示相关词，这种相关词的收益效果如何量化，给出一个方案。
- 部门介绍与反问。

垂直搜索组

一面

介绍论文，实习，项目，说了很久。

- 探讨ctr未来可能做的方向。
- 代码题：划分一个整形数组，把负数放在左边，零都放在中间，正数放在右边，要求空间复杂度 $O(1)$ 。

快排思想来做，如下：

```
1 def solution(nums):
2     low = 0 # 负数区域，指向负数区域后的第一个位置
3     curr = 0 # 0区域，指向0区域后的第一个位置，也是当前遍历位置
4     high = len(nums) - 1 # 正数区域，指向区域前的第一个位置
5
```

```
6     while curr <= high:
7         if nums[curr] < 0: # 当前元素添加到负数区域，并前进一格
8             nums[low], nums[curr] = nums[curr], nums[low] # 有可能和自身交换
9             curr += 1
10            low += 1
11        elif nums[curr] == 0:
12            curr += 1
13        else:
14            nums[curr], nums[high] = nums[high], nums[curr]
15            low -= 1
16    return nums
```

二面

介绍印象最深的一个项目，顺着项目问了很久。

- 特征分析和预处理的方法。
- LR, SVM, GBDT, NN的介绍和区别。
- GBDT和NN的使用场景以及优缺点。
- NN如何解决过拟合。
- Graph Embedding如何联合属性进行训练。

答：参见：<https://zhuanlan.zhihu.com/p/70198918>，使用属性的emb + pooling代替随机初始化的GE

- 搜索中的position bias问题（展现在前面的item点击率会偏高）如何解决？

答：有两种可以尝试的方案：

1. 在训练阶段将位置作为一个特征加入到模型中，而在预测阶段置为0或者一个统一的常数
2. 通过一个**shallow tower**（比较轻量的模型）来预测位置偏置信息，输入的特征主要是一些和位置偏置相关的特征。在多任务模型的子任务最后的sigmoid前，将shallow tower的输出结果加入进去。而在预测阶段，不考虑shallow tower的结果。

- Graph Embedding如何联合属性进行训练。
- 如何做query和doc的搜索，简单说一下思路，以及该思路存在的问题。

答：参考DSSM，用 DNN 把 query 和 doc 表达为低维语义向量，并通过 cosine 距离来计算两个语义向量的距离，最终训练出语义相似度模型。

- ctr中的selection bias问题（推荐破坏了用户真实兴趣分布）如何解决？
- 最近在读的论文，三句话介绍其观点。
- 自我评价-优缺点。

三面

自我介绍。

- 介绍自己的项目，遇到的问题。
- 开放题：推荐/搜索中的bias问题如何解决。
- 开放题：在线/离线场景，特征表现的一致性如何评估？
- 自我评估。

总结

百度的面试难度比较友好，主要是对机器学习基础，实习情况以及一些推荐场景的常见开放性问题进行查询。

京东|广告质量部定向组

京东6月22日开启提前批（详情见“京东招聘”公众号），找了相关匹配部门的学长内推，但面试安排的较晚，到7月27日才安排面试。

面试流程为 两轮技术面，最终**拿到了offer**。

由于hc紧张，发放的是意向offer，即如果明确去的话，会给sp及以上offer；不能确定的话，可以先秋招其他公司，随时沟通进展。

一面

介绍比赛，实习，很详细。

- 代码题：leetcode 1：两数之和；leetcode 169：Boyer-Moore 投票算法；
Leetcode 382：蓄水池抽样算法 + **推导**。

推导是用数学归纳法：

举例说明：1 - 10

遇到1，概率为1，保留第一个数。遇到2，概率为1/2，这个时候，1和2各1/2的概率被保留 遇到3，3被保留的概率为1/3，(之前剩下的数假设1被保留)，2/3的概率 1 被保留，(此时1被保留的总概率为 $2/3 * 1/2 = 1/3$) 遇到4，4被保留的概率为1/4，(之前剩下的数假设1被保留)，3/4的概率 1 被保留，(此时1被保留的总概率为 $3/4 * 2/3 * 1/2 = 1/4$)

...

