

# 17|作业帮算法岗武功秘籍

## 1 作业帮面经汇总资料

第一节  
作业帮面经  
汇总资料  
(整理: 江大白)  
[www.jiangedabai.com](http://www.jiangedabai.com)

- 1.1 面经汇总参考资料
- 1.2 面经涉及招聘岗位
- 1.3 面试流程时间安排
- 1.4 作业帮面经整理心得

### 1.1 面经汇总参考资料

#### ① 参考资料:

- (1) 牛客网: 作业帮面经-41 篇, [网页链接](#)
- (2) 知乎面经: [点击进入查看](#)
- (3) 面试圈: [点击进入查看](#)

#### ② 面经参考答案:

- (1) 面经答案: [点击进入查看](#)

### 1.2 面经涉及招聘岗位

#### (1) 全职岗位类

【机器学习算法工程师】、【自然语言算法工程师】、【机器学习数据挖掘工程师】、【广告推荐系统工程师】、【nlp 算法工程师】、【图像视觉算法工程师】

## 1.3 面试流程时间安排

作业帮面试流程-整理：江大白			
	面试类型	面试流程	备注（侧重点）
第一面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	很看重项目及知识点细节 算法编程能力也看重
第二面	技术面	自我介绍+项目/实习经验 +技术问答+算法编程	/
第三面	技术Leader面	自我介绍+项目经验+公司发展	学习的思路，解决问题的方法 以及对行业的深度宽度的思考
第四面	HR面	基础人力问题	/

**PS：**以上流程为大白总结归纳所得，以供参考。

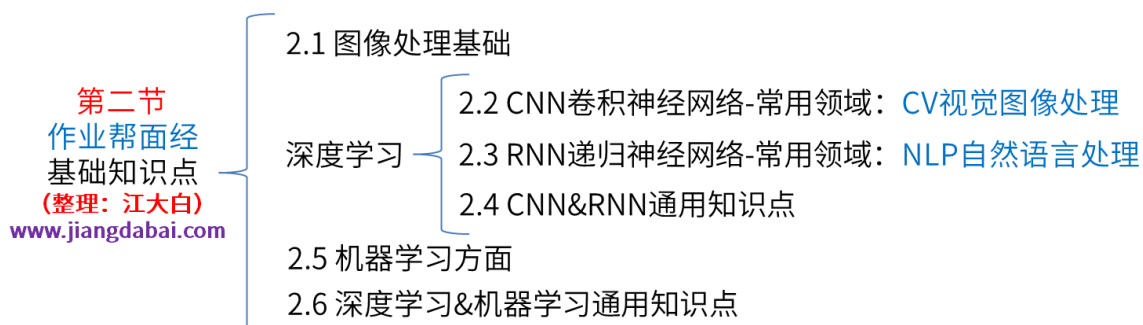
**其他注意点：**

- 有的人是三面，有的人是四面

## 1.4 作业帮面试心得汇总

- ★ 非常抠细节，非常看重码代码的能力
- ★ 有些人一面上来就是三道编程题
- ★ 算法会问很多，看重代码能力
- ★ 整体来说，面试难度中规中矩，主要看重的是实习，可能是实习做过相关的业务，所以问业务比较多，机器学习基础是回答实习问题的过程中带出来的
- ★ 面试官都挺好的，不会刻意为难，不会的我就坦白说没有了解过，面试官都会说ok，那我们换一个，也没有因此就挂掉
- ★ 虽然是算法岗，也会问一些计算机网络的问题

## 2 作业帮面经涉及基础知识点



### 2.1 图像处理基础

- canny 算子的原理说一下?

### 2.2 深度学习: CNN 卷积神经网络方面

#### 2.2.1 讲解相关原理

##### 2.2.1.1 池化方面

- 有哪些 pooling? 答了 maxpooling, averagepooling 和 global average polling。

##### 2.2.1.2 网络结构方面

- resnet v2 和 resnet v1 的区别?
- inception 各个版本的讲解;
- mobileNet v2 的实现?

##### 2.2.1.3 其他方面

- CNN 的 Batch Normalization 工作原理, 作用是什么? 预测的时候怎么工作的?

#### 2.2.2 数学计算

- 计算传统卷积和 depthwise - pointwise 的计算量?

