

最近的面试经历 @sumnous

我的简历背景是主要做社区发现算法，之前做同构网络的社区发现算法，现在做异构网络的（学术网络为例），每位面试官都会让我介绍我做的内容。我写了图像处理的经历，百度和滴滴的面试官对这一经历很感兴趣。

百度

职位是机器学习。

第一面：

介绍我的社区发现论文，并给出论文中我认为创新的地方。我说了三点。

另外：

1. 如何帮助超市定特价商品。
2. 一个黑盒里是一个图，结构未知，只知道点的个数是 n ，边的个数是 m ，写公式给出两个点相连的概率。

第二面：（女面试官！）

写代码：给一个文档，以及stopwords，统计除掉停用词之后的词频统计。

之后问到给一个查询query，如何匹配相应地文档。

由于当时并不知道TF-IDF，向量空间模型（VSM），以及sphinx全文检索技术这一块，这一面减分很多。其实之后看了下资料，比较简单。

第三面：

隔了一周部门的leader面的。问得跟项目相关。给我提得建议是系统学习下机器学习，这样再做数据挖掘的时候会有更好的全局观。虽悲剧但受教。

一道概率题，54张牌，平均分成三堆，大小王在同一堆的概率？（ $1/3 * 17/53$ ）

搜狗

职位是数据挖掘算法工程师。笔试题，有选择20+，包括单选、多选，填空3道，编程题两道。涉及范围很广，是精心设计的题目。

涉及到的内容，聚类算法，互信息，频繁项挖掘算法，分类算法，树的前中后序遍历。很基础的选择題。我遇到的面试官在做完之后跟我一起过了一遍。错得让我 $\odot___\odot$ 汗。问我会不会Hadoop，然后我说用得少。估计因此悲剧了。

能记得的选择題：

买饮料，三个瓶盖可以换一瓶，请问要买100瓶饮料，最少需要买多少瓶？（67）

填空题：

10个试管，两种液体，A是好液体，B是坏液体，只有一个试管里是B，且A、B混合会发生反

应，判断哪个试管里是坏液体，至少需要多少个试管？（4，二进制）

编程题：

1. 已知随机数生成函数 $f()$ ，返回0的概率是60%，返回1的概率是40%。根据 $f()$ 求随机数函数 $g()$ ，使返回0和1的概率是50%，不能用已有的随机生成库函数。[我整理的答案](#)
2. 给你一个数组 $A[1..n]$ ，请你在 $O(n)$ 的时间里构造一个新的数组 $B[1..n]$ ，使得 $B[i]=A[1] * A[2] * \dots * A[n] / A[i]$ 。你不能使用除法运算。[分析可参考](#)

腾讯滴滴打车

职位是算法工程师。滴滴的面试官和我都非常敬业的从上午11点面到下午2点，都木有吃饭，555。。。去之前专门找做地理位置算法的同学给我科普了他们做的路径规划算法和基于社交的location推荐系统，所以面试的开放性问题都答得很好。

第一面：

要求介绍K-means算法，并给出找起始点的策略。

开放性问题：给用户推荐一定距离内他可能去的餐馆，请给出算法策略。我想的是已知距离和餐馆的访问次数，套用TF-IDF模型。

写代码：两个字符串，判断一个字符串是否是另一个字符串的子串，并返回子串的起始位置。

第二面：

什么样的数据结构可以满足多次插入删除，取最小数，给出时间复杂度。我的回答是小顶堆， $O(n\log n)$ 。

写代码：链表逆序。（练了很多遍，写得非常顺利）

开放性问题：已知每天的呼叫订单，在线司机，和成交订单记录，如何判断异常数据。我的回答是用卷积函数来检测异常。（实际可能他们真会用到卷积）看到我做过图像处理就问了我傅里叶变换，我说我不记得了。。。

描述Dijkstra最短路径算法。（去滴滴面试之前专门复习了这个算法，讲得很顺利）

问了一个超有意思的智力题：

100个人排队，每个人只能看到自己之前的人的帽子的颜色（假设只有黑白两色），每个人都得猜自己帽子的颜色，只能说一次，说错就死掉，别人可以听到之前的人的答案以及是否死掉。请问用什么策略说死掉的人最少。

我的回答是看哪个颜色的多就说哪个颜色。后来就想不出来了。

接着面试官简单介绍了一下他们目前的工作。然后问我有没有问题问他。然后我恳求他告诉我这道题的答案。他说了他的策略。

假设只有3个人，假设true = 白，false = 黑，用这个公式 $x3 = (x1 == x2)$ ，用人话就是1和2的帽子颜色一样的话就说白，不一样的话就说黑。这个策略第一个人死的概率是1/2，剩下的两个都不会死。

他让我推广到4个人，也就是 $x4 = (x3 == (x1 == x2))$ ，照理可以推广到100人。但问题就是人很难判断，只能靠计算机来算。

豆瓣

职位是算法工程师。豆瓣的面试官会看我的Github和豆瓣主页，这点是其他公司没有的。豆瓣的面试官对我的社交网络研究非常感兴趣，他们说他们在这一块的尝试一直没成功。

第一面：

主要介绍我做的社区发现的课题，他们对这个比较感兴趣。豆瓣主要在做的东西是user-profile的建立，以及用户的兴趣匹配。提到了LDA，聚类算法。

第二面：

开放性问题：请给出给用户推荐他感兴趣的内容的算法策略。我的回答：第一种，用话题模型计算用户的兴趣分布，并找到该兴趣下的top-n，总体兴趣下排序再取top-n。第二种，通过用户友邻的兴趣来推荐。第三种，先社区划分，利用社区内用户的兴趣做推荐。

写了一段代码，二叉树，判断一个二叉树是否是另一个二叉树的子树。我写的是先比较根节点，在比较左孩子和右孩子。面试官提示我这道题更好地解法是递归比较，先比较根节点，不同的话递归比较左子树和右子树。

问了一个机器学习的问题，关于线性回归和梯度下降。让我对线性回归的公式求导。

PS: 请勿直接公开此文档，这是写给你看的版本~ 谢谢July~

sumnous

2014.05.23