二面

介绍论文，实习，很详细。

- 说说ESMM的实践为什么没收益，分析原因。
- 概率题：感冒发烧概率为0.6，不感冒发烧概率为0.1，感冒概率0.2，问发烧感冒的概率。
- 代码题：快排。
- 介绍部门 + 反问。

总结

京东的提前批面试流程比较快，一面code考察较多，整体还是考察基础。

快手|商业化广告算法

快手是8月初投递的提前批，8月20号左右安排面试。

面试流程为 三轮技术面+hr口头offer+加面(加面是为了申更高级别offer)，最终**拿到offer**意向书。

一面

介绍论文，实习。

- 代码题：字符串abc，输出所有子集。

[a,b,c,ab,ac,bc,abc]

- 开放题：百度搜索的联想词搜索功能如何实现？如果建表/建树的话，如何估算内存？

二面

- 介绍实习，说说实习训练模型具体调参调了哪些参？

- ctr预估偏高会带来的问题。
- 广告场景ctr预估为什么要保距？
- 加了bn和ly以后学习率为什么可以上升？
- hash emb的作用：增加泛化。
- l1和l2的区别，作用。
- 梯度爆炸和梯度消失。
- 代码题：快排。

三面

介绍论文。

- resNet为什么能把网络做深。

从梯度的角度来回答

- CNN中卷积和池化设计的原因/好处。
- 训练集验证集表现好，测试集表现差是什么原因？
- 训练集验证集表现好，测试集表现差是什么原因？

特征穿越/过分拟合验证集

- 线上线下表现不一致原因。

特征穿越/突然新增用户/广告导致的信息分布不一致

- OCPM和CPA的区别。
- MMOE和share bottom的区别。
- 推荐和广告的区别。
- 不使用online learning，如何捕捉用户近期的兴趣变化？

构建近期交互的统计特征，emb依时间衰减的加权特征

- 开放题：din/dien在某些公司的效果为什么不好。
- 开放题：如何利用旧数据迁移学习新的广告位场景。

滴滴|金融风控组

滴滴是6月8日参与投递的牛客网SP专场（详情见“[牛客网](#)”公众号）。

牛客网SP专场：牛客网帮忙汇总了大厂的一些SP offer，省去了填网申的麻烦，可以一键投递。

面试流程为 三轮技术面+交叉面，最终**拿到了offer**意向书。

滴滴是同一天连着三轮面试的，面试通过直接转流程到下一面试官，无需等待，这一点好评。

一面

介绍论文。

- GCN与GraphSAGE的区别，说明其本质思想。
- 介绍lr，手推lr损失和导数。
- 代码题：leetcode198-打家劫舍（该题的类似变形）。

二面

介绍实习，论文。

- FM与GBDT构建组合特征的区别。
- FM与GBDT+LR的差异。
- 开放题：金融风控场景，如何建模用户关系，并利用其来提升放贷收益？
- 概率题：圆上三点，构成锐角三角形的概率
- 代码题：剑指offer14-剪绳子

三面

介绍实习，DeepCTR项目。

- 智商题：【100囚犯问题】1, 2, 3, ..., 100共100个人，每个人在1个房间，不能查看其他房间状态，房间什么都没，每个人都有概率随机被抽中，被抽中的人到101房间一趟，101房间有一个灯，一个控制灯的开关，设计一个方案，让这100个人中能有人知道100个人都已经去过101房间。

答：让1号控制关灯，其他人控制开灯 (仅每个人第一次出去的时候开灯)，1号关灯99次，则说明此时所有人都去过101。

该方案的数学期望次数：

- 答：分步考虑如下：

1. 先考虑1号出去1次这个事件，概率为 $\frac{1}{100}$ ；

如果记期望次数为E，则有 $E = 1 \times \frac{1}{100} + (1 + E) \times \frac{99}{100}$ ，可得期望为100次。

2. 再考虑第i个未出去过的囚犯出去1次这个事件，概率为 $\frac{100-i}{100}$ ；

同理可以求得期望为 $\frac{100}{100-i}$ 次。

3. 那期望的总天数为：

$$E(X) = \sum_{i=1}^{99} (100 + \frac{100}{100-i}) = 9900 + 100 \ln 99 \approx 10359$$

参考：<https://wenku.baidu.com/view/49102f2558fb770bf78a5565.html#>

交叉面

- 介绍论文，实习，项目，比赛。
- 代码题：k种硬币(无限个)，n个碗，每个碗放1个硬币，最多连续不超过2个碗内的硬币类别相同，有多少种方法。

答：用动态规划，转移方程推导：

$$dp[0]=k$$

$$dp[1]=k*k$$

$$dp[2]=(k-1)*k+k*(k-1)*k$$

...

$$dp[i]=(k-1)*dp[i-2]+(k-1)dp[i-1]$$

总结

滴滴的一天三连面好评，这样极大的缩短了面试等待时间，而且上一轮面试官面过的可以及时反馈到下一轮，不会重复提问。