### 2.2.3 激活函数类

- 写一下 softmax 的公式

## 2.3 深度学习：RNN 递归神经网络方面

### 2.3.1 讲解相关原理

- bilstm-crf 原理，讲 crf 原理？
- LSTM 与 RNN 的不同？

### 2.3.2 手绘网络原理

- LSTM 画图

## 2.4 深度学习：CNN&RNN 通用的问题

### 2.4.1 基础知识点

- CNN 为什么比 NLP 能够取得好的效果？

### 2.4.2 模型评价

- 模型的评价指标 AUC 是如何计算的，ROC 曲线的横纵坐标代表了什么含义？
- 如果把不平衡的训练集（正负样本 1：3）通过降采样平衡后，那么对于平衡后的 AUC 值和预测概率值有怎样的变化？

## 2.5 传统机器学习方面

### 2.5.1 讲解相关原理

#### 2.5.1.1 数据准备

无

### 2.5.1.2 特征工程

#### ① 特征降维

无

#### ② 特征选择

- 说了一些特征处理，特征工程的手段？

### 2.5.1.3 有监督学习-分类和回归方面

#### ① 分类回归树（集成学习）

##### A.基于 bagging: 随机森林

- 对比 Random Forest 和 XGBoost?

##### B.基于 boosting: Adaboost、GDBT、XGBoost

- XGBoost 模型里面参数有哪些？XGBoost 模型中对数据进行采样的好处？
- GBDT 分类与回归的原理？GBDT 为什么能用残差拟合？
- xgboost 对比 gbdt 的不同，xgboost 最大的缺点是什么？
- xgboost 如何分裂，分裂时的依据，如何控制它是否分裂的，xgboost 和 GBDT 区别？GBDT 能不能预排序？
- GBDT 的 GB 和 DT 从头说一下它的过程，在进行分类和回归的时候有什么不同？
- xgboost 防止过拟合的方法，xgboost 过拟合怎么办？
- XGBoost 的损失函数是什么，节点划分准则是什么；
- GBDT 算法时间复杂度？
- 讲讲 xgboost，每个叶子节点的值是怎么确定的，如果换一个 loss 函数还是一样吗？
- xgboost 如何处理的缺失值？
- xgboost, lightgbm 的区别？

#### ② 逻辑回归 LR

- Logistic 回归能处理浮点数吗？

- LR 除了可以用梯度下降法求还可以用什么方法求？
- LR 怎么加入非线性？
- LR 中的连续值特征是如何处理的
- 为什么 LR 要先对数据进行归一化处理
- LR 用了 sigmoid 函数，那么 LR 是线性模型还是非线性模型，为什么？
- 对于传统的 LR，SVM 和树模型，对于他们的特征，你分别是怎么处理的？
- LR 损失函数，极大似然思想

### ③ SVM（支持向量机）

- SVM 为什么要转成对偶问题进行求解，为什么对偶问题的解是原问题的解？
- SVM 如何进行多分类，多分类 hinge loss 什么形式？
- SVM 的核函数都有啥，为什么转换成对偶问题，svm 对缺失值敏感吗？
- 树模型的损失函数是什么？

### ④ 决策树（DT）

- 决策树的构成，损失函数，剪枝，特征分裂？
- 决策树过拟合怎么办，决策树剪枝方法？
- 决策树对连续值和离散值特征是否会重复利用作为分割特征？
- 如何用回归树做分类问题？
- 为什么树模型对于稀疏特征效果不好？

#### 2.5.1.4 无监督学习-聚类方面

- 聚类算法中遇到哪些问题？（kmeans 可能存在空类）

#### 2.5.2 手推算法及代码

- 推导 LR？
- 推导 SVM？

## 2.6 深度学习&机器学习面经通用知识点

### 2.6.1 损失函数方面

- 问图像识别常用的损失函数？交叉熵损失
- 问逻辑回归模型，损失函数和交叉熵的区别？
- 写一下交叉熵的公式？

### 2.6.2 激活函数方面

- 常见的激活函数有哪些，然后比较一下 tanh 和 sigmoid、relu、likelyrelu？

### 2.6.3 网络优化梯度下降方面

- 梯度下降的优化算法有哪些？

### 2.6.4 正则化方面

- 正则化为什么有效果？

### 2.6.5 过拟合&欠拟合方面

- 过拟合是什么，为什么发生过拟合，怎么解决？
- 防止过拟合，dropout 实际怎么操作？
- dropout 和 BN

## 3 作业帮面经涉及项目知识点

第三节  
作业帮面经  
项目知识点  
(整理: 江大白)  
[www.jiangdabai.com](http://www.jiangdabai.com)

- 3.1 深度学习：CNN卷积神经网络方面
- 3.2 深度学习：RNN递归神经网络方面
- 3.3 强化学习方面
- 3.4 机器学习方面

## 3.1 深度学习：CNN 卷积神经网络方面

### 3.1.1 目标检测方面

- IOU 如何计算？
- 说一下 Faster rcnn 框架？

## 3.2 深度学习：RNN 递归神经网络方面

### 3.2.1 自然语言处理 NLP

#### ① Bert

- Bert 应用于有空格丢失或者单词拼写错误等数据是否有效，有何改进办法？

#### ② Word2vec

- 讲解了 word2vec，损失函数，怎么优化？
- 词袋到 word2vec 改进了什么、word2vec 到 bert 又改进了什么？

#### ③ 其他

- 如果 label 不是 one-hot 而是一个服从某个分布的向量，那么该如何设置损失函数？

## 3.3 强化学习

- 问了简历相关的项目，问 GAN 是怎么组成的，问生成器用了哪些损失函数？

## 3.4 机器学习方面

### 3.4.1 推荐系统

- 说一说 FM，FM 参数初始化是什么样子的呢？
- 介绍了一下 deepFM，还有 deepFM 的改进 deep cross network 和 xdeepfm？



## 4 数据结构与算法分析相关知识点

第四节  
作业帮面经  
数据结构与算法分析  
(整理: 江大白)  
[www.jiangdabai.com](http://www.jiangdabai.com)

- 4.1 数据结构与算法分析：线性表、属、散列表、图等
- 4.2 算法思想实战及智力题
- 4.3 其他方面：数论、计算几何、矩阵运算等
- 4.4 Leetcode&剑指offer原题

### 4.1 数据结构与算法分析

#### 4.1.1 线性表

##### 4.1.1.1 数组

- 寻找数组的最短连续子数组的长度，使得子数组的和大于等于  $t$ ?
- 旋转数组找最大值
- 旋转数组求给定数的下标?
- 找出数组中最大的  $k$  个元素;

想法:

(1) 快排，堆排或者归并排序进行排序，然后输出;

(2) 优先队列实现;

(3) 定义一个长度为  $k$  的 queue，如果大于则弹出队头，维护一个递增队列，每次都需要对比队尾元素和当前元素，在做相应操作，时间复杂度过高。

- 找出有序数组中第一个小于  $k$  的数?
- 找出乱序数组中的第  $k$  大的数?
- 顺时针打印数组?
- 连续子数组乘积不大于  $k$  的可行解个数?
- 1,5,9

10,11,13

12,13,15

数组每行有序，每列有序，求第 k 小的数？

- 求正方形数组的 90 度旋转？
- 给定一个包括 n 个整数的数组 nums 和一个目标值 target。找出 nums 中的三个整数，使得它们的和与 target 最接近。返回这三个数的和。假定每组输入只存在唯一答案。
- 数组，链表的区别

#### 4.1.1.2 链表

- 给定一个链表，划分成子链表反转。例如，10 个节点，前 4 个反转，4-8 反转，剩下两个不足 4，不用反转？

#### 4.1.1.3 栈

- 给定两个队列，实现一个栈的功能？

#### 4.1.1.4 字符串

- 字符串中第一个不重复的字符？
- 输入一个 string，例如 'aacab'，求所有可能的分割使得每一个 substring 都是回文数？

例如 [a aca b] [a a c a b] [aa c a b]

- [2、1、6、3、100、200、4] 找出最长连续序列 1、2、3、4。要 O(n) 复杂度，lz 只写了 O(nlogn) 的。然后换了个简单的，最长连续子字符串？
- 给你一个字符串 S、一个字符串 T，请在字符串 S 里面找出：包含 T 所有字母的最小子串？

字符串分割：(1) 传统方法，O(n^2) 时间复杂度较高；(2) 队列实现，时间复杂度仍旧较高；

- 最长递增子序列（要求复杂度  $n\log(n)$ ），最长不重复连续字符串（要求复杂度  $n$ ）

#### 4.1.2 树

- 输入一颗多叉树的根节点，求叶节点的值？
- 给出一串树的先序遍历字符串，\*代表空节点，要求（1）写出节点定义（2）构建树并返回根节点（3）打印中序遍历
- 已知二叉树的先序和中序遍历序列，如何还原二叉树的结构？
- 给出前序遍历和中序遍历，重建二叉树？
- 给定二叉树的前序和中序序列，重构二叉树？
- 层次遍历建树、后序打印(非递归)
- 给定  $n$ ，用 1 到  $n$  作为二叉搜索树的节点值，返回  $n$  个点所能组成的二叉搜索树的个数，比如  $n=3$ ？
- 验证是平衡二叉树？
- 完全二叉树的定义

#### 4.1.3 排序

- 排序算法最小的时间复杂度是多少？什么算法？
- 对于 1 亿个整数进行排序最小的时间复杂度是多少？（基数排序时间复杂度为  $O(n)$ ），但实际中数据量大的时候其实快排比基数排序要快？这是为什么？那对 1 亿个浮点数进行排序最小的时间复杂度呢？为什么？
- 如何从 200 亿数字中，找到最大的 100 个？
- 手写快排，快速排序
- 写一下堆排序
- 合并  $k$  个有序数组，思路（堆排和归并）

## 4.2 算法思想实战及智力题

### 4.2.1 算法思想实战

- 两个人乒乓球，A 比 B 厉害，那么 A 想赢的话，3 局 2 胜好还是 5 局 3 胜好？这个题是可以用二项分布计算的。
- 两个年级同学的身高数据，正态的，怎么判断数据是相同的分布？
- 合法括号的题目
- 青蛙跳台阶的题目
- Top k 个高频元素
- 数学题，a, b, c 三个连续的数，其中有两个质数，证明剩下的那个数一定可以被 6 整除

### 4.2.2 智力题

- N3 的木板，用 13 的块铺满，有多少种方案？
- 给定一个大小为  $n \times 3$  的木板，问用大小为  $1 \times 3$  的木板进行填充，有多少种填充方法？  
(斐波那契数列，动态规划)
- 机器人从左上角走到右下角的步数？

## 4.3 其他方面

### 4.3.1 数论

- 三个连续的数，首先它们都大于 6，并且它们有两个质数，证明另一个一定能被 6 整除？
- 设计一个随机算法，将一个整数 M 分成 N 份 要求每份在区间  $[\min V, \max V]$  之间？
- 求无符号 int 型中转为二进制有多少 1，并考虑简便算法减少位运算次数？

### 4.3.2 计算几何

- 如何判断平面内的线段是否相交？

### 4.3.3 概率分析

- 比如五个硬币有两个两面都是花，两个两个都是字，一个一花一字，问给定正面为字的情况下背面为字的概率？
- 两个随机数产生器，R1 以 0.7 的概率产生 1，以 0.3 的概率产生 0，R2 以 0.3 的概率产生 1，0.7 的概率产生 0。问如何组合这两种产生器，使新得到的随机数产生器，以 0.5 的概率产生 1，0.5 的概率产生 0，随机数产生器可复用？

### 4.3.4 矩阵运算

- 给定一个矩阵，以及两个左边，返回以两个左边为左上角和右下角的矩形中所有数字之和。要求每次调用时间复杂度尽可能低，意思就是不能遍历相加。

（先用动态规划的思路，求出以每一个点为右下角的大矩形的所有数字的和，然后根据传入的两个坐标，将矩形面积相减就可以了）

### 4.3.5 其他

- 求  $x$  的平方根，精度 0.01？
- 100 以内的奇数以及质数，将其排列，需要满足相邻两个数不互质？
- 给定一个长度为  $N$  个整数的 list，用最少的比较次数，得出 list 中的最大值和最小值？

## 4.4 Leetcode&剑指 offer 原题

- Leetcode 29 题
- Leetcode 原题：二叉树最大路径和
- Leetcode 原题：动态规划，求递增数组的最长子序列长度，要求子序列满足斐波那契数列？

## 5 编程高频问题：Python&C/C++方面

第五节  
作业帮面经  
编程高频问题  
(整理: 江大白)  
[www.jiangdabai.com](http://www.jiangdabai.com)

- 5.1 Python方面：网络框架、基础知识、手写代码相关
- 5.2 C/C++ 方面：基础知识、手写代码相关

### 5.1 python 方面

#### 5.1.1 基础知识

##### 5.1.1.1 区别比较

- 元组和列表的区别？如何遍历元组？

##### 5.1.1.2 讲解原理

- python 装饰器讲一下原理？

##### 5.1.1.3 讲解应用

- Python 中如何导入包，以及如何导包的执行顺序
- 如何深拷贝？

### 5.2 C/C++方面

无

## 6 操作系统高频问题：数据库&线程&常用命令等

第六节  
作业帮面经  
操作系统高频问题  
(整理: 江大白)  
[www.jiangdabai.com](http://www.jiangdabai.com)

- 6.1 数据库方面：基础知识、手写代码相关
- 6.2 操作系统方面：TCP、线程&进程、常用命令相关

## 6.1 数据库方面

无

## 6.2 操作系统方面

- TCP/IP 模型
- 三次握手

# 7 技术&产品&开放性问题

## 7.1 技术方面

- alpha go 的 19x19 的棋盘学习的时候，会用 pooling 吗？你猜它一共学了多少的特征？
- 有一个森林里面树木年龄、高度的数据，数据可能会有录入错误，怎么设计方案来挑出错误的数据？
- 大数据算法岗：布隆过滤器，URL 去重（一个网站有 100 亿 url 要存入一个黑名单中，每条 url 平均 50 字节。这个黑名单要怎么存？若此时随便输入一个 url，你如何快速判断该 url 是否在这个黑名单中？）

## 7.2 产品方面

- 你的流量预测模型能否应用到作业帮的场景中，预测暑期多少学生吗？